# 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

$$\mathbf{y} = |\mathbf{A}\mathbf{x}| + \mathbf{n} \ \in \mathbb{R}^M \tag{1}$$

Email: qianc@umn.edu (C. Qian), xfu@umn.edu (X. Fu), nikos@umn.edu (N. D. Sidiropoulos), dr.lei.huang@ieee.org (L. Huang). The work of N. Sidiropoulos was supported by NSF CIF-1525194. C. Qian is currently an exchange PhD student with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA, supported in part by NSFC 61271420 and the Chinese Scholarship Council.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

where $|\cdot|$ denotes the absolute value operator, $\mathbf{A}^H = [\mathbf{a}_1 \cdots \mathbf{a}_M] \in \mathbb{C}^{N\times N}$ is the measurement matrix with $H$ being the conjugate transpose, and $\mathbf{n}$ is the noise vector which in this paper, is assumed to be heavy-tailed. Our aim is to recover $\mathbf{x}$ from $\mathbf{y}$. Without noise, by introducing an auxiliary vector $\mathbf{u} = e^{j\angle(\mathbf{Ax})}$ where $\angle$ takes the phase of its argument, (1) can also be written as [5][26]

$$\mathbf{Yu} = \mathbf{Ax} \tag{2}$$

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

$$\min_{|\mathbf{u}|=\mathbf{1},\mathbf{x}} \sum_{m=1}^{M} \left(|y_m u_m - \mathbf{a}_m^H \mathbf{x}|^2 + \epsilon\right)^{p/2} \tag{3}$$

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 $\mathbf{u}$ and update $\mathbf{x}$, and then we do the same to $\mathbf{u}$.

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

$$\mathbf{x}^{(r+1)} = \arg\min_{\mathbf{x}} \sum_{m=1}^{M} \left(|y_m u_m^{(r)} - \mathbf{a}_m^H \mathbf{x}|^2 + \epsilon\right)^{p/2} \tag{4}$$

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 $\phi_p(w) := \frac{2-p}{2}\left(\frac{2}{p}w\right)^{\frac{p}{p-2}} + \epsilon w$. Then, we have*

$$\left(x^2 + \epsilon\right)^{p/2} = \min_{w\geq 0}\ wx^2 + \phi_p(w), \tag{5}$$

*and the unique minimizer is*

$$w_{\mathrm{opt}} = \frac{p}{2}\left(x^2 + \epsilon\right)^{\frac{p-2}{2}}. \tag{6}$$

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

$$\sum_{m=1}^{M}\left(|y_m u_m^{(r)} - \mathbf{a}_m^H\mathbf{x}|^2 + \epsilon\right)^{p/2}$$
$$\leq \sum_{m=1}^{M}\left(w_m^{(r)}\left|y_m u_m^{(r)} - \mathbf{a}_m^H\mathbf{x}^{(r)}\right|^2 + \phi_p\left(w_m^{(r)}\right)\right) \tag{7}$$

where

$$w_m^{(r)} := \frac{p}{2}\left(\left|y_m u_m^{(r)} - \mathbf{a}_m^H\mathbf{x}^{(r)}\right|^2 + \epsilon\right)^{\frac{p-2}{2}}. \tag{8}$$

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 $\mathbf{x}$. Notice that the RHS of (7) is convex w.r.t. $\mathbf{x}$ and can be solved in closed-form:

$$\mathbf{x}^{(r+1)} = (\mathbf{W}^{(r)}\mathbf{A})^{\dagger}\mathbf{W}^{(r)}\mathbf{Yu}^{(r)}. \tag{9}$$

where

$$\mathbf{W}^{(r)} = \mathrm{diag}\left(\sqrt{w_1^{(r)}} \cdots \sqrt{w_M^{(r)}}\right). \tag{10}$$

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

$$\mathbf{u}^{(r+1)} = \arg\min_{|\mathbf{u}|=\mathbf{1}} \sum_{m=1}^{M}\left(|y_m u_m - \mathbf{a}_m^H\mathbf{x}^{(r+1)}|^2 + \epsilon\right)^{p/2}. \tag{11}$$

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

$$u_m^{(r+1)} = e^{j\angle\left(\mathbf{a}_m^H\mathbf{x}^{(r+1)}\right)}, \quad m = 1,\ldots,M. \tag{12}$$

