Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 18, No. 1, pp. 213–232, 2022

$L_{1/2}$-REGULARIZED LEAST ABSOLUTE DEVIATION METHOD FOR SPARSE PHASE RETRIEVAL*

Lijiao Kong, Ailing Yan†, Yan Li and Jun Fan

Abstract: Sparse phase retrieval aims to recover a sparse signal from the magnitudes of its linear measurements. However, in real applications, the measurements are often corrupted by outliers and asymmetrical distribution noise. In this paper, we introduce a novel method which consists of Least absolute deviation loss function and an $L_{1/2}$ regularizer. It is a nonconvex, nonsmooth, and non-Lipschitz optimization problem. We design an efficient alternating direction method of multipliers(ADMM) to solve the problem and establish its convergence. Extensive numerical experiments demonstrate that the proposed method can recover sparse signal with less measurements and is robust to dense bounded noise as well as Laplace noise.

Key words: *nonconvex optimization, phase retrieval, alternating direction method of multipliers,* $L_{1/2}$ *regularization, asymmetrical distribution noise, robustness*

Mathematics Subject Classification: *90C26, 90C30, 90C90*

1 Introduction

Phase retrieval is the problem of recovering the unknown signal $x \in \mathbb{C}^p$ from the following model:

$$y = \mid Ax \mid^2 + \epsilon, \tag{1.1}$$

where $A \in \mathbb{C}^{n\times p}$ is the known measurements matrix, n is the number of measurements, $y \in \mathbb{R}^n$ is the squared-magnitude measurements, $\epsilon \in \mathbb{R}^n$ is noise or outliers [45] and $|\cdot|^2$ denotes the element-wise absolute-squared value. Specially, when both A and x belong to the real field, the problem is called real-valued phase retrieval [4, 30]. Phase retrieval problem has many important applications, including X-ray crystallography [21], optics [37], astronomy [14], and blind ptychography [9].

A classical kind of phase retrieval algorithm is based on the use of alternate projections between different constraints including the seminal work Gerchberg-Saxton algorithm [19] and Fienup method [18]. These methods require the prior information to be very precise, and convergence is not guaranteed since the projections are generally onto nonconvex sets. Another class of algorithms without exploring or employing the prior information of the signal

*This work was supported by the National Natural Science Foundation of China(NSFC) (No. 11801130), the Natural Science Foundation of Hebei Province (No. A2019202135) and the Ministry of Education of Humanities and Social Science Project (17YJC910005).

†Corresponding author

includes semidefinite programming (SDP) [5, 8, 36], polarization [1], alternating minimization [29], gradient methods, such as the Wirtinger flow method [7], trust region [35], etc. In many applications, the signals are naturally sparse. For example, astronomical imaging deals with the locations of stars in the sky [22], optics imaging [33], user requirement survey and so on. Recently, many methods are explored to solve sparse phase retrieval problem, which can be mainly categorized as convex and nonconvex ones. The convex methods either rely on semidefinite relaxation based methods [23, 24], or use basis pursuit methods [15, 20]. The nonconvex methods include sparse Fienup method [27], greedy method [32], Wirtinger Flow variants [4, 25, 34, 43], and $L_p(0 \leq p \leq 1)$ regularization methods [10, 16, 17, 31, 38].

In recent years, $L_p(0 \leq p \leq 1)$ regularization methods have a good performance in dealing with the sparse phase retrieval, which don't need support set information and can get a sparse solution. To get the sparsest solution, one naturally proposes to use L_0 regularization [16]. However, the corresponding optimization is generally NP hard (see [28]). In order to overcome such difficulty, researchers proposed many relevant relaxation methods, such as, L_1 regularization [10, 31, 38] and $L_q(0 < q < 1)$ regularization [17]. Although L_1 regularization is convenient to be calculated, it may yield inconsistent selections in variable selection and cannot recover a signal with the least measurements [41]. Compared to L_1 regularization, L_q regularization always uses less measurements and can generate sparser solution in compressed sensing [11]. Specially, in phase diagram study, [42] shows the following results: 1)As the value of q decreases, the L_q regularization generates sparser solution. 2)When $1/2 \leq q < 1$, $L_{1/2}$ regularization always yields the sparsest solution and when $0 < q < 1/2$, the performance of the L_q regularization takes no significant difference. Accordingly, it is concluded that $L_{1/2}$ regularization can be taken as a representative of $L_q(0 < q < 1)$ regularizations.

Regarding phase retrieval with asymmetrical noise or outliers, some stability results have been established. For example, [46] developed for minimizing the least squares empirical loss and designed a two stages algorithm, which starts with a weighted maximal correlation initialization and then follows by the reweighted gradient iterations. [12] and [45] used mean truncation and median truncation rule to weaken the influence of arbitrary outliers, respectively. [47] proposed the median-MRAF algorithm which combined median truncation rule and reweighted method. Different from these literatures, [40] took Least absolute deviation(LAD) criterion for phase retrieval to enhance the robustness against outliers. For sparse case, [16, 38] respectively used L_0 regularized and L_1 regularized LAD method, and employed alternating direction method of multipliers(ADMM) [3] to solve corresponding optimization problem.

From the studies above, the special importance of $L_{1/2}$ regularization is highlighted. Motivated by [41], we present a new $L_{1/2}$ regularized phase retrieval method to get a sparse solution and enhance the robustness, we call our method $L_{1/2}$ regularized Least absolute deviation phase retrieval($L_{1/2}$LAD PR). Since the $L_{1/2}$ regularization problem is nonconvex, nonsmooth, and non-Lipschitz, it is hard to solve. Inspired by [16], we consider efficient algorithm based on ADMM method. Fortunately, all subproblems have closed-form solutions and convergence can be guaranteed. Numerical experiments show that our method can recover sparse signal with less measurements and is robust to asymmetrical distribution noise, such as dense bounded noise and Laplace noise.

The remainder of the paper is organized as follows. In Section 2, we formulate the problem, describe our proposed method in detail and establish its convergence. Extensive numerical experiments illustrating the effectiveness and robustness of our algorithm are presented in Section 3. Conclusion and future work are given in Section 4.

2 Optimization Algorithm

A. The problem formulation

The optimization problem that we consider is minimizing the problem as follow:

$$\min_{x\in\mathbb{C}^p} \frac{1}{n}\| \mid Ax \mid^2 - y\|_1 + \lambda\|x\|_{1/2}^{1/2}, \tag{2.1}$$

where x, y, A, n have been described in (1.1), $(\cdot)^T$ represents the transpose, $A = [a_1, a_2, \ldots, a_n]^T$, $a_i \in \mathbb{C}^p$, $i = 1, 2, \ldots, n$, a_i are Gaussian random vectors, the regularization penalty parameter $\lambda > 0$ controls the tradeoff between the sparsity and data fidelity. Unless otherwise stated, for any vector V, V_i is its i-th element. $\|x\|_{1/2}^{1/2} = \sum_{i=1}^{p} \mid x_i \mid^{1/2}$, $x = (x_1, x_2, ..., x_p)^T$.

We should establish the existence of solution to problem (2.1). For the convenience of following description, we introduce $g(x) = \frac{1}{n}\| \mid Ax \mid^2 - y\|_1 + \lambda\|x\|_{1/2}^{1/2}$. Apparently, the second part of $g(x)$ is coercive, the first part of $g(x)$ is non-negative, then we can readily derive that $g(x)$ is coercive. Finally, we obtain the existence of solution to (2.1).

For the convenience of following calculation, we give two lemmas.

Lemma 2.1 (see [41, 44]). *Let b^* be the global solution of following problem*

$$arg \min_{b\in\mathbb{C}} \mid b - y_0 \mid^2 + \lambda \mid b \mid^{1/2},$$

where $y_0 \in \mathbb{C}$, the constant $\lambda > 0$. *It can be specified by*

$$b^* = \begin{cases} f_{\lambda,1/2}(y_0), & \mid y_0 \mid > \frac{\sqrt[3]{54}}{4}(\lambda)^{2/3}, \\ 0, & \text{otherwise}, \end{cases}$$

with

$$f_{\lambda,1/2}(y_0) = \frac{2}{3}y_0\left(1 + cos\left(\frac{2\pi}{3} - \frac{2}{3}\varphi_\lambda(y_0)\right)\right),$$

$$\varphi_\lambda(y_0) = arccos\left(\frac{\lambda}{8}\left(\frac{|y_0|}{3}\right)^{-\frac{3}{2}}\right).$$

