Proceedings Paper:

Li, Q., Liu, W. orcid.org/0000-0003-2968-2888, Huang, L. et al. (2 more authors) (2018) An undersampled phase retrieval algorithm via gradient iteration. In: 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM). 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), 08-11 Jul 2018, Sheffield, UK. IEEE , pp. 228-231. ISBN 978-1-5386-4752-3

https://doi.org/10.1109/SAM.2018.8449000

An Undersampled Phase Retrieval Algorithm via Gradient Iteration

Qiang Li, Wei Liu, *Senior* Member, IEEE, Lei Huang, *Senior* Member, IEEE, Weize Sun, Peichang Zhang

***Abstract*—This work addresses the issue of undersampled phase retrieval using the gradient framework and proximal regularization theorem. It is formulated as an optimization problem in terms of least absolute shrinkage and selection operator (LASSO) form with $(\ell_2+\ell_1)$ norms minimization in the case of sparse incident signals. Then, inspired by the compressive phase retrieval via majorization-minimization technique (C-PRIME) algorithm, a gradient-based PRIME algorithm is proposed to solve a quadratic approximation of the original problem. Moreover, we also proved that the C-PRIME method can be regarded as a special case of the proposed algorithm. As demonstrated by simulation results, both the magnitude and phase recovery abilities of the proposed algorithm are excellent. Furthermore, the experimental results also show the mean square error (MSE) performance of the proposed algorithm versus iterative step.**

***Index Terms*—Undersampled phase retrieval, gradient iteration, sparse signal, majorization-minimization**

I. Introduction

Algorithmically, most of the popular techniques for phase retrieval can be divided into two main categories [1]. The first one is based on the scheme of alternating minimization [2] to recover the original signal. One classic approach is to use the Gerchberg-Saxton algorithm [3], which introduces a phase information variable of linear measurements. The second and more recent class is based on the semidefinite programming (SDP) technique and the rank-1 matrix recovery framework [4], [5]. However, in the presence of high dimensional signals, the "matrix-lifting" problem will be suffered [6]. Furthermore, a Wirtinger Flow algorithm was proposed to solve the phase retrieval problem by using steepest descent method with a heuristic step [7]. More recently, a solution following similar updating rule but with a specified step size was introduced in [8].

Phase retrieval is an inherently non-convex ill-posed inverse problem. Normally, the objective of phase retrieval is to recover an original signal with relatively large probability, where the number of measurements M needs to be greater than the dimensions of incident signal N. Theoretically, it has been established that $M \geq 2N-1$ or $M \geq 4N-4$ generic measurements suffice for uniquely determining an N-dimensional real-valued or complex-valued signal, respectively [9]. In practice, however, the undersampled phase retrieval problem is often encountered, which refers to the case of $M < N$. Existing approaches attempt to solve the underdetermined problem by introducing a sparsity assumption on incident signals [10]–[12]. According to [10], a P-sparse complex signal can be recovered successfully with $M \geq 8P-2$ in the case of Gaussian measurement vectors. A GESPAR algorithm is proposed associating the damped Gauss-Newton method with the 2-opt local search approach [11], which only needs rough prior information on the sparsity level of incident signals. However, the matrix-lifting technique is required, leading to higher computational complexity. Using the convex ℓ_1 norm penalty term encouraging a sparse solution, the phase retrieval problem is formulated into the LASSO form [12], [13]. However, it requires information about the exact sparsity level P [12] and the convergence rate of compressive phase retrieval via majorization-minimization technique (C-PRIME) is usually slow or not convergent at all [13].

On the basis of the C-PRIME algorithm, a simple and efficient undersampled phase retrieval algorithm is proposed in this paper, which is called gradient-PRIME algorithm (G-PRIME for short) based on the gradient framework and the proximal regularization theorem. It is interesting that the proposed G-PRIME algorithm turns out to have a similar closed-form solution as that of the C-PRIME approach, but our G-PRIME algorithm is based on the derivation of the gradient framework. The simulation results prove the phase recovery ability and mean square error (MSE) performance of the proposed G-PRIME algorithm.

II. Problem formulation

The problem of estimating an N-dimensional complex signal $\boldsymbol{x}$ from M magnitude-only linear measurements $\boldsymbol{y}$ is called phase retrieval. A basic phase retrieval model with intensity measurements is

$$y_i = |(\boldsymbol{A}\boldsymbol{x})_i|^2 + n_i, \quad i = 1, \cdots, M \tag{1}$$

where $|\cdot|$ is the element-wise magnitude, y_i and complex measurement matrix $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ are known beforehand and $\boldsymbol{n} = [n_1, \cdots, n_M]^T$ denotes noise.

