

# Robust Phase Retrieval by Alternating Minimization

Seonho Kim, *Student Member, IEEE*, and Kiryung Lee, *Senior Member, IEEE*

**Abstract**—We consider a least absolute deviation (LAD) approach to the robust phase retrieval problem that aims to recover a signal from its absolute measurements corrupted with sparse noise. To solve the resulting non-convex optimization problem, we propose a robust alternating minimization (Robust-AM) derived as an unconstrained Gauss-Newton method. To solve the inner optimization arising in each step of Robust-AM, we adopt two computationally efficient methods for linear programs. We provide a non-asymptotic convergence analysis of these practical algorithms for Robust-AM under the standard Gaussian measurement assumption. These algorithms, when suitably initialized, are guaranteed to converge linearly to the ground truth at an order-optimal sample complexity with high probability while the support of sparse noise is arbitrarily fixed and the sparsity level is no larger than  $1/4$ . Additionally, through comprehensive numerical experiments on synthetic and image datasets, we show that Robust-AM outperforms existing methods for robust phase retrieval offering comparable theoretical performance guarantees.

**Index Terms**—phase retrieval, outliers, least absolute deviation, linear program, convex optimization

## I. INTRODUCTION

Phase retrieval refers to the recovery of unknown signals  $\mathbf{x}_* \in \mathbb{R}^d$  (or  $\mathbb{C}^d$ ) from the magnitudes of its linear measurements, which are formulated as

$$b_i = |\langle \mathbf{a}_i, \mathbf{x}_* \rangle|, \quad i = 1, \dots, m, \quad (1)$$

where  $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^d$  (or  $\mathbb{C}^d$ ) and are known measurement vectors. Solving the set of nonlinear equations in (1) arises in numerous applications including X-ray crystallography, diffraction and array imaging, and optics (e.g. [2]–[5]). We consider the robust phase retrieval from the noisy amplitude measurements in (1) corrupted with sparse noise, i.e.

$$b_i = \begin{cases} \xi_i & \text{if } i \in I_{\text{out}} \\ |\langle \mathbf{a}_i, \mathbf{x}_* \rangle| & \text{if } i \in I_{\text{in}} \end{cases} \quad (2)$$

where  $I_{\text{out}} \subset [m]$  and  $I_{\text{in}} = [m] \setminus I_{\text{out}}$  collect the unknown indices of outliers and inliers respectively, and  $\{\xi_i\}_{i \in I_{\text{out}}}$  is an arbitrary sequence in  $\mathbb{R}$ . For example, such a scenario arises in phase retrieval imaging applications [6] due to various reasons including detection failures and recording errors.

A suite of methods designed for the plain phase retrieval [7] has been adapted to address the outliers. These methods provide not only empirically successful performances but also theoretical analyses under random measurement models. For instance, anchored regression [8] and PhaseMax [9] formulate

phase retrieval given an initial estimate as a linear program. RobustPhaseMax [10] modifies these methods to offer robust estimation by introducing an auxiliary variable to describe the outliers. In another example, Reshaped Wirtinger Flow (RWF) [11] and Amplitude Flow [12] follow a subgradient descent approach for a least squares estimator (LSE). Median-RWF [13] is a variant of these methods tailored to robust phase retrieval. Specifically, Median-RWF uses a truncation type of regularization that identifies and excludes outliers in each iteration by median-based thresholding on the consistency of the current estimate to the measurements. Median-RWF significantly improves the empirical performance of RobustPhaseMax by tolerating a higher fraction of outliers. However, the regularization of Median-RWF involves algorithmic parameters that have been tuned specifically for the Gaussian measurement model. However, it has not been discussed how to generalize the tuning parameters to other measurement models.

A recent work proposed an approach to robust phase retrieval in the classical robust regression framework in statistics [14]. Instead of the least squares, they adopted the *least absolute deviation* (LAD) [15] to enforce the consistency to the squared amplitude measurements with outliers. The parameter estimation is then cast as a nonconvex optimization problem. They proposed a prox-linear method that updates the estimate iteratively through local linearization of the forward model. This algorithm can be viewed as a variant of the Gauss-Newton method that regularizes the updates with the proximity to the previous iterate. The prox-linear algorithm iteratively refines the estimate through a sequence of quadratic programs for prox-linear problems and provides comparable performance to Median-RWF. Importantly, the Gauss-Newton method does not involve any tuning parameter. However, for large-scale applications such as those in astronomical or medical imaging, further acceleration of this iterative method is desired. They developed the proximal operator graph splitting (POGS) solver for this purpose.

In this paper, we propose an optimization approach to robust phase retrieval that shares strong theoretical guarantees (high tolerance of outlier ratio and no tuning parameters) with the prox-linear algorithm and further improves its computational cost. The objective is achieved by a simple unconstrained Gauss-Newton method for LAD. The resulting optimization is equivalent to an alternating minimization algorithm for LAD, as described in [16], which is solved by a sequence of linear programs. Since this alternating minimization approach is robust in the presence of outliers, we refer to the optimization as Robust-AM. Since this alternating minimization is a robust estimator in the presence of outliers, we refer to the optimization as Robust-AM. Our main theoretical result demonstrates that a suitably initialized Robust-AM converges

Seonho Kim and Kiryung Lee are with the Department of ECE, The Ohio State University, Columbus, OH 43210 USA (e-mail: kim.7604@osu.edu). This work was supported in part by NSF CAREER Award CCF-1943201. A preliminary version of this work will be presented at the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [1]

to the ground-truth signal linearly from  $m = \mathcal{O}(d)$  random amplitude-only measurements including up to 25% outliers. The desired initialization can be obtained by the existing robust spectral estimators [13], [14]. We verified through comprehensive numerical simulations that Robust-AM empirically outperforms the existing methods for robust phase retrieval. Particularly, it can tolerate a higher fraction of outliers and provide exact recovery with fewer observations. Furthermore, due to its unconstrained optimization formulation with the absolute amplitude measurement model, Robust-AM admits a computationally efficient ADMM algorithm, which runs faster than POGS for the prox-linear method. As shown in Figure 1, ADMM for Robust-AM converges faster than POGS for the prox-linear method. In this experiment, the fraction of outliers is set to  $\eta := |I_{\text{out}}|/m = 0.3$ , with outlier entries generated following zero and a Cauchy distribution with median 0 and mean-absolute-deviation 1. The convergence is measured by the metric  $\text{dist}(\mathbf{x}, \mathbf{x}_*) := \min_{\alpha \in \{\pm 1\}} \|\mathbf{x} - \alpha \mathbf{x}_*\|_2$  for  $\mathbf{x}, \mathbf{x}_* \in \mathbb{R}^d$ . Figure 1 shows that the unconstrained Gauss-Newton method, without any explicit control over the proximity to previous iterates, converges to the ground truth signal  $\mathbf{x}_*$  without overshooting.

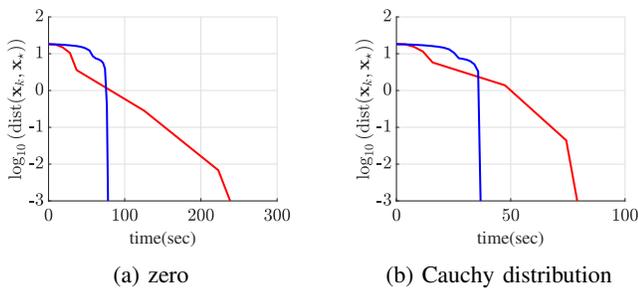


Fig. 1: Convergence of Robust-AM by ADMM [17] (blue) and prox-linear by POGS (red) in run time ( $d = 1,000$ ,  $m = 10,000$ , and  $\eta = 0.3$ ).

**Notations :** Boldface lowercase letters denote column vectors. We use  $\|\cdot\|_1$  and  $\|\cdot\|_2$  to denote the  $\ell_1$  norm and the Euclidean norm respectively. For brevity, the shorthand notation  $[n]$  denotes the set  $\{1, \dots, n\}$  for  $n \in \mathbb{N}$ . We adopt the big-O notation so that  $q \lesssim p$  is alternatively written as  $q = \mathcal{O}(p)$ . With a notation  $\mathcal{O}$ , we ignore logarithmic factors.

## II. ROBUST ALTERNATING MINIMIZATION

We consider the minimization of the composite function  $\ell = h \circ F$  where  $h : \mathbb{R}^m \rightarrow \mathbb{R}$  is a convex function and  $F : \mathbb{R}^d \rightarrow \mathbb{R}^m$  is a nonlinear mapping. In the special case when  $F$  is differentiable, Burke and Ferris [20] proposed a constrained Gauss-Newton method where the amount of the update is upper-bounded by a threshold. Duchi and Ruan [14] considered a variant where the constraint on the proximity on consecutive iterates is substituted by regularization with an additive penalty. We consider a more challenging case where  $F$  is non-differentiable and propose an unconstrained Gauss-Newton method where the variable sequence  $(\mathbf{x}_k)_{k \in \mathbb{N} \cup \{0\}}$  is iteratively updated by

$$\mathbf{x}_{k+1} \in \underset{\mathbf{x} \in \mathbb{R}^d}{\text{argmin}} h(F(\mathbf{x}_k) + F'(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)) \quad (3)$$

where  $F'(\mathbf{x}_k) \in \mathbb{R}^{m \times d}$  denotes the Clarke's generalized Jacobian matrix at  $\mathbf{x}_k$  [21]. Due to the local linear approximation of  $F$  at  $\mathbf{x}_k$  in (3),  $\mathbf{x}_{k+1}$  is obtained as a solution to a convex program. In a special case where  $h : \mathbb{R}^m \rightarrow \mathbb{R}$  and  $F : \mathbb{R}^d \rightarrow \mathbb{R}^m$  are respectively given by

$$h(\mathbf{z}) = \|\mathbf{z}\|_1 \quad (4)$$

and

$$F(\mathbf{x}) = (|\langle \mathbf{a}_i, \mathbf{x} \rangle| - b_i)_{i=1}^m, \quad (5)$$

their composition reduces to

$$\ell(\mathbf{x}) := \frac{1}{m} \sum_{i=1}^m ||\langle \mathbf{a}_i, \mathbf{x} \rangle| - b_i|. \quad (6)$$

Then the minimization of  $\ell$  corresponds to the LAD approach to robust phase retrieval with the absolute amplitude measurement model. Furthermore, given  $h$  and  $F$  as in (4) and (5), the update rule in (3) is explicitly written as

$$\mathbf{x}_{k+1} \in \underset{\mathbf{x} \in \mathbb{R}^d}{\text{argmin}} \sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{x} \rangle - \text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \cdot b_i|. \quad (7)$$

The resulting algorithm (7), derived from an unconstrained Gauss-Newton method of robust phase retrieval, is equivalent to an alternating minimization approach to the LAD formulation of robust phase retrieval when noisy measurements with a negative sign are discarded. An analogous alternating minimization for least-squares phase retrieval has been studied in the literature [16], [22]. Due to the robustness of LAD, we refer to the iterative algorithm by (7) as a robust alternating minimization (Robust-AM).

