LEAST SQUARES PHASE RETRIEVAL USING FEASIBLE POINT PURSUIT

Cheng Qian[‡] *Nicholas D. Sidiropoulos*[§] *Kejun Huang*[§] *Lei Huang*[⋆] *H. C. So*[†]

[‡]Dept. of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China
[§]Dept. of ECE, University of Minnesota, MN, USA
[⋆]College of Information Engineering, Shenzhen University, Shenzhen, China
[†]Dept. of Electronic Engineering, City University of Hong Kong, Hong Kong, China

ABSTRACT

Phase retrieval has recently attracted renewed interest. It is revisited here through a new approach based on nonconvex quadratically constrained quadratic programming (QCQP). A least-squares (LS) formulation is adopted, and a recently developed non-convex QCQP approximation technique called *feasible point pursuit (FPP)* is tailored to obtain a new LS-FPP phase retrieval algorithm. The Cramér-Rao bound (CRB) is also derived for phase retrieval under additive white Gaussian noise. We demonstrate through simulations that the LS-FPP method outperforms the prior art and its mean square error approaches the CRB.

Index Terms— Phase retrieval, quadratically constrained quadratic programming (QCQP), semidefinite programming (SDP), feasible point pursuit (FPP), Cramér-Rao bound (CRB).

1. INTRODUCTION

Phase retrieval (PR) is a fundamental problem in many different areas of science, e.g., crystallography [1], diffraction imaging [2] and microscopy [3], where it is easier to measure the magnitude than the phase. Specifically, PR seeks to recover an unknown complex-valued signal (up to a global phase factor) or real-valued signal (up to a sign ambiguity) from its magnitude measurements of the form

$$y_i = |\mathbf{a}_i^H \mathbf{x}|^2 + n_i,\ i = 1, \cdots, M \tag{1}$$

where $|\cdot|$ is the magnitude of a complex number, $(\cdot)^H$ is the conjugate transpose, $\mathbf{a}_i \in \mathbb{C}^N$ is a known measurement vector and n_i denotes additive white Gaussian noise with variance σ_n^2.

Email: qianc@umn.edu (C. Qian), nikos@umn.edu (N. D. Sidiropoulos), huang663@umn.edu (K. Huang), dr.lei.huang@ieee.org (L. Huang), hcso@ee.cityu.edu.hk (H. C. So). The work of N. Sidiropoulos was supported by NSF CIF-1525194. K. Huang was supported by a UMII dissertation fellowship. C. Qian is now as an exchanging PhD student with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA, supported in part by the Chinese Scholarship Council.

Early works for PR were mainly based on alternating optimization. Among them, Gerchberg-Saxton [4] and Fienup [5] are the most well-known and widely used methods in practice, where the unknown $\mathbf{x}$ is iteratively estimated by imposing Fourier and real-space magnitude constraints. However, the alternating optimization approach can converge to a local minimum, and often fails to recover to the true solution. Recently, an alternative based on semidefinite relaxation (SDR), referred to as *PhaseLift* [6], has been developed for PR. Instead of estimating $\mathbf{x}$ directly, PhaseLift tries to find a rank-1 matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^H$ that satisfies the set of linear equalities $y_i = \text{trace}(\mathbf{A}_i\mathbf{X}), \forall i \in \{1, \cdots, M\}$ where trace$(\cdot)$ stands for the trace operator and $\mathbf{A}_i = \mathbf{a}_i\mathbf{a}_i^H$, and then picks the principal eigenvector of $\mathbf{X}$ as the estimate of $\mathbf{x}$. It has been shown in [7] that if $M \sim \mathcal{O}(N \log N)$ i.i.d. Gaussian measurements are used, PhaseLift can accurately recover $\mathbf{x}$ with high probability. However, in the presence of noise, the constraints above are at best approximately satisfied, so there is no guarantee that PhaseLift will yield a rank-1 solution [8]. Another semidefinite programming based algorithm is *PhaseCut* [9], which takes a similar approach as PhaseLift but it estimates the phase of y_i first. Since PhaseCut can be read as the classical *MaxCut* problem in networks, it allows fast SDR algorithms developed for MaxCut to be applied to PhaseCut.