Lemma 2.2. *The global solution $\hat{v}$ of following problem has analytic expression,*

$$\hat{v} = arg \min_{v\in\mathbb{R}} \frac{t}{2}(v - y_1)^2 + \mid v^2 - y_2 \mid,$$

where $y_1, y_2 \in \mathbb{R}$, $t > 0$ is a constant. It can be derived

$$\hat{v} = h_t(y_1, y_2),$$

when $0 < t \leq 2$,

$$h_t(y_1, y_2) = \begin{cases} \frac{y_1}{1+\frac{2}{t}}, & \text{if } y_2 \leq 0 \text{ or } y_2 > 0 \text{ and } \mid y_1 \mid > \sqrt{y_2}(1 + \frac{2}{t}), \\ \sqrt{y_2}, & \text{if } y_2 > 0 \text{ and } 0 < y_1 \leq \sqrt{y_2}(1 + \frac{2}{t}), \\ -\sqrt{y_2}, & \text{if } y_2 > 0 \text{ and } -\sqrt{y_2}(1 + \frac{2}{t}) \leq y_1 \leq 0, \end{cases}$$

when $t > 2$,

$$h_t(y_1, y_2) = \begin{cases} \frac{y_1}{1+\frac{2}{t}}, & \text{if } y_2 \leq 0 \text{ or } y_2 > 0 \text{ and } \mid y_1 \mid > \sqrt{y_2}(1+\frac{2}{t}), \\ \frac{y_1}{1-\frac{2}{t}}, & \text{if } y_2 > 0 \text{ and } \mid y_1 \mid < \sqrt{y_2}(1-\frac{2}{t}), \\ \sqrt{y_2}, & \text{if } y_2 > 0 \text{ and } \sqrt{y_2}(1-\frac{2}{t}) \leq y_1 \leq \sqrt{y_2}(1+\frac{2}{t}), \\ -\sqrt{y_2}, & \text{if } y_2 > 0 \text{ and } -\sqrt{y_2}(1+\frac{2}{t}) \leq y_1 \leq -\sqrt{y_2}(1-\frac{2}{t}). \end{cases}$$

Proof. We discuss the minimizer of the problem in two cases as follows.
1) If $y_2 \leq 0$, the problem can be written as

$$\begin{aligned} \hat{v} &= arg \min_{v \in \mathbb{R}} \frac{t}{2}(v-y_1)^2 + v^2 - y_2 \\ &= arg \min_{v \in \mathbb{R}} \left(\frac{t}{2}+1\right) v^2 - ty_1 v + \frac{t}{2} y_1^2 - y_2. \end{aligned}$$

It is clear that $\hat{v} = \frac{y_1}{1+\frac{2}{t}}$.
2) If $y_2 > 0$, let

$$h(v) = \begin{cases} h_1(v), & \text{if } \mid v \mid \geq \sqrt{y_2}, \\ h_2(v), & \text{otherwise}, \end{cases}$$

where $h_1(v) = \frac{t}{2}(v-y_1)^2 + v^2 - y_2$ and $h_2(v) = \frac{t}{2}(v-y_1)^2 - v^2 + y_2$.
The derivative and second derivative of the functions $h_1(v)$, $h_2(v)$ are computed as follows,

$$\begin{aligned} h_1'(v) &= (t+2)v - ty_1, \\ h_1^{''}(v) &= t+2 > 0, \end{aligned}$$

and

$$\begin{aligned} h_2'(v) &= (t-2)v - ty_1, \\ h_2^{''}(v) &= t-2. \end{aligned}$$

For the convenience of following analysis, we define $\ell_1^* = \frac{y_1}{1+\frac{2}{t}}$ and $\ell_2^* = \frac{y_1}{1-\frac{2}{t}}$ (when $t \neq 2$).

Next, we discuss the global minimization $\hat{v}$ of $h(v)$. When $0 < t \leq 2$, we can get $h_2^{''}(v) = t - 2 \leq 0$. Then $h_2(v)$ is concave when $\mid v \mid < \sqrt{y_2}$. By analyzing the second derivative of function $h(v)$, the position relations of ℓ_1^*, ℓ_2^* and $\pm\sqrt{y_2}$, we have the global minimization

$$\hat{v} = \begin{cases} \ell_1^*, & \mid \ell_1^* \mid > \sqrt{y_2}, \\ -\sqrt{y_2}, & -\sqrt{y_2} \leq \ell_1^* \leq 0, \\ \sqrt{y_2}, & 0 < \ell_1^* \leq \sqrt{y_2}. \end{cases}$$

When $t > 2$, we obtain $h_2^{''}(v) = t - 2 > 0$. Hence $h_2(v)$ is convex when $\mid v \mid < \sqrt{y_2}$. Notice that

$$\mid \ell_1^* \mid = \left| \frac{y_1}{1+\frac{2}{t}} \right| < \left| \frac{y_1}{1-\frac{2}{t}} \right| = \mid \ell_2^* \mid .$$

A similar analysis of the above case shows that the global minimization can be obtained by

$$\hat{v} = \begin{cases} \ell_1^*, & \mid \ell_1^* \mid > \sqrt{y_2}\ , \\ \ell_2^*, & \{-\sqrt{y_2} \leq \ell_1^* \leq 0 \text{ and } -\sqrt{y_2} < \ell_2^* \leq 0\} \text{ or} \\ & \quad \{0 < \ell_1^* \leq \sqrt{y_2} \text{ and } 0 < \ell_2^* < \sqrt{y_2}\}, \\ \sqrt{y_2}, & 0 < \ell_1^* \leq \sqrt{y_2} \text{ and } \ell_2^* \geq \sqrt{y_2}, \\ -\sqrt{y_2}, & -\sqrt{y_2} \leq \ell_1^* \leq 0 \text{ and } \ell_2^* \leq -\sqrt{y_2}. \end{cases}$$

To sum up, we can get the solution $\hat{v}$. □

Further, for the convenience of description in the following, we define an operator $\mathcal{H}_t(c, w)$ for any $c = (c_1, c_2, \ldots, c_p)^T,\ w = (w_1, w_2, \ldots, w_p)^T \in \mathbb{R}^p$, where

$$\mathcal{H}_t(c, w) = (h_t(c_1, w_1), h_t(c_2, w_2), \ldots, h_t(c_p, w_p))^T.$$

B. Solving the objective with ADMM

We apply the ADMM algorithm [3] to solve the proposed model (2.1), which is equivalent to

$$\begin{aligned} &\min_{x,q,z} \tfrac{1}{n} \left\| | z |^2 - y \right\|_1 + \lambda \|q\|_{1/2}^{1/2} \\ &s.t. \quad \begin{pmatrix} A \\ I \end{pmatrix} x - \begin{pmatrix} z \\ q \end{pmatrix} = 0, \end{aligned} \tag{2.2}$$

where $z \in \mathbb{C}^n,\ q \in \mathbb{C}^p,\ I$ is the $p \times p$ identity matrix. Then, the augmented Lagrangian $\mathcal{L}_r(x, q, z; \Lambda)$ is introduced in order to settle the following saddle-point problem

$$\begin{aligned} \max_{\Lambda} \min_{x,q,z} \mathcal{L}_r(x, q, z; \Lambda) = \tfrac{1}{n} \left\| | z |^2 - y \right\|_1 + \lambda \|q\|_{1/2}^{1/2} + \\ \Re \left(\left\langle \Lambda, \begin{pmatrix} A \\ I \end{pmatrix} x - \begin{pmatrix} z \\ q \end{pmatrix} \right\rangle \right) + \tfrac{r}{2} \left\| \begin{pmatrix} A \\ I \end{pmatrix} x - \begin{pmatrix} z \\ q \end{pmatrix} \right\|_2^2, \end{aligned} \tag{2.3}$$

where $\Lambda \in \mathbb{C}^{n+p}$ is Lagrangian multiplier, $r > 0$ is augmented Lagrangian penalty parameter, $\langle \cdot, \cdot \rangle$ denotes the complex inner product of two vectors, $\Re(\cdot)$ is the real part. For the convenience of following calculation, we rewrite (2.3) to its equivalent form:

$$\begin{aligned} \max_{\Lambda_1, \Lambda_2} \min_{x,q,z} \mathcal{L}_r(x, q, z; \Lambda_1, \Lambda_2) = \frac{1}{n} \left\| | z |^2 - y \right\|_1 + \lambda \|q\|_{1/2}^{1/2} + \Re \left(\langle \Lambda_2, x - q \rangle \right) + \\ \frac{r}{2} \|x - q\|_2^2 + \Re \left(\langle \Lambda_1, Ax - z \rangle \right) + \frac{r}{2} \|Ax - z\|_2^2, \end{aligned} \tag{2.4}$$

where $\Lambda_1 \in \mathbb{C}^n,\ \Lambda_2 \in \mathbb{C}^p$ are Lagrangian multipliers.

