The Estimation Performance of Nonlinear Least Squares for Phase Retrieval

Meng Huang and Zhiqiang Xu

***Abstract*— Suppose that $\boldsymbol{y} = |A\boldsymbol{x}_0| + \boldsymbol{\eta}$ where $\boldsymbol{x}_0 \in \mathbb{R}^d$ is the target signal and $\eta \in \mathbb{R}^m$ is a noise vector. The aim of phase retrieval is to estimate $\boldsymbol{x}_0$ from $\boldsymbol{y}$. A popular model for estimating $\boldsymbol{x}_0$ is the nonlinear least squares $\widehat{x} := \mathrm{argmin}_{\boldsymbol{x}} \||A\boldsymbol{x}| - \boldsymbol{y}\|_2$. One has already developed many efficient algorithms for solving the model, such as the seminal error reduction algorithm. In this paper, we present the estimation performance of the model with proving that $\|\widehat{x} - \boldsymbol{x}_0\| \lesssim \|\boldsymbol{\eta}\|_2/\sqrt{m}$ under the assumption of A being a Gaussian random matrix. We also prove the reconstruction error $\|\boldsymbol{\eta}\|_2/\sqrt{m}$ is sharp. For the case where $\boldsymbol{x}_0$ is sparse, we study the estimation performance of both the nonlinear Lasso of phase retrieval and its unconstrained version. Our results are non-asymptotic, and we do not assume any distribution on the noise $\boldsymbol{\eta}$. To the best of our knowledge, our results represent the first theoretical guarantee for the nonlinear least squares and for the nonlinear Lasso of phase retrieval.**

***Index Terms*— Phase retrieval, estimation performance, nonlinear least squares, nonlinear Lasso.**

I. INTRODUCTION

A. Phase Retrieval

SUPPOSE that $\boldsymbol{x}_0 \in \mathbb{F}^d$ with $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ is the target signal. The information that we gather about $\boldsymbol{x}_0$ is

$$\boldsymbol{y} = |A\boldsymbol{x}_0| + \eta,$$

where $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)^\top \in \mathbb{F}^{m\times d}$ is the known measurement matrix and $\eta \in \mathbb{R}^m$ is a noise vector. Throughout this paper, we often assume that $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix with entries $a_{jk} \sim N(0,1)$ and $m \gtrsim d$. In addition, we also assume that η is a fixed or random vector independent of A.

The aim of phase retrieval is to estimate $\boldsymbol{x}_0$ from $\boldsymbol{y}$. Phase retrieval is raised in numerous applications such as X-ray crystallography [10], [16], microscopy [15], astronomy [5], coherent diffractive imaging [8], [21] and optics [27] etc.

Manuscript received May 28, 2019; revised October 4, 2019; accepted March 16, 2020. Date of publication March 26, 2020; date of current version November 20, 2020. The work of Zhiqiang Xu was supported in part by NSFC under Grant 91630203 and Grant 11688101 and in part by the Beijing Natural Science Foundation under Grant Z180002. *(Corresponding author: Zhiqiang Xu.)*

The authors are with the School of Mathematical Sciences, Institute of Computational Mathematics, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100049, China (e-mail: hm@lsec.cc.ac.cn; xuzq@lsec.cc.ac.cn).

Communicated by M. Davenport, Associate Editor for Statistical Learning.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.2983562

A popular model for recovering $\boldsymbol{x}_0$ is

$$\underset{\boldsymbol{x}\in\mathbb{F}^d}{\mathrm{argmin}} \, \||A\boldsymbol{x}| - \boldsymbol{y}\|_2. \tag{I.1}$$

If $\boldsymbol{x}_0$ is sparse, both the constrained nonlinear Lasso model

$$\min_{\boldsymbol{x}\in\mathbb{F}^d} \||A\boldsymbol{x}| - \boldsymbol{y}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_1 \le R, \tag{I.2}$$

and its unconstrained version

$$\min_{\boldsymbol{x}\in\mathbb{F}^d} \||A\boldsymbol{x}| - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{x}\|_1, \tag{I.3}$$

have been considered for recovering $\boldsymbol{x}_0$. As we will see later, one has already developed many efficient algorithms to solve (I.1). The aim of this paper is to study the performance of (I.1) as well as of (I.2) and (I.3) from the theoretical viewpoint. Particularly, we focus on the question: *how well can one recover $\boldsymbol{x}_0$ by solving these above three models?*

B. Algorithms for Phase Retrieval

1) Algorithms for (I.1): Although the objective function in (I.1) is non-convex, many computational algorithms turn to be successful actually with a good initialization. One of the oldest algorithms for phase retrieval is the error-reduction algorithm which is raised in [6], [8]. The error-reduction algorithm is to solve the following model

$$\min_{\boldsymbol{x}\in\mathbb{F}^d, C\in\mathbb{F}^{m\times m}} \|A\boldsymbol{x} - C\boldsymbol{y}\|_2, \tag{I.4}$$

where $C = \mathrm{diag}(c_1, \ldots, c_m)$ with $|c_j| = 1, j = 1, \ldots, m$. The error-reduction is an alternating projection algorithm that iterates between C and $\boldsymbol{x}$. A simple observation is that $\boldsymbol{x}^\#$ is a solution to (I.1) if and only if $(\boldsymbol{x}^\#, \mathrm{diag}(\mathrm{sign}(A\boldsymbol{x}^\#)))$ is a solution to (I.4). Hence, the error-reduction algorithm can be used to solve (I.1). The convergence property of the error-reduction algorithm is studied in [18], [26]. Beyond the error-reduction algorithm, one also develops many algorithms to solve (I.1). For example, in [28], Wang, Giannakis and Eldar develop the TAF method for solving the model with proving TAF converges linearly to the global optimal solution. In [32], Zhang, Zhou, Liang and Chi also develop the Reshaped WF algorithm and prove the linear convergence to global solution of (I.1). In [31], Wei proposed the Kaczmarz algorithms which is exactly stochastic gradient descent for (I.1) (see [23]). The convergence property of the Kaczmarz algorithm was studied in [14], [23]. For the sparse phase retrieval, a standard ℓ_1 norm term is added to the above objective functions to obtain the models for sparse phase retrieval, such as (I.2) and (I.3). The theoretical framework for the recovery of sparse signals from the magnitude of the measurements is built up in [30]. The gradient descent method with thresholding can be used to solve those models successfully [1], [29]. A simple two-stage sparse phase retrieval strategy is also studied in [13].

2) Other Algorithms: An alternative model for phase retrieval is

$$\min_{\boldsymbol{x}\in\mathbb{F}^d} \sum_{i=1}^{m} \left(|\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2 - y_i^2\right)^2. \tag{I.5}$$

Note that the object function in (I.5) is smooth and one develops many algorithms for solving it, such as Gauss-Newton algorithms [7], and trust-region methods [22]. A gradient descent method is applied to solve (I.5), which provides the Wirtinger Flow (WF) [2] and Truncated Wirtinger Flow (TWF) [4] algorithms. It has been proved that both WF and TWF algorithms linearly converge to the true solution when the measurement vectors are random Gaussian measurements. For the case where the measurement vectors are Fourier measurements, in [11], [12], a gradient descent method is proposed for solving (I.5).

One convex method to handle phase retrieval problem is PhaseLift [3] which lifts the quadratic system to recover a rank-1 positive semi-definite matrix by solving a semi-definite programming. An alternative convex method is PhaseMax [9] which recasts this problem as a linear programming by an anchor vector. In [17], a numerical comparison between PhaseLift and a gradient method for solving (I.5) is presented.

C. Our Contributions

The aim of this paper is to study the estimation performance of the nonlinear least squares for phase retrieval. We obtain the measurement vector $\boldsymbol{y} = |A\boldsymbol{x}_0| + \eta$, where $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^\top$ is the measurement matrix with $\boldsymbol{a}_j \in \mathbb{R}^d, \boldsymbol{x}_0 \in \mathbb{R}^d$ and $\eta \in \mathbb{R}^m$ is a noise vector. We would like to estimate $\boldsymbol{x}_0$ from $\boldsymbol{y}$.

Firstly, we consider the following nonlinear least squares model:

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \quad \||A\boldsymbol{x}| - \boldsymbol{y}\|_2^2. \tag{I.6}$$

Though one has already developed many algorithms for finding the solution to (I.6), to our knowledge, there is no result concerning the reconstruction error of (I.6) so far. One of main results in this paper is the following theorem which shows that the reconstruction error of model (I.6) can be reduced proportionally to $\|\boldsymbol{\eta}\|_2/\sqrt{m}$ and it becomes quite small when $\|\eta\|_2$ is bounded and m is large. Throughout the paper, to state conveniently, we use $A \lesssim B$ to denote $A \le C_0\, B$ for any $A, B \in \mathbb{R}$, where $C_0 \in \mathbb{R}_+$ is an absolute constant and the value varies with the context. The notion $\gtrsim$ can be defined similarly.

Theorem I.1: Suppose that $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix whose entries are independent Gaussian random variables. We assume that $m \gtrsim d$. The following holds with probability at least $1-3\exp(-cm)$ where $c>0$ is an absolute constant. For any fixed vector $\boldsymbol{x}_0 \in \mathbb{R}^d$, suppose that $\widehat{\boldsymbol{x}} \in \mathbb{R}^d$ is any global solution to (I.6). Then

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \lesssim \frac{\|\eta\|_2}{\sqrt{m}}. \tag{I.7}$$

The next theorem implies that the reconstruction error in Theorem I.1 is sharp in the power of m.

Theorem I.2: Let $m \gtrsim d$. Assume that $\boldsymbol{x}_0 \in \mathbb{R}^d$ is a fixed vector. Assume that $\eta \in \mathbb{R}^m$ is a fixed vector which satisfies $\sqrt{2/\pi}\cdot|\sum_{i=1}^m \eta_i|/m \ge \delta_0$ and $\|\eta\|_2/\sqrt{m} \le \delta_1$ for some constants $\delta_0 > 0$ and $\delta_1 > 0$. Suppose that $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix whose entries are independent Gaussian random variables. Let $\widehat{\boldsymbol{x}}$ be any global solution to (I.6). Then there exists an $\epsilon_0 > 0$ and constants $c > 0$, $c_{\delta_0,\boldsymbol{x}_0} > 0$ such that the following holds with probability at least $1 - 6\exp(-c\epsilon_0^2\, m)$:

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \ge c_{\delta_0,\boldsymbol{x}_0}. \tag{I.8}$$

Here, the constant $c_{\delta_0,\boldsymbol{x}_0}$ only depends on δ_0 and $\|\boldsymbol{x}_0\|_2$.