Duchi and Ruan [14] considered a similar robust phase retrieval with the squared amplitude measurement model via their regularized Gauss-Newton method.

## III. OPTIMIZATION ALGORITHMS

This section discusses numerical algorithms for Robust-AM. First, we note that the optimization in (7) is equivalent to a linear program

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^d, (t_i)_{i=1}^m}{\text{minimize}} \langle \mathbf{t}, \mathbf{1}_m \rangle \\ & \text{subject to } t_i \geq \langle \mathbf{a}_i, \mathbf{x} \rangle - \text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \cdot b_i, \\ & \quad t_i \geq -\langle \mathbf{a}_i, \mathbf{x} \rangle + \text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \cdot b_i, \quad \forall i \in [m] \end{aligned} \quad (8)$$

where  $\mathbf{1}_m = [1, \dots, 1]^T \in \mathbb{R}^m$ . There exist various computationally efficient numerical methods to solve linear programs. For example, the derandomized algorithm by van den Brand [19] finds an exact solution to a linear program with  $d$  variables and  $m$  constraints at the cost of  $\mathcal{O}((m+d)^c)$  multiplications where  $c \approx 2.38$ .

To further accelerate the convergence of Robust-AM, we also adopt iterative numerical algorithms that provide an approximate solution to the inner optimization in (7). In particular, we consider two alternating direction method of multipliers (ADMM) algorithms and a subgradient descent algorithm for the inner optimization. We refer to the Robust-AM with approximate solutions to the inner optimization by these ADMM algorithms as *fast Robust-AM* since they provide a significantly lower computational cost for the entire convergence of Robust-AM to an  $\epsilon$ -accurate estimate.

TABLE I: Comparison of RobustPhaseMax [10], Median-RWF [13], Prox-linear [14] and Robust-AM for robust phase retrieval in terms of computational cost to obtain  $\epsilon$ -accurate solution and sparse noise assumptions for the performance guarantees.

Method	Computational cost	Algorithm type	Support model	Tolerable sparsity level
RobustPhaseMax	$\mathcal{O}(m^3 + (m+d)^2 \log(1/\epsilon))$ $\tilde{\mathcal{O}}((m+d)^{2.38} \log(1/\epsilon))$	ADMM for LP [18] Deterministic LP [19]	adversarial	unspecified
Median-RWF	$\mathcal{O}(md \log(1/\epsilon))$	truncated gradient descent	arbitrary fixed	unspecified
Prox-linear	$\mathcal{O}(\log \log(1/\epsilon)(md^2 + md \log(1/\epsilon)))^1$	regularized Gauss-Newton (POGS)	arbitrary fixed	1/4
Robust-AM (Theorem IV.1)	$\mathcal{O}(m^3 + (m+d)^2 \log^2(1/\epsilon))$ $\tilde{\mathcal{O}}((m+d)^{2.38} \log^2(1/\epsilon))$	unconstrained Gauss-Newton via [18] unconstrained Gauss-Newton via [19]	arbitrary fixed	1/4

<sup>1</sup>We establish this computational cost under the assumption that the POGS linear converges to the solution for the inner optimization of prox-linear. However, to the best of our knowledge, the convergence rate of POGS has not been shown. Thus, this computational cost is a conjecture.

### A. ADMM for LAD

Given  $\mathbf{x}_k$ , the optimization in (7) is viewed as LAD for linear regression and one can use an ADMM algorithm for LAD [17, Chapter 6.1]. To describe the update rule of the ADMM algorithm, we introduce shorthand notations for the sake of brevity. Let  $\mathbf{A} \in \mathbb{R}^{m \times d}$  be a matrix whose  $i$ -th row is  $\mathbf{a}_i^\top$  for  $i \in [m]$ ,  $\mathbf{b} := (b_1, \dots, b_m) \in \mathbb{R}^m$ , and  $\mathbf{\Lambda}_k = \text{diag}(\text{sign}(\langle \mathbf{a}_1, \mathbf{x}_k \rangle), \dots, \text{sign}(\langle \mathbf{a}_m, \mathbf{x}_k \rangle))$ . By following [17, Chapter 6.1] with an auxiliary variable  $\mathbf{y}^t \in \mathbb{R}^d$  and dual variable  $\phi^t \in \mathbb{R}^m$ , the update rules are given in a closed form as follows:

$$\mathbf{x}^{t+1} = \mathbf{A}^+ \left( \mathbf{y}^t - \frac{1}{\rho} \phi^t \right), \quad (9a)$$

$$\begin{aligned} \mathbf{y}^{t+1} &= \mathbf{\Lambda}_k \mathbf{b} \\ &+ \text{sign} \left( \mathbf{A} \mathbf{x} + \frac{1}{\rho} \phi - \mathbf{\Lambda}_k \mathbf{b} \right) \odot \left[ \left[ \mathbf{A} \mathbf{x} + \frac{1}{\rho} \phi - \mathbf{\Lambda}_k \mathbf{b} \right] - \frac{1}{\rho} \right]_+, \end{aligned} \quad (9b)$$

$$\phi^{t+1} = \phi^t + \rho(\mathbf{A} \mathbf{x}^{t+1} - \mathbf{y}^{t+1}), \quad (9c)$$

where  $\odot$  denotes the Hadamard product. The most expensive step in (9) is the least squares problem in (9a). Since it repeats with the same  $\mathbf{A}$ , the pseudo inverse  $\mathbf{A}^+$  of  $\mathbf{A}$  can be pre-computed as  $\mathbf{A}^+ = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$  with cost  $\mathcal{O}(d^3 + d^2 m)$  and be used on memory over iterations. For faster convergence, we adopt the varying step size strategy for  $\rho$  [17, Section 3.4.1]. Importantly,  $\mathbf{A}$  remains the same over the outer iteration of Robust-AM, the pseudo inverse is computed only once. The POGS algorithm [23] for the prox-linear [14, Section 5] involves a similar matrix inversion. However, since their matrix evolves over the outer iteration, unlike the fast Robust-AM with ADMM, it is necessary for POGS to repeat the matrix inversion. Recall that we wanted to adopt ADMM for the inner iteration of Robust-AM to accelerate the convergence with approximate solutions. Therefore, the convergence rate in the inner optimization is crucial. However, to the best of our knowledge, the convergence rate has not been shown for the above ADMM algorithm and the POGS algorithm. Below we will present another ADMM algorithm and a subgradient descent method for (7) with proven linear convergence in the next section. Despite their theoretical convergence results, the ADMM by (9) empirically outperformed the other methods. In our numerical studies, we found that the fast Robust-AM with ADMM by (9) provides faster empirical convergence than POGS (see Figure 1).

### B. ADMM for linear program with linear convergence

Wang and Shroff [18] proposed the ADMM approach for a linear program and showed that their ADMM approach solves a linear program significantly faster than standard software such as CPLEX [24] and Gurobi [25]. Moreover, they showed the linear convergence result for their ADMM approach. To apply their approach to our linear program (8), we reformulate it into the standard form of a linear program (only with equality constraints) [18, Equation 1] by introducing  $2m$  auxiliary variables  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$  as

$$\begin{aligned} &\text{minimize} && \langle \mathbf{c}, \mathbf{w} \rangle \\ &\text{subject to} && \mathbf{B} \mathbf{w} = \mathbf{p}_k, \quad \mathbf{u}, \mathbf{s} \geq \mathbf{0}_m, \end{aligned} \quad (10)$$

where  $\mathbf{0}_m := [0, \dots, 0]^\top \in \mathbb{R}^m$ ,  $\mathbf{0}_{m,d} := [\mathbf{0}_m, \dots, \mathbf{0}_m] \in \mathbb{R}^{m \times d}$ , and

$$\mathbf{c} := [\mathbf{0}_d; \mathbf{1}_m; \mathbf{0}_m; \mathbf{0}_m] \in \mathbb{R}^{d+3m}$$

$$\mathbf{w} := [\mathbf{x}; \mathbf{t}; \mathbf{u}; \mathbf{s}] \in \mathbb{R}^{d+3m}$$

$$\mathbf{p}_k := [\mathbf{\Lambda}_k \mathbf{b}; \mathbf{\Lambda}_k \mathbf{b}] \in \mathbb{R}^{2m}$$

$$\mathbf{B} := \begin{bmatrix} \mathbf{A} & -\mathbf{I}_m & \mathbf{0}_{m,m} & \mathbf{I}_m \\ \mathbf{A} & \mathbf{I}_m & -\mathbf{I}_m & \mathbf{0}_{m,m} \end{bmatrix} \in \mathbb{R}^{2m \times (d+3m)}.$$

