Phase Retrieval via Second-Order Nonsmooth Optimization

Zhong Zhuang [1] **Gang Wang** [1] **Yash Travadi** [2] **Ju Sun** [3]

Abstract

We propose a new formulation for Fourier phase retrieval, and develop a second-order optimization method for solving it. The new method is remarkably stable and enjoys substantially improved convergence compared to popular Fourier phase retrieval methods.

1. Introduction

Phase retrieval (PR) is the problem of recovering a complex-valued matrix $\boldsymbol{X} \in \mathbb{C}^{n \times n}$ from its Fourier magnitudes $\boldsymbol{Y} = |\mathcal{F}(\boldsymbol{X})|^2 \in \mathbb{R}_+^{m \times m}$ (typically $m \geq n$, i.e., oversampled Fourier transform and $|\cdot|^2$ is applied element-wise). PR arises in diverse areas of scientific imaging, where the imaging process can be modeled as the Fourier transform. Due to physical limitations, however, practical imaging detectors can only record the Fourier magnitudes but not the phases (Bendory et al., 2017; Fannjiang & Strohmer, 2020).

At best, $\boldsymbol{X}$ can only be recovered from $\boldsymbol{Y}$ up to: i) a global phase, i.e., $\boldsymbol{X}e^{i\theta}$ for an unknown $\theta \in [0, 2\pi)$; ii) shift of $\boldsymbol{X}$ (if there are zero boundary rows or columns); and, iii) 2D flipping of $\boldsymbol{X}$, all due to properties of the Fourier transform. Recovery up to these intrinsic symmetries is possible for generic $\boldsymbol{X}$'s when $m \geq 2n - 1$ (Hayes, 1982; Bendory et al., 2017), which we always assume in this paper.

For actual computation, PR is often set up in $\mathbb{C}^{m \times m}$, instead of $\mathbb{C}^{n \times n}$. Oversampled Fourier transform on $\boldsymbol{X}$ amounts to padding $\boldsymbol{X}$ with $\mathbf{0}$ blocks to make it size $m \times m$ and then performing order $m \times m$ 2D Fourier transform (see Section 2.1 for details). Denote the padded matrix as $\boldsymbol{Z} \in \mathbb{C}^{m \times m}$, then $|\mathcal{F}(\boldsymbol{Z})|^2 = \boldsymbol{Y}$ and projection of $\boldsymbol{Z}$ on the padded blocks are $\mathbf{0}$, inducing two constraints on $\boldsymbol{Z}$. Popular methods for PR, including hybrid input-output (HIO) (Fienup, 1982), relaxed averaged alternating reflections (RAAR) (Bauschke et al., 2002; Luke, 2004), as well as difference map (DM) (Elser et al., 2007), are all variants of proximal (projection) splitting methods to find an intersection point of the two, sometimes with one or two additional, constraint sets (Marchesini, 2006; Combettes & Pesquet, 2011; Luke, 2020).

Unfortunately, these popular methods are sensitive to optimization hyper-parameters, and bad choices may lead to stagnation or sluggish convergence. Even if the parameter setting is appropriate, these methods typically require a large number of iterations to converge to an acceptable solution.[1] To address these challenges, a two-stage procedure is often implemented in practice: one of these methods is first run for a limited number of iterations to find a good initialization, and then local descent methods such as the celebrated error-reduction method (Gerchberg & Saxton, 1972; Fienup, 1982) are deployed for rapid refinement to the initial solution. Overall, the resulting hybrid algorithm tends to converge faster, but deciding a reliable switching point is tricky in practice and requires expert knowledge or extra computations—the quality of the initialization is critical to the success of the algorithm.

In this paper, we propose a new method for PR. Our method is based on a novel constrained formulation for PR, which is solved using the augmented Lagrangian method (ALM). The resulting nonsmooth primal subproblem is solved by an efficient quasi-Newton method, which leads to the fast convergence of our method in practice. Preliminary comparison with HIO demonstrates that our new method requires less parameter tuning but produces much more reliable and consistent recovery performance. Remarkably, our method exhibits order-of-magnitude faster convergence than HIO, on par with another second-order algorithm based on a saddle point formulation which tends to be more involved technically (Marchesini, 2007; Tripathi et al., 2015).