Remark I.3: According to the proof of Theorem I.2, we can deduce that

$$c_{\delta_0,\boldsymbol{x}_0} = \min\{\delta_0/36, \|\boldsymbol{x}_0\|_2 \sin\theta_2\}, \tag{I.9}$$

where $\theta_2 := f^{-1}(\delta_0/(4\|\boldsymbol{x}_0\|_2)) > 0$ and $f(\theta) := 2/\pi\cdot(\sin\theta + (\pi/2-\theta)\cos\theta) - |\cos\theta|$ is a monotonically increasing function for $\theta \in [0,\pi/2]$. For the case where $\delta_0/(4\|\boldsymbol{x}_0\|_2) \notin \{f(\theta) : \theta \in [0,\pi/2]\}$, we can choose $c_{\delta_0,\boldsymbol{x}_0} = \delta_0/36$.

Remark I.4: We next explain the reason why the error bound in Theorem I.1 is sharp in the power of m. To derive a contradiction, we assume that there exists an $\alpha > 0$ such that

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \lesssim \frac{\|\eta\|_2}{m^{1/2+\alpha}} \quad \text{for } m \gtrsim d, \tag{I.10}$$

holds for any fixed $\boldsymbol{x}_0 \in \mathbb{R}^d$ with high probability. Here, $\widehat{\boldsymbol{x}} \in \mathbb{R}^d$ is any solution to (I.6) which depends on $\boldsymbol{x}_0$ and η. We assume

$$\underline{\lim}_{m\to\infty} |\sum_{i=1}^m \eta_i/m| \ge \delta_0 \quad \text{and} \quad \overline{\lim}_{m\to\infty} \|\eta\|_2/\sqrt{m} \le \delta_1$$

where $\delta_0, \delta_1 > 0$. For example, if we take $\eta = (1,\ldots,1)^\top \in \mathbb{R}^m$, then $\delta_0 = \delta_1 = 1$. For a fixed $\boldsymbol{x}_0 \in \mathbb{R}^d$, Theorem I.2 implies the following holds with high probability

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \ge c_{\delta_0,\boldsymbol{x}_0}, \quad \text{for } m \gtrsim d, \tag{I.11}$$

where $c_{\delta_0,\boldsymbol{x}_0} > 0$. However, (I.10) implies that

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \lesssim \frac{\delta_1}{m^\alpha} \to 0, \quad m\to\infty,$$

which contradicts with (I.11). Hence, (I.10) does not hold.

We next turn to the phase retrieval for sparse signals. Here, we assume that $\boldsymbol{x}_0 \in \mathbb{R}^d$ is s-sparse, which means that there are at most s nonzero entries in $\boldsymbol{x}_0$. We first consider the estimation performance of the following constrained nonlinear Lasso model

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \||A\boldsymbol{x}| - \boldsymbol{y}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_1 \le R, \tag{I.12}$$

where R is a parameter which specifies a desired sparsity level of the solution. The following theorem presents the estimation performance of model (I.12):

Theorem I.5: Suppose that $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix whose entries are independent Gaussian random variables. If $m \gtrsim s\log(ed/s)$, then the following holds with probability at least $1-3\exp(-cm)$ where $c>0$ is a constant. For any fixed s-sparse vector $\boldsymbol{x}_0 \in \mathbb{R}^d$, suppose that $\widehat{\boldsymbol{x}} \in \mathbb{R}^d$ is any global solution to (I.12) with parameter $R := \|\boldsymbol{x}_0\|_1$ and $\boldsymbol{y} = |A\boldsymbol{x}_0| + \eta$. Then

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \lesssim \frac{\|\eta\|_2}{\sqrt{m}}.$$

The unconstrained Lagrangian version of (I.12) is

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \quad \||A\boldsymbol{x}| - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{x}\|_1, \tag{I.13}$$

where $\lambda > 0$ is a parameter which depends on the desired level of sparsity. The following theorem presents the estimation performance of model (I.13):

Theorem I.6: Suppose that $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix whose entries are independent Gaussian random variables. If $m \gtrsim s\log(ed/s)$, then the following holds with probability at least $1 - \exp(-cm) - 1/d^2$ where $c > 0$ is a constant. For any fixed s-sparse vector $\boldsymbol{x}_0 \in \mathbb{R}^d$, suppose that $\widehat{\boldsymbol{x}} \in \mathbb{R}^d$ is any global solution to (I.13) with the positive parameter $\lambda \gtrsim \|\eta\|_1 + \|\eta\|_2\sqrt{\log d}$ and $\boldsymbol{y} = |A\boldsymbol{x}_0| + \eta$. Then

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \lesssim \frac{\lambda\sqrt{s}}{m} + \frac{\|\eta\|_2}{\sqrt{m}}. \tag{I.14}$$

We can use a similar method in Remark I.4 to show that the reconstruction error in Theorem I.5 is sharp. In Theorem I.6, if η satisfies $\|\eta\|_1/\|\eta\|_2 \lesssim \sqrt{m/s}$ then $\lambda\sqrt{s}/m + \|\eta\|_2/\sqrt{m} \lesssim \|\eta\|_2/\sqrt{m}$. This implies the reconstruction error in Theorem I.6 is tight provided $\|\eta\|_1/\|\eta\|_2 \lesssim \sqrt{m/s}$. For the case where η is a general vector, we conjecture that Theorem I.6 still holds provided $\lambda \gtrsim \|\eta\|_2\sqrt{\log d}$. Under this conjecture, we can take $\lambda \approx \|\eta\|_2\sqrt{\log d}$ and replace (I.14) by

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \lesssim \frac{\|\eta\|_2}{\sqrt{m}}.$$

Numerical experiments in Example I.9 also support this conjecture.

D. Comparison to Related Works

1) Least Squares: We first introduce the estimation of signals from the noisy linear measurements. Suppose that $\boldsymbol{x}_0 \in \mathbb{R}^d$ is the target signal. Set

$$\boldsymbol{y}' = A\boldsymbol{x}_0 + \boldsymbol{\eta},$$

where $A \in \mathbb{R}^{m\times d}$ is the measurement matrix and $\boldsymbol{\eta} \in \mathbb{R}^m$ is a noise vector. We suppose that A is a Gaussian random matrix with entries $a_{jk} \sim N(0,1)$ and we also suppose that $m \gtrsim d$. A popular method for recovering $\boldsymbol{x}_0$ from $\boldsymbol{y}'$ is the least squares:

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \|A\boldsymbol{x} - \boldsymbol{y}'\|_2^2. \tag{I.15}$$

Then the solution of model (I.15) is $\widehat{\boldsymbol{x}'} = (A^\top A)^{-1}A^\top \boldsymbol{y}'$, which implies that

$$\widehat{\boldsymbol{x}'} - \boldsymbol{x}_0 = (A^\top A)^{-1}A^\top\eta.$$

Thus with probability at least $1 - 4\exp(-cd)$ one has

$$\begin{aligned} \|\widehat{\boldsymbol{x}'} - \boldsymbol{x}_0\|_2 &= \|(A^\top A)^{-1}A^\top\eta\|_2 \\ &\le \|(A^\top A)^{-1}\|_2\|A^\top\eta\|_2 \\ &\lesssim \frac{\sqrt{d}}{m}\|\eta\|_2, \end{aligned}$$

where the last inequality follows from the fact that $\|A^\top\eta\|_2 \le 3\sqrt{d}\|\eta\|_2$ and $\lambda_{\min}(A) \ge O(\sqrt{m})$ hold with probability at least $1 - 4\exp(-cd)$ for any Gaussian random matrix [24, Theorem 7.3.3]. Then the following holds with high probability

$$\|\widehat{\boldsymbol{x}'} - \boldsymbol{x}_0\|_2 \lesssim \frac{\sqrt{d}\|\eta\|_2}{m}, \tag{I.16}$$

where $\widehat{\boldsymbol{x}'}$ is the solution of (I.15).

For nonlinear least squares with phaseless measurement $\boldsymbol{y} = |A\boldsymbol{x}_0| + \eta$, we consider

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \||A\boldsymbol{x}| - \boldsymbol{y}\|_2. \tag{I.17}$$

Theorem I.1 implies that

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \lesssim \frac{\|\eta\|_2}{\sqrt{m}} \tag{I.18}$$

where $\widehat{\boldsymbol{x}}$ is any solution to (I.17). Remark I.4 implies that the upper bound is sharp. Note that the error order about m for nonlinear least squares is $O(1/\sqrt{m})$ while one for least squares is $O(1/m)$. Hence, the result in Theorem I.1 highlights an essential difference between linear least square model (I.15) and the nonlinear least square model (I.17).

2) Lasso: If assume that the signal $\boldsymbol{x}_0$ is s-sparse and $\boldsymbol{y}' = A\boldsymbol{x}_0 + \boldsymbol{\eta}$, one turns to the Lasso

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \|A\boldsymbol{x} - \boldsymbol{y}'\|_2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_1 \le R. \tag{I.19}$$

If $m \gtrsim s\log d$, then the solution $\widehat{\boldsymbol{x}'}$ of (I.19) satisfies

$$\|\widehat{\boldsymbol{x}'} - \boldsymbol{x}_0\|_2 \lesssim \|\eta\|_2\sqrt{s\log d}/m \tag{I.20}$$

with high probability (see [24]).

For the nonlinear Lasso, Theorem I.5 shows that any solution $\widehat{\boldsymbol{x}}$ to $\min_{\|\boldsymbol{x}\|_1 \le \|\boldsymbol{x}_0\|_1} \||A\boldsymbol{x}| - \boldsymbol{y}\|$ with $\boldsymbol{y} = |A\boldsymbol{x}_0| + \eta$ satisfies

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \lesssim \|\eta\|_2/\sqrt{m} \tag{I.21}$$

with high probability. Comparing (I.20) with (I.21), we find that the reconstruction error of Lasso is similar to that of nonlinear Lasso when $m = O(s\log d)$, while Lasso has a better performance than the nonlinear Lasso provided $m \gg s\log d$.

3) Unconstrained Lasso: We next turn to the unconstrained Lasso

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \|A\boldsymbol{x}-\boldsymbol{y}'\|_2^2+\lambda\|\boldsymbol{x}\|_1 \tag{I.22}$$

where $\boldsymbol{y}' = A\boldsymbol{x}_0+\eta$ and $\boldsymbol{x}_0$ is a s-sparse vector. If the parameter $\lambda \gtrsim \|\eta\|_2\sqrt{\log d}$, then $\widehat{\boldsymbol{x}}'$ satisfies

$$\|\widehat{\boldsymbol{x}}'-\boldsymbol{x}_0\|_2 \lesssim \frac{\lambda\sqrt{s}}{m}$$

with high probability (see [24]) where $\widehat{\boldsymbol{x}}'$ is the solution of (I.22).