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

**Algorithm 1** AIRLS for phase retrieval

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

*Remark:* The computational burden of AIRLS mainly lies in the calculation of $\mathbf{x}$ at each iteration, which has complexity $\mathcal{O}(MN^2)$. Thus, if the algorithm converges after, say, $r$th iterations, the total complexity is $\mathcal{O}(rMN^2)$. Compared to semidefinite relaxation based methods which have worst-case complexity of $\mathcal{O}(N^{6.5})$, 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 $\mathbf{x}$ [7],[15]-[18], we provide a closed-form CRB formula corresponding to the phase and amplitude of $\mathbf{x}$. 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* $\mathbf{x}$ *is bounded below by*

$$\mathrm{CRB}_{Laplacian,|\mathbf{x}|} = \sum_{i=1}^{N} d_i \tag{13}$$

*and the variance of any unbiased estimate corresponding to the phase of complex-valued* $\mathbf{x}$ *is bounded below by*

$$\mathrm{CRB}_{Laplacian,\angle(\mathbf{x})} = \sum_{i=N+1}^{2N} d_i \tag{14}$$

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

$$\mathbf{F} = \frac{4}{\sigma_n^2}\mathbf{G}\,\mathrm{diag}(|\mathbf{A}\mathbf{x}|)^{-2}\,\mathbf{G}^T \tag{15}$$

*with*

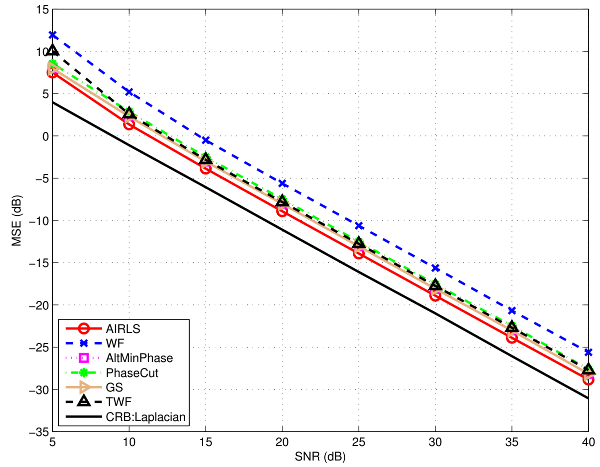
$$\mathbf{G} = \begin{bmatrix} \mathrm{diag}(|\mathbf{x}|)^{-1} & \\ & \mathbf{I}_N \end{bmatrix} \begin{bmatrix} Re\left\{\mathrm{diag}\left(\mathbf{x}^*\right)\mathbf{A}^H\mathrm{diag}(\mathbf{A}\mathbf{x})\right\} \\ Im\{\mathrm{diag}(\mathbf{x}^*)\mathbf{A}^H\mathrm{diag}(\mathbf{A}\mathbf{x})\} \end{bmatrix}. \tag{16}$$

Here, $(\cdot)^{\dagger}$, $(\cdot)^{-1}$, $(\cdot)^T$ and $(\cdot)^*$ represent the pseudo-inverse, inverse, transpose and conjugate operators, respectively. When $\mathbf{A}$ is nontrivial, we find that $\mathbf{F}$ 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:

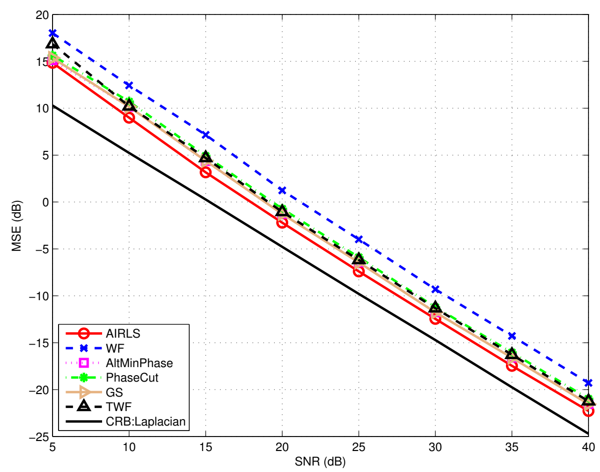
**Theorem 2** *In Gaussian noise, the CRB is four times larger than the CRB in Theorem 1, i.e.,*

$$\mathrm{CRB}_{Gaussian} = 4\mathrm{CRB}_{Laplacian} = 4\mathrm{trace}\left(\mathbf{F}^{\dagger}\right) \tag{17}$$