Another motivation for developing alternative, and hopefully more robust, PR methods is theoretical. As discussed above, popular methods for PR all fall under projection methods that tackle feasibility formulations of PR. This is surprisingly narrow when compared to the wide variety of effective formulations and numerical algorithms for other practical

[1]Department of Electrical and Computer Engineering, University of Minnesota [2]School of Statistics, University of Minnesota [3]Department of Computer Science and Engineering, University of Minnesota. Correspondence to: Zhong Zhuang <zhuan143@umn.edu>, Ju Sun <jusun@umn.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

[1]When $\boldsymbol{X}$ is known to be real-valued and have non-negative entries, an additional non-negative constraint is often enforced which can substantially improve the convergence.

problems, and also sets a considerably high entry point for theoretical development (Fannjiang & Zhang, 2020; Levin & Bendory, 2019; Barnett et al., 2018). Our vastly different formulation and method offer a new opportunity here.

2. New Formulation and Method for PR

2.1. Background

Let $\boldsymbol{X} \in \mathbb{C}^{n\times n}$. Our PR model is $\boldsymbol{Y} = |\mathcal{F}(\boldsymbol{X})|^2 \in \mathbb{R}^{m\times m}$, where $|\cdot|^2$ is applied element-wise. We assume $m \geq 2n-1$ and $\mathcal{F}$ denotes the $m\times m$ 2D discrete Fourier transform, i.e., for any $\boldsymbol{Z} \in \mathbb{C}^{m\times m}$, $\mathcal{F}(\boldsymbol{Z}) = \boldsymbol{F}\boldsymbol{Z}\boldsymbol{F}^\top$, where $\boldsymbol{F} = [e^{-i\frac{2\pi}{m}(i-1)(j-1)}]_{ij}/\sqrt{m}$ is the $m\times m$ normalized discrete Fourier matrix. With slight abuse of notation, we write $\mathcal{F}(\boldsymbol{X})$ to mean

$$\mathcal{F}(\boldsymbol{X}) = \boldsymbol{F}\begin{bmatrix}\boldsymbol{X} & \boldsymbol{0}_{n\times(m-n)} \\ \boldsymbol{0}_{(m-n)\times n} & \boldsymbol{0}_{(m-n)\times(m-n)}\end{bmatrix}\boldsymbol{F}^\top \tag{1}$$

i.e., $\boldsymbol{X}$ is padded with $\boldsymbol{0}$ blocks to have size $m\times m$ before the transform is performed.

In view of the oversampling operator Eq. (1), we define a linear operator $\mathcal{A}: \mathbb{C}^{m\times m} \to \mathbb{C}^{m^2-n^2}$ so that for any $\boldsymbol{Z} \in \mathbb{C}^{m\times m}$, $\mathcal{A}(\boldsymbol{Z})$ collects elements of $\boldsymbol{Z}$—except for the top-left $n\times n$ block—by the column order into a length-(m^2-n^2) vector. So PR can be naturally phrased as a feasibility problem[2]

$$\text{find } \boldsymbol{Z} \in \mathbb{C}^{m\times m} \text{ s.t. } \underbrace{|\mathcal{F}(\boldsymbol{Z})|^2 = \boldsymbol{Y}}_{\mathcal{M}},\ \underbrace{\mathcal{A}(\boldsymbol{Z}) = \boldsymbol{0}}_{\mathcal{S}} \tag{2}$$

where $\mathcal{M}$ and $\mathcal{S}$ are the magnitude and support constraints, respectively.

Projections onto $\mathcal{M}$ and onto $\mathcal{S}$ both admit simple solutions

$$\mathcal{P}_{\mathcal{M}}(\boldsymbol{Z}) = \mathcal{F}^{-1}\left(\sqrt{\boldsymbol{Y}} \odot \frac{\mathcal{F}(\boldsymbol{Z})}{|\mathcal{F}(\boldsymbol{Z})|}\right) \tag{3}$$

$$\mathcal{P}_{\mathcal{S}}(\boldsymbol{Z}) = \begin{cases} Z_{ij} & i\leq n \text{ and } j \leq n \\ 0 & \text{otherwise}\end{cases} \tag{4}$$

where $\odot$ denotes the pointwise (Hadamard) product. This makes projection-type methods practical. Although $\mathcal{S}$ is a convex set, $\mathcal{M}$ is a nonconvex set. In practice, the naive version of alternating projection does not work for PR (Gerchberg & Saxton, 1972; Fienup, 1982), motivating development of the modified versions: HIO, RAAR, and DM that work reasonably well on natural images (Marchesini, 2006).