Then, by following [18, Algorithm 1], the update rule is given as a closed form with auxiliary variable  $\mathbf{y}^t = [\mathbf{y}_1^t; \mathbf{y}_2^t] \in \mathbb{R}^{d+3m}$  and dual variable  $\mathbf{z}^t = [\mathbf{z}_1^t; \mathbf{z}_2^t] \in \mathbb{R}^{d+5m}$  for  $\mathbf{y}_1 \in \mathbb{R}^{d+m}$ ,  $\mathbf{y}_2, \mathbf{z}_1 \in \mathbb{R}^{2m}$ , and  $\mathbf{z}_2 \in \mathbb{R}^{d+3m}$  as

$$\mathbf{w}^{t+1} = \frac{1}{\rho} (\mathbf{I} + \mathbf{B}^\top \mathbf{B})^{-1} (\mathbf{B}_1^\top (\mathbf{z}^t + \rho(\mathbf{B}_2 \mathbf{y}^t - \bar{\mathbf{p}}_k)) + \mathbf{c}), \quad (11a)$$

$$\mathbf{y}^{t+1} = \mathbf{w}^{t+1} + \frac{\mathbf{z}_y^t}{\rho}, \quad \mathbf{y}_2^{t+1} = [\mathbf{y}_2^{t+1}]_+, \quad (11b)$$

$$\mathbf{z}_1^{t+1} = \mathbf{z}_1^t + \rho(\mathbf{B} \mathbf{x}^{t+1} - \mathbf{p}), \quad \mathbf{z}_2^{t+1} = \mathbf{z}_2^t + \rho(\mathbf{w}^{t+1} - \mathbf{y}^{t+1}), \quad (11c)$$

where

$$\mathbf{B}_1 := \begin{bmatrix} \mathbf{B} \\ \mathbf{I}_{d+3m} \end{bmatrix}, \quad \mathbf{B}_2 := \begin{bmatrix} \mathbf{0}_{d+2m, d+3m} \\ -\mathbf{I}_{d+3m} \end{bmatrix}, \quad \bar{\mathbf{p}}_k := \begin{bmatrix} \mathbf{p}_k \\ \mathbf{0}_{3m} \end{bmatrix},$$

and  $[\cdot]_+$  takes the positive part of each entry of the input vector. The most expensive step is the matrix inversion given in (11a). It is calculated via the matrix-inversion lemma as

$$(\mathbf{I}_{d+3m} + \mathbf{B}^\top \mathbf{B})^{-1} = \mathbf{I}_{d+3m} - \mathbf{B}^\top (\mathbf{I}_{2m} + \mathbf{B} \mathbf{B}^\top)^{-1} \mathbf{B}$$

with cost  $\mathcal{O}(m^3)$ . Since this step does not depend on previous outer iterations, one can use a pre-computed result on memory

over the inner and outer iterations. Hence, by the linear convergence result [18, Theorem 1], the cost for an  $\epsilon_k$ -accurate solution to (10) is  $\mathcal{O}(m^3 + (m+d)^2 \log(1/\epsilon_k))$ . However, due to more auxiliary variables in (10) compared to (7), in our numerical studies, the ADMM algorithm by (11) showed slower convergence in the run time relative to the algorithm by (9).

### C. Subgradient descent for LAD

Yang and Lin [26] proposed a restarted subgradient (RSG) for non-smooth optimization. The specification of their subgradient descent to LAD in (7) is written as

$$\mathbf{x}^{t+1} = \mathbf{x}^t - \frac{\eta_t}{m} \sum_{i=1}^m \text{sign}(\langle \mathbf{a}_i, \mathbf{x}^t \rangle - \text{sign}(\langle \mathbf{a}_i, \mathbf{x}_i \rangle) \cdot b_i) \cdot \mathbf{a}_i, \quad (12)$$

where  $\eta_t > 0$  denotes a step size. The step size remains the same for  $T$  consecutive iterations and then decreases by half. They showed that the subsequence of iterates sampled at every  $T$  indices converges at a linear rate for a sufficiently large  $T$ . Therefore, the cost for an  $\epsilon$ -accurate solution to (7) is  $\mathcal{O}(mdT \log(1/\epsilon))$ . However, in our numerical studies, RSG did not provide the fastest convergence in the run time compared with the other ADMM algorithms.

## IV. THEORETICAL RESULTS

In this section, we present the convergence analysis of the Robust-AM algorithms under the following assumptions. First, we adopt the standard random linear measurements and outliers with arbitrary support and adversarial values [14].

**Assumption 1:** The measurement vectors  $(\mathbf{a}_i)_{i=1}^m$  are independent copies of  $\mathbf{a} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_d)$ .

**Assumption 2:** The outliers are supported on an arbitrarily fixed set  $I_{\text{out}}$  with  $|I_{\text{out}}| = \eta m$  for  $\eta \in [0, 1/4]$  and their magnitudes  $|\xi_i|$  can be adversarial.

Additionally, to provide the convergence analysis of the fast Robust-AM, we introduce an extra assumption that quantifies the suboptimality of solving (13) by ADMM.

**Assumption 3:** There exists a bounded sequence  $(\epsilon_k)_{k \in \mathbb{N}}$  such that  $\mathbf{x}_k$  is an inexact minimizer up to the sub-optimality level  $\epsilon_k$  for all  $k \in \mathbb{N}$ , i.e.

$$\begin{aligned} & \sum_{i=1}^m |\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \mathbf{x}_{k+1} \rangle - b_i| \\ & \leq \epsilon_k + \min_{\mathbf{x} \in \mathbb{R}^d} \sum_{i=1}^m |\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \mathbf{x} \rangle - b_i|. \end{aligned} \quad (13)$$

We denote the highest sub-optimality level as  $\epsilon_{\max}$ , i.e.

$$\epsilon_{\max} := \max_{k \in \mathbb{N}} \epsilon_k.$$

**Theorem IV.1.** *Suppose that Assumptions 1, 2, and 3 hold. Then there exist absolute constants  $C, c > 0$  and constants  $\nu_\eta \in (0, 1), \lambda_\eta > 0$  depending only on  $\eta$ , for which the following statement holds for all  $\mathbf{x}_* \in \mathbb{R}^d$  with probability at least  $1 - \exp(-cd)$ : If  $m \geq Cd$  and*

$$\max(\text{dist}(\mathbf{x}_0, \mathbf{x}_*), \lambda_\eta \epsilon_{\max}) \leq \sin(2/25) \|\mathbf{x}_*\|_2, \quad (14)$$

*then the sequence  $(\mathbf{x}_k)_{k \in \mathbb{N} \cup \{0\}}$  by the fast Robust-AM algorithm satisfies*

$$\text{dist}(\mathbf{x}_k, \mathbf{x}_*) \leq \nu_\eta^k \cdot \text{dist}(\mathbf{x}_0, \mathbf{x}_*) + \lambda_\eta \epsilon_{\max} \quad (15)$$

*for all  $k \in \mathbb{N}$ , where  $\text{dist}(\mathbf{x}, \mathbf{x}_*) := \min_{\alpha \in \{\pm 1\}} \|\mathbf{x} - \alpha \mathbf{x}_*\|_2$ .*

The Robust-AM algorithm updates iterates with an exact solution to (7). Therefore, setting  $\epsilon_{\max}$  to 0 in Theorem IV.1 provides a sufficient condition for the exact recovery of  $\mathbf{x}_*$  by Robust-AM. We compare the specification of Theorem IV.1 to this scenario to the analogous results for competing methods: RobustPhaseMax [10], Median-RWF [13], and prox-linear [14]. Theorem IV.1 as well as the previous results achieve the exact recovery when the number of observations  $m$  exceeds a multiple of the signal dimension  $d$ . Earlier theoretical results on RobustPhaseMax and Median-RWF showed that there exists an unspecified numerical constant so that the algorithms provide the exact recovery if the outlier fraction is below this constant. In contrast, the analyses of the prox-linear [14] and Robust-AM (Theorem IV.1) demonstrate that these methods can tolerate outliers up to 1/4 of the total observations. These theoretical guarantees consider different degrees of adversary for their outlier models. The performance guarantee of RobustPhaseMax by Hand [10] assumed the highest adversary so that both the support and values of sparse noise are adversarial. The performance guarantees of Median-RWF by Zhang et al. [13] considered the same outlier model as in Assumption 2, but they also introduced additive noise of a bounded norm in addition to sparse noise. Duchi and Ruan [14] used the lowest adversary so that the support of sparse noise is random but the nonzero values of sparse noise can depend on the measurements. Despite providing performance guarantees under the highest adversary, as shown in Section V, RobustPhaseMax showed significantly inferior empirical performance relative to the other methods in terms of the tolerable outlier ratio.

Theorem IV.1 establishes a local linear convergence of the Robust-AM algorithms. As discussed in Section II, Robust-AM has no explicit control over the amount of the update in each iteration unlike the constrained or regularized versions of the Gauss-Newton method [14], [20]. However, despite its simple form, Robust-AM provides the monotone decrease of the estimation error toward zero without any overshooting for robust phase retrieval in the setting of Theorem IV.1. All convergence analyses by Theorem IV.1 and previous work [13], [14] require an initialization within a neighborhood of the ground truth. The size of the basin of convergence was determined with an explicit numerical constant only in [10] and Theorem IV.1. Various initialization methods with theoretical performance guarantees have been developed to obtain the desired initial estimate [13], [14]. The sample complexity for these initialization methods does not exceed those for the subsequent estimators in order.