For the sparse phase retrieval model

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \||A\boldsymbol{x}|-\boldsymbol{y}\|_2^2+\lambda\|\boldsymbol{x}\|_1 \tag{I.23}$$

with $\boldsymbol{y}=|A\boldsymbol{x}_0|+\eta$, Theorem I.6 shows that

$$\min\{\|\widehat{\boldsymbol{x}}-\boldsymbol{x}_0\|_2,\|\widehat{\boldsymbol{x}}+\boldsymbol{x}_0\|_2\} \lesssim \frac{\lambda\sqrt{s}}{m}+\frac{\|\eta\|_2}{\sqrt{m}} \tag{I.24}$$

where the parameter $\lambda \gtrsim \|\eta\|_1+\|\eta\|_2\sqrt{\log d}$ and $\widehat{\boldsymbol{x}}$ is any solution to (I.23). Our result requires that the parameter λ in nonlinear Lasso model is larger than linear case.

4) The Generalized Lasso With Nonlinear Observations: In [20], Y. Plan and R. Vershynin consider the following nonlinear observations

$$y_j = f_j(\langle \boldsymbol{a}_j,\boldsymbol{x}_0\rangle),\quad j=1,\dots,m$$

where $f_j:\mathbb{R}\to\mathbb{R}$ are independent copies of an unknown random or deterministic function f and $\boldsymbol{a}_j\in\mathbb{R}^d, j=1,\dots,m$, are Gaussian random vectors. The K-Lasso model is employed to recover $\boldsymbol{x}_0$ from $y_j, j=1,\dots,m$:

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \|A\boldsymbol{x}-\boldsymbol{y}\|_2^2 \quad \text{s.t.}\quad \boldsymbol{x}\in K, \tag{I.25}$$

where $K\subset\mathbb{R}^d$ is some known set. Suppose that $\widehat{\boldsymbol{x}}$ is the solution to (I.25). Y. Plan and R. Vershynin [20] show that $\|\widehat{\boldsymbol{x}}-\mu\cdot\boldsymbol{x}_0\|$ tends to 0 with m tending to infinity, where $\mu=\mathbb{E}(f(g)g)$ with g being a Gaussian random variable. Unfortunately, applying the result to phase retrieval problem, it gives that $\mu=\mathbb{E}(|g|\cdot g)=0$ and hence $\|\widehat{\boldsymbol{x}}\|$ tends to 0 with m tending to infinity where $\widehat{\boldsymbol{x}}$ is the solution to the least square mode (I.25) with $K=\mathbb{R}^d$ and $y_j=|\langle \boldsymbol{a}_j,\boldsymbol{x}_0\rangle|$. This means that the generalized Lasso does not work for phase retrieval. Hence, one has to employ the nonlinear Lasso (or nonlinear least squares) for solving phase retrieval. This is also our motivation for this project.

E. Numerical Experiments

The purpose of numerical experiments is to verify our results given in Subsection I-C. In our experiments, the measurement vectors $\boldsymbol{a}_1,\dots,\boldsymbol{a}_m$ are independent and identically generated from Gaussian random distribution and the noise vector $\eta\in\mathbb{R}^m$ is generated from Poisson distribution with parameter 1, i.e., the entries $\eta_i\sim\text{Pois}(1)$.

Example I.7: In this example, we show the reconstruction error $O(\|\eta\|_2/\sqrt{m})$ presented in Theorem I.1 is sharp. We choose a standard Gaussian random vector $\boldsymbol{x}_0\in\mathbb{R}^d$ with $d=100$ as the target signal and adopt the TAF [28] to solve the nonlinear least squares (I.6). We vary m within the range $[2d,40d]$. For each fixed m, we run 50 times trials and calculate the average ratio ρ_m:

$$\rho_m := \frac{\min\{\|\widehat{\boldsymbol{x}}-\boldsymbol{x}_0\|_2,\|\widehat{\boldsymbol{x}}+\boldsymbol{x}_0\|_2\}}{\|\eta\|_2/\sqrt{m}}. \tag{I.26}$$

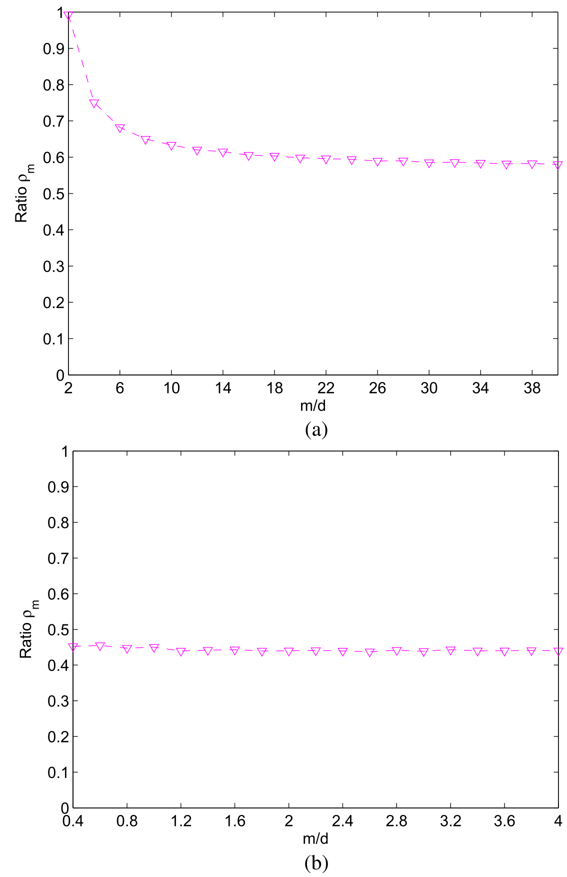

Fig. 1. Numerical experiments for verifying the reconstruction error in Theorem I.1 and Theorem I.5, respectively. (a) The ratio ρ_m against the measurement number m for nonlinear least squares. (b) The ratio ρ_m against the number of measurement m for constrained nonlinear Lasso.

Figure 1(a) depicts the ρ_m against the measurement number m. The numerical results show that ρ_m tends to be a constant around 0.58 which verifies the bound $O(\|\eta\|_2/\sqrt{m})$ in Theorem I.1 is sharp.

Example I.8: The purpose of this numerical experiment is to verify the tightness of the error bound $O(\|\eta\|_2/\sqrt{m})$ presented in Theorem I.5. We take the sparsity level $s=10$. The target sparse vector $\boldsymbol{x}_0\in\mathbb{R}^d$ is chosen randomly in standard normal distribution with $d=1000$. The support set of $\boldsymbol{x}_0$ is drawn from the uniform distribution over the set of all subsets of $[1,d]$ of size $s=10$. To solve the constrained nonlinear Lasso (I.12), we use the initialization method introduced in [1] to obtain a good guess and then update it by the projection gradient decent algorithm (see [29]). We vary m within the range $[0.4d,4d]$. For each fixed m, we run 50 times trials and calculate the average ratio ρ_m defined in (I.26). The result is plotted in Figure 1(b). Numerical results show that ρ_m tends to be a constant around 0.44 which verifies the bound $O(\|\eta\|_2/\sqrt{m})$ presented in Theorem I.5 is sharp.

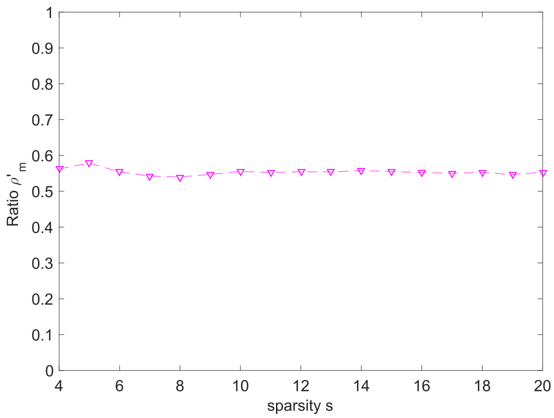

Fig. 2. Numerical experiments for verifying the the conjecture stated after Theorem I.6. The graph shows a plot of the ratio ρ'_m against the number of measurement m for unconstrained nonlinear Lasso.

Example I.9: The purpose of this numerical experiment is to present numerical evidences for the conjecture which says Theorem I.6 still holds for $\lambda \gtrsim \|\eta\|_2\sqrt{\log d}$. We vary the sparsity level s within the range $[4, 20]$. For each sparsity level s, the target sparse vector $\boldsymbol{x}_0 \in \mathbb{R}^d$ with $d = 1000$ is chosen randomly from standard normal distribution and the number of measurements $m = \lceil 2s\log(ed/s)\rceil$. The unconstrained nonlinear Lasso (I.13) is solved by combining gradient decent algorithm and soft thresholding with the regularization parameter $\lambda = 0.1\cdot\sqrt{m\log d} \approx \|\eta\|_2\sqrt{\log d}$. We run 50 times trials for each sparsity s and calculate the average ratio ρ'_m:

$$\rho'_m := \frac{\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\}}{\lambda\sqrt{s}/m + \|\eta\|_2/\sqrt{m}}.$$

The results are depicted in Figure 2. According to the numerical results, we can see that ρ'_m tends to be a constant around 0.55. This verifies the conjecture which is stated after Theorem I.6.

F. Organization

The paper is organized as follows. In Section II, we introduce some notations and lemmas which are used in this paper. We provide the proofs of main results in Section III. Some discussions are given in IV.

II. Preliminaries

The aim of this section is to introduce some definitions and lemmas which play a key role in our paper.

A. Gaussian Width

For a subset $T \subset \mathbb{R}^d$, the Gaussian width is defined as

$$w(T) := \mathbb{E}\sup_{\boldsymbol{x}\in T}\langle g, \boldsymbol{x}\rangle \quad \text{where} \quad g \sim N(0, I_d).$$

The Gaussian width $w(T)$ is one of the basic geometric quantities associated with the subset $T \subset \mathbb{R}^d$ (see [24]). We now give several examples about Gaussian width. The first example is Euclidean unit ball $\mathbb{S}^{d-1}$, where a simple calculation leads to

$$w(\mathbb{S}^{d-1}) = O(\sqrt{d}).$$

Another example is the unit ℓ_1 ball B_1^d in $\mathbb{R}^d$. It can be showed that (see e.g. [24])

$$w(B_1^d) = O(\sqrt{\log d}).$$

In this paper, we often use the following set

$$K_{d,s} := \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \le 1, \quad \|\boldsymbol{x}\|_1 \le \sqrt{s}\right\},$$

with the Gaussian width $w(K_{d,s}) = O(\sqrt{s\log(ed/s)})$ (see e.g. [24]).