We invoke ADMM to solve (2.4), which is sketched as Algorithm 1.

$$\begin{aligned} x^{j+1} &= arg\ \min_{x} \mathcal{L}_r(x, q^j, z^j; \Lambda_1^j, \Lambda_2^j), \\ q^{j+1} &= arg\ \min_{q} \mathcal{L}_r(x^{j+1}, q, z^j; \Lambda_1^j, \Lambda_2^j), \\ z^{j+1} &= arg\ \min_{z} \mathcal{L}_r(x^{j+1}, q^{j+1}, z; \Lambda_1^j, \Lambda_2^j), \\ \Lambda_1^{j+1} &= \Lambda_1^j + r(Ax^{j+1} - z^{j+1}), \\ \Lambda_2^{j+1} &= \Lambda_2^j + r(x^{j+1} - q^{j+1}). \end{aligned}$$

In the following, we discuss the solution to each sub-minimization problem with respect to(w.r.t.) $x,\ q,\ z$.

1) x-subproblem. The x-subproblem is

$$\min_{x \in \mathbb{C}^p} \Re \left(\langle \Lambda_1, Ax \rangle \right) + \Re \left(\langle \Lambda_2, x \rangle \right) + \tfrac{r}{2} \|Ax - z\|_2^2 + \tfrac{r}{2} \|x - q\|_2^2. \tag{2.5}$$

We compute the first order optimality condition for (2.5) w.r.t. the complex-valued variable x by separating the real and complex parts. Detailed derivation is given in the following, where the sign re(or Re) stands for the real part, im(or Im) for the imaginary part, $(\cdot)^H$ represents the complex conjugate transpose, $i = \sqrt{-1}$ is the imaginary unit.

We denote

$$x = x^{re} + ix^{im};\ q = q^{re} + iq^{im};\ z = z^{re} + iz^{im};$$
$$A = A_{re} + iA_{im};\ \Lambda_1 = \Lambda_1^{re} + i\Lambda_1^{im}\ ;\ \Lambda_2 = \Lambda_2^{re} + i\Lambda_2^{im}.$$

Further, we can get

$$Ax = (A_{re} + iA_{im})(x^{re} + ix^{im}) = (A_{re}x^{re} - A_{im}x^{im}) + i(A_{re}x^{im} + A_{im}x^{re}),$$

$$\begin{aligned}\langle \Lambda_1, Ax\rangle &= \langle \Lambda_1^{re} + i\Lambda_1^{im}, (A_{re}x^{re} - A_{im}x^{im}) + i(A_{re}x^{im} + A_{im}x^{re})\rangle \\ &= (\Lambda_1^{re} + i\Lambda_1^{im})^H[(A_{re}x^{re} - A_{im}x^{im}) + i(A_{re}x^{im} + A_{im}x^{re})] \\ &= [(\Lambda_1^{re})^T(A_{re}x^{re} - A_{im}x^{im}) + (\Lambda_1^{im})^T(A_{re}x^{im} + A_{im}x^{re})] + \\ &\quad i[-(\Lambda_1^{im})^T(A_{re}x^{re} - A_{im}x^{im}) + (\Lambda_1^{re})^T(A_{re}x^{im} + A_{im}x^{re})],\end{aligned}$$

$$\begin{aligned}\langle \Lambda_2, x\rangle &= \langle \Lambda_2^{re} + i\Lambda_2^{im}, x^{re} + ix^{im}\rangle \\ &= (\Lambda_2^{re} + i\Lambda_2^{im})^H(x^{re} + ix^{im}) \\ &= [(\Lambda_2^{re})^T x^{re} + (\Lambda_2^{im})^T x^{im}] + i[(\Lambda_2^{re})^T x^{im} - (\Lambda_2^{im})^T x^{re}],\end{aligned}$$

$$\begin{aligned}\|Ax - z\|_2^2 &= \|(A_{re}x^{re} - A_{im}x^{im} - z^{re}) + i(A_{re}x^{im} + A_{im}x^{re} - z^{im})\|_2^2 \\ &= \|A_{re}x^{re} - A_{im}x^{im} - z^{re}\|_2^2 + \|A_{re}x^{im} + A_{im}x^{re} - z^{im}\|_2^2,\end{aligned}$$

$$\|x - q\|_2^2 = \|(x^{re} - q^{re}) + i(x^{im} - q^{im})\|_2^2 = \|x^{re} - q^{re}\|_2^2 + \|x^{im} - q^{im}\|_2^2.$$

Let

$$D(x) = \Re\left(\langle \Lambda_1, Ax\rangle\right) + \Re\left(\langle \Lambda_2, x\rangle\right) + \tfrac{r}{2}\left\|Ax - z\right\|_2^2 + \tfrac{r}{2}\left\|x - q\right\|_2^2. \tag{2.6}$$

Take the derivations above into (2.6),

$$\begin{aligned}D(x^{re}, x^{im}) &= (\Lambda_1^{re})^T(A_{re}x^{re} - A_{im}x^{im}) + (\Lambda_1^{im})^T(A_{re}x^{im} + A_{im}x^{re}) + (\Lambda_2^{re})^T x^{re} \\ &\quad + (\Lambda_2^{im})^T x^{im} + \frac{r}{2}\left(\|A_{re}x^{re} - A_{im}x^{im} - z^{re}\|_2^2 + \|A_{re}x^{im} + A_{im}x^{re} - z^{im}\|_2^2\right) \\ &\quad + \frac{r}{2}\left(\|x^{re} - q^{re}\|_2^2 + \|x^{im} - q^{im}\|_2^2\right).\end{aligned} \tag{2.7}$$

Compute the partial derivatives w.r.t. x^{re}, x^{im} and let them be equal to 0. By the first order optimality condition, we can get the optimal solution,

$$\begin{aligned}\tfrac{\partial D}{\partial x^{re}} &= A_{re}^T\Lambda_1^{re} + A_{im}^T\Lambda_1^{im} + \Lambda_2^{re} + r[A_{re}^T(A_{re}x^{re} - A_{im}x^{im} - z^{re}) \\ &\quad + A_{im}^T(A_{re}x^{im} + A_{im}x^{re} - z^{im})] + r(x^{re} - q^{re}) = 0,\end{aligned} \tag{2.8}$$

$$\begin{aligned}\tfrac{\partial D}{\partial x^{im}} &= -A_{im}^T\Lambda_1^{re} + A_{re}^T\Lambda_1^{im} + \Lambda_2^{im} + r[-A_{im}^T(A_{re}x^{re} - A_{im}x^{im} - z^{re}) \\ &\quad + A_{re}^T(A_{re}x^{im} + A_{im}x^{re} - z^{im})] + r(x^{im} - q^{im}) = 0.\end{aligned} \tag{2.9}$$

For the convenience of observation, we rewrite (2.8) and (2.9),

$$\begin{aligned}&r(A_{re}^T A_{re} + A_{im}^T A_{im} + I)x^{re} + r(-A_{re}^T A_{im} + A_{im}^T A_{re})x^{im} + \\ &r(-A_{re}^T z^{re} - A_{im}^T z^{im} - q^{re}) + A_{re}^T\Lambda_1^{re} + A_{im}^T\Lambda_1^{im} + \Lambda_2^{re} = 0,\end{aligned} \tag{2.10}$$

$$\begin{aligned}&r(-A_{im}^T A_{re} + A_{re}^T A_{im})x^{re} + r(A_{re}^T A_{re} + A_{im}^T A_{im} + I)x^{im} + \\ &r(A_{im}^T z^{re} - A_{re}^T z^{im} - q^{im}) - A_{im}^T\Lambda_1^{re} + A_{re}^T\Lambda_1^{im} + \Lambda_2^{im} = 0.\end{aligned} \tag{2.11}$$

Since

$$A_{re}^T A_{re} + A_{im}^T A_{im} = Re(A^H A),$$
$$A_{re}^T A_{im} - A_{im}^T A_{re} = Im(A^H A),$$

then (2.10) and (2.11) can be further simplified,

$$\begin{aligned} &r[(Re(A^H A) + I)x^{re} - Im(A^H A)x^{im}] \\ &\qquad = rq^{re} - (A_{re}^T \Lambda_1^{re} + A_{im}^T \Lambda_1^{im}) + r(A_{re}^T z^{re} + A_{im}^T z^{im}) - \Lambda_2^{re}, \end{aligned} \tag{2.12}$$

$$\begin{aligned} &r[Im(A^H A)x^{re} + (Re(A^H A) + I)x^{im}] \\ &\qquad = rq^{im} + (A_{im}^T \Lambda_1^{re} - A_{re}^T \Lambda_1^{im}) + r(-A_{im}^T z^{re} + A_{re}^T z^{im}) - \Lambda_2^{im}. \end{aligned} \tag{2.13}$$

By analysis and comparison, (2.12) and (2.13) are the real part and the imaginary part of following equation, respectively,

$$r(I + A^H A)x = rq - A^H \Lambda_1 + rA^H z - \Lambda_2. \tag{2.14}$$

So we can calculate the x-subproblem directly by (2.14).