Next, we discuss the computational costs for the robust estimators. First, RobustPhaseMax is formulated as a linear program and thus it can be exactly solved with  $\mathcal{O}((m+d)^{2.38} \log(1/\epsilon))$  multiplications by derandomized algorithm [19]. Furthermore, as we discussed in Section III-B, there

exists an ADMM algorithm for the linear program that costs  $\mathcal{O}(m^3 + (m + d)^2 \log(1/\epsilon))$  for an  $\epsilon$ -accurate solution. Due to the term  $\log(1/\epsilon)$ , if the desired accuracy decreases in proportion to the size of the problem, it is preferable to use ADMM. Otherwise, the derandomized algorithm will be computationally efficient. The other estimators are given as an iterative algorithm with a proven convergence rate. Therefore, we compare their computational costs to obtain an  $\epsilon$ -accurate solution. Median-RWF is a truncated gradient descent with the per-iteration cost of  $\mathcal{O}(md)$ . Since the linear convergence of Median-RWF has been established, the total cost is  $\mathcal{O}(md \log(1/\epsilon))$ . Unlike Median-RWF, the updates in prox-linear and Robust-AM involve a nontrivial inner optimization, respectively cast as a quadratic program and a linear program. One may use an exact solver for these sub-problems. For example, there exists an interior point method for quadratic programs with the cost  $\mathcal{O}((m + d)^4)$  [27]. Since it has been shown that prox-linear converges quadratically, the total cost with this exact inner solver is  $\mathcal{O}((m + d)^4 \log \log(1/\epsilon))$ . The inner optimization in Robust-AM can be exactly solved at the cost  $\tilde{\mathcal{O}}((m + d)^{2.38} \log(1/\epsilon))$  by the derandomized algorithm [19]. Due to its linear convergence, the total cost of Robust-AM is  $\tilde{\mathcal{O}}((m + d)^{2.38} \log(1/\epsilon))$ . However, as shown in Theorem IV.1, the linear convergence of Robust-AM remains valid when the inner optimization problems are solved only approximately. The fast Robust-AM with the ADMM solver for linear programs has the per-iteration cost of  $\mathcal{O}(m^3 + (m + d)^2 \log(1/\epsilon_{\max}))$  as shown in Section III. Due to its linear convergence in Theorem IV.1, the total cost to obtain the  $\epsilon + \lambda_\eta \epsilon_{\max}$  accuracy is  $\mathcal{O}(m^3 + (m + d)^2 \log(1/\epsilon_{\max}) \log(1/\epsilon))$ . In contrast, the convergence rate of POGS for the inner optimization in prox-linear has not been established. We summarize the comparison for the computational costs of algorithms in Table I.

Lastly, we elaborate on the dependence of the parameters  $\nu_\eta$  and  $\lambda_\eta$  in Theorem IV.1 on the outlier ratio  $\eta$ . The linear convergence parameter  $\nu_\eta$  in (15) is explicitly specified as an increasing function of  $\eta$  in the proof of Theorem IV.1 and illustrated in Figure 2a. Therefore, smaller  $\eta$  implies faster convergence. The final error bound by (15) with  $k$  going to infinity is given as the amplification of the sub-optimality parameter  $\epsilon_{\max}$  in the inner optimization by a factor of  $\lambda_\eta$ . First, similar to  $\nu_\eta$ , the parameter is also explicitly given as an increasing function of  $\eta$  in the proof (see Figure 2b). However, the final estimation can still be sufficiently small, as one can set the accuracy parameter to a sufficiently low value (less than  $10^{-10}$ ) using linear program packages in readily available software such as CPLEX and Gurobi. Hence, the assumption on  $\{\epsilon_i\}_{i=1}^k$  in Theorem IV.1 is easily satisfied.

## V. NUMERICAL RESULTS

This section compares the empirical performances of Robust-AM to its theoretical analysis in Theorem IV.1. Robust-AM is also compared against the competing methods for robust phase retrieval, which are RobustPhaseMax, Median-RWF, and the prox-linear. Recall that all these methods require an initial estimate. For this purpose, we adopt the spectral method by Zhang et al. [13].

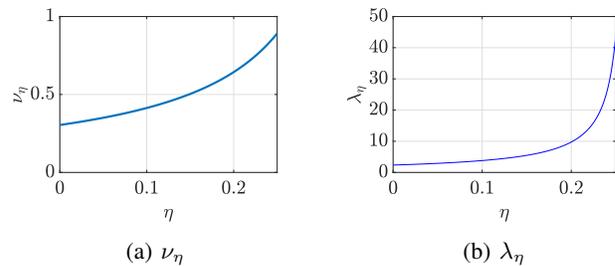


Fig. 2: The dependence of parameters  $\nu_\eta$  and  $\lambda_\eta$  in Theorem IV.1 on the outlier fraction  $\eta$ .

### A. Synthetic data experiments

First, through experiments on synthetic data, we show that the numerical results corroborate our theoretical findings in Theorem IV.1 and Robust-AM outperforms the competing methods. In this experiment, the measurement vectors are generated so that  $\{\mathbf{a}_i\}_{i=1}^m \stackrel{i.i.d.}{\sim} \text{Normal}(\mathbf{0}, \mathbf{I}_d)$  by following the assumptions in Theorem IV.1 and analogous theoretical analyses of the other methods. The ground-truth signal is generated as  $\mathbf{x}_* \sim \text{Normal}(\mathbf{0}, \mathbf{I}_d)$  independently from the measurement vectors. The outlier support is randomly selected following the uniform distribution on all possible subsets  $I_{\text{out}} \subset [m]$  of size  $\eta m$ .

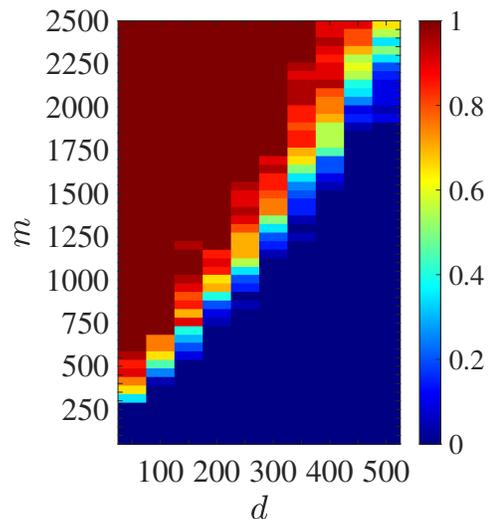
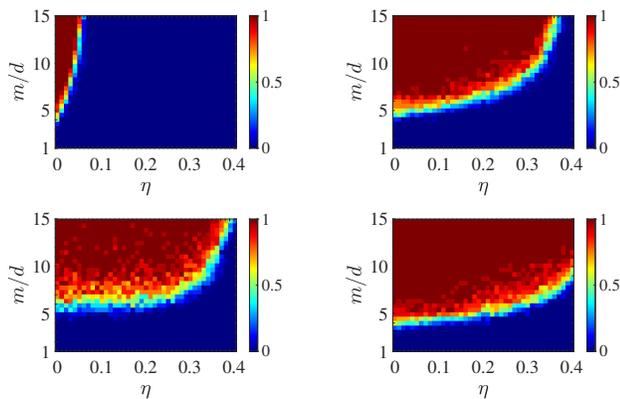
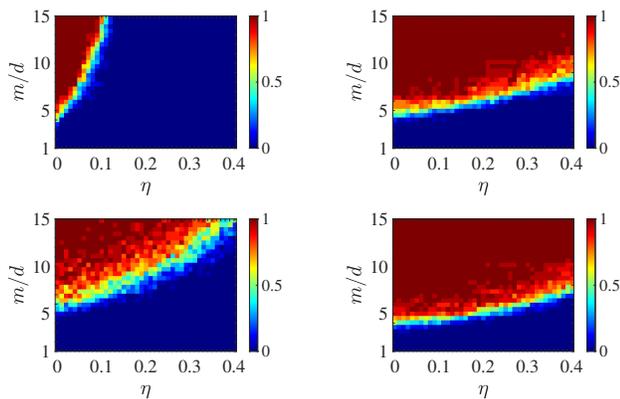


Fig. 3: Phase transition of empirical success rate by Robust-AM per the number of measurements  $m$  and the dimension  $d$ .

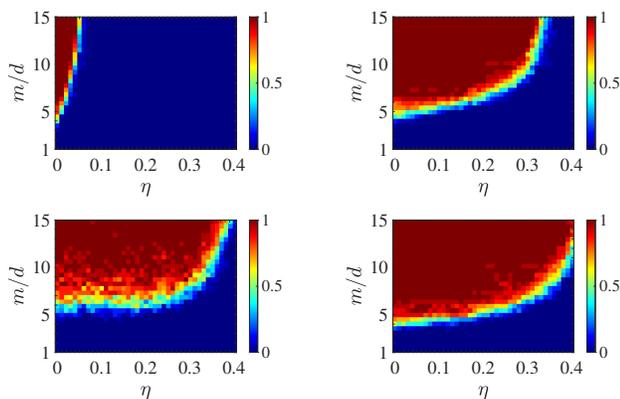
Figure 3 shows the phase transition of the empirical success rate by Robust-AM through Monte Carlo simulations, where the outlier values are i.i.d. following the Cauchy distribution with median 0 and mean-absolute-deviation 1. The fraction of outliers is fixed to  $\eta = 0.25$ . Recall that the performance guarantee in Theorem IV.1 applies uniformly to all ground-truth signals. To observe the empirical performance in an analogous setting, we design the experiment as follows: 1) Generate 20 sets of random measurement vectors  $\{\mathbf{a}_i\}_{i=1}^m$ . Generate 30 sets of random ground-truth  $\mathbf{x}_*$ ; 2) For each fixed  $\{\mathbf{a}_i\}_{i=1}^m$ , success is declared if the estimator recovers all 30 ground-truth signals by satisfying  $\text{dist}(\hat{\mathbf{x}}, \mathbf{x}_*) \leq 10^{-3}$ .