B. Gaussian Concentration Inequality

Lemma II.1 [24]: Consider a random vector $X \sim N(0, I_d)$ and a Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$ with constant $\|f\|_{\text{Lip}}$: $|f(X) - f(Y)| \le \|f\|_{\text{Lip}} \cdot \|X - Y\|_2$. Then for every $t \ge 0$, we have

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \ge t\} \le 2\exp\left(-\frac{ct^2}{\|f\|_{\text{Lip}}}\right).$$

C. Strong RIP

To study the phaseless compressed sensing, Voroninski and Xu introduce the definition of strong restricted isometry property (SRIP) (see [25]).

Definition II.2 [25]: The matrix $A \in \mathbb{R}^{m\times d}$ satisfies the Strong Restricted Isometry Property of order s and constants $\theta_-, \theta_+ \in (0, 2)$ if the following holds

$$\theta_-\|\boldsymbol{x}\|_2^2 \le \min_{\substack{I\subset[m],\\ |I|\ge m/2}} \|A_I\boldsymbol{x}\|_2^2 \le \max_{\substack{I\subset[m],\\ |I|\ge m/2}} \|A_I\boldsymbol{x}\|_2^2 \le \theta_+\|\boldsymbol{x}\|_2^2 \tag{II.1}$$

for all $\boldsymbol{x} \in K_{d,s}$. Here, A_I denotes the submatrix of A where only *rows* with indices in I are kept, $[m] := \{1, \ldots, m\}$ and $|I|$ denotes the cardinality of I.

The following lemma shows that Gaussian random matrices satisfy SRIP with high probability for some non-zero universal constants $\theta_-, \theta_+ > 0$.

Lemma II.3 [25, Theorem 2.1]: Suppose that $t > 1$ and that $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix with entries $a_{jk} \sim N(0, 1)$. Let $m = O(ts\log(ed/s))$ where $s \in [1, d] \cap \mathbb{Z}$ and $t \ge 1$ is a constant. Then there exist constants θ_-, θ_+ with $0 < \theta_- < \theta_+ < 2$, independent with t, such that $A/\sqrt{m}$ satisfies SRIP of order $t\cdot s$ and constants θ_-, θ_+ with probability at least $1 - \exp(-cm)$, where $c > 0$ is an absolute constant.

Remark II.4: In [25], the authors just present the proof of Lemma II.3 for the case where $\boldsymbol{x}$ is s-sparse. Note that the set $K_{d,s}$ has covering number $N(K_{d,s}, \varepsilon) \le \exp(Cs\log(ed/s)/\varepsilon^2)$ [19, Lemma 3.4]. It is easy to extend the proof in [25] to the case where $\boldsymbol{x} \in K_{d,s}$.

III. Proof of the Main Results

To give the reconstruction errors of our three optimization models, as we will see next, in each proof it needs to present a lower bound of $\inf_{\boldsymbol{h}\in K}\|A_T\boldsymbol{h}\|_2^2$ and control the size of $\sup_{\boldsymbol{h}\in K,\, \eta\in S}\langle\boldsymbol{h}, A^\top\eta\rangle$, where $T \subset \{1, \ldots, m\}$ and K, S are some subsets of $\mathbb{R}^d$ and $\mathbb{R}^m$, respectively. For the first term, we use Strong RIP to deduce a lower bound; and for the other term we combine the Gaussian width with probability concentration inequality to handle it.

A. Proof of Theorem I.1

We begin with a simple lemma.

Lemma III.1: Suppose that $m \geq d$. Let $A \in \mathbb{R}^{m\times d}$ be a Gaussian matrix whose entries are independent Gaussian random variables. Then the following holds with probability at least $1 - 2\exp(-cm)$

$$\sup_{\substack{\boldsymbol{h}\in\mathbb{R}^d\\ \eta\in\mathbb{R}^m}} \langle \boldsymbol{h}, A^\top \eta\rangle \leq 3\sqrt{m}\|\boldsymbol{h}\|_2\|\eta\|_2.$$

Proof: Since $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix, we have $\|A\|_2 \leq 3\sqrt{m}$ with probability at least $1 - 2\exp(-cm)$ [24, Theorem 7.3.3]. We obtain that

$$\langle \boldsymbol{h}, A^\top\eta\rangle \leq \|\boldsymbol{h}\|_2\|A^\top\eta\|_2 \leq \|\boldsymbol{h}\|_2\|A^\top\|_2\|\eta\|_2 \leq 3\sqrt{m}\|\boldsymbol{h}\|_2\|\eta\|_2$$

holds with probability at least $1 - 2\exp(-cm)$, where $\|\cdot\|_2$ denotes the operator norm of the matrix. We arrive at the conclusion. □

Proof of Theorem I.1: Set $\boldsymbol{h}^- := \widehat{\boldsymbol{x}} - \boldsymbol{x}_0$ and $\boldsymbol{h}^+ := \widehat{\boldsymbol{x}} + \boldsymbol{x}_0$. Since $\widehat{\boldsymbol{x}}$ is the solution of (I.6), we have

$$\||A\widehat{\boldsymbol{x}}| - \boldsymbol{y}\|_2^2 \leq \||A\boldsymbol{x}_0| - \boldsymbol{y}\|_2^2. \tag{III.1}$$

For any index set $T \subset \{1,\ldots,m\}$, we let $A_T := [\boldsymbol{a}_j : j \in T]^\top$ be the submatrix of A. Denote

$$\begin{aligned}
T_1 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = 1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = 1\}\\
T_2 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = -1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = -1\}\\
T_3 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = 1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = -1\}\\
T_4 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = -1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = 1\}.
\end{aligned}$$

Without loss of generality, we assume that $\#(T_1 \cup T_2) = \beta m \geq m/2$ (otherwise, we can assume that $\#(T_3 \cup T_4) \geq m/2$). Then we have

$$\||A\widehat{\boldsymbol{x}}| - \boldsymbol{y}\|_2^2 \geq \|A_{T_1}\boldsymbol{h}^- - \eta_{T_1}\|_2^2 + \|A_{T_2}\boldsymbol{h}^- + \eta_{T_2}\|_2^2.$$

The (III.1) implies that

$$\|A_{T_1}\boldsymbol{h}^- - \eta_{T_1}\|_2^2 + \|A_{T_2}\boldsymbol{h}^- + \eta_{T_2}\|_2^2 \leq \|\eta\|^2$$

and hence

$$\|A_{T_{12}}\boldsymbol{h}^-\|_2^2 \leq 2\langle \boldsymbol{h}^-, A_{T_1}^\top\eta_{T_1} - A_{T_2}^\top\eta_{T_2}\rangle + \|\eta_{T_{12}^c}\|^2 \tag{III.2}$$

where $T_{12} := T_1 \cup T_2$. Choosing $t = 2$ and $s := d/2$ in Lemma II.3, we can obtain that

$$\|A_{T_{12}}\boldsymbol{h}^-\|_2^2 \geq c_0 m\|\boldsymbol{h}^-\|_2^2 \tag{III.3}$$

holds with probability at least $1 - \exp(-cm)$. On the other hand, Lemma III.1 states that with probability at least $1 - 2\exp(-cm)$ the following holds:

$$\langle \boldsymbol{h}^-, A_{T_1}^\top\eta_{T_1} - A_{T_2}^\top\eta_{T_2}\rangle \leq 6\sqrt{m}\|\boldsymbol{h}^-\|_2\|\eta\|_2. \tag{III.4}$$

Putting (III.3) and (III.4) into (III.2), we obtain

$$c_0 m\|\boldsymbol{h}^-\|_2^2 \leq 12\sqrt{m}\|\boldsymbol{h}^-\|_2\|\eta\|_2 + \|\eta_{T_{12}^c}\|_2^2 \tag{III.5}$$

with probability at least $1 - 3\exp(-cm)$, which implies that

$$\|\boldsymbol{h}^-\|_2 \lesssim \frac{\|\eta\|_2}{\sqrt{m}}.$$

For the case where $\#(T_3 \cup T_4) \geq m/2$, we can obtain that

$$\|\boldsymbol{h}^+\|_2 \lesssim \frac{\|\eta\|_2}{\sqrt{m}}$$

by a similar method to above. □

B. Proof of Theorem I.2

To this end, we present the following lemmas.

Lemma III.2: Suppose that $\widehat{\boldsymbol{x}}$ is any global solution of model (I.6). Then $\widehat{\boldsymbol{x}}$ satisfies the following fixed-point equation:

$$\widehat{\boldsymbol{x}} = (A^\top A)^{-1}A^\top(\boldsymbol{y} \odot \mathrm{s}(A\widehat{\boldsymbol{x}})), \tag{III.6}$$

where $\odot$ denotes the Hadamard product and $\mathrm{s}(A\widehat{\boldsymbol{x}}) := \left(\frac{\langle \boldsymbol{a}_1, \widehat{\boldsymbol{x}}\rangle}{|\langle \boldsymbol{a}_1, \widehat{\boldsymbol{x}}\rangle|}, \ldots, \frac{\langle \boldsymbol{a}_m, \widehat{\boldsymbol{x}}\rangle}{|\langle \boldsymbol{a}_m, \widehat{\boldsymbol{x}}\rangle|}\right)$ for any $\widehat{\boldsymbol{x}} \in \mathbb{R}^d$. Here, $\frac{\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle}{|\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle|} = 1$ is adopted if $\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle = 0$.