Here, we highlight HIO, which is application of the famous Douglas-Rachford splitting method (Lindstrom & Sims, 2020; Bauschke et al., 2002) to a reformulation of Eq. (2)

$$\min_{\boldsymbol{Z}\in\mathbb{C}^{m\times m}} \delta_{\mathcal{M}}(\boldsymbol{Z}) + \delta_{\mathcal{S}}(\boldsymbol{Z}) \tag{5}$$

where $\delta_\Pi(\cdot)$ is the set indicator function which assumes 0 for input inside Π and $+\infty$ otherwise. One can also perform a variable-splitting to Eq. (5) and obtain a constrained form:

$$\min_{\boldsymbol{Z},\boldsymbol{W}\in\mathbb{C}^{m\times m}} \delta_{\mathcal{M}}(\boldsymbol{W}) + \delta_{\mathcal{S}}(\boldsymbol{Z}) \text{ s.t. } \boldsymbol{Z} = \boldsymbol{W}. \tag{6}$$

Applying the alternating direction method of multiplier (ADMM) to Eq. (6) also reproduces HIO (Wen et al., 2012).[3]

2.2. New formulation & second-order methods

The constraint $\mathcal{A}(\boldsymbol{Z}) = 0$ is relatively simple, and so we leave it there and arrive at an alternative form of Eq. (2)

$$\min_{\boldsymbol{Z}\in\mathbb{C}^{n\times n}} \delta_{\mathcal{M}}(\boldsymbol{Z}) \text{ s.t. } \mathcal{A}(\boldsymbol{Z}) = \boldsymbol{0}.$$

As will be clear later, removing an indicator function relative to Eq. (5) allows us to employ ALM. Introducing $\boldsymbol{W}$ such that $\boldsymbol{W} = \boldsymbol{Z}$ and turning it into a penalty term, we obtain

$$\min_{\boldsymbol{Z},\boldsymbol{W}} \frac{1}{2}\|\boldsymbol{Z}-\boldsymbol{W}\|_F^2 + \delta_{\mathcal{M}}(\boldsymbol{W}) \text{ s.t. } \mathcal{A}(\boldsymbol{Z}) = \boldsymbol{0} \tag{7}$$

which is our new formulation for PR. Now we apply ALM to Eq. (7). We first form the augmented Lagrangian

$$\mathcal{L}_\rho(\boldsymbol{Z},\boldsymbol{W},\boldsymbol{\Lambda}) \doteq \frac{1}{2}\|\boldsymbol{Z}-\boldsymbol{W}\|_F^2 + \delta_{\mathcal{M}}(\boldsymbol{W}) + \frac{\rho}{2}\left\|\mathcal{A}(\boldsymbol{Z}) + \frac{\boldsymbol{\Lambda}}{\rho}\right\|_F^2. \tag{8}$$

The key to deploying ALM is solving the primal subproblem

$$\min_{\boldsymbol{Z},\boldsymbol{W}} \mathcal{L}_\rho(\boldsymbol{Z},\boldsymbol{W},\boldsymbol{\Lambda}) \tag{9}$$

for a fixed $\boldsymbol{\Lambda}$, for which many algorithm choices are available. Here, we focus on second-order methods, hoping to achieve fast convergence for the ALM. Since

$$\min_{\boldsymbol{Z},\boldsymbol{W}} \mathcal{L}(\boldsymbol{Z},\boldsymbol{W}) = \min_{\boldsymbol{Z}}\left(\min_{\boldsymbol{W}} \mathcal{L}(\boldsymbol{Z},\boldsymbol{W})\right) \doteq \min_{\boldsymbol{Z}} \mathcal{G}(\boldsymbol{Z})$$

we have that

$$\mathcal{G}(\boldsymbol{Z}) \doteq \frac{1}{2}\|\boldsymbol{Z}-\mathcal{P}_{\mathcal{M}}(\boldsymbol{Z})\|_F^2 + \frac{\rho}{2}\left\|\mathcal{A}(\boldsymbol{Z}) + \frac{\Lambda}{\rho}\right\|_F^2.$$

[2] One may wonder why the problem is not directly formulated in $\mathbb{C}^{n\times n}$, then only a single constraint $|\mathcal{F}(\boldsymbol{X})|^2 = \boldsymbol{Y}$ is involved. The reason is that no effective methods have been discovered for solving it in $\mathbb{C}^{n\times n}$ to date.