2) q-subproblem. The q-subproblem can be written as

$$\min_{q \in \mathbb{C}^p} \lambda \|q\|_{1/2}^{1/2} + \frac{r}{2} \left\| q - x - \frac{\Lambda_2}{r} \right\|_2^2. \tag{2.15}$$

The minimization w.r.t. q is a variable-splitting problem, updating the auxiliary vector q can be performed element-by-element. We just need to consider the minimization problem w.r.t. q_d, $d = 1, 2, \ldots, p$,

$$q_d^* = arg \min_{q_d \in \mathbb{C}} \frac{2\lambda}{r} \mid q_d \mid^{1/2} + \left| q_d - \left(x + \frac{\Lambda_2}{r}\right)_d \right|^2, \tag{2.16}$$

where q^* is the optimal solution of q-subproblem (2.15).

The minimization problem (2.16) can be solved by Lemma 2.1. Let $u_d = \left(x + \frac{\Lambda_2}{r}\right)_d$, the optimal solution is

$$q_d^* = \begin{cases} f_{2\lambda/r,1/2}(u_d), & \mid u_d \mid > \frac{\sqrt[3]{54}}{4} \left(\frac{2\lambda}{r}\right)^{2/3}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.17}$$

3) z-subproblem. The z-subproblem is given as follow

$$\min_{z \in \mathbb{C}^n} \frac{1}{n} \left\| \mid z \mid^2 - y \right\|_1 + \frac{r}{2} \left\| z - \left(Ax + \frac{\Lambda_1}{r}\right) \right\|_2^2. \tag{2.18}$$

For convenience, let $W = Ax + \frac{\Lambda_1}{r} \in \mathbb{C}^n$.

The z-subproblem is also variable-separable, the minimization problem w.r.t. z_m is

$$\min_{z_m \in \mathbb{C}} \frac{1}{n} \left| \mid z_m \mid^2 - y_m \right| + \frac{r}{2} \mid z_m - W_m \mid^2, \tag{2.19}$$

where $m = 1, 2, \ldots, n$. Inspired by [16], by a geometric interpretation, let $z_m = k \cdot W_m$, $k \geq 0 \in \mathbb{R}$. We transform the complex field minimization problem about variable z_m into the real field problem about k and consider the following equivalent problem

$$arg \min_{k \in \mathbb{R}} \frac{nr}{2} (k-1)^2 + \left| k^2 - \frac{y_m}{|W_m|^2} \right|.$$

The problem above can be solved by Lemma 2.2, hence the optimal solution z^* of problem (2.18) is

$$z_m^* = h_{nr}\left(1, \frac{y_m}{\mid W_m \mid^2}\right) \cdot W_m,\ m = 1, 2, \ldots, n.$$

That is,

$$z^* = \mathcal{H}_{nr}\left(\mathbf{1}, y/\mid W \mid^2\right) \odot W, \tag{2.20}$$

where $\mathbf{1} = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$, $/$ and $\odot$ denote the componentwise division and multiplication, respectively.

Algorithm 1 $L_{1/2}$LAD PR: ADMM method for solving (2.4)

Input: Parameters $r = 10^{-2}$, $iter = 300$, $\epsilon_{ADMM} = 10^{-4}$, $\lambda = 10^{-4}$.
A spectral initialization q^0, $z^0 = Aq^0$, $\Lambda_1^0 = 0$, $\Lambda_2^0 = 0$.
Output: The iterative sparse solution q.
for $j = 0 : iter$ **do**

$$x^{j+1} = \frac{1}{r}(I + A^H A)^{-1}(rq^j - A^H \Lambda_1^j + rA^H z^j - \Lambda_2^j).$$

for $d = 1 : p$ **do**

$$q_d^{j+1} = \begin{cases} f_{2\lambda/r,1/2}(u_d), & \mid u_d \mid > \frac{\sqrt[3]{54}}{4}(\frac{2\lambda}{r})^{2/3}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.21}$$

with $u_d = \left(x^{j+1} + \frac{\Lambda_2^j}{r}\right)_d$.

end for

$$z^{j+1} = \mathcal{H}_{nr}\left(\mathbf{1}, y/\mid W \mid^2\right) \odot W, \tag{2.22}$$

with $W = \left(Ax^{j+1} + \frac{\Lambda_1^j}{r}\right)$.

Update multipliers

$$\begin{aligned} \Lambda_1^{j+1} &= \Lambda_1^j + r(Ax^{j+1} - z^{j+1}), \\ \Lambda_2^{j+1} &= \Lambda_2^j + r(x^{j+1} - q^{j+1}). \end{aligned} \tag{2.23}$$

if $\|x^{j+1} - x^j\| < \epsilon_{ADMM}$ **then**
break;
end if
end for

C. Convergence Analysis

For the convenience of analysis, we introduce the following definition. Considering q-sub-problem

$$\min_{q \in \mathbb{C}^p} \lambda \|q\|_{1/2}^{1/2} + \frac{r}{2}\|q - u\|_2^2, \tag{2.24}$$

where $u = x + \frac{\Lambda_2}{r}$. According to [26], the first-order stationary point definition of (2.24) is given as follow.

Definition 2.3. Let $\hat{q}$ be a vector in $\mathbb{C}^p$ and $Q = Diag(\hat{q})$. $\hat{q}$ is a first-order stationary point of (2.24) if

$$rQ(\hat{q} - u) + \frac{\lambda}{2} \mid \hat{q} \mid^{1/2} = 0,$$

where $Diag(\hat{q})$ denotes a $p \times p$ diagonal matrix whose diagonal is formed by the vector $\hat{q}$, $\mid \hat{q} \mid^{1/2}$ denotes a p-dimensional vector whose dth component is $\mid \hat{q}_d \mid^{1/2}$.

The convergence proof is inspired by [10]. We show that our proposed algorithm converges to a saddle point by satisfying Karush-Kuhn-Tucker(KKT) conditions, which is a typical method for nonconvex problems. The KKT conditions of the Lagrangian $\mathcal{L}_r(x, q, z; \Lambda_1, \Lambda_2)$ in (2.4) are defined as follows:

$$\begin{cases} \partial_x \mathcal{L}_r(\tilde{x}, \tilde{q}, \tilde{z}; \tilde{\Lambda}_1, \tilde{\Lambda}_2) = 0, \\ r\tilde{Q}(\tilde{q} - \tilde{x} - \frac{\tilde{\Lambda}_2}{r}) + \frac{\lambda}{2} \mid \tilde{q} \mid^{1/2} = 0, \\ \partial_z \mathcal{L}_r(\tilde{x}, \tilde{q}, \tilde{z}; \tilde{\Lambda}_1, \tilde{\Lambda}_2) = 0, \\ \partial_{\Lambda_1} \mathcal{L}_r(\tilde{x}, \tilde{q}, \tilde{z}; \tilde{\Lambda}_1, \tilde{\Lambda}_2) = 0, \\ \partial_{\Lambda_2} \mathcal{L}_r(\tilde{x}, \tilde{q}, \tilde{z}; \tilde{\Lambda}_1, \tilde{\Lambda}_2) = 0, \end{cases} \tag{2.25}$$

where $(\tilde{x}, \tilde{q}, \tilde{z}; \tilde{\Lambda}_1, \tilde{\Lambda}_2)$ is a saddle point, ∂ represents the partial derivative, $\tilde{Q} = Diag(\tilde{q})$, $\tilde{q}$ is a first-order stationary point of q-subproblem (2.15).

Remark 2.4. Specially, for complex variable z, the first-order optimality condition associated with the z-subproblem (2.18) is

$$0 \in \tfrac{1}{n} \partial_z \| \mid z \mid^2 - y \|_1 - \Lambda_1 - r(Ax - z).$$

By reference [39](Lemma 4.1), for variable $\bar{z} \in \mathbb{C}$,

$$\partial \mid \bar{z} \mid = \begin{cases} \frac{\bar{z}}{|\bar{z}|}, & \text{If } \bar{z} \neq 0, \\ \{e | e \in \mathbb{C}, \mid e \mid \leq 1\}, & \text{otherwise.} \end{cases}$$

Combining the two equations above, we obtain the first-order optimality condition of z-subproblem

$$0 \in \left(\frac{2}{nr} sign\left(\mid z_m \mid^2 - y_m \right) + 1 \right) z_m - \left(\frac{\Lambda_1}{r} + Ax \right)_m, \; m = 1, 2, \ldots, n,$$

where, for any $\bar{x} \in \mathbb{R}$,

$$sign(\bar{x}) = \begin{cases} 1, & \bar{x} > 0, \\ -1, & \bar{x} < 0, \\ 0, & \bar{x} = 0. \end{cases}$$

Now, we detail the KKT conditions corresponding to the three variables x, q, z :

$$A^H \tilde{\Lambda}_1 + \tilde{\Lambda}_2 = 0, \tag{2.26}$$

$$\tfrac{\lambda}{2} \mid \tilde{q} \mid^{1/2} - \tilde{Q}\tilde{\Lambda}_2 = 0, \tag{2.27}$$

$$\left(\frac{2}{nr} sign\left(\mid \tilde{z}_m \mid^2 - y_m \right) + 1 \right) \tilde{z}_m - \left(\frac{\tilde{\Lambda}_1}{r} + A\tilde{x} \right)_m \ni 0, \; m = 1, 2, \ldots, n, \tag{2.28}$$

$$\tilde{z} = A\tilde{x}, \tag{2.29}$$

$$\tilde{x} = \tilde{q}. \tag{2.30}$$

Theorem 2.5. *Assuming that the successive differences of the two multipliers* $\{\Lambda_1^j - \Lambda_1^{j-1}\}$, $\{\Lambda_2^j - \Lambda_2^{j-1}\}$ *converge to zero and* $\{x^j\}$ *is bounded, then there exists a subsequence of iterative sequence of Algorithm 1 converging to an accumulation point that satisfies the KKT conditions of the saddle point problem* (2.4).