Proof: Let

$$L(\boldsymbol{x}) := \||A\boldsymbol{x}| - \boldsymbol{y}\|_2^2.$$

Consider the smooth function

$$G(\boldsymbol{x}, \boldsymbol{u}) := \|A\boldsymbol{x} - \boldsymbol{u} \odot \boldsymbol{y}\|_2^2$$

with $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{u} \in U := \{\boldsymbol{u} = (u_1, \ldots, u_m) \in \mathbb{R}^m : |u_i| = 1,\ i = 1, \ldots, m\}$. Recall that $L(\boldsymbol{x})$ has a global minimum at $\widehat{\boldsymbol{x}}$. Then $G(\boldsymbol{x}, \boldsymbol{u})$ has a global minimum at $(\widehat{\boldsymbol{x}}, s(A\widehat{\boldsymbol{x}}))$. Indeed, if there exists $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{u}})$ such that $G(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{u}}) < G(\widehat{\boldsymbol{x}}, s(A\widehat{\boldsymbol{x}}))$, then

$$\begin{aligned}
L(\widetilde{\boldsymbol{x}}) = \||A\widetilde{\boldsymbol{x}}| - \boldsymbol{y}\|_2^2 &\leq \|A\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{u}} \odot \boldsymbol{y}\|_2^2\\
&= G(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{u}})\\
&< G(\widehat{\boldsymbol{x}}, s(A\widehat{\boldsymbol{x}})) = L(\widehat{\boldsymbol{x}}).
\end{aligned}$$

This contradicts the assumption that $L(\boldsymbol{x})$ has a global minimum at $\widehat{\boldsymbol{x}}$. Thus we have

$$G(\widehat{\boldsymbol{x}}, s(A\widehat{\boldsymbol{x}})) \leq G(\boldsymbol{x}, s(A\widehat{\boldsymbol{x}})) \quad \text{for any} \quad \boldsymbol{x} \in \mathbb{R}^d,$$

i.e., the function $G(\boldsymbol{x}, \mathrm{s}(A\widehat{\boldsymbol{x}}))$ has a global minimum at $\widehat{\boldsymbol{x}}$. Here, we consider $G(\boldsymbol{x}, \mathrm{s}(A\widehat{\boldsymbol{x}}))$ as a function about $\boldsymbol{x}$ since $\mathrm{s}(A\widehat{\boldsymbol{x}})$ is a fixed vector. Note that $G(\boldsymbol{x}, \mathrm{s}(A\widehat{\boldsymbol{x}}))$ is differentiable and

$$\nabla G(\boldsymbol{x}, \mathrm{s}(A\widehat{\boldsymbol{x}})) = 2A^\top(A\boldsymbol{x} - \boldsymbol{y} \odot \mathrm{s}(A\widehat{\boldsymbol{x}})).$$

And $G(\boldsymbol{x}, \mathrm{s}(A\widehat{\boldsymbol{x}}))$ has a global minimum at $\widehat{\boldsymbol{x}}$, we have

$$\nabla G(\widehat{\boldsymbol{x}}, \mathrm{s}(A\widehat{\boldsymbol{x}})) = 2A^\top(A\widehat{\boldsymbol{x}} - \boldsymbol{y} \odot \mathrm{s}(A\widehat{\boldsymbol{x}})) = 0$$

which implies the conclusion. □

Lemma III.3: Let $m \gtrsim d$. Suppose that $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix whose entries are independent Gaussian random variables. For a fixed vector $\boldsymbol{x}_0 \in \mathbb{R}^d$ and a fixed noise vector $\eta \in \mathbb{R}^m$, let $\widehat{\boldsymbol{x}}$ be the solution of model (I.6). For any fixed $\epsilon > 0$, set

$$\beta_\epsilon := \left|\|\boldsymbol{x}_0\|_2 \cdot f(\theta) + \sqrt{2/\pi} \cdot \sum_{i=1}^m \eta_i/m\right| - (\|\boldsymbol{x}_0\|_2 + \|\eta\|_2/\sqrt{m})\epsilon,$$

where $f(\theta) := 2/\pi \cdot (\sin\theta + (\pi/2 - \theta)\cos\theta) - |\cos\theta|$ and θ is the angle between $\widehat{\boldsymbol{x}}$ and $\boldsymbol{x}_0$. Then the following holds with probability at least $1 - 6\exp(-c\epsilon^2\ m)$:

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \geq \beta_\epsilon/9.$$

Proof: According to Lemma III.2, we have

$$\widehat{\boldsymbol{x}} = (A^\top A)^{-1} A^\top (\boldsymbol{y} \odot \mathrm{s}(A\widehat{\boldsymbol{x}})). \tag{III.7}$$

Without loss of generality, we can assume $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2 \le \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2$, which implies that $0 \le \theta \le \pi/2$. From (III.7), we have

$$\widehat{\boldsymbol{x}} - \boldsymbol{x}_0 = (A^\top A)^{-1} A^\top (\boldsymbol{y} \odot \mathrm{s}(A\widehat{\boldsymbol{x}}) - A\boldsymbol{x}_0),$$

which implies that

$$\begin{aligned} \|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2 &\ge \sigma_{\min}((A^\top A)^{-1}) \|A^\top (\boldsymbol{y} \odot \mathrm{s}(A\widehat{\boldsymbol{x}}) - A\boldsymbol{x}_0)\|_2 \\ &\ge \frac{1}{9m} \|A^\top (\boldsymbol{y} \odot \mathrm{s}(A\widehat{\boldsymbol{x}}) - A\boldsymbol{x}_0)\|_2. \end{aligned}$$

Here, we use the fact that $\|A\|_2 \le 3\sqrt{m}$ holds with probability at least $1 - 2\exp(-cm)$ [24, Theorem 7.3.3] since $A \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix.

Without loss of generality, we can assume $\widehat{\boldsymbol{x}} \ne 0$. Indeed, (III.7) implies $A^\top \boldsymbol{y} = 0$ provided $\widehat{\boldsymbol{x}} = 0$, which gives that $\boldsymbol{x}_0 = 0$ and $\eta = 0$. Thus our conclusion holds. By the unitary invariance of Gaussian random vectors, we can take $\widehat{\boldsymbol{x}} = \|\widehat{\boldsymbol{x}}\|_2 \boldsymbol{e}_1$ and $\boldsymbol{x}_0 = \|\boldsymbol{x}_0\|_2 (\cos\theta \cdot \boldsymbol{e}_1 + \sin\theta \cdot \boldsymbol{e}_2)$, where θ is the angle between $\widehat{\boldsymbol{x}}$ and $\boldsymbol{x}_0$. Thus,

$$\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2 \ge \frac{1}{9m} \|A^\top (\boldsymbol{y} \odot \mathrm{s}(A\mathbf{e}_1) - A\boldsymbol{x}_0)\|_2 = \frac{1}{9m} \|\boldsymbol{z}\|_2,$$

where $\boldsymbol{z} := (z_1, \ldots, z_d)^\top := A^\top (\boldsymbol{y} \odot \mathrm{s}(A\mathbf{e}_1) - A\boldsymbol{x}_0)$. Note that the first entry of $\boldsymbol{z}$ is

$$z_1 = \sum_{i=1}^m \left(|a_{i,1}| (|a_i^\top \boldsymbol{x}_0| + \eta_i) - a_{i,1} \cdot a_i^\top \boldsymbol{x}_0 \right).$$

This implies that

$$\begin{aligned} \|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2 \ge& \frac{|z_1|}{9m} \\ =& \Bigg| \|\boldsymbol{x}_0\|_2 \cdot \frac{1}{9m} \sum_{i=1}^m \left| a_{i,1} (a_{i,1} \cos\theta + a_{i,2} \sin\theta) \right| \\ & - \|\boldsymbol{x}_0\|_2 \cdot \frac{1}{9m} \sum_{i=1}^m a_{i,1} (a_{i,1} \cos\theta + a_{i,2} \sin\theta) \\ & + \frac{1}{9m} \sum_{i=1}^m \eta_i |a_{i,1}| \Bigg| \\ =& \left| \frac{\|\boldsymbol{x}_0\|_2}{9m} \sum_{i=1}^m (|\xi_i| - \xi_i) + \frac{1}{9m} \sum_{i=1}^m \eta_i |a_{i,1}| \right|, \end{aligned} \tag{III.8}$$

where $\xi_i := a_{i,1}(a_{i,1}\cos\theta + a_{i,2}\sin\theta)$. It is clear that ξ_i is a subexponential random variable with $\mathbb{E}\xi_i = \cos\theta$. We claim that $\mathbb{E}|\xi_i| = 2/\pi \cdot (\sin\theta + (\pi/2 - \theta)\cos\theta)$. Then the Bernstein's inequality implies that, for any fixed $\epsilon > 0$,

$$\left| \frac{1}{m} \sum_{i=1}^m (|\xi_i| - \xi_i) - \frac{2}{\pi} \cdot (\sin\theta + (\frac{\pi}{2} - \theta)\cos\theta) + \cos\theta \right| \le \epsilon \tag{III.9}$$

holds with probability at least $1 - 2\exp(-c\epsilon^2\ m)$. We next consider $\frac{1}{m}\sum_{i=1}^m \eta_i |a_{i,1}|$. Note that $\mathbb{E}|a_{i,1}| = \sqrt{2/\pi}$. Then by Hoeffding's inequality we can obtain that

$$\left| \frac{1}{m} \sum_{i=1}^m \eta_i |a_{i,1}| - \sqrt{\frac{2}{\pi}} \cdot \frac{1}{m} \sum_{i=1}^m \eta_i \right| \le \frac{\|\eta\|_2}{\sqrt{m}} \epsilon \tag{III.10}$$

holds with probability at least $1 - 2\exp(-c\epsilon^2\ m)$ for any $\epsilon > 0$. Substituting (III.9) and (III.10) into (III.8), we obtain that

$$\begin{aligned} \|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2 \ \ge\ & \frac{1}{9} \cdot \Bigg(\left| \|\boldsymbol{x}_0\|_2 f(\theta) + \sqrt{\frac{2}{\pi}} \cdot \frac{1}{m} \sum_{i=1}^m \eta_i \right| \\ & - \left(\|\boldsymbol{x}_0\|_2 + \frac{\|\eta\|_2}{\sqrt{m}} \right) \epsilon \Bigg) \end{aligned}$$

holds with probability at least $1 - 6\exp(-c\epsilon^2\ m)$. Thus we arrive at the conclusion.

It remains to argue that $\mathbb{E}|\xi_i| = 2/\pi \cdot (\sin\theta + (\pi/2 - \theta)\cos\theta)$. By spherical coordinates integral,

$$\begin{aligned} \mathbb{E}|\xi_i| &= \mathbb{E}\left| a_{i,1}(a_{i,1}\cos\theta + a_{i,2}\sin\theta) \right| \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty r^3\ e^{-r^2/2} |\cos\phi \cos(\theta - \phi)| dr d\phi \\ &= \frac{1}{2\pi} \int_0^{2\pi} |\cos\theta + \cos(2\phi - \theta)| d\phi \\ &= \frac{1}{\pi} \int_0^{\pi} |\cos\theta + \cos\phi| d\phi \\ &= \frac{2}{\pi} (\sin\theta + (\pi/2 - \theta)\cos\theta) \end{aligned}$$

where we use the identities $2\cos\phi\cos(\theta - \phi) = \cos\theta + \cos(2\phi - \theta)$ in second line. □

Proof of Theorem I.2: From Lemma III.3, it is easy to prove that (I.8) holds for $\boldsymbol{x}_0 = 0$. Then it suffices to prove the theorem for $\boldsymbol{x}_0 \ne 0$. Since $\|\eta\|_2/\sqrt{m} \le \delta_1$ with $\delta_1 \ge 0$, there exists an $\epsilon_0 > 0$ so that

$$(\|\boldsymbol{x}_0\|_2 + \|\eta\|_2/\sqrt{m})\epsilon_0 \le \delta_0/2.$$

Set

$$\overline{\eta} := \sqrt{2/\pi} \cdot \sum_{i=1}^m \eta_i/m,$$

and

$$f(\theta) := 2/\pi \cdot (\sin\theta + (\pi/2 - \theta)\cos\theta) - |\cos\theta|, \quad 0 \le \theta \le \pi.$$

Note that $f(\theta)$ is a monotonically increasing function for $\theta \in [0, \pi/2]$.

