INEXACT ALTERNATING OPTIMIZATION FOR PHASE RETRIEVAL WITH OUTLIERS

Cheng Qian† Xiao Fu‡ Nicholas D. Sidiropoulos§ Lei Huang*

†Dept. of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China
‡Dept. of ECE, University of Minnesota, Minneapolis, MN, USA
§College of Information Engineering, Shenzhen University, Shenzhen, China

ABSTRACT

Most of the available phase retrieval algorithms were explicitly or implicitly developed under a Gaussian noise model, using least squares (LS) formulations. However, in some applications of phase retrieval, an unknown subset of the measurements can be seriously corrupted by outliers, where LS is not robust and will degrade the estimation performance severely. This paper presents an Alternating Iterative Reweighted Least Squares (AIRLS) method for phase retrieval in the presence of such outliers. The AIRLS employs two-block alternating optimization to retrieve the signal through solving an \( \ell_p \)-norm minimization problem, where \( 0 < p < 2 \). The Cramér-Rao bound (CRB) for Laplacian as well as Gaussian noise is derived for the measurement model considered, and simulations show that the proposed approach outperforms state-of-the-art algorithms in heavy-tailed noise.

Index Terms — Phase retrieval, iterative reweighted least squares, Cramér-Rao bound (CRB).

1. INTRODUCTION

The problem of recovering a signal from the magnitude of linear measurements, referred to as phase retrieval, appears in many areas including optical imaging, X-ray crystallography, coherent diffraction imaging, and astronomy, where the detector only records intensity information.

The most well-known algorithms for phase retrieval are Gerchberg-Saxton (GS) [1] and Fienup hybrid input-output (HIO) [2], which are alternating projection methods. These can work well in practice, despite lacking of theoretical performance guarantees. Over the past five years, modern optimization techniques have been suggested to solve this difficult problem, e.g., semidefinite relaxation (SDR) approaches such as PhaseLift [3]-[4] and PhaseCut [5], and a gradient descent type method named Wirtinger-Flow (WF) [6]. It has been proved that exact recovery is possible with these latter methods, with high probability, in the noiseless setting with random Gaussian measurements. In our recent work, we proposed a least-squares feasible point pursuit (LS-FPP) approach [7] for solving the same optimization problem as LS PhaseLift, but instead of relaxing it using SDR, we employ the FPP approach proposed in [8] for general nonconvex quadratically constrained problems.

Most of the available phase retrieval algorithms like PhaseLift, WF and LS-FPP, were explicitly or implicitly developed under a Gaussian noise model, building upon LS formulations. However, in some practical applications of phase retrieval, an unknown subset of the measurements may be corrupted by outliers [9]-[13]. One representative example is high energy coherent X-ray imaging using a charge-coupled device (CCD) where the impulsive noise originated from X-ray radiation might effect on the CCD [13]. Under such a circumstance, LS criterion is no longer robust, and may result in severe performance degradation. Several robust phase retrieval algorithms, e.g., [9]-[12], have been carried out to handle heavy-tailed noise but with assumption that the signal is sparse. Thus, to implement them to retrieve dense signals seems not so straightforward. This promotes us to devise robust algorithms for dense signals.

In this paper, we focus on designing a robust algorithm that is resistant to outliers and nearly optimal for Gaussian-type noise. We develop a two-block inexact alternating optimization based algorithm, which is effective in dealing with heavy-tailed noise by employing the \( \ell_p \)-fitting criterion with \( 0 < p \leq 2 \). In order to assist in performance comparison, we also derive the Cramér-Rao bound (CRB) for the considered phase retrieval model with unknown parameters being defined as the phase and amplitude of the input signal, which, to the best of our knowledge, has not yet been addressed in the literature, such as [7], [14]-[18].

2. PROPOSED ALGORITHM

Consider a phase retrieval model with the form of

\[
y = |Ax| + n \in \mathbb{R}^M
\]  

(1)
where $| \cdot |$ denotes the absolute value operator, $A^H = [a_1 \cdots a_M] \in \mathbb{C}^{N \times N}$ is the measurement matrix with $H$ being the conjugate transpose, and $n$ is the noise vector which in this paper, is assumed to be heavy-tailed. Our aim is to recover $x$ from $y$. Without noise, by introducing an auxiliary vector $u = e^{j\mathcal{L}(Ax)}$ where $\mathcal{L}$ takes the phase of its argument, (1) can also be written as [5][26]

$$Y_u = Ax$$

(2)

where $Y = \text{diag}(y)$ is a diagonal matrix. In the presence of impulsive noise, it is more appealing to employ the following criterion for phase recovery:

$$\min_{|u|_1,x} \sum_{m=1}^M (|y_m u_m - a_m^H x|^2 + \epsilon)^{p/2}$$

(3)

where $|| \cdot ||_2$ is the 2-norm, $0 < p < 2$ is chosen to down-weigh outliers, and $\epsilon > 0$ is a small regularization parameter that keeps the cost function within its differentiable domain when $p < 1$, which will prove handy in devising an effective algorithm later. Particularly, when $1 < p < 2$, we may simply set $\epsilon = 0$.