It is easy to observe that the intensity measurements are non-convex and not linear with regard to $\boldsymbol{x}$ due to the magnitude operator. We consider undersampled phase retrieval in this paper, which is an ill-posed inverse problem. Without loss of

The work described in this paper was supported by the National Natural Science Foundation of China under Grants U1501253, 61601304, 61501300, 61601300, 61501485, the Natural Science Foundation of Guangdong Province under Grant 2015A030311030 and by the Foundation of Shenzhen Government under Grants ZDSYS201507081625213, KC2015ZDYF0023A and JCYJ20160520165659418, the Foundation of China Postdoctoral Science under Grant 2015M582414, 2017M610547, 2016057 and the China Scholarship Council.

Q. Li, L. Huang, W. Sun and P. Zhang are with the College of Information Engineering, Shenzhen University, Shenzhen 518060, China. (Corresponding author: Lei Huang, lhuang8sasp@hotmail.com).

W. Liu is with the Department of Electronic and Electrical Engineering, University of Sheffield, S1 3JD Sheffield, United Kingdom.

generality, we assume that the incident signal is sparse, which can be found in various areas, such as optical imaging [5] and astronomy [1].

Because using modulus information $\{\sqrt{y_i}\}_{i=1}^{M}$ has a smaller variance value of additive noise than that of intensity information $\{y_i\}_{i=1}^{M}$ in the case of $|(\boldsymbol{A}\boldsymbol{x})_i| > 0.5$ [13], we formulate the undersampled phase retrieval problem as the following optimization model

$$\min_{\boldsymbol{x}} \sum_{i=1}^{M} \left(\sqrt{y_i} - |(\boldsymbol{A}\boldsymbol{x})_i|\right)^2 + \lambda \|\boldsymbol{x}\|_1 \tag{2}$$

where the parameter $\lambda > 0$ is a regularization penalty factor and $\|\boldsymbol{x}\|_1$ denotes ℓ_1 norm of vector $\boldsymbol{x}$, which is used to regularize the ill-posed phase retrieval problem and promote sparsity in $\boldsymbol{x}$.

Due to the magnitude operator, (2) is not a convex problem either, which can not be directly solved by CVX and other standard convex optimization approaches. Employing the majorization-minimization (MM) technique, in [13], an efficient C-PRIME method was proposed to solve a convex surrogate problem instead. The surrogate optimization problem is convex with regard to $\boldsymbol{x}$ and equivalent to the following problem

$$\boldsymbol{x} = \arg\min_{\boldsymbol{x}} \left[C \|\boldsymbol{x} - \boldsymbol{c}\|_2^2 + \lambda \|\boldsymbol{x}\|_1\right] \tag{3}$$

where C is a constant satisfying $C \geq \lambda_{\max}(\boldsymbol{A}^H\boldsymbol{A})$ and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix.

The optimization problem (3) has a simple closed-form solution at the k iteration using the soft thresholding method, i.e.,

$$\boldsymbol{x}^k = e^{j\,\mathrm{ang}(\boldsymbol{c})} \odot \max\left\{|\boldsymbol{c}| - \frac{\lambda}{2C}, 0\right\} \tag{4}$$

where $\mathrm{ang}(\cdot)$ denotes the phase angle and $\odot$ denotes the Hadamard (element-wise) product of two vectors. The vector $\boldsymbol{c}$ is a constant independent of the variable $\boldsymbol{x}$:

$$\boldsymbol{c} = \boldsymbol{x}^{k-1} - \frac{1}{C}\boldsymbol{A}^H\left(\boldsymbol{A}\boldsymbol{x}^{k-1} - \sqrt{\boldsymbol{y}} \odot e^{j\,\mathrm{ang}(\boldsymbol{A}\boldsymbol{x}^{k-1})}\right) \tag{5}$$

The C-PRIME method solves the surrogate optimization problem in (3) with a simple closed-form solution at every iteration.