Choosing $\epsilon = \epsilon_0$ in Lemma III.3, with probability at least $1 - 6\exp(-c\epsilon_0^2\ m)$, we have

$$\min\{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}} + \boldsymbol{x}_0\|_2\} \ge \left(\left| \|\boldsymbol{x}_0\|_2 \cdot f(\theta_0) + \overline{\eta} \right| - \delta_0/2 \right)/9, \tag{III.11}$$

where θ_0 is the angle between $\widehat{\boldsymbol{x}}$ and $\boldsymbol{x}_0$. Without loss of generality, we can assume $0 \le \theta_0 \le \pi/2$ and hence $f(\theta_0) \ge f(0) = 0$.

Noting $|\overline{\eta}| \ge \delta_0$, we divide the rest of the proof into three cases.

Case 1: $\overline{\eta} \geq \delta_0$.

In this case, (III.11) implies that

$$\min\{\|\widehat{\boldsymbol{x}}-\boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}}+\boldsymbol{x}_0\|_2\} \geq \big(\overline{\eta}-\delta_0/2\big)/9 \geq \delta_0/18$$

holds with probability at least $1-6\exp(-c\epsilon_0^2\, m)$.

Case 2: $\overline{\eta} \leq -\delta_0$ and $|\overline{\eta}| \leq \|\boldsymbol{x}_0\|_2 \cdot f(\theta_0)$.

In this case, we have $f(\theta_0) \geq \delta_0/\|\boldsymbol{x}_0\|_2$. Since the function $f(\theta)$ is monotonicity, we have $\theta_0 \geq \theta_1 := f^{-1}(\delta_0/\|\boldsymbol{x}_0\|_2) > 0$, which implies that

$$\min\{\|\widehat{\boldsymbol{x}}-\boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}}+\boldsymbol{x}_0\|_2\} \geq \|\boldsymbol{x}_0\|_2 \sin\theta_1.$$

Case 3: $\overline{\eta} \leq -\delta_0$ and $|\overline{\eta}| > \|\boldsymbol{x}_0\|_2 \cdot f(\theta_0)$.

We claim that there exists a constant $c_{\delta_0,\boldsymbol{x}_0}$ such that the following holds with probability at least $1-6\exp(-c\epsilon_0^2\, m)$

$$\min\{\|\widehat{\boldsymbol{x}}-\boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}}+\boldsymbol{x}_0\|_2\} \geq c_{\delta_0,\boldsymbol{x}_0} \tag{III.12}$$

where $c_{\delta_0,\boldsymbol{x}_0}$ only depends on δ_0 and $\|\boldsymbol{x}_0\|_2$. Indeed, if $|\overline{\eta}| - \|\boldsymbol{x}_0\|_2 f(\theta_0) \geq 3/4 \cdot |\overline{\eta}|$, then (III.11) implies

$$\begin{aligned} &\min\{\|\widehat{\boldsymbol{x}}-\boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}}+\boldsymbol{x}_0\|_2\} \\ \geq\ & \big(|\overline{\eta}| - \|\boldsymbol{x}_0\|_2 f(\theta_0) - \delta_0/2\big)/9 \\ \geq\ & \delta_0/36. \end{aligned}$$

If $|\overline{\eta}| - \|\boldsymbol{x}_0\|_2 f(\theta_0) < 3/4 \cdot |\overline{\eta}|$, then $f(\theta_0) \geq \delta_0/(4\|\boldsymbol{x}_0\|_2)$. It can also give that

$$\min\{\|\widehat{\boldsymbol{x}}-\boldsymbol{x}_0\|_2, \|\widehat{\boldsymbol{x}}+\boldsymbol{x}_0\|_2\} \geq \|\boldsymbol{x}_0\|_2 \cdot \sin\theta_2,$$

where $\theta_2 := f^{-1}(\delta_0/(4\|\boldsymbol{x}_0\|_2)) > 0$. Choosing $c_{\delta_0,\boldsymbol{x}_0} := \min\{\delta_0/36, \|\boldsymbol{x}_0\|_2 \sin\theta_2\}$, we arrive at the conclusion. □

C. Proof of Theorem I.5

We first extend Lemma III.1 to sparse case.

Lemma III.4: For any fixed $s > 0$, let $m \gtrsim s\log(ed/s)$. Suppose that $A \in \mathbb{R}^{m\times d}$ is a Gaussian matrix whose entries are independent Gaussian random variables. Set

$$K_{d,s} := \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \leq 1, \|\boldsymbol{x}\|_1 \leq \sqrt{s}\right\}.$$

Then for any fixed $\eta \in \mathbb{R}^m$, the following holds with probability at least $1-2\exp(-cm)$

$$\sup_{\substack{\boldsymbol{h}\in K_{d,s}\\ T\subset\{1,\ldots,m\}}} \langle \boldsymbol{h}, A^\top \eta_T\rangle \lesssim \sqrt{m}\cdot\|\eta\|_2\cdot\|\boldsymbol{h}\|_2, \tag{III.13}$$

where η_T denotes the vector generated by η with entries in T are themselves and others are zeros.

Proof: For any fixed $T \subset \{1,\ldots,m\}$, we have

$$\begin{aligned} \mathbb{E}\sup_{\boldsymbol{h}\in K_{d,s}} \langle \boldsymbol{h}, A^\top\eta_T\rangle &= \|\eta_T\|_2 \cdot w(K_{d,s}) \\ &\leq C\sqrt{s\log(ed/s)}\|\eta\|_2 \\ &\leq C\sqrt{m}\|\eta\|_2, \end{aligned}$$

where the first inequality follows from the fact of the Gaussian width $w(K_{d,s}) \leq C\sqrt{s\log(ed/s)}$ and the second inequality follows from $m \geq c_0 s\log(ed/s)$. We next use Lemma II.1 to give a tail bound for $\sup_{\boldsymbol{h}\in K_{d,s}}\langle \boldsymbol{h}, A^\top\eta_T\rangle$. To this end, we set

$$f(A) := \sup_{\boldsymbol{h}\in K_{d,s}} \langle \boldsymbol{h}, A^\top\eta_T\rangle.$$

We next show that $f(A)$ is a Lipschitz function on $\mathbb{R}^{m\times d}$ and its Lipschitz constant is $\|\eta\|_2$. Indeed, for any matrices $A_1, A_2 \in \mathbb{R}^{m\times d}$, it holds that

$$\begin{aligned} &\Big|\sup_{\boldsymbol{h}\in K_{d,s}} \langle \boldsymbol{h}, A_1^\top\eta_T\rangle - \sup_{\boldsymbol{h}\in K_{d,s}} \langle \boldsymbol{h}, A_2^\top\eta_T\rangle\Big| \\ \leq\ & \Big|\sup_{\boldsymbol{h}\in K_{d,s}} \langle (A_1-A_2)\boldsymbol{h}, \eta_T\rangle\Big| \\ \leq\ & \|\eta\|_2\|A_1-A_2\|_F. \end{aligned}$$

Then Lemma II.1 implies that

$$\begin{aligned} &\mathbb{P}\left\{\sup_{\boldsymbol{h}\in K_{d,s}} \langle \boldsymbol{h}, A^\top\eta_T\rangle \geq \mathbb{E}\sup_{\boldsymbol{h}\in K_{d,s}} \langle \boldsymbol{h}, A^\top\eta_T\rangle + t\right\} \\ \leq\ & 2\exp\Big(-\frac{ct^2}{\|\eta\|_2^2}\Big). \end{aligned} \tag{III.14}$$

Suppose that $C_1 > 0$ is a constant satisfying $C_1^2 \cdot c > 1$. Choosing $t = C_1\sqrt{m}\|\eta\|_2$ in (III.14), we obtain that the following holds with probability at least $1-2\exp(-c\cdot C_1^2\cdot m)$

$$\sup_{\boldsymbol{h}\in K_{d,s}} \langle \boldsymbol{h}, A^\top\eta_T\rangle \leq C_0\sqrt{m}\|\eta\|_2$$

for any fixed $T \subset \{1,\ldots,m\}$.

Finally, note that the number of all subset $T \subset \{1,\ldots,m\}$ is 2^m. Taking a union bound over all the sets gives

$$\sup_{\substack{\boldsymbol{h}\in K_{d,s}\\ T\subset\{1,\ldots,m\}}} \langle \boldsymbol{h}, A^\top\eta_T\rangle \leq C_0\sqrt{m}\|\eta\|_2$$

with probability at least $1-2\exp(-\tilde{c}m)$. Here, we use the fact of $C_1^2\cdot c > 1$. We arrive at the conclusion. □

Proof of Theorem I.5: Set $\boldsymbol{h}^- := \widehat{\boldsymbol{x}} - \boldsymbol{x}_0$, $\boldsymbol{h}^+ := \widehat{\boldsymbol{x}} + \boldsymbol{x}_0$ and set

$$\begin{aligned} T_1 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = 1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = 1\} \\ T_2 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = -1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = -1\} \\ T_3 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = 1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = -1\} \\ T_4 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = -1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = 1\}. \end{aligned}$$

Without loss of generality, we can assume that $\#(T_1\cup T_2) = \beta m \geq m/2$. Using an argument similar to one for (III.2), we obtain that

$$\|A_{T_{12}}\boldsymbol{h}^-\|_2^2 \leq 2\langle \boldsymbol{h}^-, A_{T_1}^\top\eta_{T_1} - A_{T_2}^\top\eta_{T_2}\rangle + \|\eta_{T_{12}^c}\|^2. \tag{III.15}$$

To this end, we first need to show $\|\boldsymbol{h}^-\|_1 \leq 2\sqrt{s}\|\boldsymbol{h}^-\|_2$. Indeed, let $S := \mathrm{supp}(\boldsymbol{x})$ and note that

$$\begin{aligned} \|\widehat{\boldsymbol{x}}\|_1 = \|\boldsymbol{x}_0 + \boldsymbol{h}^-\|_1 &= \|\boldsymbol{x}_0 + \boldsymbol{h}_S^-\|_1 + \|\boldsymbol{h}_{S^c}^-\|_1 \\ &\geq \|\boldsymbol{x}_0\|_1 - \|\boldsymbol{h}_S^-\|_1 + \|\boldsymbol{h}_{S^c}^-\|_1. \end{aligned}$$

Here $\boldsymbol{h}_S^-$ denotes the restriction of the vector $\boldsymbol{h}^-$ onto the set of coordinates S. Then the constrain condition $\|\widehat{\boldsymbol{x}}\|_1 \leq R := \|\boldsymbol{x}_0\|_1$ implies that $\|\boldsymbol{h}_{S^c}^-\|_1 \leq \|\boldsymbol{h}_S^-\|_1$. Using Hölder inequality, we obtain that

$$\|\boldsymbol{h}^-\|_1 = \|\boldsymbol{h}_S^-\|_1 + \|\boldsymbol{h}_{S^c}^-\|_1 \leq 2\|\boldsymbol{h}_S^-\|_1 \leq 2\sqrt{s}\|\boldsymbol{h}^-\|_2.$$

We next give a lower bound for the left hand of inequality (III.15). Set

$$K := \left\{\boldsymbol{h} \in \mathbb{R}^d : \|\boldsymbol{h}\|_2 \leq 1,\ \|\boldsymbol{h}\|_1 \leq 2\sqrt{s}\right\}.$$

Note that $\boldsymbol{h}^-/\|\boldsymbol{h}^-\|_2 \in K$. Since $A/\sqrt{m}$ satisfies strong RIP (see Lemma II.3), we obtain that

$$\|A_{T_{12}}\boldsymbol{h}^-\|_2^2 \;\geq\; c_0 m\|\boldsymbol{h}^-\|_2^2 \tag{III.16}$$

holds with probability at least $1-\exp(-cm)$, provided $m \gtrsim s\log(ed/s)$.