More recently, a new approach named *Wirtinger Flow (WF)* [10] which relies on a smart initialization followed by relatively simple first-order refinement has been proposed. Although it has been theoretically proved that when a sufficiently large number of i.i.d. Gaussian measurements are employed, WF can yield the desired solution with high probability, reliable recovery cannot be guaranteed if the number of measurements is small or when the measurement vectors are not random - mainly because the principal eigenvector used for initialization is not a good approximation of $\mathbf{x}$ in such cases.

In this paper, we present a new method for PR from measurements of type (1). This method, which we call *LS-FPP*, is obtained using a non-convex quadratically constrained quadratic programming (QCQP) formulation of *least-squares (LS)* PR, and an approximation technique called *feasible point* *pursuit (FPP)* that was recently introduced in [11]. LS-FPP is designed for i.i.d. Gaussian measurement errors, where LS is equivalent to maximum likelihood, resulting in optimal estimates that have excellent statistical properties. The performance of LS-FPP is evaluated against state-of-the-art algorithms and the general Cramér-Rao bound (CRB) for PR from magnitude measurements in additive Gaussian noise, which is also derived here. Derivations and additional results can be found in the journal version [12], which also includes another algorithm for a different measurement model.

2. PROPOSED ALGORITHM

The LS formulation of PR has been recently considered in [10], but the WF approach does not always work well, as we will show in our simulations in Section 4. This is not surprising, of course, since we are dealing with an NP-hard problem. Our contribution here is to recast LS PR as a non-convex quadratic-plus-linear problem, and then approximate it using FPP. As we will demonstrate, our approach gives consistently better approximation results, especially in challenging scenarios, at the cost of additional computational complexity.

The LS formulation for PR is [10]

$$\min_{\mathbf{x}} \sum_{i=1}^{M} (y_i - \mathbf{x}^H \mathbf{A}_i \mathbf{x})^2. \tag{2}$$

Define

$$w_i = y_i - \mathbf{x}^H \mathbf{A}_i \mathbf{x},\ \forall i \tag{3}$$

which leads to

$$\mathcal{P}_0 \begin{cases} \min\limits_{\mathbf{w},\mathbf{x}} & ||\mathbf{w}||_2^2 \\ \text{s. t.} & \mathbf{x}^H \mathbf{A}_i \mathbf{x} + w_i = y_i,\ \forall i \end{cases}$$

where $\mathbf{w} = [w_1 \ \cdots \ w_M]^T$. Let $\mathbf{e}_i$ be the ith column of $\mathbf{I}_M$ where $\mathbf{I}_M$ denotes a $M \times M$ identity matrix, such that $w_i = \mathbf{e}_i^T \mathbf{w}$. Then the constraints in $\mathcal{P}_0$ become

$$\mathbf{x}^H \mathbf{A}_i \mathbf{x} + \mathbf{e}_i^T \mathbf{w} = y_i \tag{4}$$

which can be rewritten as

$$\mathbf{x}^H \mathbf{A}_i \mathbf{x} + \mathbf{e}_i^T \mathbf{w} \leq y_i \tag{5a}$$

$$\mathbf{x}^H \mathbf{A}_i \mathbf{x} + \mathbf{e}_i^T \mathbf{w} \geq y_i. \tag{5b}$$

It is clear that the constraints in (5b) are non-convex, so $\mathcal{P}_0$ belongs to a class of non-convex QCQP problems, which is NP-hard in its general form. To approximately solve $\mathcal{P}_0$, we follow [11] to deal with (5b) using successive convex approximation. Since $\mathbf{A}_i$ is positive semidefinite with only one positive eigenvalue, for any $\mathbf{z}$ and $\mathbf{x}$, we have

$$(\mathbf{x} - \mathbf{z})^H \mathbf{A}_i (\mathbf{x} - \mathbf{z}) \geq 0 \tag{6}$$

which, after expanding the left-hand side of (6), yields

$$\mathbf{x}^H \mathbf{A}_i \mathbf{x} \geq 2\mathrm{Re}\{\mathbf{z}^H \mathbf{A}_i \mathbf{x}\} - \mathbf{z}^H \mathbf{A}_i \mathbf{z} \tag{7}$$