III. Proposed Algorithm based on the Gradient Framework

A. The G-PRIME Algorithm

The optimisation in (3) can be cast as a second order cone programming problem. We first consider the following general formulation

$$\boldsymbol{x} = \arg\min_{\boldsymbol{x}} \left[F(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})\right] \tag{6}$$

where f is a smooth convex function and g is a continuous convex function which is possibly nonsmooth.

Specifically, for the optimization problem (3), let $f(\boldsymbol{x}) = C\|\boldsymbol{x} - \boldsymbol{c}\|_2^2$ and $g(\boldsymbol{x}) = \lambda\|\boldsymbol{x}\|_1$. One of the most popular methods for solving the problem is the iterative shrinkage-thresholding algorithm (ISTA) [15]. The iterative procedure of ISTA is

$$\boldsymbol{x}^k = \frac{\boldsymbol{a}}{|\boldsymbol{a}|} \odot \max\left\{|\boldsymbol{a}| - \lambda\mu, 0\right\} \tag{7}$$

where μ denotes an appropriate step size and the vector $\boldsymbol{a}$ is

$$\boldsymbol{a} = \boldsymbol{x}_{k-1} - 2\mu C\left(\boldsymbol{x}_{k-1} - \boldsymbol{c}\right) \tag{8}$$

Analyzing the above equation, we can find that the ISTA algorithm has the same solution with the C-PRIME algorithm in the case of $\mu = \frac{1}{2C}$.

Similar to the C-PRIME algorithm, the update of $\boldsymbol{x}_k$ in the ISTA method is employed at the previous value $\boldsymbol{x}_{k-1}$. In the following section, we will consider another given quantity $\boldsymbol{\eta}$ which may or may not be equal to $\boldsymbol{x}_{k-1}$. According to Taylor series expansion and the proximal regularization theorem [14], for a given point $\boldsymbol{\eta}$, a quadratic approximation of $F(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$ can be written as

$$Q_L(\boldsymbol{x}, \boldsymbol{\eta}) = f(\boldsymbol{\eta}) + \langle \boldsymbol{x} - \boldsymbol{\eta}, \nabla f(\boldsymbol{\eta})\rangle + \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{\eta}\|^2 + g(\boldsymbol{x}) \tag{9}$$

where L plays the role of a step and $\nabla f(\cdot)$ is the complex gradient vector. Then, we have

$$\boldsymbol{x}^k = \arg\min_{\boldsymbol{x}} \left\{Q_L(\boldsymbol{x}, \boldsymbol{\eta})\right\} \tag{10}$$

Discarding the constant term about $\boldsymbol{\eta}$, the optimization function (10) is simplified as

$$\boldsymbol{x}^k = \arg\min\left\{g(\boldsymbol{x}) + \frac{L}{2}\left\|\boldsymbol{x} - \left(\boldsymbol{\eta} - \frac{1}{L}\nabla f(\boldsymbol{\eta})\right)\right\|^2\right\} \tag{11}$$

As mentioned above, we know that $f(\boldsymbol{x}) = C\|\boldsymbol{x} - \boldsymbol{c}\|_2^2$ and then can get

$$\nabla f(\boldsymbol{\eta}) = 2C(\boldsymbol{\eta} - \boldsymbol{c}) \tag{12}$$

After that, $\boldsymbol{x}^k$ can be represented as

$$\boldsymbol{x}^k = \arg\min\left\{\lambda\|\boldsymbol{x}\|_1 + \frac{L}{2}\left\|\boldsymbol{x} - \left[\boldsymbol{\eta} - \frac{2C}{L}(\boldsymbol{\eta} - \boldsymbol{c})\right]\right\|^2\right\} \tag{13}$$

Furthermore, according to the soft thresholding method [16], we have

$$\boldsymbol{x}^k = e^{j\,\mathrm{ang}(\boldsymbol{b})} \odot \max\left\{|\boldsymbol{b}| - \frac{\lambda}{L}, 0\right\} \tag{14}$$

where

$$\boldsymbol{b} = \boldsymbol{\eta} - \frac{2C}{L}(\boldsymbol{\eta} - \boldsymbol{c}) \tag{15}$$

Then, if $\boldsymbol{\eta} = \boldsymbol{x}^{k-1}$, substituting (5) into (15) and simplifying it, we have

$$\boldsymbol{b} = \boldsymbol{x}^{k-1} - \frac{2}{L}\boldsymbol{A}^H\left(\boldsymbol{A}\boldsymbol{x}^{k-1} - \sqrt{\boldsymbol{y}} \odot e^{j\,\mathrm{ang}(\boldsymbol{A}\boldsymbol{x}^{k-1})}\right) \tag{16}$$