[3] The equivalence of Douglas-Rachford splitting and ADMM on convex problems has been known for decades (Gabay, 1983). This result as proved in (Wen et al., 2012) builds on a specific argument. Very recently, a formal equivalence between the Douglas-Rachford splitting and ADMM has been established for general nonconvex problems (Themelis & Patrinos, 2020).

As a distance function, $\|\boldsymbol{Z} - \mathcal{P}_{\mathcal{M}}(\boldsymbol{Z})\|_F^2$ is globally Lipschitz and hence Clarke subdifferentiable (Clarke, 1990). It can be verified that the Wirtinger (as $\boldsymbol{Z}$ is complex-valued) subdifferentiable

$$\frac{1}{2}(\boldsymbol{Z} - \mathcal{P}_{\mathcal{M}}(\boldsymbol{Z})) \subset \partial_{\boldsymbol{Z}} \frac{1}{2}\|\boldsymbol{Z} - \mathcal{P}_{\mathcal{M}}(\boldsymbol{Z})\|_F^2. \tag{10}$$

Thus, it holds that

$$\frac{1}{2}(\boldsymbol{Z} - \mathcal{P}_{\mathcal{M}}(\boldsymbol{Z}) + \rho\mathcal{A}^*\mathcal{A}(\boldsymbol{Z}) + \mathcal{A}^*(\boldsymbol{\Lambda})) \subset \partial_{\boldsymbol{Z}}\mathcal{G}(\boldsymbol{Z}).$$

We are ready to solve the nonsmooth $\min_{\boldsymbol{Z}} \mathcal{G}(\boldsymbol{Z})$ using the L-BFGS quasi-Newton method;[4] for implementation, we use the off-the-shelf package TensorLab 3.0[5].

Quasi-Newton methods were designed for solving smooth problems. However, numerous recent works have empirically demonstrated its power in solving nonsmooth problems where gradients are replaced by subgradients (Lewis & Overton, 2012; Curtis et al., 2016; Asl & Overton, 2020). Our success here contributes another example.

We also draw inspiration from recent breakthroughs in applying semi-smooth Newton methods to solve large-scale SDPs and certain convex machine learning problems (Yang et al., 2015; Li et al., 2018). There, under the ALM framework, the primal problem is solved by a semi-smooth Newton method based on surrogate Hessian derived from generalized Jacobian and the second-order information helps achieve super-linear convergence on certain structured convex problems. Here, our investigation is purely empirical and our Hessian approximation is formed using the L-BFGS mechanism, but we are solving the difficult nonconvex PR problem.

2.3. Saddle-point optimization & others

To the best of our knowledge, the only available second-order method for PR was developed in Marchesini (2007), which considers the following objective function

$$\mathcal{L}(\boldsymbol{Z}) = \|\mathcal{P}_{\mathcal{M}}\boldsymbol{Z} - \boldsymbol{Z}\|_F^2 - \|\mathcal{P}_{\mathcal{S}}\boldsymbol{Z} - \boldsymbol{Z}\|_F^2.$$

There $\mathcal{L}(\boldsymbol{Z})$ is minimized over $\boldsymbol{Z}_{\mathcal{S}}$, i.e., the part of $\boldsymbol{Z}$ on the support $\mathcal{S}$, and maximized over $\boldsymbol{Z}_{\mathcal{S}^c}$, i.e., the part of $\boldsymbol{Z}$ off the support, leading to a minimax saddle-point optimization problem. The saddle problem is solved via a gradient descent-ascent framework, and the step sizes are optimized by a second-order method; see also Tripathi et al. (2015).

Another possibility is combining Douglas-Rachford with quasi-Newton methods (Themelis & Patrinos, 2020); we are not aware of this new method applied to PR yet.