(a) Cauchy distribution



(b) uniform distribution



(c) zero

Fig. 4: Phase transition of success rate per the measurement ratio  $m/d$  and the fraction of outliers  $\eta$  for various outlier magnitude models. Subfigures are displayed according to RobustPhaseMax (top-left), Median-RWF (top-right), prox-linear method (bottom-left), and Robust-AM (bottom-right).

where  $\hat{\mathbf{x}}$  denotes the estimate; 3) The empirical success rate is calculated on the outcomes from 20 distinct sets of measurement vectors. The transition occurs at the boundary where the number of measurements is proportional to the ambient dimension (signal length). This empirical result corroborates our theoretical finding in Theorem IV.1. Next, we repeat the same experiment on RobustPhaseMax, Median-RWF, and the

prox-linear. Figure 4a compares the empirical performance of Robust-AM against RobustPhaseMax, Median-RWF, and the prox-linear by displaying the phase transition of these methods for a range of the outlier fraction  $\eta$  in this setting. The ambient dimension is set to  $d = 100$ . Figure 4a shows that Robust-AM outperforms all the other methods with a significantly lower threshold for the phase transition. We further expand the comparison to other models for outlier values. The second scenario draws  $\xi_i$  from the uniform distribution on  $(-d\|\mathbf{x}_*\|_2^2/2, d\|\mathbf{x}_*\|_2^2/2)$ . The third scenario sets  $\xi_i$  to 0. As observed in Figures 4b and 4c, similar trends appear in the other outlier models. RobustPhaseMax, while providing the strongest theoretical performance guarantee, shows the worst empirical performance in the comparison. There is no consistent dominance between Median-RWF and the prox-linear algorithm. Median-RWF outperforms the prox-linear in the second scenario, but the other way around in the other scenarios.

Next, we compare the convergence speed of Robust-AM and the prox-linear algorithm. In this experiment, the dimension parameters are set to  $m = 1,500$  and  $d = 200$  where the values of outliers are zero. The outlier ratio varies over  $\eta \in \{0.1, 0.2, 0.3\}$ . Figure 5 illustrates how the log of  $\text{dist}(\mathbf{x}_k, \mathbf{x}_*)$  decays over the iteration index  $k$ . The median over 10 trials is plotted. In their theoretical analyses, the prox-linear algorithm converges faster at a quadratic rate than the linear convergence of Robust-AM in Theorem IV.1. However, as shown in Figure 5, Robust-AM empirically converges faster than the prox-linear algorithm in the iteration count for all considered  $\eta$ . Moreover, Figure 5 illustrates that the number of iterations for Robust-AM increases as  $\eta$  increases. This implies that for each iteration, the convergence rate of Robust-AM is proportional to  $\eta$ . This supports our theoretical finding that the convergence parameter  $\nu_\eta$  in Theorem IV.1 is an increasing function of  $\eta$  as shown in Figure 2a.

### B. Real image experiments

We further apply Robust-AM to a set of image data to show that Robust-AM continues outperforming the other competing methods for non-Gaussian measurement models. We adopt the structured random measurement model in the experimental setting in [14, Section 6.3] given by

$$\mathbf{A}_H = (\mathbf{I}_k \otimes \mathbf{H}_n)[\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_k]^\top \in \mathbb{R}^{kn \times n}, \quad (16)$$

where  $\mathbf{H}_n \in \mathbb{R}^{n \times n}$  denotes the normalized Hadamard matrix and  $\mathbf{S}_1, \dots, \mathbf{S}_k \in \mathbb{R}^{n \times n}$  are diagonal matrices whose diagonal entries are independently drawn uniformly random from  $\{\pm 1\}$ . The measurement vector  $\mathbf{a}_i$  is the  $i$ -th column of  $\mathbf{A}_H^\top$  for  $i \in [m]$ , where  $m = kn$ . The linear measurement operator in (16) applies to the vectorized version of a 2D input image  $\mathbf{X}_* \in \mathbb{R}^{n_1 \times n_2}$  denoted by  $\mathbf{x}_* := \text{Vec}(\mathbf{X}_*) \in \mathbb{R}^n$  with  $n = n_1 \times n_2$ . The measurements corresponding to outliers are substituted by zero in the experiment.

Robust-AM and the competing algorithms are tested on the collection of 50 images of handwritten digits<sup>1</sup> Figure 7

<sup>1</sup><https://hastie.su.domains/ElemStatLearn/datasets/zip.digits>.

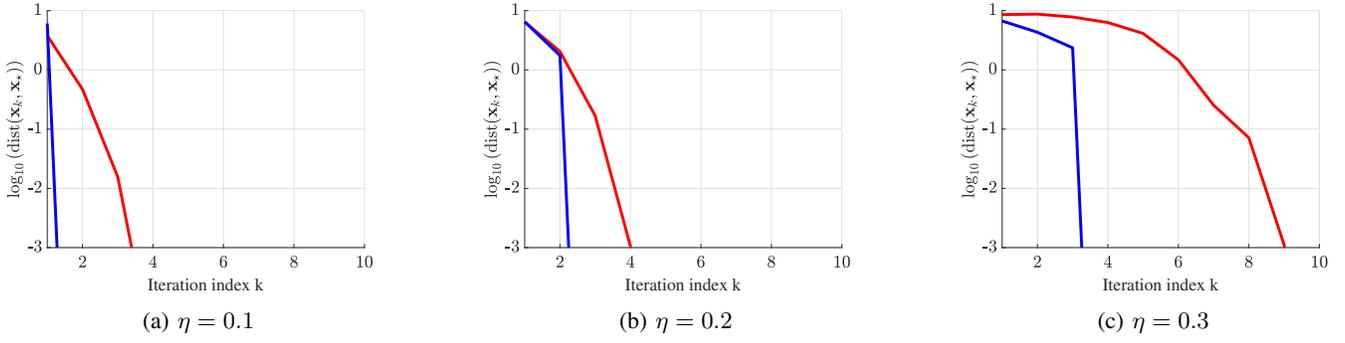


Fig. 5: Convergence of Robust-AM (blue) and the prox-linear (red) in the iteration count.

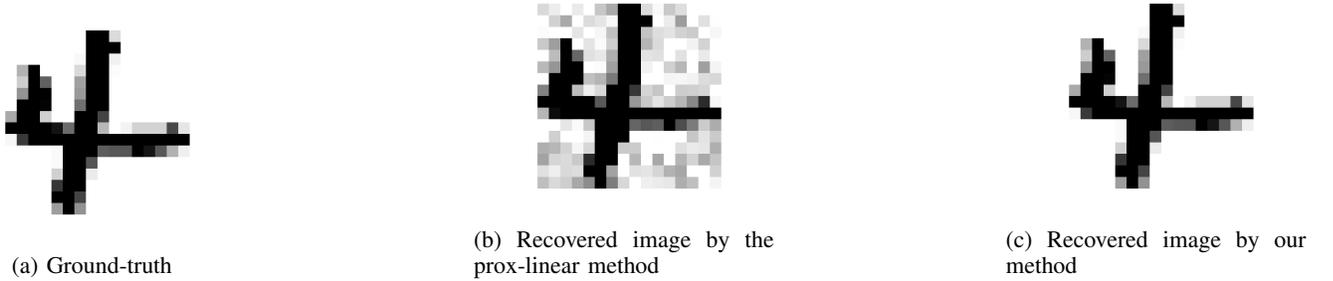


Fig. 6: Example of recovery for an image data.

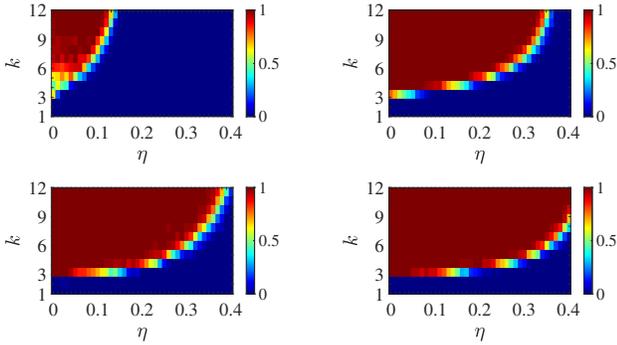


Fig. 7: Phase transition of success rate per  $k$  and the fraction of outliers  $\eta$  for zero outlier magnitude models. Subfigures are displayed according to RobustPhaseMax (top-left), Median-RWF (top-right), prox-linear method (bottom-left), and Robust-AM (bottom-right).

compares the two methods in the empirical success rate over 50 images, where the number of random modulations  $k$  and the outlier fraction  $\eta$  respectively vary over  $k \in \{1, \dots, 12\}$  and  $\eta \in [0, 0.4]$ . Similar to the previous experiments on synthetic data, Figure 7 demonstrates that Robust-AM outperforms the competing algorithms by providing recovery with smaller  $k$  for each observed  $\eta$ . Since the algorithmic parameters of Median-RWF are specifically selected for Gaussian measurements in [13], we heuristically tuned the step size to 0.2 so that Median-RWF performs for the measurement setting (16).