In this case, the solution $\boldsymbol{x}$ depends on step L rather than parameter C. Here, we call the algorithm as G-PRIME. It is interesting that we obtain the same solution of the problem (3) as the C-PRIME algorithm in the case of $L = 2C$ but from a totally different gradient theorem. Moreover, the C-PRIME method can be regarded as a special case of the proposed G-PRIME in the case of $\boldsymbol{\eta} = \boldsymbol{x}^{k-1}$. It can be seen from the above analysis that the update of $\boldsymbol{x}^k$ is only employed on $\boldsymbol{x}^{k-1}$ in the case of $\boldsymbol{\eta} = \boldsymbol{x}^{k-1}$, which is the same as the C-PRIME method. The G-PRIME is tabulated in Algorithm 1.

Algorithm 1: G-PRIME algorithm

Input: $\boldsymbol{A}, \boldsymbol{y}, \lambda, K$

Step 1. Initial $\boldsymbol{x}^0 \leftarrow$ random complex vector
Choose $L = 2 * \lambda_{max}(\boldsymbol{A}^H \boldsymbol{A})$

for $k = 1, \cdots, K$ **do**

Step 2. Determine $\boldsymbol{b}$ by (15)

Step 3. Update $\boldsymbol{x}^k$ by

$$\boldsymbol{x}^k = e^{j\,\mathrm{ang}(\boldsymbol{b})} \odot \max\left\{|\boldsymbol{b}| - \tfrac{\lambda}{L}, 0\right\}$$

end for

Output: $\boldsymbol{x}^K$.

IV. Simulation Results

In this section, we investigate the performance of the proposed algorithm and compare it with the existing ones including C-PRIME [13] and ISTA [15]. It should be noted that the original ISTA algorithm in [15] is used to tackle the general linear inverse problem. In this paper, combining the model of the C-PRIME algorithm, the ISTA technique can solve the phase retrieval problem, which is abbreviated as the ISTA-PRIME algorithm.

In the following experiments, we assume that the measurement matrix $\boldsymbol{A}$ is standard complex Gaussian distributed, corrupted with real-valued additive white Gaussian noise and the original complex signal is generated randomly. The length N of the original complex signal is set as 128 with sparsity level $P = 8$ and the number of measurements is 120. In following simulations, the signal-to-noise ratio is SNR=25dB. The parameter C and regularization penalty factor λ in all tested methods are set as $C = \lambda_{\max}(\boldsymbol{A}^H \boldsymbol{A})$ and $\lambda = 0.1$, respectively. We assign step size $L = 2C$ for our proposed G-PRIME algorithm unless specified otherwise. For the ISTA-PRIME algorithm, the iterative step size μ should satisfy $\mu \in (0, 1/||\boldsymbol{A}^H \boldsymbol{A}||]$ [15]. The other parameters are initialized as in Algorithm 1.

Firstly, for the proposed G-PRIME algorithm, in order to compare the magnitudes of the recovered signal and the original signal intuitively, the magnitude curves are shown in Fig. 1 at the 200th iteration. It is observed that the nonzero values in the recovered signal are almost the same as those in the original signal, which proves that the G-PRIME algorithm can recover the magnitude information successfully. Furthermore, to verify the recovering ability of phase information, Fig. 2 plots the recovered signal and the original signal at iteration numbers $k = 1, 20, 200$, in which we can observe the iteration process of the proposed G-PRIME algorithm. It can be seen that the recovered signal is a random complex vector at the first iteration and the position of recovered signal is already close to that of the original signal after 200 iterations, which confirms that the phase recovery ability of the G-PRIME algorithm is also excellent.

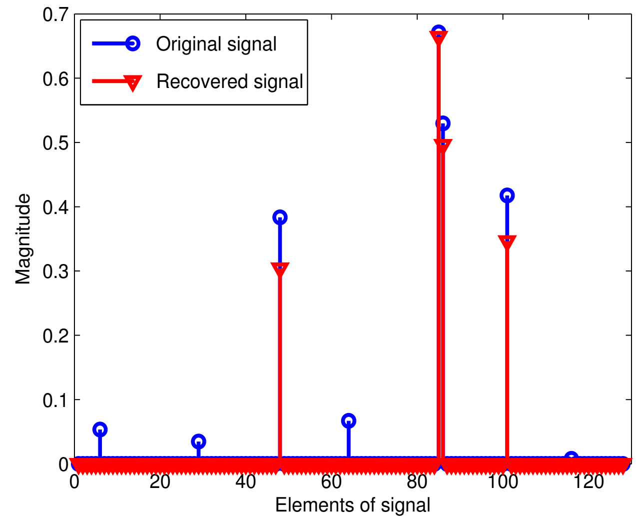

Fig. 1. Magnitudes of original and recovered signals.