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




(a) MSE on amplitude



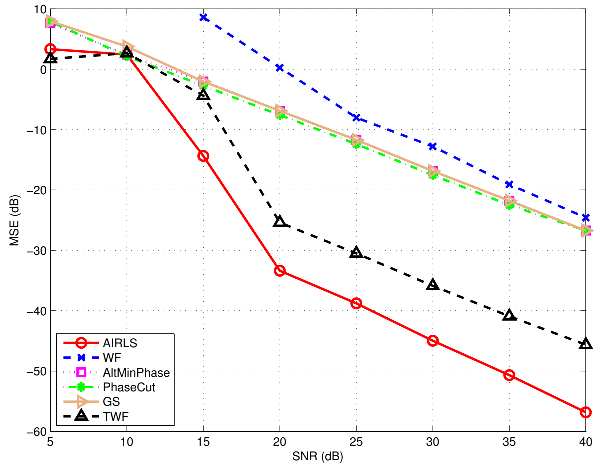

(b) MSE on phase

**Fig. 1.** MSE versus SNR in Laplacian noise.

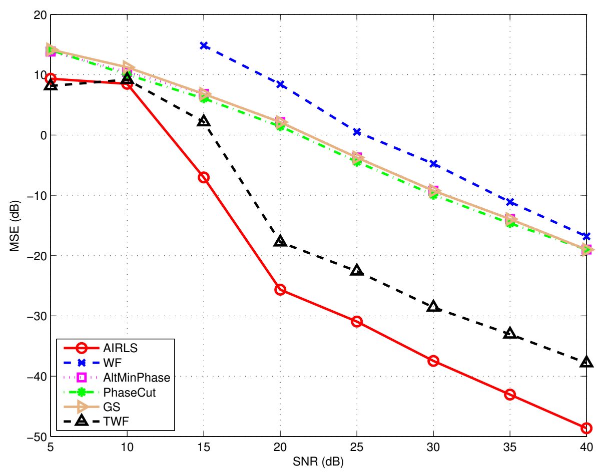
## 4. SIMULATION RESULTS

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 globe phase ambiguity between the true and estimated $\mathbf{x}$. The signal $\mathbf{x}$ with length $N = 64$ is generated from a complex Gaussian distribution with mean zero and covariance matrix $\mathbf{I}_N$. 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




(a) MSE on amplitude




(b) MSE on phase

**Fig. 2**. MSE versus SNR in S$\alpha$S noise.

the AltMinPhase and GS algorithms, and achieves the best performance, while the TWF and PhaseCut perform similarly. Note that since WF admits a model where noise is added after $|\mathbf{a}_i^H\mathbf{x}|^2$ , we use $\{y_i^2\}_{i=1}^M$ for WF, which results in multiplicated noise. Therefore, WF works in a much more severe scenario than the other competitors, and it produces very large MSE.

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

$$\phi(t;\alpha,\gamma) = \exp\left(-\gamma^{\alpha}|t|^{\alpha}\right) \tag{18}$$

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 $\alpha < 2$, $\phi(t;\alpha,\gamma)$ 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 S$\alpha$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 S$\alpha$S noise scenario.

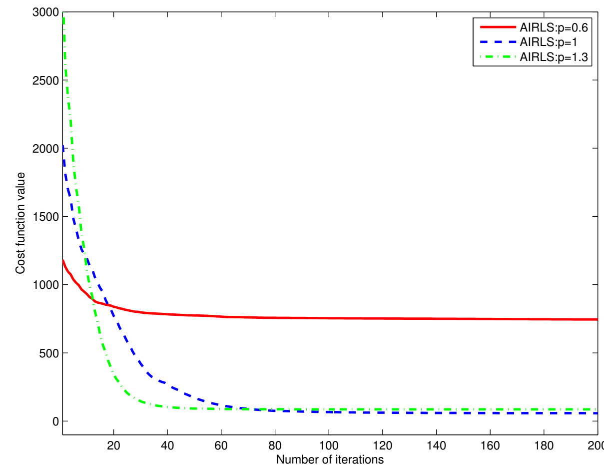



**Fig. 3**. Convergence behavior of AIRLS.

Finally, Fig. 3 plots the convergence behavior of AIRLS algorithm with $p = 0.6$, 1 and 1.3, which illustrates that the AIRLS converges in terms of cost function value after 100 iterations.