## VI. PROOF OF THEOREM IV.1

We first prove by the induction on the iteration index  $j$  that

$$\text{dist}(\mathbf{x}_j, \mathbf{x}_*) \leq \nu_\eta \cdot \text{dist}(\mathbf{x}_{j-1}, \mathbf{x}_*) + \frac{\epsilon_{j-1}}{C_\eta} \quad (17)$$

holds for all  $j \in \mathbb{N}$  for some numerical constant  $\nu_\eta \in (0, 1)$  and  $C_\eta > 0$  depending only on  $\eta$ . Let  $k \in \mathbb{N}$  be arbitrarily fixed. Suppose that  $\mathbf{x}_j$  satisfies (17) for all  $j \leq k$ . Note that the distance between  $\mathbf{x}$  and  $\mathbf{x}_*$  is written as

$$\text{dist}(\mathbf{x}, \mathbf{x}_*) = \|\mathbf{x} - \varphi(\mathbf{x})\mathbf{x}_*\|_2, \quad (18)$$

where

$$\varphi(\mathbf{x}) := \underset{\alpha \in \{\pm 1\}}{\text{argmin}} \|\mathbf{x} - \alpha \mathbf{x}_*\|_2.$$

Then we have  $\text{dist}(\mathbf{x}_{k+1}, \mathbf{x}_*) \leq \|\mathbf{x}_{k+1} - \varphi(\mathbf{x}_k)\mathbf{x}_*\|_2$  and  $\text{dist}(\mathbf{x}_k, \mathbf{x}_*) = \|\mathbf{x}_k - \varphi(\mathbf{x}_k)\mathbf{x}_*\|_2$ . Therefore, it follows that

$$\|\mathbf{x}_{k+1} - \varphi(\mathbf{x}_k)\mathbf{x}_*\|_2 \leq \nu_\eta \|\mathbf{x}_k - \varphi(\mathbf{x}_k)\mathbf{x}_*\|_2 + \frac{\epsilon_k}{C_\eta} \quad (19)$$

implies (17) for  $j = k + 1$ . This completes the induction argument.

Therefore, it suffices to show that the hypothesis of the theorem implies (19). For the sake of brevity, we denote the objective function of the optimization formulation in (7) by

$$f_{\mathbf{x}_k}(\mathbf{x}) = \frac{1}{m} \sum_{i=1}^m |\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \mathbf{x} \rangle - b_i|.$$

Then (13) provides

$$\underbrace{f_{\mathbf{x}_k}(\mathbf{x}_{k+1})}_{(A)} \leq \underbrace{f_{\mathbf{x}_k}(\varphi(\mathbf{x}_k)\mathbf{x}_*)}_{(B)} + \epsilon_k. \quad (20)$$

Next, we derive a lower bound (resp. an upper bound) on (A) (resp. (B)) of (20). By from the definition of  $b_i$  in (2), (A) is written as

$$\begin{aligned}
 (A) &= \frac{1}{m} \sum_{i=1}^m |\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \mathbf{x}_{k+1} \rangle - b_i| \\
 &= \frac{1}{m} \sum_{i \in I_{\text{in}}} \underbrace{|\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \mathbf{x}_{k+1} \rangle - \langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle|}_{(a)} \\
 &\quad + \frac{1}{m} \sum_{i \in I_{\text{out}}} |\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \mathbf{x}_{k+1} \rangle - \xi_i|. \tag{21}
 \end{aligned}$$

To simplify the partial summation over  $I_{\text{in}}$ , we introduce the spherical wedge [28] defined by

$$W_{\mathbf{x}, \mathbf{z}} := \{\mathbf{v} \in \mathbb{S}^{d-1} \mid \text{sign}(\langle \mathbf{v}, \mathbf{x} \rangle) \neq \text{sign}(\langle \mathbf{v}, \mathbf{z} \rangle)\}. \tag{22}$$

Then it follows that  $\langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle$  and  $\langle \mathbf{a}_i, \mathbf{x}_k \rangle$  have the opposite sign if and only if  $\mathbf{a}_i \in W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}$ . Therefore, the summand in (a) is rewritten as

$$\begin{aligned}
 (a) &= \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \in W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}\}} |\langle \mathbf{a}_i, \mathbf{x}_{k+1} + \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle| \\
 &\quad + \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \notin W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}\}} |\langle \mathbf{a}_i, \mathbf{x}_{k+1} - \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle|.
 \end{aligned}$$

The second summand on the right-hand side provides a valid lower bound on (a) since the other summand is nonnegative. Combining the above results, we obtain that

$$\begin{aligned}
 (A) &\geq \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \notin W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}\}} |\langle \mathbf{a}_i, \mathbf{x}_{k+1} - \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle| \\
 &\quad + \frac{1}{m} \sum_{i \in I_{\text{out}}} |\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \mathbf{x}_{k+1} \rangle - \xi_i|. \tag{23}
 \end{aligned}$$

Similarly, (B) is written as

$$\begin{aligned}
 (B) &= \frac{1}{m} \sum_{i \in I_{\text{in}}} \underbrace{|\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle - \langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle|}_{(b)} \\
 &\quad + \frac{1}{m} \sum_{i \in I_{\text{out}}} |\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle - \xi_i|.
 \end{aligned}$$

If  $\mathbf{a}_i \in W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}$ , then  $\langle \mathbf{a}_i, \mathbf{x}_k \rangle$  and  $\langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle$  have the opposite sign and hence (b) satisfies

$$(b) = 2 |\langle \mathbf{a}_i, \mathbf{x}_* \rangle| \leq 2 |\langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* - \mathbf{x}_k \rangle|.$$

Otherwise, if  $\mathbf{a}_i \notin W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}$ , then (b) = 0. Therefore, we have

$$\begin{aligned}
 (B) &\leq \frac{2}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \in W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}\}} |\langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* - \mathbf{x}_k \rangle| \\
 &\quad + \frac{1}{m} \sum_{i \in I_{\text{out}}} |\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle - \xi_i|. \tag{24}
 \end{aligned}$$

By plugging in the bounds of (23) and (24) into (20), we obtain that (20) implies

$$\begin{aligned}
 &\frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \notin W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}\}} |\langle \mathbf{a}_i, \mathbf{x}_{k+1} - \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle| \\
 &\quad + \frac{1}{m} \sum_{i \in I_{\text{out}}} \underbrace{|\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \mathbf{x}_{k+1} \rangle - \xi_i|}_{(*)} \\
 &\quad - \frac{1}{m} \sum_{i \in I_{\text{out}}} \underbrace{|\text{sign}(\langle \mathbf{a}_i, \mathbf{x}_k \rangle) \langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle - \xi_i|}_{(**)} \\
 &\leq \frac{2}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \in W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}\}} |\langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* - \mathbf{x}_k \rangle| + \epsilon_k. \tag{25}
 \end{aligned}$$

By applying the triangle inequality to the summands in (\*) and (\*\*), we obtain a necessary condition of (25) given by

$$\begin{aligned}
 &\frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \notin W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}\}} |\langle \mathbf{a}_i, \mathbf{x}_{k+1} - \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle| \\
 &\quad - \frac{1}{m} \sum_{i \in I_{\text{out}}} |\langle \mathbf{a}_i, \mathbf{x}_{k+1} - \varphi(\mathbf{x}_k) \mathbf{x}_* \rangle| \\
 &\leq \frac{2}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \in W_{\mathbf{x}_k, \varphi(\mathbf{x}_k) \mathbf{x}_*}\}} |\langle \mathbf{a}_i, \varphi(\mathbf{x}_k) \mathbf{x}_* - \mathbf{x}_k \rangle| + \epsilon_k. \tag{26}
 \end{aligned}$$

We have shown that (20) implies (26). In the remainder of the proof, we demonstrate that if (26) is satisfied, then (19) holds with high probability. This is achieved by applying a probabilistic lower bound on (c) and probabilistic upper bounds on (d) and (e), using concentration inequalities.

To this end, note that the measurement vectors  $\{\mathbf{a}_i\}_{i=1}^m$  depend not only on the current iterate  $\mathbf{x}_k$  and the next iterate  $\mathbf{x}_{k+1}$ , but also on the indication functions within the spherical wedge in (c) and (e). Therefore, we consider the uniform bounds for all iterates and the collection of spherical wedges with the largest angle less than  $\theta \in (0, \pi)$ . We introduce the corresponding lemmas below.

**Lemma VI.1.** *Let  $\theta \in (0, \pi)$ ,  $\eta \in (0, 1/2)$  and  $\delta > 0$ . Suppose that  $\{\mathbf{a}_i\}_{i=1}^m$  are independent copies of  $\mathbf{g} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_d)$ . Let*

$$W_\theta := \{W_{\mathbf{x}, \mathbf{z}} : \mathbf{x}, \mathbf{z} \in \mathbb{R}^d, \angle(\mathbf{x}, \mathbf{z}) \leq \theta\}, \tag{27}$$

where  $W_{\mathbf{x}, \mathbf{z}}$  is defined in (22). Then there exists an absolute

constant  $C$  such that

$$\inf_{\substack{W \in \mathcal{W}_\theta \\ \mathbf{z} \in \mathbb{S}^{d-1}}} \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \notin W\}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| \geq (1 - \eta) \sqrt{\frac{2}{\pi}} - \frac{2\theta}{\pi} \left( \sqrt{\frac{2}{\pi}} + \sqrt{2 \log \left( \frac{e\pi}{2\theta} \right)} \right) - \frac{\theta}{20} \left( \sqrt{\frac{2\theta}{\pi}} + 1 \right), \quad (28)$$

$$\sup_{\mathbf{z} \in \mathbb{S}^{d-1}} \frac{1}{m} \sum_{i \in I_{\text{out}}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| \leq \eta \sqrt{\frac{\pi}{2}} + \sqrt{\eta} \frac{\theta}{20}, \quad (29)$$

and

$$\sup_{\substack{W \in \mathcal{W}_\theta \\ \mathbf{z} \in \mathbb{S}^{d-1}}} \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \in W\}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| \leq \frac{2\theta}{\pi} \left( \sqrt{\frac{2}{\pi}} + \sqrt{2 \log \left( \frac{e\pi}{2\theta} \right)} \right) + \sqrt{\frac{2\theta}{\pi}} \cdot \frac{\theta}{20} \quad (30)$$