Proof. We complete the proof in two steps. First, we show the boundedness of all variables. Because of the update of two multipliers (2.23) and the assumption that their successive differences converge, we can derive that

$$\lim_{j\to\infty} Ax^j - z^j = 0, \ \lim_{j\to\infty} x^j - q^j = 0, \tag{2.31}$$

with the assumption that $\{x^j\}$ is bounded, which implies the boundedness of $\{z^j\}$ and $\{q^j\}$. By the iterative expressions (2.21) and (2.22), which demonstrate the $\{\Lambda_1^j\}$ and $\{\Lambda_2^j\}$ is bounded, respectively.

The boundedness of all variables guarantees that there is a subsequence $\{(x^{j_t}, q^{j_t}, z^{j_t}; \Lambda_1^{j_t}, \Lambda_2^{j_t})\} \subset \{(x^j, q^j, z^j; \Lambda_1^j, \Lambda_2^j)\}$ and $(\tilde{x}, \tilde{q}, \tilde{z}; \tilde{\Lambda}_1, \tilde{\Lambda}_2)$, such that

$$\lim_{j_t\to\infty} (x^{j_t}, q^{j_t}, z^{j_t}; \Lambda_1^{j_t}, \Lambda_2^{j_t}) = (\tilde{x}, \tilde{q}, \tilde{z}; \tilde{\Lambda}_1, \tilde{\Lambda}_2). \tag{2.32}$$

We then prove that the accumulation point $(\tilde{x}, \tilde{q}, \tilde{z}; \tilde{\Lambda}_1, \tilde{\Lambda}_2)$ satisfies the KKT conditions. It follows from (2.31) that the KKT conditions w.r.t. z and q, i.e., (2.29) and (2.30) are satisfied. Since A is a linear operator in a finite-dimensional space, (2.26) is satisfied. We can verify that the KKT condition (2.28) of z is equivalent to (2.22) when $z = \tilde{z}$ and $W = A\tilde{x} + \frac{\tilde{\Lambda}_1}{r}$. Finally, we can obtain that

$$\tilde{q}_d = \begin{cases} f_{2\lambda/r,\ 1/2}(\tilde{u}_d), & |\ \tilde{u}_d\ | > \frac{\sqrt[3]{54}}{4}(\frac{2\lambda}{r})^{2/3}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.33}$$

where $\tilde{u}_d = \left(\tilde{x} + \frac{\tilde{\Lambda}_2}{r}\right)_d$, $d = 1, 2, \ldots, p$, which is just the solution of (2.27) by Theorem 1 in [41]. Hence the proof is completed. □

Although the assumption for the convergence of successive difference of multipliers in Theorem 2.5 seems strong, we observe through our numerical results that the proposed algorithm is always convergent.

3 Numerical Experiments

All simulations were performed on a 64-bit laptop computer running Windows 7 system with an AMD A8-6410 APU and 4GB of RAM.

A. Experimental parameters and Initialization

We take $p = 128$ in all experiments, generating the true signal as Gaussian random sparse vector. The measurements matrix A satisfies $a_i \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}/2) + i\mathcal{N}(\mathbf{0}, \mathbf{I}/2)$. The choice of parameters in our model is quite easy. The regularization parameter $\lambda > 0$ controls the level of sparsity in the reconstructed signal, by manual tuning, λ is given by fixed 10^{-4}. The penalty parameter r impacts the convergence rate, we take 10^{-2}.

For nonconvex problems, ADMM can converge to different (and in particular, nonoptimal) points, depending on the initial values and the penalty parameter [3]. We take Wirtinger flow [7] initial q^0, which obeys $dist(q^0, x) \le \frac{1}{8}\|x\|$, where $dist(q, x) = \min_{\phi\in[0,2\pi]} \|q - e^{i\phi}x\|_2$. Hence the algorithm proposed converges from the neighborhood of the global minimizer.

Since phase retrieval is nonconvex, like many phase retrieval methods, we also employ multiple initializations to realize a better recovery. To control the tradeoff between the successful recovery and computation complexity, we take at most 10 initializations in the following trials.

Table 1: Comparison of reconstruction methods

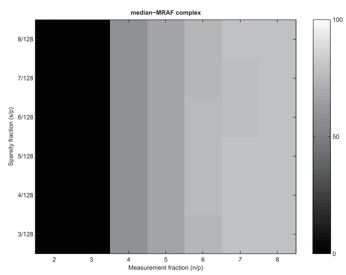
Method	Implementation	Measurements matrix	Robustness	Sparse solution
L0L1PR [16]	ADMM	Fourier related	noise	✓
LAD-ADMM [40]	ADMM	Gaussian	Gaussian mixture noises	
Median-RWF [45]	gradient descent	Gaussian	noise outliers	
Median-MRAF [47]	gradient descent	Gaussian	noise outliers	
$L_{1/2}$LAD PR	ADMM	Gaussian	noise outliers	✓

B. Monte Carlo Comparisons (1D)

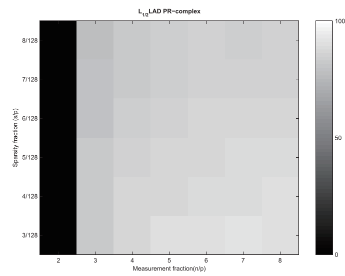
In this section, we report numerical simulation results to demonstrate how the peak-signal-to-error ratio (PSER) which is used in [38] depends on the measurements ratio $n/p = 2:6$ and the sparsity $s = 3:8$ via 50-trial Monte Carlo simulations. Define PSER$= -10log_{10}$(median squared error), where median squared error is the median of the squared errors relative to the true signal over the set of trials. In our phase retrieval algorithms study, we find some Fourier phase retrieval methods can't solve the Gaussian phase retrieval problem directly, like [38], and vice-versa, such as [45]. Since L0L1PR [16] listed in Table 1 applies to the Fourier related phase retrieval in complex case and the running time of LAD-ADMM [40] is too long. Moreover, median-MRAF [47] is the development of median-RWF [45], so we only compare with the latest method median-MRAF in real case and complex case in Figure 1, respectively.

For the convenience of comparison, we limit the range of PSER value in [0,100] as follow. As is shown in Figure 1, the darker the color, which means PSER value is smaller and the error is larger; the lighter the color, the recovery is more successful. It is observed that the required number of measurements and the median squared errors of $L_{1/2}$LAD PR are significantly better than median-MRAF in both cases. It is known that in real case $x \in \mathbb{R}^p$, we need at least $n \geq 2p-1$ measurements to have the phase retrieval property [2], in complex case $x \in \mathbb{C}^p$, the same question result is $n \geq 4p-4$ in [13], so the proposed $L_{1/2}$LAD PR algorithm accords with the theoretical sample complexity.

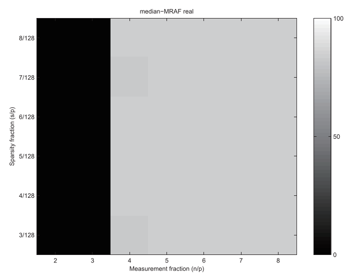
$$\text{PSER} = \begin{cases} 0, & \text{PSER} < 0, \\ 100, & \text{PSER} > 100, \\ \text{PSER}, & \text{otherwise.} \end{cases}$$



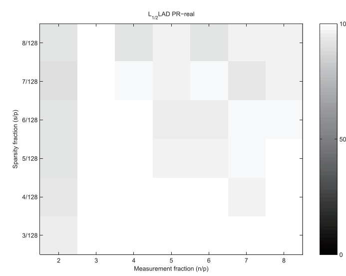
(a) Median-MRAF complex



(b) $L_{1/2}$LAD PR complex



(c) Median-MRAF real



(d) $L_{1/2}$LAD PR real

Figure 1: The PSER of 50 trials reconstructed using Median-MRAF, $L_{1/2}$LAD PR for a range of measurement (n/p) and sparsity fractions (s/p) in the complex case and real case.