## 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

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

[2] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758-2769, 1982.

[3] E. J. Candès, T. Strohmer, and V. Voroninski. "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241-1274, 2013.

[4] E. J. Candès, Y. C. Eldar, T. Strohmer and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Review*, vol. 57, no. 2 pp. 225-251, 2015.

[5] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47-81, 2015.

[6] E. J. Candès, X. Li and M. Soltanolkotabi, "Phase retrieval via Wirtinger Flow: Theory and algorithms," *IEEE Trans. Information Theory*, vol. 61, no. 4, pp. 1985-2007, 2015.

[7] C. Qian, N. D. Sidiropoulos, K. Huang, L. Huang and H. C. So, "Phase retrieval using feasible point pursuit: Algorithms and Cramér-Rao bound," *IEEE Trans. Signal Process.*, submitted, 2015.

[8] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar and N. S. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Process. Letters*, vol. 22, no. 7, pp. 804-808, 2015.

[9] J. Sigl, "Nonlinear residual minimization by iteratively reweighted least squares," *arXiv preprint arXiv*, 1504.06815, 2015.

[10] T. T. Cai, X. Li and Z. Ma, "Optimal rates of convergence of noisy sparse phase retrieval via thresholded wirtinger flow," *arXiv preprint arXiv*, 1504.03385, 2015.

[11] D. S. Weller, A. Pnueli, O. Radzyner, G. Divon, Y. C. Eldar and J. A. Fessler, "Phase retrieval of sparse signals using optimization transfer and ADMM," *Proc. of 2014 Internat. Conf. on Image Process. (ICIP)*, pp. 1342-1346, Paris, 2014.

[12] D. S. Weller, A. Pnueli, G. Divon, O. Radzyner, Y. C. Eldar and J. A. Fessler, "Undersampled phase retrieval with outliers," *IEEE Trans. Computational Imaging*, vol. 1, no. 4, pp. 247-258, 2015.

[13] H. M. L. Faulkner and J. M. Rodenburg, "Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm," *Physical review letters*, vol. 93, no. 2, pp. 023903, 2004

[14] J. N. Cederquist and C. C. Wackerman, "Phase-retrieval error: A lower bound," *Journal of the Optical Society of America A*, vol. 4, no. 9 pp. 1788-1792, 1987.

[15] R. Balan, "The Fisher information matrix and the CRLB in a non-AWGN model for the phase retrieval problem," *Proc. of 2015 Internat. Conf. on Sampl. Theory and Applications (SampTA)*, pp. 178-182, Washington, DC, 2015.

[16] R. Balan, "Reconstruction of signals from magnitudes of redundant representations," *arXiv preprint arXiv*, 1207.1134, 2012.

[17] R. Balan, "Reconstruction of signals from magnitudes of redundant representations: The complex case," *Foundations of Computational Mathematics*, pp. 1-45, 2013.

[18] A. S. Bandeira, J. Cahill, D. G. Mixon and A. A. Nelson, "Saving phase: Injectivity and stability for phase retrieval," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 106-125, 2014.

[19] X. Fu, K. Huang, W.-K. Ma, N. D. Sidiropoulos, and R. Bro, "Joint tensor factorization and outlying slab suppression with applications," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6315-6328, 2015.

[20] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87-90, 2001.

[21] A. O. Hero, III, J. A. Fessler and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. Signal Process.*, vol. 44, pp. 2026-2041, Aug. 1996

[22] C. R. Rao, *Linear statistical inference and its applications*, 2nd ed. New York: Wiley, 1973.

[23] K. Huang and N. D. Sidiropoulos, "Putting nonnegative matrix factorization to the test: A tutorial derivation of pertinent Cramér-Rao bounds and performance benchmarking," *IEEE Signal Processing Magazine, Special Issue on Source Separation and Applications*, vol. 31, no. 3, pp. 76-86, 2014.

[24] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2678-2689, 2005.

[25] Y. Chen and E. Candés, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Advances in Neural Information Processing System*, pp. 739-747, 2015.

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