hold with probability at least  $1 - \delta$  provided that

$$m \geq C \cdot \theta^{-2} (d \log(m/d) \vee \log(1/\delta)). \quad (31)$$

*Proof:* See Section VIII.  $\blacksquare$

Now we derive the largest angle for the spherical wedge  $W_{\mathbf{x}_k, \varphi(\mathbf{x}_k)\mathbf{x}_*}$ . Since the angle between  $\mathbf{x}_k$  and  $\varphi(\mathbf{x}_k)\mathbf{x}_*$  is always acute, we have

$$\begin{aligned} \sin(\angle(\mathbf{x}_k, \varphi(\mathbf{x}_k)\mathbf{x}_*)) &= \left\| \left( \mathbf{I}_d - \frac{\mathbf{x}_k \mathbf{x}_k^\top}{\|\mathbf{x}_k\|_2^2} \right) \frac{\varphi(\mathbf{x}_k)\mathbf{x}_*}{\|\mathbf{x}_*\|_2} \right\| \\ &\leq \left\| \left( \mathbf{I}_d - \frac{\mathbf{x}_k \mathbf{x}_k^\top}{\|\mathbf{x}_k\|_2^2} \right) \frac{\varphi(\mathbf{x}_k)\mathbf{x}_* - \mathbf{x}_k}{\|\mathbf{x}_*\|_2} \right\| \\ &\stackrel{(i)}{\leq} \frac{\|\mathbf{x}_k - \varphi(\mathbf{x}_k)\mathbf{x}_*\|_2}{\|\mathbf{x}_*\|_2} = \frac{\text{dist}(\mathbf{x}_k, \mathbf{x}_*)}{\|\mathbf{x}_*\|_2} \\ &\stackrel{(ii)}{\leq} \sin\left(\frac{2}{25}\right), \end{aligned} \quad (32)$$

where (i) holds since the project operator is non-expansive; (ii) follows since the induction hypothesis implies

$$\begin{aligned} \text{dist}(\mathbf{x}_k, \mathbf{x}_*) &\leq \nu_\eta^k \cdot \text{dist}(\mathbf{x}_0, \mathbf{x}_*) + \frac{\max_{i \in [0:k-1]} \epsilon_i}{C_\eta} \sum_{t=0}^{k-1} \nu_\eta^t \\ &\leq \nu_\eta^k \cdot \text{dist}(\mathbf{x}_0, \mathbf{x}_*) + (1 - \nu_\eta) \sin\left(\frac{2}{25}\right) \|\mathbf{x}_*\|_2 \sum_{t=0}^{k-1} \nu_\eta^t \\ &\leq \sin\left(\frac{2}{25}\right) \|\mathbf{x}_*\|_2, \end{aligned}$$

where the second and the last inequalities follow from (14).

Hence, in Lemma VI.1, we plug in  $\theta = 2/25$ . Then the sample complexity in Theorem IV.1 invokes Lemma VI.1, (28), (29), and (30) hold with probability at least  $1 - \delta$  simultaneously. The remainder of the proof is conditioned on the events that (28), (29), and (30) hold.

By applying (28) and (29) to (c) and (d) of (26) and (30) to (e) of (26) with the choice of  $\theta = 2/25$ , we obtain

$$\|\mathbf{x}_{k+1} - \varphi(\mathbf{x}_k)\mathbf{x}_*\|_2 \leq \nu_\eta \|\mathbf{x}_k - \varphi(\mathbf{x}_k)\mathbf{x}_*\|_2 + \frac{\epsilon_k}{C_\eta}$$

for

$$\nu_\eta := \frac{c_0}{C_\eta} \quad \text{and} \quad C_\eta := (1 - 2\eta) \sqrt{\frac{2}{\pi}} - c_0 - \frac{1}{250} (1 + \sqrt{\eta}), \quad (33)$$

where

$$c_0 := \frac{4}{25\pi} \left( \sqrt{\frac{2}{\pi}} + \sqrt{2 \log \left( \frac{25e\pi}{4} \right)} \right) + \frac{1}{625\sqrt{\pi}}.$$

Since  $\nu_\eta$  satisfies

$$\frac{d\nu_\eta}{d\eta} = \frac{c_0 \left( 2\sqrt{\frac{2}{\pi}} + \frac{1}{500\sqrt{\eta}} \right)}{\left( (1 - 2\eta) \sqrt{\frac{2}{\pi}} - c_0 - \frac{1}{250} (1 + \sqrt{\eta}) \right)^2} > 0$$

for all  $\eta \in [0, 1/4]$ , it is monotonically increasing in  $\eta$  and upper-bounded as  $\nu_\eta \leq \nu_{1/4} < 9/10$ . This implies  $\nu_\eta < 1$  uniformly over  $\eta \in [0, 1/4]$ . This completes the proof of (19).

## VII. SUPPORTING LEMMAS

**Lemma VII.1.** *Let  $\mathbf{g} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_d)$  and  $\theta \in (0, \pi)$ . Let  $W_\theta$  be defined as in (27). Then we have*

$$\sup_{W \in \mathcal{W}_\theta} \mathbb{P}(\mathbf{g} \in W) \leq \frac{\theta}{\pi}.$$

*Proof:* Let  $W \in \mathcal{W}_\theta$  be arbitrarily fixed. It follows from the definitions in (27) and (22) that  $W$  is a cone. Therefore,  $\mathbf{g} \in W$  if and only if  $\mathbf{g}/\|\mathbf{g}\|_2 \in W$ . Furthermore, note that  $\mathbf{g}/\|\mathbf{g}\|_2$  is uniformly distributed in  $\mathbb{S}^{d-1}$ . Then we have

$$\mathbb{P}(\mathbf{g} \in W) = \mathbb{P}\left(\frac{\mathbf{g}}{\|\mathbf{g}\|_2} \in W\right) \leq \frac{\theta}{\pi}. \quad (34)$$

The assertion follows since  $W$  was arbitrary.  $\blacksquare$

**Lemma VII.2** ([29, Lemma 2.1]). *Let  $\delta \in (0, 1)$  and  $\{\mathbf{a}_i\}_{i=1}^m$  be independent copies of  $\mathbf{g} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_d)$ . Then it holds with probability at least  $1 - \delta$  that*

$$\sup_{\mathbf{z} \in \mathbb{S}^{d-1}} \left| \frac{1}{m} \sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{z} \rangle| - \sqrt{\frac{2}{\pi}} \right| \leq 4\sqrt{\frac{d}{m}} + \sqrt{\frac{2 \log(2/\delta)}{m}}. \quad (35)$$

**Lemma VII.3** ([30, Lemma 6.4]). *Let  $\delta \in (0, 1)$  and  $\{\mathbf{a}_i\}_{i=1}^m$  be independent copies of  $\mathbf{g} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_d)$ . Let  $s \in \mathbb{N}$  satisfy  $s < m$ . Then it holds with probability at least  $1 - \delta$  that*

$$\begin{aligned} \sup_{\substack{\mathbf{z} \in \mathbb{S}^{d-1} \\ T: |T| \leq s}} \frac{1}{s} \sum_{i \in T} |\langle \mathbf{a}_i, \mathbf{z} \rangle| \\ \leq \sqrt{\frac{2}{\pi}} + 4\sqrt{\frac{d}{s}} + \sqrt{2 \log \left( \frac{em}{s} \right)} + \sqrt{\frac{2}{s}} \cdot \log \left( \frac{2}{\delta} \right). \end{aligned} \quad (36)$$

**Lemma VII.4** ([28, Lemma 5.1]). *Let  $\delta \in (0, 1)$  and an acute angle  $\theta > 0$ . Suppose  $\{\mathbf{a}_i\}_{i=1}^m$  be independent copies of a random variable  $\mathbf{a} \in \mathbb{R}^d$  and we consider the set  $W_\theta$  given by (27). Then, if*

$$m \geq (4\pi/\theta)^2 (2d \log(2em/d) + \log(2/\delta)),$$

we have

$$\sup_{W \in \mathcal{W}_\theta} \frac{1}{m} \sum_{i=1}^m \mathbb{1}_{\{\mathbf{a}_i \in W\}} \leq \frac{2\theta}{\pi}. \quad (37)$$

holds with probability at least  $1 - \delta$ .

## VIII. PROOF OF LEMMA VI.1

We proceed with the proof under the following four events, each of which holds with probability at least  $1 - \delta/4$ . The first event is defined as

$$\begin{aligned} \sup_{\mathbf{z} \in \mathbb{S}^{d-1}} \left| \frac{1}{m} \sum_{i \in I_{\text{in}}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| - (1 - \eta) \sqrt{\frac{2}{\pi}} \right| \\ \leq 4\sqrt{\frac{d}{m}} + \sqrt{\frac{2 \log(8/\delta)}{m}}, \end{aligned} \quad (38)$$

which holds with probability at least  $1 - \delta/4$ . Since by the assumption on outliers, we have a set  $|I_{\text{in}}|$  with  $|I_{\text{in}}| = (1 - \eta)m$  and the outliers are independent of  $\{\mathbf{a}_i\}_{i=1}^m$ . Hence, (38) is a direct result of (35) in Lemma VII.2. By following the same argument, we also have that

$$\sup_{\mathbf{z} \in \mathbb{S}^{d-1}} \left| \frac{1}{m} \sum_{i \in I_{\text{out}}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| - \eta \sqrt{\frac{2}{\pi}} \right| \leq 4\sqrt{\frac{\eta d}{m}} + \sqrt{\frac{2\eta \log(8/\delta)}{m}} \quad (39)$$

holds with probability at least  $1 - \delta/4$ .