We follow the rationale of alternating optimization to deal with Problem (3), i.e., we first fix $u$ and update $x$, and then we do the same to $u$.

Assume that the current solution at iteration $r$ is $(x^{(r)}, u^{(r)})$. At step $(r + 1)$, the subproblem w.r.t. $x$ becomes

$$x^{(r+1)} = \arg \min_x \sum_{m=1}^M (|y_m u_m^{(r)} - a_m^H x|^2 + \epsilon)^{p/2}$$

(4)

which is very difficult to handle. Particularly, when $p < 1$, the subproblem itself is non-convex. To circumvent this, we propose to employ the following lemma [19]:

**Lemma 1** Assume $0 < p < 2$, $\epsilon \geq 0$, and $\phi_p(w) := \frac{2-2^p}{2^p} \left(\frac{2}{p}\right)^{\frac{p}{p-2}} |w|^2 + \epsilon w$. Then, we have

$$(x^2 + \epsilon)^{p/2} = \min_{w \geq 0} w(x^2 + \phi_p(w)),$$

(5)

and the unique minimizer is

$$w_{opt} = \frac{p}{2} \left(x^2 + \epsilon\right)^{\frac{p-2}{2}}.$$

(6)

By Lemma 1, an upper bound of $\sum_{m=1}^M (|y_m u_m^{(r)} - a_m^H x|^2 + \epsilon)^{p/2}$ that is tight at the current solution $x^{(r)}$ can be easily found:

$$\leq \sum_{m=1}^M \left(\frac{p}{2} \left(y_m^2 u_m^{(r)} - a_m^H x^{(r)}\right)^2 + \phi_p\left(u_m^{(r)}\right)\right)^{p/2}$$

(7)

where

$$w_m^{(r)} := \frac{p}{2} \left(\left|y_m u_m^{(r)} - a_m^H x^{(r)}\right|^2 + \epsilon\right)^{\frac{p-2}{2}}.$$

(8)

Instead of directly dealing with Problem (4), we solve a surrogate problem using the right hand side (RHS) of (7) at each iteration to update $x$. Notice that the RHS of (7) is convex w.r.t. $x$ and can be solved in closed-form:

$$x^{(r+1)} = (W^{(r)}A)^T W^{(r)} Y_u^{(r)}.$$

(9)

where

$$W^{(r)} = \text{diag} \left(\sqrt{w^{(r)}_1}, \ldots, \sqrt{w^{(r)}_M}\right).$$

(10)

We note that the update of $u$ is followed by the solution of $x^{(r+1)}$ rather than $x^{(r)}$. Thus, the conditional problem w.r.t. $u$ is

$$u^{(r+1)} = \arg \min_{|u|_1} \sum_{m=1}^M (|y_m u_m^{(r)} - a_m^H x^{(r+1)}|^2 + \epsilon)^{p/2}.$$

(11)

Although the problem is non-convex, it can be easily solved to optimality. It is observed from (11) that given a fixed $x$, for any $p > 0$, the solutions w.r.t. $u$ are identical, i.e., simply aligning the angle of $y_m u_m$ to that of $a_m^H x^{(r+1)}$, which is exactly

$$u_m^{(r+1)} = e^{j\mathcal{L}(a_m^H x^{(r+1)})}, \quad m = 1, \ldots, M.$$

(12)

The explicit steps of the proposed method are summarized in Algorithm 1:

**Algorithm 1** AIRLS for phase retrieval

1: function AIRLS$(y, A, x^{(0)})$
2: Initialize $u^{(0)} = \exp(j\mathcal{L}(Ax^{(0)}))$ and $W^{(0)} = \sqrt{\frac{p}{2}} \text{diag}(|yu^{(0)} - Ax^{(0)}|^2 + \epsilon 1_M)^{\frac{p-2}{2}}$, where $1_M$ denotes a one vector of length $M$.
3: while Until some stopping criterion is reached do
4: $x^{(r)} = (W^{(r-1)}A)^T W^{(r-1)} Y_u^{(r-1)}$
5: $u^{(r)} = \exp(j\mathcal{L}(Ax^{(r)}))$
6: $W^{(r)} = \sqrt{\frac{p}{2}} \text{diag}(|Yu^{(r)} - Ax^{(r)}|^2 + \epsilon 1_M)^{\frac{p-2}{2}}$
7: end while
8: end function

Remark: The computational burden of AIRLS mainly lies in the calculation of $x$ at each iteration, which has complexity $O(MN^2)$. Thus, if the algorithm converges after, say, $r$th iterations, the total complexity is $O(rMN^2)$. Compared to semidefinite relaxation based methods which have worst-case complexity of $O(N^{6.5})$, our approach is much, much simpler.
3. CRAMÉR-RAO BOUND

It is well-known that the CRB provides the lowest achievable bound of the variance of any unbiased estimators, and it usually serves as a benchmark for mean squares error (MSE) performance comparison. Over the past few years, several CRBs have been derived for different phase retrieval models, e.g., two dimensional Fourier-based measurements [14], noise added prior to taking the magnitude [15] and quadratic model [7], [16]-[18]. However, there is no such a CRB been studied for the model in (1). Unlike those existing lower bounds derived with unknown parameters composed by real and imaginary parts of $x$ [7], [15]-[18], we provide a closed-form CRB formula corresponding to the phase and amplitude of $x$. The following theorem presents the CRB for a particular type of heavy-tailed noise: Laplacian noise.