where $\mathrm{Re}\{\cdot\}$ takes the real part of its argument. Following the rationale in [11], we replace (5b) by

$$2\mathrm{Re}\{\mathbf{z}^H \mathbf{A}_i \mathbf{x}\} + \mathbf{e}_i^T \mathbf{w} + s_i \geq y_i + \mathbf{z}^H \mathbf{A}_i \mathbf{z} \tag{8}$$

where $s_i \geq 0$ is a slack variable. This leads to the following formulation:

$$\mathcal{P}_1 \begin{cases} \min\limits_{\mathbf{x},\mathbf{w},\mathbf{s}} & ||\mathbf{w}||^2 + \lambda \sum\limits_{i=1}^{M} s_i \\ \text{s. t.} & 2\mathrm{Re}\{\mathbf{z}^H \mathbf{A}_i \mathbf{x}\} + \mathbf{e}_i^T \mathbf{w} + s_i \geq y_i + \mathbf{z}^H \mathbf{A}_i \mathbf{z} \\ & \mathbf{x}^H \mathbf{A}_i \mathbf{x} + \mathbf{e}_i^T \mathbf{w} \leq y_i, \\ & s_i \geq 0, \forall i. \end{cases}$$

where $\mathbf{s} = [s_1 \ \cdots \ s_M]^T$ and λ is the regularization parameter that balances the objective function and slack penalty term. Starting from an initial $\mathbf{z}$, we obtain $(\mathbf{x}, \mathbf{w}, \mathbf{s})$ iteratively by solving a sequence of problems of type $\mathcal{P}_1$. Since the cost function in $\mathcal{P}_1$ is fixed throughout the iterations, and the solution of the kth iteration is also feasible for the $(k+1)$th iteration, it follows that the optimal cost sequence generated this way will be non-increasing, and since it is bounded from below it is guaranteed to converge. The steps for LS-FPP are summarized in **Algorithm 1**.

Algorithm 1 LS-FPP Algorithm for Phase Retrieval

function $\hat{\mathbf{x}} = \text{LS-FPP}(\mathbf{A}, \mathbf{y}, \lambda, \mathbf{z})$
    repeat
        $\hat{\mathbf{x}} \leftarrow$ solution of $\mathcal{P}_1$
        $\mathbf{z} = \hat{\mathbf{x}}$
    until a stopping criterion on the cost function of $\mathcal{P}_1$ is satisfied
end function

Remark: The problem in $\mathcal{P}_1$ is convex and can be solved via interior point methods [13]-[14]. The worst-case complexity of solving $\mathcal{P}_1$ is $\mathcal{O}\big((N + 3M)^{3.5}\big)$. Moreover, few outer iterations of LS-FPP are usually needed, so that the overall complexity is often manageable for moderate N.

3. CRAMÉR-RAO BOUND

Balan [15, 16] has derived the Fisher Information Matrix (FIM) for the model in (1), and noted in [16] that it is singular. Attributing this to the lack of global phase identifiability, he proposed using side information about $\mathbf{x}$ to compute the CRB. Identifiability neither implies nor is implied by a finite FIM [17], so we make no such assumption, and use the pseudoinverse instead, as detailed later. In the case of real $\mathbf{x}$, Balan's result is valid only for real measurement vectors. Balan also derived [18] the FIM for white Gaussian noise added *prior* to taking the magnitude square, i.e., $y_i = |\mathbf{a}_i^H\mathbf{x} + n_i|^2$, which is different from our model in (1). We also note [19], where the CRB has been derived for a 2-D phase retrieval model with 2-D Fourier measurements.

The following theorem presents a closed-form expression for the general CRB for complex-valued $\mathbf{x}$ under white Gaussian noise.