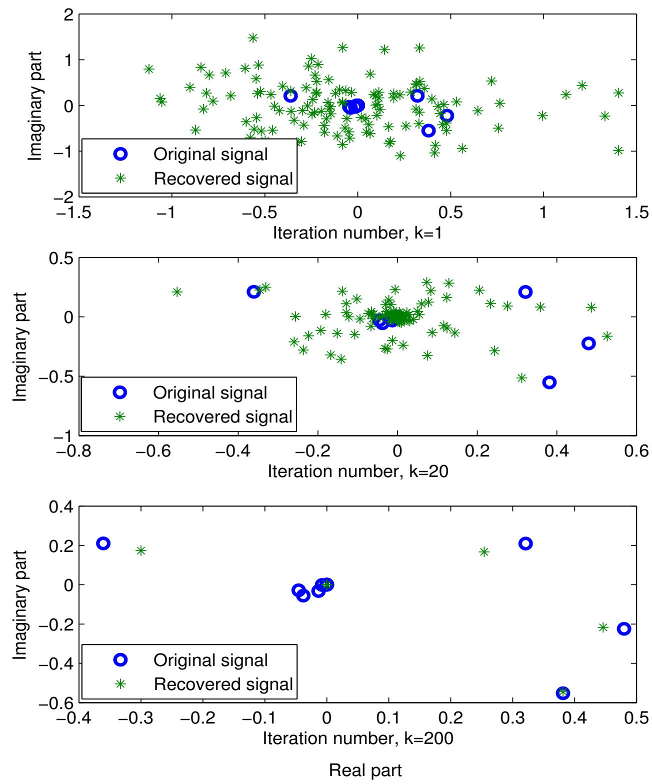

Fig. 2. Original and recovered signals.

Now we compare the MSE performance of the proposed algorithm with those of the ISTA-PRIME ($\mu = 0.1, 1/||\boldsymbol{A}^H \boldsymbol{A}||$) and C-PRIME algorithms and the MSE results are shown in Fig. 3, where all the MSE curves decrease gradually. As mentioned in Section III, the G-PRIME algorithm has the same solution as the C-PRIME algorithm in the case of $L = 2C$. So the MSE curve of C-PRIME is not shown in Fig. 3. It is obvious that the G-PRIME ($L = 2C$), C-PRIME and ISTA-PRIME algorithms have the same steady-state value

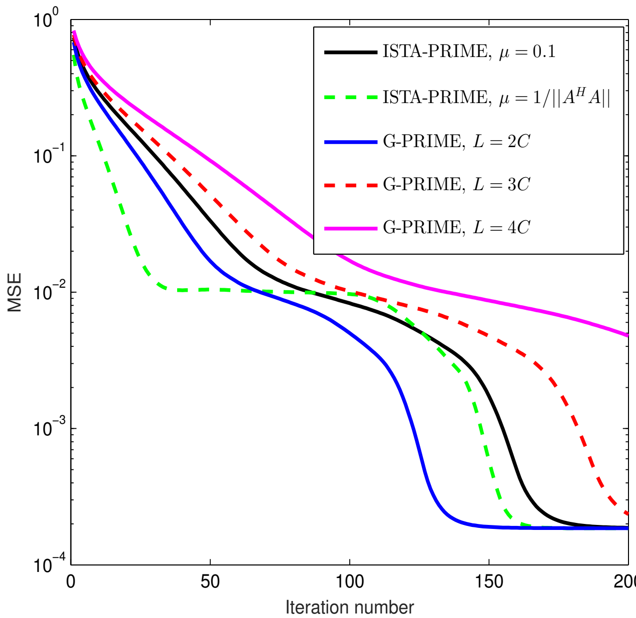

Fig. 3. MSE versus iteration number.