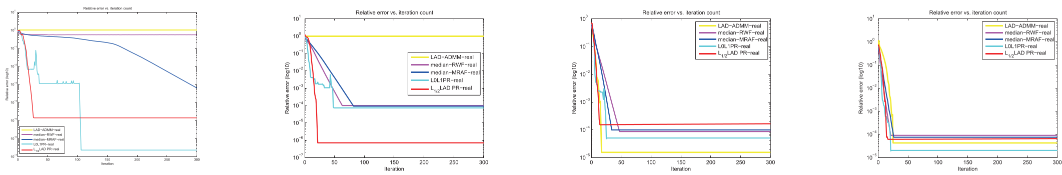
In the following sections, we fix the sparsity as $s = 8$ and reconstruct the signal corrupted by noise or outliers using $L_{1/2}$LAD PR and competing algorithms listed in Table 1. Both LAD-ADMM and L0L1PR use L_1 norm loss function, and L0L1PR adds a L_0 regularization. Median-RWF and Median-MRAF which use L_2 norm loss function are highly robust to outliers by heuristic truncated rules. We compare the relative errors with respect to the iteration count t at different measurement fractions (n/p), where relative error is $\min_{|c|=1} \frac{\|cx-x^{(t)}\|_2}{\|x\|_2}$, x is the true signal, $x^{(t)}$ is the tth iteration point.

Remark 3.1. We take $n = 2p,\ 3p,\ 4p,\ 6p$ in real case and $n =\ 4p,\ 5p,\ 6p,\ 8p$ in complex case. In details, $n = 2p$ and $n = 4p$ is the approximate theoretical sample complexity. $n = 8p$ is the number of measurements used for LAD-ADMM in [40], we take $n = 6p$ in real case when LAD-ADMM has a stable recovery.

C. Exact Recovery for Noise-Free Data

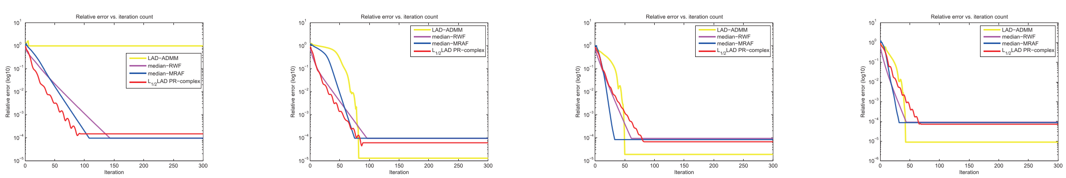
In the noise-free case, Figure 2 shows in real case, when $n = 2p$, only L0L1PR and $L_{1/2}$LAD PR have good recovery performance. When $n = 3p,\ 4p,\ 6p$, $L_{1/2}$LAD PR is slightly better than other 4 algorithms. In Figure 3, it shows in complex case, when $n = 4p$, Median-RWF, Median-MRAF and $L_{1/2}$LAD PR can recover the signal. When $n = 5p,\ 6p,\ 8p$, $L_{1/2}$LAD

PR has similar performance as compared methods, and LAD-ADMM is slightly better than other 3 algorithms.



(a) n=2p (b) n=3p (c) n=4p (d) n=6p

Figure 2: The relative error with respect to the iteration count for LAD-ADMM, median-RWF, median-MRAF, L0L1PR, $L_{1/2}$LAD PR with noise-free data in real case.

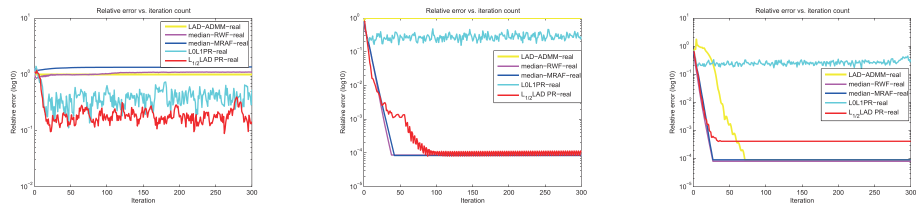


(a) n=4p (b) n=5p (c) n=6p (d) n=8p

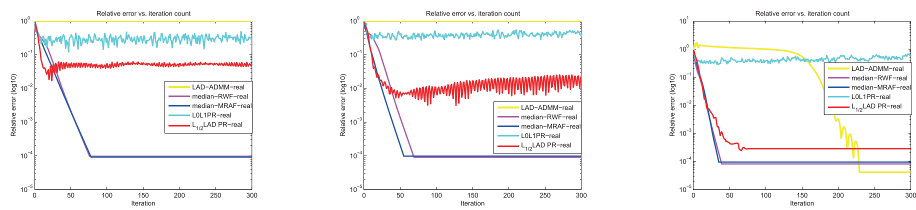
Figure 3: The relative error with respect to the iteration count for LAD-ADMM, median-RWF, median-MRAF, $L_{1/2}$LAD PR with noise-free data in complex case.

D. Stable Recovery With Sparse Outliers

We next examine the performance of $L_{1/2}$LAD PR in presence of sparse outliers. The outlier value is generated from a uniform distribution $\mathcal{U}(0, \omega_{max})$, where $\omega_{max} = 0.1 y_{max}$, y_{max} is the largest measurement in y. The entries of the sparse outliers are generated from Bernoulli(0.1) or Bernoulli(0.2). It can be seen from Figure 4, L0L1PR is not robust to outliers in real case. Combining with Figure 5, When $n \geq 3p$, Median-RWF and Median-MRAF can realize exact recovery with sparse outliers both in real and complex cases. When $n \geq 4p$ in real case and $n \geq 5p$ in complex case, the relative error of $L_{1/2}$LAD PR can decrease to about 5×10^{-4}, which shows that $L_{1/2}$LAD PR has certain robustness to outliers. When $n \geq 5p$, LAD-ADMM has a satisfactory recovery. We also find that the outliers ratio(0.1 or 0.2) has little effect on the experiments.

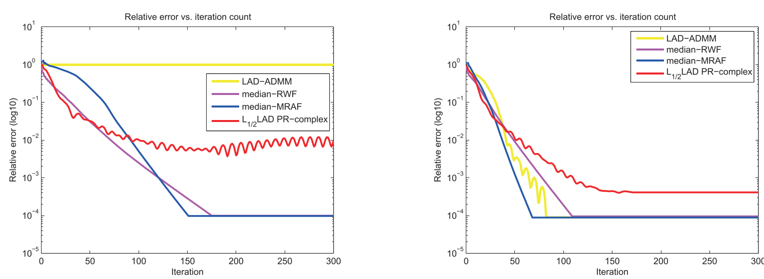


(a) n=2p; $\delta = 0.1$; (b) n=4p; $\delta = 0.1$; (c) n=6p; $\delta = 0.1$;

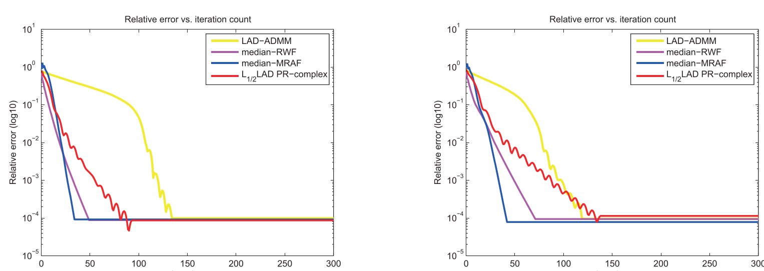


(d) n=3p; $\delta = 0.1$; (e) n=4p; $\delta = 0.2$; (f) n=6p; $\delta = 0.2$;

Figure 4: The relative error with respect to the iteration count for LAD-ADMM, median-RWF, median-MRAF, L0L1PR, $L_{1/2}$LAD PR with sparse outliers in real case. ($\omega = \frac{\omega_{max}}{y_{max}} = 0.1$, outliers ratio $\delta = 0.1,\ 0.2.$)



(a) n=4p; $\delta = 0.1$; (b) n=5p; $\delta = 0.1$;

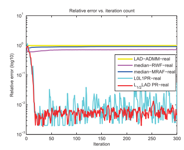


(c) n=8p; $\delta = 0.1$; (d) n=8p; $\delta = 0.2$;

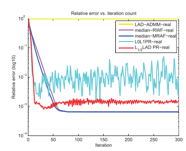
Figure 5: The relative error with respect to the iteration count for LAD-ADMM, median-RWF, median-MRAF, $L_{1/2}$LAD PR with sparse outliers in complex case. ($\omega = \frac{\omega_{max}}{y_{max}} = 0.1$, outliers ratio $\delta = 0.1,\ 0.2.$)

E. Stable Recovery With Dense Bounded Noise

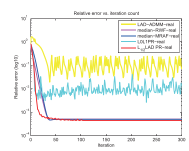
Now, we consider the existence of dense bounded noise. The entries of the dense bounded noise are generated independently from $\mathcal{U}(0, \eta_{max})$, where $\eta_{max}/\|x\|^2 = 0.001, 0.01$. It can be seen from Figure 6 and 7, $L_{1/2}$LAD PR shows great robustness to dense bounded noise in all cases, while L0L1PR shows poor performance. Median-RWF and Median-MRAF have similar performance when $n \geq 3p$ both in complex and real cases. LAD-ADMM is also robust with noise when the number of measurements satisfies $n = 6p$ in real case and $n = 8p$ in complex case. Another reasonable observation, we find the relative reconstruction error has 10 times increase as η shrinks by a factor of 10 for all algorithms .