On the other hand, Lemma III.4 implies that

$$\langle \boldsymbol{h}^-, A_{T_1}^\top \eta_{T_1} - A_{T_2}^\top \eta_{T_2}\rangle \;\leq\; 2C\sqrt{m}\|\eta\|_2\|\boldsymbol{h}^-\|_2 \tag{III.17}$$

holds with probability at least $1-2\exp(-cm)$. Putting (III.17) and (III.16) into (III.15), we obtain that

$$c_0 m\|\boldsymbol{h}^-\|_2^2 \leq 4C\sqrt{m}\|\eta\|_2\|\boldsymbol{h}^-\|_2 + \|\eta_{T_{12}^c}\|^2 \tag{III.18}$$

holds with probability at least $1-3\exp(-cm)$. The (III.18) implies that

$$\|\boldsymbol{h}^-\|_2 \lesssim \frac{\|\eta\|_2}{\sqrt{m}}.$$

Similarly, if $\#(T_3\cup T_4) \geq m/2$, we can obtain that

$$\|\boldsymbol{h}^+\|_2 \lesssim \frac{\|\eta\|_2}{\sqrt{m}}.$$

□

D. Proof of Theorem I.6

To this end, we introduce the following lemma.

Lemma III.5: Let $A \in \mathbb{R}^{m\times d}$ be a Gaussian matrix whose entries are independent Gaussian random variables and $\eta \in \mathbb{R}^m$ be a fixed vector. Then the following holds with probability at least $1-1/d^2$

$$\sup_{\substack{\boldsymbol{h}\in\mathbb{R}^d\\ T\subset\{1,\ldots,m\}}} \langle \boldsymbol{h}, A^\top \eta_T\rangle \lesssim (\|\eta\|_1 + \|\eta\|_2\sqrt{\log d})\|\boldsymbol{h}\|_1, \tag{III.19}$$

where η_T denotes the vector generated by η with entries in T are themselves and others are zeros.

Proof: By applying Hölder's inequality with ℓ_1 and ℓ_∞ norms, we have

$$\langle \boldsymbol{h}, A^\top\eta_T\rangle \leq \|A^\top\eta_T\|_\infty \cdot \|\boldsymbol{h}\|_1.$$

Thus it is sufficient to present an upper bound of $\sup_{T\subset\{1,\ldots,m\}}\|A^\top\eta_T\|_\infty$. We use $\tilde{\boldsymbol{a}}_j \in \mathbb{R}^m, j=1,\ldots,d$, to denote the *column* vectors of A. Then for any fixed index j and $t>0$, we have

$$\mathbb{P}\left(\sup_{T\subset\{1,\ldots,m\}} |\tilde{\boldsymbol{a}}_j^\top \eta_T| > t\right) \leq \mathbb{P}\left(\sum_{i=1}^m |\eta_i||\tilde{\boldsymbol{a}}_{j,i}| > t\right).$$

A simple calculation shows that $\mathbb{E}|\eta_i||\tilde{\boldsymbol{a}}_{j,i}| = \sqrt{2/\pi}|\eta_i|$. By Hoeffding's inequality, we obtain that

$$\mathbb{P}\left(\sum_{i=1}^m |\eta_i||\tilde{\boldsymbol{a}}_{j,i}| > C\Big(\|\eta\|_1 + \|\eta\|_2\sqrt{\log d}\Big)\right) \leq \frac{1}{d^3} \tag{III.20}$$

holds for some constant $C>0$. Taking a union bound over all indexes $j\in\{1,\ldots,d\}$, (III.20) implies

$$\sup_{T\subset\{1,\ldots,m\}} \|A^\top\eta_T\|_\infty \lesssim \|\eta\|_1 + \|\eta\|_2\sqrt{\log d}$$

with probability at least $1-1/d^2$. Thus, we arrive at the conclusion. □

Proof of Theorem I.6: Set $\boldsymbol{h}^- := \widehat{\boldsymbol{x}} - \boldsymbol{x}_0$ and $\boldsymbol{h}^+ := \widehat{\boldsymbol{x}} + \boldsymbol{x}_0$. Without loss of generality, we assume that $\|\boldsymbol{h}^-\|_1 \leq \|\boldsymbol{h}^+\|_1$. Since $\widehat{\boldsymbol{x}}$ is the solution of (I.13), we have

$$\begin{aligned} \||A\widehat{\boldsymbol{x}}| - \boldsymbol{y}\|^2 + \lambda\|\widehat{\boldsymbol{x}}\|_1 &\leq \||A\boldsymbol{x}_0| - \boldsymbol{y}\|^2 + \lambda\|\boldsymbol{x}_0\|_1 \\ &= \|\eta\|_2^2 + \lambda\|\boldsymbol{x}_0\|_1. \end{aligned} \tag{III.21}$$

For any index set $T \subset \{1,\ldots,m\}$, we set $A_T := [\boldsymbol{a}_j : j\in T]^\top$ which is a submatrix of A. Set

$$\begin{aligned} T_1 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = 1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = 1\} \\ T_2 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = -1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = -1\} \\ T_3 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = 1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = -1\} \\ T_4 &:= \{j : \mathrm{sign}(\langle \boldsymbol{a}_j, \widehat{\boldsymbol{x}}\rangle) = -1,\ \mathrm{sign}(\langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle) = 1\}. \end{aligned}$$

Then a simple calculation leads to

$$\begin{aligned} \||A\widehat{\boldsymbol{x}}| - \boldsymbol{y}\|_2^2 &= \|A_{T_1}\boldsymbol{h}^- - \eta_{T_1}\|_2^2 + \|A_{T_2}\boldsymbol{h}^- + \eta_{T_2}\|_2^2 \\ &\quad + \|A_{T_3}\boldsymbol{h}^+ - \eta_{T_3}\|_2^2 + \|A_{T_4}\boldsymbol{h}^+ + \eta_{T_4}\|_2^2. \end{aligned} \tag{III.22}$$

Substituting (III.22) into (III.21), we obtain that

$$\begin{aligned} \|A_{T_{12}}\boldsymbol{h}^-\|_2^2 + \|A_{T_{34}}\boldsymbol{h}^+\|_2^2 &\leq 2\langle \boldsymbol{h}^-, A_{T_1}^\top\eta_{T_1} - A_{T_2}^\top\eta_{T_2}\rangle \\ &\quad + 2\langle \boldsymbol{h}^+, A_{T_3}^\top\eta_{T_3} - A_{T_4}^\top\eta_{T_4}\rangle \\ &\quad + \lambda(\|\boldsymbol{x}_0\|_1 - \|\boldsymbol{h}^+ - \boldsymbol{x}_0\|_1), \end{aligned} \tag{III.23}$$

where $T_{12} := T_1\cup T_2$ and $T_{34} := T_3\cup T_4$. We claim that $\|\boldsymbol{h}^-\|_1 \leq 4\sqrt{s}\|\boldsymbol{h}^-\|_2$ and $\|\boldsymbol{h}^+\|_1 \leq 4\sqrt{s}\|\boldsymbol{h}^+\|_2$ hold with high probability. Indeed, let $S := \mathrm{supp}(\boldsymbol{x}_0) \subset \{1,\ldots,d\}$. Then

$$\begin{aligned} \|\boldsymbol{h}^+ - \boldsymbol{x}_0\|_1 &= \|\boldsymbol{h}_S^+ - \boldsymbol{x}_0\|_1 + \|\boldsymbol{h}_{S^c}^+\|_1 \\ &\geq \|\boldsymbol{x}_0\|_1 - \|\boldsymbol{h}_S^+\|_1 + \|\boldsymbol{h}_{S^c}^+\|_1, \end{aligned} \tag{III.24}$$

where the inequality follows from triangle inequality. According to Lemma III.5, we obtain that

$$\langle \boldsymbol{h}^-, A_{T_1}^\top\eta_{T_1} - A_{T_2}^\top\eta_{T_2}\rangle \leq \frac{\lambda}{8}\|\boldsymbol{h}^-\|_1 \tag{III.25}$$

and

$$\langle \boldsymbol{h}^+, A_{T_3}^\top\eta_{T_3} - A_{T_4}^\top\eta_{T_4}\rangle \leq \frac{\lambda}{8}\|\boldsymbol{h}^+\|_1 \tag{III.26}$$

hold with probability at least $1-1/d^2$, where $\lambda \gtrsim \|\eta\|_1 + \|\eta\|_2\sqrt{\log d}$. Putting (III.24), (III.25) and (III.26) into (III.23) and using the fact $\|\boldsymbol{h}^-\|_1 \leq \|\boldsymbol{h}^+\|_1$, we can obtain that

$$\|A_{T_{12}}\boldsymbol{h}^-\|_2^2 + \|A_{T_{34}}\boldsymbol{h}^+\|_2^2 \leq \frac{\lambda}{2}\|\boldsymbol{h}^+\|_1 + \lambda(\|\boldsymbol{h}_S^+\|_1 - \|\boldsymbol{h}_{S^c}^+\|_1) \tag{III.27}$$

holds with probability at least $1-1/d^2$. The (III.27) implies that

$$\frac{\lambda}{2}\|\boldsymbol{h}^+\|_1 + \lambda(\|\boldsymbol{h}_S^+\|_1 - \|\boldsymbol{h}_{S^c}^+\|_1) \geq 0,$$

which gives $\|\boldsymbol{h}_{S^c}^+\|_1 \leq 3\|\boldsymbol{h}_S^+\|_1$ and hence $\|\boldsymbol{h}^+\|_1 \leq 4\|\boldsymbol{h}_S^+\|_1$. By the Hölder's inequality, we obtain that

$$\|\boldsymbol{h}^+\|_1 \leq 4\sqrt{s}\|\boldsymbol{h}^+\|_2.$$

On the other hand, note that

$$\|\boldsymbol{h}_S^+\|_1 = \|\widehat{\boldsymbol{x}}_S + \boldsymbol{x}_0\|_1, \quad \|\boldsymbol{h}_S^-\|_1 = \|\widehat{\boldsymbol{x}}_S - \boldsymbol{x}_0\|_1$$

and

$$\|\boldsymbol{h}_{S^c}^+\|_1 = \|\boldsymbol{h}_{S^c}^-\|_1.$$

Combining with $\|\boldsymbol{h}^-\|_1 \le \|\boldsymbol{h}^+\|_1$, we can obtain that $\|\boldsymbol{h}^-\|_1 \le 4\sqrt{s}\|\boldsymbol{h}^-\|_2$.