Theorem 3.1 *For $\mathbf{x} \in \mathbb{C}^N$, the CRB for the PR model in* (1) *is*

$$\mathrm{CRB}_c = \mathrm{trace}\left(\mathbf{F}_c^\dagger\right) \tag{9}$$

where $(\cdot)^\dagger$ represents the pseudo-inverse and the FIM is given by

$$\mathbf{F}_c = \frac{4}{\sigma_n^2}\mathbf{G}_c\mathbf{G}_c^T \tag{10}$$

with

$$\mathbf{G}_c = \begin{bmatrix} \mathrm{Re}\{\mathbf{A}_1\mathbf{x}\} & \cdots & \mathrm{Re}\{\mathbf{A}_M\mathbf{x}\} \\ \mathrm{Im}\{\mathbf{A}_1\mathbf{x}\} & \cdots & \mathrm{Im}\{\mathbf{A}_M\mathbf{x}\} \end{bmatrix}. \tag{11}$$

Here, $\mathrm{Im}\{\cdot\}$ denotes the imaginary part of its argument. It is worth mentioning that $\mathbf{F}_c$ is always singular with rank-1 deficiency. Therefore, we replace the inverse of $\mathbf{F}_c$ by its pseudo-inverse to generate the CRB. It has been pointed out in [20]-[22] that $\mathrm{trace}(\mathbf{F}_c^\dagger)$ is a valid lower bound which is often attainable in practice and thus predictive of optimal estimator performance [21]. Also note that, when $\mathbf{x}$ is real-valued, the corresponding CRB can be viewed as a special case of the complex-valued one, see the following Theorem 3.2.

Theorem 3.2 *For $\mathbf{x} \in \mathbb{R}^N$, the CRB for the PR model in* (1) *is*

$$\mathrm{CRB}_r = \mathrm{trace}\left(\mathbf{F}_r^{-1}\right) \tag{12}$$

where $(\cdot)^{-1}$ denotes the inverse and the FIM is given by

$$\mathbf{F}_r = \frac{4}{\sigma_n^2}\mathbf{G}_r\mathbf{G}_r^T \tag{13}$$

with

$$\mathbf{G}_r = \begin{bmatrix} \mathrm{Re}\{\mathbf{A}_1\}\mathbf{x} & \cdots & \mathrm{Re}\{\mathbf{A}_M\}\mathbf{x} \end{bmatrix}. \tag{14}$$

We expect a reduced bound when the number of measurements is increased (keeping the signal-to-noise ratio (SNR) fixed). The following result confirms this.

Proposition 3.3 *For given $\mathbf{x}$ and fixed σ_n, the CRBs for both complex- and real-valued $\mathbf{x}$ decrease as more measurements are made available:*

$$\begin{aligned} \mathrm{CRB}_c(\mathbf{A}(:,1:M+1)) &\le \mathrm{CRB}_c(\mathbf{A}(:,1:M)) \\ \mathrm{CRB}_r(\mathbf{A}(:,1:M+1)) &\le \mathrm{CRB}_r(\mathbf{A}(:,1:M)) \end{aligned} \tag{15}$$

where $\mathbf{A}(:,\ell:r)$ is (Matlab notation for) the submatrix of $\mathbf{A}$ comprising columns from ℓ to r inclusive.

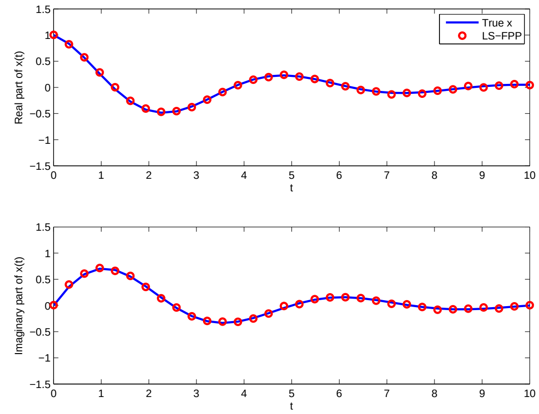


Fig. 1. Signal recovery of $\mathbf{x}$ by LS-FPP.