Next, we describe the following event: for an arbitrary fixed  $\alpha \in (0, 1)$ , it holds with probability at least  $1 - \delta/4$  that

$$\begin{aligned} \sup_{\substack{T: |T| \leq \alpha m \\ \mathbf{z} \in \mathbb{S}^{d-1}}} \frac{1}{m} \sum_{i \in T \cap I_{\text{in}}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| \leq \\ \alpha \sqrt{\frac{2}{\pi}} + 4\sqrt{\frac{\alpha d}{m}} + \alpha \sqrt{2 \log\left(\frac{e}{\alpha}\right)} + \sqrt{\frac{2\alpha \log(8/\delta)}{m}}. \end{aligned} \quad (40)$$

Again, since by the Assumption 1, we have a fixed set  $|I_{\text{in}}|$  with  $|I_{\text{in}}| = (1 - \eta)m$  and the outliers are independent of  $\{\mathbf{a}_i\}_{i=1}^m$ , (40) holds by (36) in Lemma VII.3.

Since (31) invokes Lemma VII.4 with probability at least  $1 - \delta/4$ , it holds with probability at least  $1 - \delta/4$  that

$$\sup_{W \in \mathcal{W}_\theta} \sum_{i=1}^m \mathbb{1}_{\{\mathbf{a}_i \in W\}} \leq \frac{2\theta m}{\pi}. \quad (41)$$

Since we have shown that (38),(39),(40) and (41) hold with probability at least  $1 - \delta$ , we will move forward with the remainder of the proof by assuming those conditions are satisfied.

We first show (28). We observe that for an arbitrary  $W \in \mathcal{W}_\theta$  and  $\mathbf{z} \in \mathbb{S}^{d-1}$ , it holds deterministically that

$$\begin{aligned} \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \notin W\}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| = \\ \frac{1}{m} \sum_{i \in I_{\text{in}}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| - \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \in W\}} |\langle \mathbf{a}_i, \mathbf{z} \rangle|. \end{aligned}$$

Hence, by taking infimum on both sides over sets  $\mathcal{W}_\theta$  and  $\mathbb{S}^{d-1}$ , we have

$$\begin{aligned} \inf_{\substack{W \in \mathcal{W}_\theta \\ \mathbf{z} \in \mathbb{S}^{d-1}}} \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \notin W\}} |\langle \mathbf{a}_i, \mathbf{z} \rangle| \\ \geq \underbrace{\inf_{\mathbf{z} \in \mathbb{S}^{d-1}} \frac{1}{m} \sum_{i \in I_{\text{in}}} |\langle \mathbf{a}_i, \mathbf{z} \rangle|}_{(A)} - \underbrace{\sup_{\substack{W \in \mathcal{W}_\theta \\ \mathbf{z} \in \mathbb{S}^{d-1}}} \frac{1}{m} \sum_{i \in I_{\text{in}}} \mathbb{1}_{\{\mathbf{a}_i \in W\}} |\langle \mathbf{a}_i, \mathbf{z} \rangle|}_{(B)}. \end{aligned} \quad (42)$$

We first obtain a lower bound on (A) and an upper bound on (B). We have a lower bound on (A) by (38):

$$(A) \geq (1 - \eta) \sqrt{\frac{2}{\pi}} - 4\sqrt{\frac{d}{m}} - \sqrt{\frac{2 \log(8/\delta)}{m}}. \quad (43)$$

By taking  $m$  (31) in (43) for a sufficiently large  $C > 0$ , we have

$$(A) \geq (1 - \eta) \sqrt{\frac{2}{\pi}} - \frac{\theta}{20}. \quad (44)$$

It remains to show an upper bound on (B). Under the event (41), we have

$$(B) \leq \sup_{\substack{T: |T| \leq 2\theta m/\pi \\ \mathbf{z} \in \mathbb{S}^{d-1}}} \frac{1}{m} \sum_{i \in T \cap I_{\text{in}}} |\langle \mathbf{a}_i, \mathbf{z} \rangle|.$$

Therefore, by letting  $\alpha = 2\theta/\pi$  in (40), (40) gives an upper bound on (B):

$$(B) \leq \frac{2\theta}{\pi} \sqrt{\frac{2}{\pi}} + 4\sqrt{\frac{2\theta d}{\pi m}} + \frac{2\theta}{\pi} \sqrt{2 \log\left(\frac{e\pi}{2\theta}\right)} + \sqrt{\frac{4\theta \log(8/\delta)}{\pi m}}. \quad (45)$$

Taking  $m$  according to (31) yields

$$(B) \leq \frac{2\theta}{\pi} \left( \sqrt{\frac{2}{\pi}} + \sqrt{2 \log\left(\frac{e\pi}{2\theta}\right)} \right) + \frac{\theta}{20} \sqrt{\frac{2\theta}{\pi}}. \quad (46)$$

Hence, putting the results (44) and (46) into (42) completes the proof of the statement (28).

For the proofs of remaining statements in (29) and (30), the upper bound in (29) is a direct consequence of (39) with choosing  $n$  according (31). Lastly, (30) is the result of the upper bound of (B) in (46). These complete the proof of (29) and (30).

## IX. CONCLUSION

The least absolute deviation (LAD) has been a popular statistical method for regression in the presence of outliers. We consider the LAD approach to robust phase retrieval with the magnitude-only measurement model. To solve the resulting non-convex optimization, we derive a robust alternating minimization method (Robust-AM) as an unconstrained Gauss-Newton method. Furthermore, we propose fast Robust-AM by exploiting efficient solvers and show that Robust-AM by ADMM converges faster than a similar approach known as the prox-linear by its efficient solver POGS [14].

We established a local convergence analysis of Robust-AM under the standard Gaussian measurement model when the support of sparse noise is arbitrarily fixed but magnitudes can be adversarial. A suitably initialized Robust-AM converges linearly to the ground truth uniformly over all ground-truth signals when the number of measurements  $m$  is proportional to the signal length  $d$  and the outlier fraction is up to  $1/4$ . This theoretical result is comparable to existing prior art in the literature. Furthermore, the numerical results show that Robust-AM outperforms the existing guaranteed methods for various outlier models in both synthetic data and real image data.

## REFERENCES

- [1] S. Kim and K. Lee, "Sequence of linear program for robust phase retrieval," *2024 IEEE International Conference on Acoustics, Speech and Signal Processing*, to appear.
- [2] A. Walthers, "The question of phase retrieval in optics," *Optica Acta: International Journal of Optics*, vol. 10, no. 1, pp. 41–49, 1963.
- [3] O. Bunk, A. Diaz, F. Pfeiffer, C. David, B. Schmitt, D. K. Satapathy, and J. F. Van Der Veen, "Diffractive imaging for periodic samples: retrieving one-dimensional concentration profiles across microfluidic channels," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 63, no. 4, pp. 306–314, 2007.
- [4] A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," *Inverse Problems*, vol. 27, no. 1, p. 015005, 2010.
- [5] 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 Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.
- [6] D. S. Weller, A. Pnueli, G. Divon, O. Radzyner, Y. C. Eldar, and J. A. Fessler, "Undersampled phase retrieval with outliers," *IEEE Transactions on Computational Imaging*, vol. 1, no. 4, pp. 247–258, 2015.
- [7] J. Dong, L. Valzania, A. Maillard, T.-a. Pham, S. Gigan, and M. Unser, "Phase retrieval: From computational imaging to machine learning: A tutorial," *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 45–57, 2023.
- [8] S. Bahmani and J. Romberg, "Phase retrieval meets statistical learning theory: A flexible convex relaxation," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 252–260.
- [9] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2675–2689, 2018.
- [10] P. Hand and V. Voroninski, "Corruption robust phase retrieval via linear programming," *arXiv preprint arXiv:1612.03547*, 2016.
- [11] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, "A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms," *Journal of Machine Learning Research*, 2017.
- [12] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 773–794, 2017.
- [13] H. Zhang, Y. Chi, and Y. Liang, "Median-truncated nonconvex approach for phase retrieval with outliers," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7287–7310, 2018.
- [14] J. C. Duchi and F. Ruan, "Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval," *Information and Inference: A Journal of the IMA*, vol. 8, no. 3, pp. 471–529, 2019.
- [15] P. Bloomfield and W. L. Steiger, *Least absolute deviations: theory, applications, and algorithms*. Springer, 1983.
- [16] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, p. 237, 1972.
- [17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [18] S. Wang and N. Shroff, "A new alternating direction method for linear programming," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [19] J. van den Brand, "A deterministic linear program solver in current matrix multiplication time," in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2020, pp. 259–278.
- [20] J. V. Burke and M. C. Ferris, "A Gauss–Newton method for convex composite optimization," *Mathematical Programming*, vol. 71, no. 2, pp. 179–194, 1995.
- [21] F. Clarke, *Optimization and Nonsmooth Analysis*, ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1990.
- [22] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *Advances in Neural Information Processing Systems*, vol. 26, 2013.
- [23] N. Parikh and S. Boyd, "Block splitting for distributed optimization," *Mathematical Programming Computation*, vol. 6, no. 1, pp. 77–102, 2014.
- [24] K. Holmström, A. O. Göran, and M. M. Edvall, "User's guide for tomlab/cplex v12. 1," *Tomlab Optim. Retrieved*, vol. 1, p. 2017, 2009.
- [25] L. Gurobi Optimization, "Gurobi optimizer reference manual," 2021.
- [26] T. Yang and Q. Lin, "Rsg: Beating subgradient method without smoothness and strong convexity," *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 236–268, 2018.
- [27] Y. Ye and E. Tse, "An extension of karmarkar's projective algorithm for convex quadratic programming," *Mathematical programming*, vol. 44, pp. 157–179, 1989.
- [28] Y. S. Tan and R. Vershynin, "Phase retrieval via randomized kaczmarz: theoretical guarantees," *Information and Inference: A Journal of the IMA*, vol. 8, no. 1, pp. 97–123, 2019.
- [29] Y. Plan and R. Vershynin, "Dimension reduction by random hyperplane tessellations," *Discrete & Computational Geometry*, vol. 51, no. 2, pp. 438–461, 2014.
- [30] —, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 482–494, 2012.