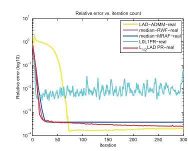
(a) n=2p; $\eta = 0.001$;



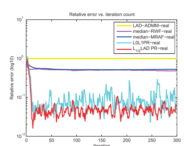
(b) n=3p; $\eta = 0.001$;



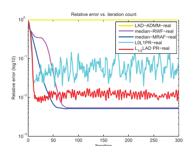
(c) n=4p; $\eta = 0.001$;



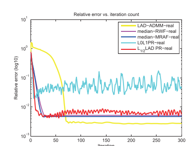
(d) n=6p; $\eta = 0.001$;



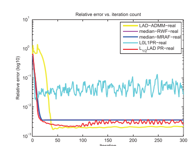
(e) n=2p; $\eta = 0.01$;



(f) n=3p; $\eta = 0.01$;

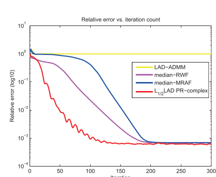


(g) n=4p; $\eta = 0.01$;

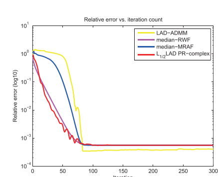


(h) n=6p; $\eta = 0.01$;

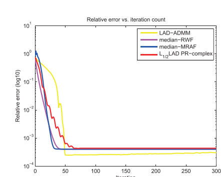
Figure 6: The relative error with respect to the iteration count for LAD-ADMM, median-RWF, median-MRAF, L0L1PR, $L_{1/2}$LAD PR with dense bounded noise in real case. ($\eta = \frac{\eta_{max}}{\|x\|^2} = 0.001, 0.01$.)



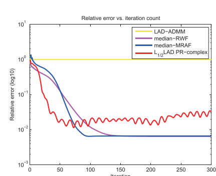
(a) n=4p; $\eta = 0.001$;



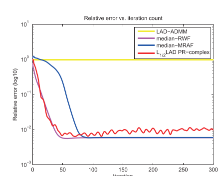
(b) n=5p; $\eta = 0.001$;



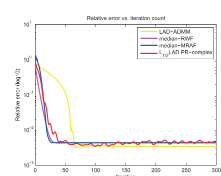
(c) n=8p; $\eta = 0.001$;



(d) n=4p; $\eta = 0.01$;



(e) n=5p; $\eta = 0.01$;

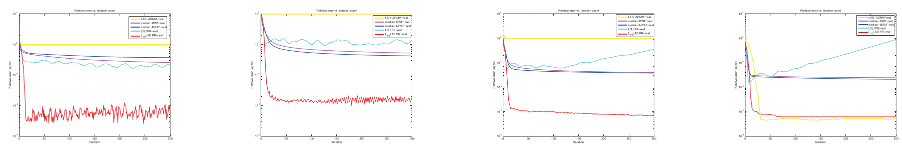


(f) n=8p; $\eta = 0.01$;

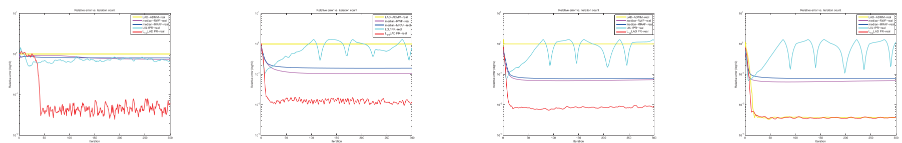
Figure 7: The relative error with respect to the iteration count for LAD-ADMM, median-RWF, median-MRAF, $L_{1/2}$LAD PR with dense bounded noise in complex case. ($\eta = \frac{\eta_{max}}{\|x\|^2} = 0.001, 0.01$.)

F. Stable Recovery With Laplace Noise

Finally, we consider the presence of Laplace noise, the entries of Laplace noise are generated from *Laplace* $(0, \mu_{max}/\sqrt{2})$, where $\frac{\mu_{max}}{\|y\|_2/\sqrt{n}} = 0.001,\ 0.01$. As can be observed in Figure 8 and 9, surprisingly, $L_{1/2}$LAD PR is very robust to Laplace noise, especially in real case, no matter when $n = 2p,\ 3p,\ 4p,\ 6p$. However, other methods show poor performance, even when $n = 6p$, L0L1PR, Median-RWF, Median-MRAF don't have satisfactory recovery. In complex case, $L_{1/2}$LAD PR, Median-RWF, Median-MRAF have similar performance, but the performance of LAD-ADMM is not stable. Another logical observation, we find the relative reconstruction error has 10 times increase as μ_{max} shrinks by a factor of 10 for all algorithms.

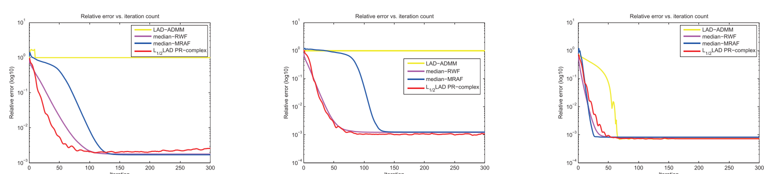


(a) n=2p; $\mu = 0.001$; (b) n=3p; $\mu = 0.001$; (c) n=4p; $\mu = 0.001$; (d) n=6p; $\mu = 0.001$;

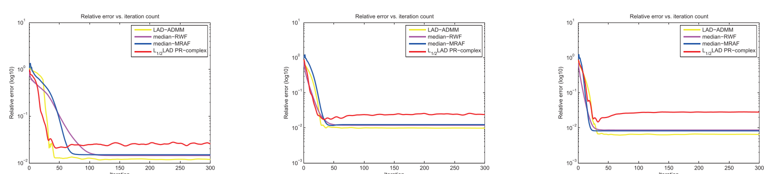


(e) n=2p; $\mu = 0.01$; (f) n=3p; $\mu = 0.01$; (g) n=4p; $\mu = 0.01$; (h) n=6p; $\mu = 0.01$;

Figure 8: The relative error with respect to the iteration count for LAD-ADMM, median-RWF, median-MRAF, L0L1PR, $L_{1/2}$LAD PR with Laplace noise in real case. ($\mu = \frac{\mu_{max}}{\|y\|_2/\sqrt{n}} = 0.001,\ 0.01$.)



(a) n=4p; $\mu = 0.001$; (b) n=5p; $\mu = 0.001$; (c) n=8p; $\mu = 0.001$;



(d) n=4p; $\mu = 0.01$; (e) n=5p; $\mu = 0.01$; (f) n=8p; $\mu = 0.01$;

Figure 9: The relative error with respect to the iteration count for LAD-ADMM, median-RWF, median-MRAF, $L_{1/2}$LAD PR with Laplace noise in complex case. ($\mu = \frac{\mu_{max}}{\|y\|_2/\sqrt{n}} = 0.001,\ 0.01$.)

Our simulation illustrates that, with the robust L_1 norm loss function and $L_{1/2}$ regularization, $L_{1/2}$LAD PR algorithm has two significant advantages over competing methods. One is its ability to recover the signal with less measurements, the other is its robustness to asymmetrical distribution noise, such as dense bounded noise and Laplace noise.

4 Conclusion

We propose and demonstrate $L_{1/2}$LAD PR algorithm for recovering a sparse signal vector from its Gaussian measurements. We show via extensive experiments that $L_{1/2}$LAD PR outperforms the comparative phase retrieval approaches in terms of the number of required measurements and robustness to dense bounded noise, Laplace noise. As the reader may notice, our algorithm is applicable to both the complex case and the real case, which is general. One interesting future study direction is to consider the non-i.i.d. measurement vectors like the Fourier basis measurements which are in the application of the coded diffraction patterns [6].

Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their insight and helpful comments and suggestions which greatly improve the quality of the paper.