4. SIMULATION RESULTS

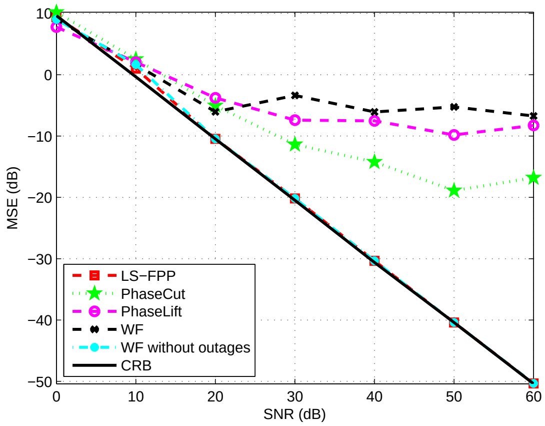
In this section, we compare the performance of LS-FPP, PhaseLift, PhaseCut and WF algorithms for PR. LS-FPP is initialized by the output of PhaseLift, and the stopping criterion is $||\mathbf{z}_k - \hat{\mathbf{x}}||_2^2 \le 10^{-5}$ or a maximum number of iterations, set to 10. The SNR is defined as

$$\mathrm{SNR} = \frac{\sum_{i=1}^M |\mathbf{a}_i^H\mathbf{x}|^4}{M\sigma_n^2}.$$

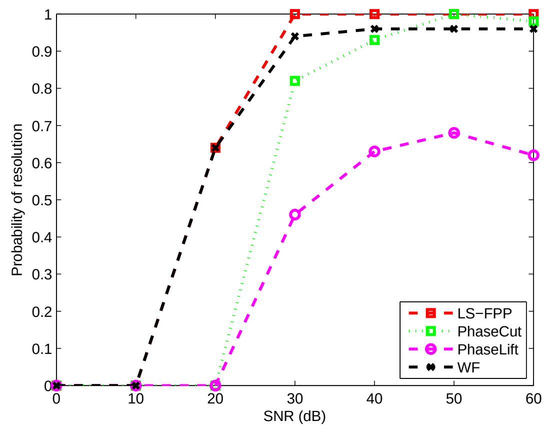
Furthermore, the signal $\mathbf{x}$ is fixed throughout all Monte-Carlo trials, and chosen as a uniformly sampled version of $\exp((j0.4\pi - 0.3)t), t \in [0, 10]$ comprising $N = 32$ samples. For illustration purposes, we remove the global phase ambiguity after signal recovery.

To begin, let us illustrate the recovery performance of LS-FPP by means of example, where we set $N = 32$, $M = 5N = 160$ and SNR = 20 dB. The measurement vectors $\{\mathbf{a}_i\}_{i=1}^M$ are generated from a complex normal distribution. It is shown in Fig. 1 that our scheme can accurately recover $\mathbf{x}$.

We now compare the performance in terms of mean square error (MSE) and probability of resolution for all methods, as a function of SNR, using 50 Monte-Carlo trials. The signal is considered to be resolved if $||\hat{\mathbf{x}} - \mathbf{x}||_2 \le 0.1$, otherwise we declare an outage. The CRB in Theorem 3.1 is also included as a benchmark. Fig. 2(a) depicts the MSE results for Gaussian measurements when $N = 32$ and $M = 160$, from which we can observe that the LS-FPP achieves the best performance and outperforms WF, PhaseLift and PhaseCut when SNR is higher than 10 dB. The relatively large MSE of WF is mainly caused by occasional outages, as can be verified from Fig. 2(b), where the probabilities of resolution for WF, PhaseCut and PhaseLift are around 95%, 98% and 60%, respectively, at high SNR. We also include a curve 'WF without outages' to show the performance of WF after discarding trials with outage. Similar results can also be found

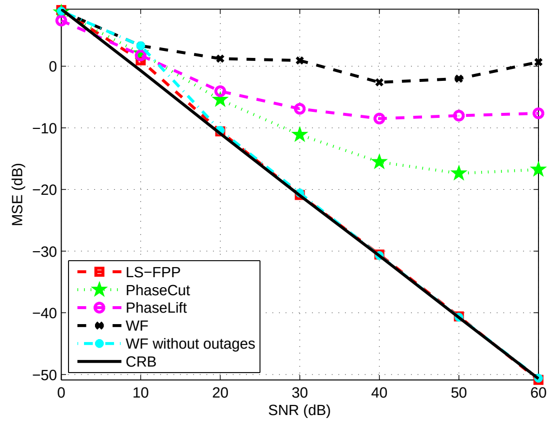


(a) MSE versus SNR.