**Theorem 1** In Laplacian noise, the variance of any unbiased estimate corresponding to the amplitude of complex-valued $x$ is bounded below by

$$\text{CRB}_{\text{Laplacian, } |x|} = \sum_{i=1}^{N} d_i$$

and the variance of any unbiased estimate corresponding to the phase of complex-valued $x$ is bounded below by

$$\text{CRB}_{\text{Laplacian, } \angle x} = \sum_{i=N+1}^{2N} d_i$$

where $d = [d_1 \cdots d_{2N}]$ contains the main diagonal elements of $F^T$ which in this case is defined as

$$F = \frac{4}{\sigma_n^2} G \text{diag}(|Ax|)^{-2} G^T$$

with

$$G = \begin{bmatrix} \text{diag}(|x|)^{-1} & I_N \end{bmatrix} \begin{bmatrix} \text{Re} \{ \text{diag} (x^*) A^H \text{diag} (Ax) \} \\ \text{Im} \{ \text{diag} (x^*) A^H \text{diag} (Ax) \} \end{bmatrix}.$$  

Here, $(\cdot)^T$, $(\cdot)^{-1}$, $(\cdot)^*$ represent the pseudo-inverse, inverse, transpose and conjugate operators, respectively. When $A$ is nontrivial, we find that $F$ is always singular with rank $(2N-1)$. There, according to [20]-[23], instead of using inverse, we adopt the pseudo-inverse to compute a looser CRB, which is also a valid bound that can be used as a benchmark to examine the efficiency of any unbiased estimators. Besides the Laplacian case, one may also cares about the CRB for Gaussian noise. The CRB for Gaussian noise can be directly computed using the following theorem:

**Theorem 2** In Gaussian noise, the CRB is four times larger than the CRB in Theorem 1, i.e.,

$$\text{CRB}_{\text{Gaussian}} = 4 \text{CRB}_{\text{Laplacian}} = 4 \text{trace} \left( F^T \right)$$

With Theorem 2, it is easy to compute the CRB for Gaussian noise. Note that similar results can also be found in [24].

4. SIMULATION RESULTS

In this section, we illustrate numerically the performance of the proposed method and compare it with GS [1], WF [6], PhaseCut [5], truncated WF (TWF) [25] and AltMinPhase [26] in terms of MSE performance, where the MSE is computed after removing the global phase ambiguity between the true and estimated $x$. The signal $x$ with length $N = 64$ is generated from a complex Gaussian distribution with mean zero and covariance matrix $I_N$. The number of measurements is $M = 10N$. All results are conducted in a computer with 3.6 GHz i7-4790 CPU and 8 GB RAM and averaged from 200 Monte-Carlo tests. In the following, we consider two types of heavy-tailed noise models: Laplacian and Gaussian mixture noise. For the former, we also include the CRB in Theorem 1 as a performance benchmark.

It is seen in Fig. 1 that our method slightly outperforms...
possesses heavy tails. In the following, we set $\alpha = 0.8$ and $\gamma = 2$. As we can see, the performance gap between AIRLS and its competitors becomes even larger in $\alpha\sigma S$ noise, where AIRLS actually gains at least 10 dB improvement to the TWF method when SNR $> 20$ dB, indicating that the proposed approach is more robust than the others in the face of heavy-tailed outliers. When SNR $< 15$ dB, there is no MSE values for WF. That is because WF frequently produces “not a number” in the $\alpha\sigma S$ noise scenario.

Fig. 2 depicts the MSE performance versus SNR in $\alpha\sigma S$ noise, where The PDF of a symmetric $\alpha$-stable ($\alpha\sigma S$) distribution is usually not available, but its characteristic function can be written in an analytically closed-form expression as

$$\phi(t; \alpha, \gamma) = \exp(-\gamma^\alpha |t|^\alpha)$$

where $\Phi(\alpha) = \tan(\alpha \pi/2)$, $0 < \alpha \leq 2$ is the stability parameter that controls the density of impulses (e.g., smaller $\alpha$ implies sparser impulses), and $\gamma > 0$ is a scale factor which measures the width of the distribution. For $\alpha < 2$, $\phi(t; \alpha, \gamma)$

5. CONCLUSION

The phase retrieval problem has been revisited via a residual minimization in the $\ell_p$-norm, resulting in an AIRLS algorithm working with $0 < p < 2$. The AIRLS is able to deal with impulsive noise effectively, without necessarily knowing the noise distribution beforehand. Furthermore, two CRBs were derived for Laplacian and Gaussian noise models. Simulations demonstrated that the proposed method outperforms the state-of-art in impulsive noise.

6. REFERENCES