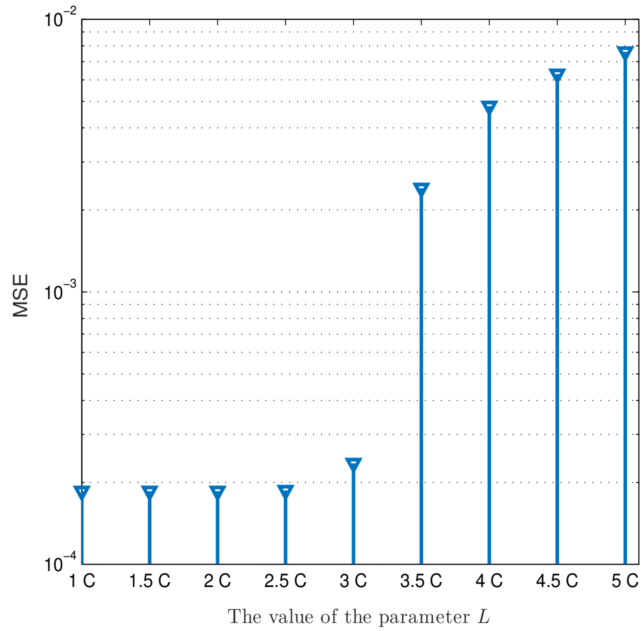

Fig. 4. MSE of the G-PRIME algorithm versus step L.

2×10^{-4}. Furthermore, the G-PRIME $(L = 2C)$ and C-PRIME algorithms converge when the iteration number is close to 150 and the ISTA-PRIME algorithm approach convergence when the iteration number reaches 170-190. Fig. 3 also depicts that the G-PRIME algorithm has the slower convergence rate in the cases of $(L = 2C$ and $L = 3C)$.

Then, we consider the MSE performance of the G-PRIME algorithm under different steps. The MSE values of G-PRIME versus parameter L from $1C$ to $5C$ are shown in Fig. 4, where the MSE values are almost the same in the interval $[1C, 2.5C]$. Specifically, the MSE values close to 2×10^{-4}. Moreover, we observe that as the L increases, the MSE values become higher rapidly when the step is in the interval $[3C, 5C]$.

V. Conclution

A undersampled phase retrieval algorithm based on the gradient framework have been proposed. To solve the non-convex ill-posed phase retrieval problem, the G-PRIME technique is employed to solve a quadratic approximation of the original problem. In this work, we have proved the ISTA algorithm has the same solution with the C-PRIME algorithm when a suitable step is choosed. Also, the C-PRIME method can be regarded as a special case of the proposed G-PRIME algorithm. Numerical results have confirmed that the proposed algorithm has excellent phase recovery ability.

References

[1] K. Huang, Y. C. Eldar, and N. D. Sidiropoulos, "Phase retrieval from 1D Fourier measurements: convexity, uniqueness, and algorithms," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6105-6117, Dec. 2016.

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

[3] R. W. Gerchberg, and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237-246, 1972.

[4] E. J. Candès, Y. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 199-225, Feb. 2013.

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

[6] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery frommagnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241-1274, Aug. 2013.

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

[8] T. Qiu, P. Babu, and D. P. Palomar, "PRIME: Phase retrieval via Majorization-Minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5174-5186, Oct. 2016.

[9] G. Wang, L. Zhang, G. B. Giannakis, M. Akcakaya, and J. Chen, Sparse phase retrieval via truncated amplitude flow," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 479-491, Jan. 2018.

[10] X. Li, and V. Voroninski, "Sparse signal recovery from quadratic measurements via convex programming," *SIAM Journal on Mathematical Analysis*, vol. 45, no. 5, pp. 3019-3033, Sep. 2013.

[11] Y. Shechtman, A. Beck, and Y. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 928-938, Feb. 2014.

[12] S. Mukherjee, and C. S. Seelamantula, "Fienup algorithm with sparsity constraints: Application to frequency-domain optical-coherence tomography," *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4659-4672, Sep. 2014.

[13] T. Qiu, and D. P. Palomar, "Undersampled sparse phase retrieval via majorization-minimization," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5957-5969, Nov. 2017.

[14] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate $O(1/k^2)$," *Dokl. Akad. Nauk SSSR*, vol. 269, pp. 543-547, 1983.

[15] A. Beck, and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183-202, Mar. 2009.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, vol. 3 no. 1, pp. 1-122, Jan. 2011.