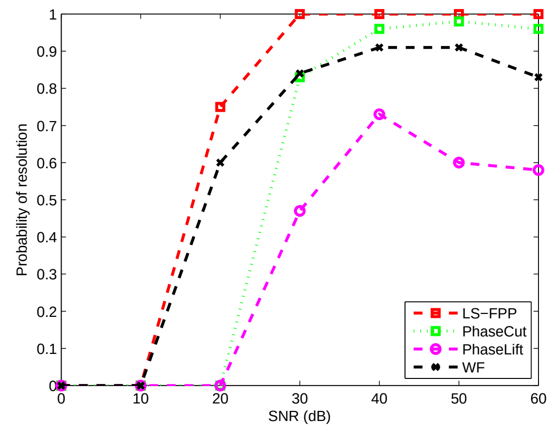


(b) Probability of resolution.

Fig. 2. Performance comparison with Gaussian measurements.



(a) MSE versus SNR.



(b) Probability of resolution.

Fig. 3. Performance comparison with masked Fourier measurements.

in Fig. 3, where 4 masked Fourier measurements are used. Here, the masked Fourier matrix has the form

$$\mathbf{A}^H = \begin{bmatrix} \mathbf{F}\mathbf{D}_1 \\ \mathbf{F}\mathbf{D}_2 \\ \mathbf{F}\mathbf{D}_3 \\ \mathbf{F}\mathbf{D}_4 \end{bmatrix} \tag{16}$$

where $\mathbf{F}$ is a $N \times N$ Fourier matrix and $\mathbf{D}_i$ is a $N \times N$ diagonal masking matrix with its diagonal entries generated by $b_1 b_2$, whereby b_1 and b_2 are independent and distributed as [10]

$$b_1 = \begin{cases} 1 & \text{with prob. } 0.25 \\ -1 & \text{with prob. } 0.25 \\ -j & \text{with prob. } 0.25 \\ j & \text{with prob. } 0.25 \end{cases} \tag{17}$$

and

$$b_2 = \begin{cases} \sqrt{2}/2 & \text{with prob. } 0.8 \\ \sqrt{3} & \text{with prob. } 0.2. \end{cases} \tag{18}$$

5. CONCLUSION

The PR problem was revisited through the LS-FPP approach, which builds upon recent work on feasible point pursuit for non-convex QCQP problems. LS-FPP is based on a LS criterion that is tailored for white Gaussian noise added after taking the magnitude square of the linear measurements. Furthermore, the relevant CRB was also derived and studied. Simulation results suggest that LS-FPP outperforms the state-of-the-art and its MSE comes very close to the CRB in the high SNR regime. The main drawback of LS-FPP is its relatively high computational complexity, especially compared to WF. Hence, finding ways of bringing down the complexity (e.g., possibly using variations of the initialization of the WF scheme) will be the subject of future work.

6. REFERENCES

[1] R. W. Harrison, "Phase problem in crystallography," *Journal of the Optical Society of America A*, vol. 10, no. 5, pp. 1046-1055, 1993.

[2] F. Pfeiffer, T. Weitkamp, O. Bunk and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources," *Nature Physics*, vol. 2, no. 4, pp. 258-261, 2006.

[3] J. Miao, T. Ishikawa, Q. Shen and T, "Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes," *Annu. Rev. Phys. Chem.*, vol. 59, pp. 387-410, 2008.

[4] 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.

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

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

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

[13] C. Helmberg, F. Rendl, R. J. Vanderbei and H. Wolkowicz, "An interior-point method for semidefinite programming," *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 342-361, 1996.

[14] S. J. Kim, K. Koh, M. Lustig, S. Boyd and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE Journal of Selected Topics in Signal Process.*, vol. 1, no. 4, pp. 606-617, 2007.

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

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

[17] S. Basu and Y. Bresler, "The stability of nonlinear least squares problems and the Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3426-3436, 2000.

[18] 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

[19] 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.

[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] 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.