We next present an upper bound of $\|\boldsymbol{h}^-\|_2$. Without loss of generality, we assume that $\#T_{12} = \beta m \ge m/2$. The (III.22) implies that

$$\||A\widehat{\boldsymbol{x}}| - \boldsymbol{y}\|_2^2 \ge \|A_{T_1}\boldsymbol{h}^- - \eta_{T_1}\|_2^2 + \|A_{T_2}\boldsymbol{h}^- + \eta_{T_2}\|_2^2. \tag{III.28}$$

Substituting (III.28) into (III.21) we obtain that

$$\begin{aligned} \|A_{T_{12}}\boldsymbol{h}^-\|_2^2 &\le 2\langle \boldsymbol{h}^-, A_{T_1}^\top \eta_{T_1} - A_{T_2}^\top \eta_{T_2}\rangle \\ &\quad + \lambda(\|\boldsymbol{x}_0\|_1 - \|\boldsymbol{h}^- + \boldsymbol{x}_0\|_1) + \|\eta_{T_{12}^c}\|^2 \\ &\le 2\langle \boldsymbol{h}^-, A_{T_1}^\top \eta_{T_1} - A_{T_2}^\top \eta_{T_2}\rangle \\ &\quad + \lambda(\|\boldsymbol{h}_S^-\|_1 - \|\boldsymbol{h}_{S^c}^-\|_1) + \|\eta_{T_{12}^c}\|^2. \end{aligned} \tag{III.29}$$

Here, we use

$$\begin{aligned} \|\boldsymbol{h}^- + \boldsymbol{x}_0\|_1 &= \|\boldsymbol{h}_S^- + \boldsymbol{x}_0\|_1 + \|\boldsymbol{h}_{S^c}^-\|_1 \\ &\ge \|\boldsymbol{x}_0\|_1 - \|\boldsymbol{h}_S^-\|_1 + \|\boldsymbol{h}_{S^c}^-\|_1. \end{aligned}$$

We consider the left side of (III.29). Recall that $\|\boldsymbol{h}^-\|_1 \le 4\sqrt{s}\|\boldsymbol{h}^-\|_2$. Then

$$\|A_{T_{12}}\boldsymbol{h}^-\|_2^2 \ge c_0 m\|\boldsymbol{h}^-\|_2^2 \tag{III.30}$$

with probability at least $1 - \exp(-cm)$, provided $m \gtrsim s\log(ed/s)$ (see Remark II.4). For the right hand of (III.29), we use (III.25) and (III.26) to obtain that

$$\begin{aligned} \|A_{T_{12}}\boldsymbol{h}^-\|_2^2 &\le \frac{\lambda}{4}\|\boldsymbol{h}^-\|_1 + \lambda(\|\boldsymbol{h}_S^-\|_1 - \|\boldsymbol{h}_{S^c}^-\|_1) + \|\eta_{T_{12}^c}\|^2 \\ &\le \frac{5\lambda}{4}\|\boldsymbol{h}_S^-\|_1 + \|\eta_{T_{12}^c}\|^2 \\ &\le \frac{5\lambda\sqrt{s}}{4}\|\boldsymbol{h}^-\|_2 + \|\eta_{T_{12}^c}\|^2 \end{aligned} \tag{III.31}$$

holds with probability at least $1 - 1/d^2$. Combining (III.30) and (III.31), we have

$$c_0 m\|\boldsymbol{h}^-\|_2^2 \le \frac{5\lambda\sqrt{s}}{4}\|\boldsymbol{h}^-\|_2 + \|\eta_{T_{12}^c}\|^2$$

with probability at least $1 - \exp(-cm) - 1/d^2$. By solving the above inequality, we arrive at the conclusion

$$\|\boldsymbol{h}^-\|_2 \lesssim \frac{\lambda\sqrt{s}}{m} + \frac{\|\eta\|_2}{\sqrt{m}}.$$

□

IV. Discussion

We have analyzed the estimation performance of the nonlinear least squares for phase retrieval. We show that the reconstruction error of the nonlinear least square model is $O(\|\eta\|_2/\sqrt{m})$ and we also prove that this recovery bound is optimal in the power of m. For sparse phase retrieval, we obtain similar results for the nonlinear Lasso. It is of interest to extend the results in this paper to complex signals. Moreover, assume that $y_i = f(|\langle \boldsymbol{a}_i, \boldsymbol{x}_0\rangle|) + \eta_i, i = 1, \dots, m$, where $f: \mathbb{R} \to \mathbb{R}$ is a continuous function. It is interesting to consider the recovery error of the model $\min_{\boldsymbol{x}} \||A\boldsymbol{x}| - \boldsymbol{y}\|_2$ under this setting, which is the subject of our future work.

References

[1] T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow," *Ann. Statist.*, vol. 44, no. 5, pp. 2221–2251, Oct. 2016.

[2] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.

[3] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, Aug. 2013.

[4] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Commun. Pure Appl. Math.*, vol. 70, no. 5, pp. 822–883, May 2017.

[5] J. C. Dainty and J. R. Fienup, "Phase retrieval and image reconstruction for astronomy," *Image Recovery, Theory Appl.*, vol. 231, p. 275, 1987.

[6] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, p. 2758, Aug. 1982.

[7] B. Gao and Z. Xu, "Phaseless recovery using the Gauss–Newton method," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5885–5896, Nov. 2017.

[8] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, May 1972.

[9] T. Goldstein and C. Studer, "PhaseMax: Convex phase retrieval via basis pursuit," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2675–2689, Apr. 2018.

[10] R. W. Harrison, "Phase problem in crystallography," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 10, no. 5, pp. 1046–1055, May 1993.

[11] T. Isernia, F. Soldovieri, G. Leone, and R. Pierri, "On the local minima in phase reconstruction algorithms," *Radio Sci.*, vol. 31, no. 6, pp. 1887–1899, Nov. 1996.

[12] T. Isernia, G. Leone, and R. Pierri, "A quadratic inverse problem: The phase retrieval," in *Inverse Methods Action*. Berlin, Germany: Springer, 1990, pp. 285–291.

[13] M. Iwen, A. Viswanathan, and Y. Wang, "Robust sparse phase retrieval made easy," *Appl. Comput. Harmon. Anal.*, vol. 42, no. 1, pp. 135–142, Jan. 2017.

[14] H. Jeong and C. S. Güntürk, "Convergence of the randomized Kaczmarz method for phase retrieval," 2017, *arXiv:1706.10291*. [Online]. Available: http://arxiv.org/abs/1706.10291

[15] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, "Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes," *Annu. Rev. Phys. Chem.*, vol. 59, no. 1, pp. 387–410, May 2008.

[16] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 7, no. 3, p. 394, Mar. 1990.

[17] R. Moretta and R. Pierri, "Performance of phase retrieval via phaselift and quadratic inversion in circular scanning case," *IEEE Trans. Antennas Propag.*, vol. 67, no. 12, pp. 7528–7537, Dec. 2019.

[18] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4814–4826, Sep. 2015.

[19] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1275–1297, Aug. 2013.

[20] Y. Plan and R. Vershynin, "The generalized Lasso with non-linear observations," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1528–1537, Mar. 2016.

[21] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.

[22] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," *Found. Comput. Math.*, vol. 18, no. 5, pp. 1131–1198, Oct. 2018.

[23] Y. S. Tan and R. Vershynin, “Phase retrieval via randomized Kaczmarz: Theoretical guarantees,” *Inf. Inference, A, J. IMA*, vol. 8, no. 1, pp. 97–123, Mar. 2019.
[24] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*. Cambridge, U.K.: Cambridge Univ. Press, 2018.
[25] V. Voroninski and Z. Xu, “A strong restricted isometry property, with an application to phaseless compressed sensing,” *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 386–395, Mar. 2016.
[26] I. Waldspurger, “Phase retrieval with random Gaussian sensing vectors by alternating projections,” *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3301–3312, May 2018.
[27] A. Walther, “The question of phase retrieval in optics,” *Optica Acta, Int. J. Opt.*, vol. 10, no. 1, pp. 41–49, Jan. 1963.
[28] G. Wang, G. B. Giannakis, and Y. C. Eldar, “Solving systems of random quadratic equations via truncated amplitude flow,” *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 773–794, Feb. 2018.
[29] G. Wang, L. Zhang, G. B. Giannakis, M. Akcakaya, and J. Chen, “Sparse phase retrieval via truncated amplitude flow,” *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 479–491, Jan. 2018.
[30] Y. Wang and Z. Xu, “Phase retrieval for sparse signals,” *Appl. Comput. Harmon. Anal.*, vol. 37, no. 3, pp. 531–544, Nov. 2014.
[31] K. Wei, “Solving systems of phaseless equations via Kaczmarz methods: A proof of concept study,” *Inverse Problems*, vol. 31, no. 12, Dec. 2015, Art. no. 125008.
[32] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, “A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms,” *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 5164–5198, 2017.

Meng Huang received the M.S. and Ph.D. degrees from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2017 and 2019, respectively. He is currently a Post-Doctoral Research Associate with the Department of Mathematics, Hong Kong University of Science and Technology. His research interests include phase retrieval, compressed sensing, and low-rank matrix recovery.

Zhiqiang Xu received the Ph.D. degree from the Department of Applied Mathematics, Dalian University of Technology, Dalian, China, in 2003. He is currently a Professor with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. His research interests include approximation theory, compressed sensing, and phase retrieval.