References

[1] B. Alexeev, A.S. Bandeira, M. Fickus and D.G. Mixon, Phase retrieval with polarization, *SIAM J. Imaging Sci.* 7 (2012) 35–66.

[2] R. Balan, P. Casazza and D. Edidin, On signal reconstruction without phase, *Appl. Comput. Harmon. Anal.* 20 (2006) 345–356.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the Alternating Direction Method of Multipliers, *Foundations & Trends in Machine Learning*, 3 (2010) 1–122.

[4] T. T. Cai, X. Li and Z. Ma, Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow, *Ann. Statist.* 44 (2016) 2221–2251.

[5] E.J. Candès and Y.C. Eldar, Phase retrieval via matrix completion, *SIAM Rev.* 57 (2015) 225–251.

[6] E.J. Candès, X. Li and M. Soltanolkotabi, Phase retrieval from coded diffraction patterns, *Appl. Comput. Harmon. Anal.* 39 (2015) 277–299.

[7] E.J. Candès, X. Li and M. Soltanolkotabi, Phase retrieval via wirtinger flow: theory and algorithms, *IEEE Trans. Inform. Theory*, 61 (2015) 1985–2007.

[8] E.J. Candès, T. Strohmer and V. Voroninski, Phaselift: exact and stable signal recovery from magnitude measurements via convex programming, *Comm. Pure Appl. Math.* 66 (2013) 1241–1274.

[9] H. Chang, P. Enfedaque and S. Marchesini, Blind ptychographic phase retrieval via convergent alternating direction method of multipliers, *SIAM J. Imaging Sci.* 12 (2019) 153–185.

[10] H. Chang, Y. Lou, Y. Duan and S. Marchesini, Total variation-based phase retrieval for Poisson noise removal, *SIAM J. Imaging Sci.* 11 (2018) 24–55.

[11] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Processing Letters* 14 (2007) 707–710.

[12] Y. Chen and E.J. Candès, Solving random quadratic systems of equations is nearly as easy as solving linear systems, *Comm. Pure Appl. Math.* 70 (2015).

[13] A. Conca, D. Edidin, M. Hering and C. Vinzant, Algebraic characterization of injectivity in phase retrieval, *Appl. Comput. Harmon. Anal.* 38 (2015) 346-356.

[14] J.C. Dainty and J.R. Fienup, Phase retrieval and image reconstruction for astronomy, *Image Recovery Theory & Application*, (1987) 231–275.

[15] O. Dhifallah and Y.M. Lu, Fundamental limits of PhaseMax for phase retrieval: a replica analysis, in: *C. 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE*, 2017.

[16] Y. Duan, C. Wu, Z.F. Pang and H. Chang, L^0-regularized variational methods for sparse phase retrieval, *arXiv:1612.02538*, 2016.

[17] J. Fan, L.C. Kong, L.Q. Wang and N.H. Xiu, Variable selection in sparse regression with quadratic measurements, *Statist. Sinica* 28 (2018) 1157–1178.

[18] J.R. Fienup, Phase retrieval algorithms: a comparison, *Applied Optics.* 21 (1982) 2758–69.

[19] R.W. Gerchberg and W.O. Saxton, A practical algorithm for the determination of phase from image and diffraction plane picture, *Optik* 35 (1972) 237–246.

[20] T. Goldstein and C. Studer, PhaseMax: convex phase retrieval via basis pursuit, *IEEE Trans. Inform. Theory* (2018) 2675–2689.

[21] H.A. Hauptman, The phase problem of X-ray crystallography: overview, *Rep. Progr. Phys.* 54 (1991) 1427–54.

[22] K. Jaganathan, S. Oymak and B. Hassibi, On robust phase retrieval for sparse signals, *C. 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL*, (2012) 794–799.

[23] K. Jaganathan, S. Oymak and B. Hassibi, Phase retrieval for sparse signals using rank minimization, in: *C. Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference*,

[24] K. Jaganathan, S. Oymak and B. Hassibi, Recovery of sparse 1-D signals from the magnitudes of their Fourier transform, in: *C. 2012 IEEE International Symposium on Information Theory Proceedings. IEEE*, 2012.

[25] M.B. Lazreg and R. Amara, A robust sparse wirtinger flow algorithm with optimal stepsize for sparse phase retrieval, in: *C. 2018 15th International Multi-Conference on Systems, Signals & Devices*, 2018.

[26] Z.S. Lu, Iterative reweighted minimization methods for l_p regularized unconstrained nonlinear programming, *Math. Program.* 147 (2014) 277–307.

[27] S. Mukherjee and C.S. Seelamantula, An iterative algorithm for phase retrieval with sparsity constraints: application to frequency domain optical coherence tomography, in: *C. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE*, 2012, pp. 553–556.

[28] B.K. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Comput.* 24 (1995) 227–234.

[29] P. Netrapalli, P. Jain and S.R. Sanghavi, Phase retrieval using alternating minimization, *IEEE Trans. Signal Process.* 63 (2015) 4814–4826.

[30] H. Ohlsson and Y.C. Eldar, On conditions for uniqueness in sparse phase retrieval, in: *C. Proceedings Icassp IEEE International Conference on Acoustics Speech* & *Signal Processing*, 2014.

[31] P. Sarangi and P. Pal, Robust sparse phase retrieval from differential measurements using reweighted L_1 minimization, in: *C. 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), Sheffield*, 2018, pp. 223–227.

[32] Y. Shechtman, A. Beck and Y.C. Eldar, GESPAR: efficient phase retrieval of sparse signals, *IEEE Trans. Signal Process.* 62 (2014) 928–938.

[33] Y. Shechtman, Y.C. Eldar, A. Szameit and M. Segev, Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing, *Optics Express* 19 (2011) 14807–14822.

[34] M. Soltanolkotabi, Structured signal recovery from quadratic measurements: breaking sample complexity barriers via nonconvex optimization, *IEEE Trans. Inform. Theory* 65 (2019) 2374–2400.

[35] J. Sun, Q. Qu and J. Wright, A geometric analysis of phase retrieval, in: *C. 2016 IEEE International Symposium on Information Theory (ISIT). IEEE*, 2016.

[36] I. Waldspurger, A.D. Aspremont and S. Mallat, Phase recovery, MaxCut and complex semidefinite programming, *Math. Program.* 149 (2019) 47–81.

[37] A. Walther, The question of phase retrieval in optics, *Optica Acta* 10 (1963), 41–9.

[38] D.S. Weller, A. Pnueli, G. Divon, O. Radzyner, Y.C. Eldar and J.A. Fessler, Undersampled phase retrieval with outliers, *IEEE Trans. Comput. Imaging* 1 (2015) 247–258.

[39] Z. Wen, C. Yang, X. Liu and S. Marchesini, Alternating direction methods for classical and ptychographic phase retrieval, *Inverse Problems* 11 (2012) 115010.

[40] J. Xue, H.C. So and X. Liu, Robust phase retrieval with outliers, in: *C. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE*, 2020.

[41] Z.B. Xu and X.Y. Chang, $L_{1/2}$ regularization: a thresholding representation theory and a fast solver, *IEEE Trans. Neural Netw. Learn. Syst.* 23 (2012) 1013–1027.

[42] Z B. Xu, H.L. Guo, Y. Wang and H. Zhang, Representative of $L_{1/2}$ regularization among $L_q(0 < q \leq 1)$ regularizations: an experimental study based on phase diagram, *Automatica Sinica* 38 (2012) 1225–1228.

[43] Z. Yuan, H. Wang and Q. Wang, Phase retrieval via sparse wirtinger Flow, *J. Comput. Appl. Math.* 355 (2019) 162–173.

[44] J.S. Zeng, J. Fang and Z.B. Xu, Sparse SAR imaging based on $L_{1/2}$ regularization, *Science China Information Sciences* 55 (2012) 1755–1775.

[45] H. Zhang, Y. Chi and Y. Liang, Median-Truncated nonconvex approach for phase retrieval with outliers, *IEEE Trans. Inform. Theory* 64 (2018) 7287–7310.

[46] 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.* 18 (2017) 1–35.

[47] Q. Zhang, D. Liu, F. Hu, A. Li and H. Cheng, Median momentum reweighted amplitude flow for phase retrieval with arbitrary corruption, *J. Modern Opt.* 3 (2021) 1–8.

Manuscript received 29 June 2021
revised 14 September 2021
accepted for publication 12 October 2021

Lijiao Kong
School of Science, Hebei University of Technology
Tianjin, China
E-mail address: lijiaokonghebut@163.com

Ailing Yan
School of Science, Hebei University of Technology
Tianjin, China
E-mail address: ailing-yan78@163.com

Yan Li
School of Insurance and Economics
University of International Business and Economics
Beijing, 100029, China
E-mail address: liyan2010@uibe.edu.cn

Jun Fan
School of Science, Hebei University of Technology
Tianjin, China
E-mail address: fanjunmath@hotmail.com