[4]Proximal-Newton methods (Lee et al., 2014) can be considered too, which may also be competitive. We prefer the current method, as it can be easily extended when there are additional regularization terms on $\boldsymbol{Z}$, e.g., $\|\boldsymbol{Z}\|_1$ for promoting sparsity.

[5]Available online: https://www.tensorlab.net/.

3. Experiments

3.1. Setup

Algorithms. We compared our method with HIO and saddle-point optimization (SPO) (Marchesini, 2007), all initialized with $\boldsymbol{Z}_0 = \mathcal{F}^{-1}(\sqrt{\boldsymbol{Y}} \odot e^{i\boldsymbol{\Theta}})$, where $\boldsymbol{\Theta}$ is a random phase matrix. The extra variable $\boldsymbol{W}$ in our method was initialized with $\boldsymbol{W}_0 = \mathcal{P}_{\mathcal{S}}(\boldsymbol{Z}_0)$.

- **HIO**: We experimented with 5 different relaxation parameters: $\beta \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$; the algorithm stops when $\|\sqrt{\boldsymbol{Y}} - |\mathcal{F}(\boldsymbol{Z}/\|\boldsymbol{Z}\|_F)|\|_F / \|\sqrt{\boldsymbol{Y}}\|_F \leq 10^{-6}$, or 10^5 iterations is reached.
- **SPO**: We used the default parameters[6]. The algorithm terminates if $\mathcal{L}_M(\boldsymbol{Z}) = \|\mathcal{P}_{\mathcal{M}}\boldsymbol{Z} - \boldsymbol{Z}\|_F^2 \leq 10^{-4}$, or the iteration count reaches 10^5.
- **Ours**: We generated ρ uniformly at random from $[0, 0.1]$ at every iteration. The algorithm terminates if $\|\boldsymbol{Z} - \boldsymbol{W}\|_F \leq 10^{-6}$, or it runs 10^4 iterations.

Data. We tested 22 natural images with varying resolutions, some of which are shown below.

3.2. Performance and robustness

Table 1. Number of PR failures on the 22 test images

IMAGE/SIZE	25	50	100	200	300
HIO $\beta = 0.5$	1	3	4	2	4
HIO $\beta = 0.6$	2	3	3	3	2
HIO $\beta = 0.7$	3	2	4	4	4
HIO $\beta = 0.8$	3	1	4	4	2
HIO $\beta = 0.9$	3	2	3	3	3
ALM	0	2	2	3	2
SPO	3	3	2	1	2

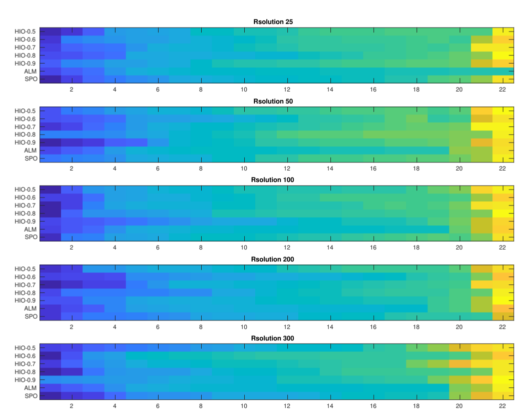

Figure 1. Estimation errors

[6]We obtained a copy of the original codes from the author.

Performance of HIO is known to be sensitive to the β parameter, and hence we tested 5 different values as specified above. We compared these 7 setting on 22 images. If the distance between the recovered image $\widehat{\boldsymbol{X}}$ and the true image $\boldsymbol{X}$ (corrected for the symmetries) are small than 0.05, a recovery success is declared; otherwise, it is counted as a failure. Table 1 reports the number of failures for the 7 methods, across 5 different image resolutions: 25×25 through 300×300. Overall, our ALM algorithm and SPO are the two most stable methods across all images and resolutions, even though we have not carefully tuned the parameters.

Relative to ALM, SPO is sensitive to initialization. Sometimes, we need to repeat the random phase initialization several times to obtain a success. The results reported above only count the ultimate successes. In contrast, we do not need to perform the repeated initialization for our method. So our method is more reliable than SPO and HIO in terms of overall recovery performance and robustness.

3.3. Convergence

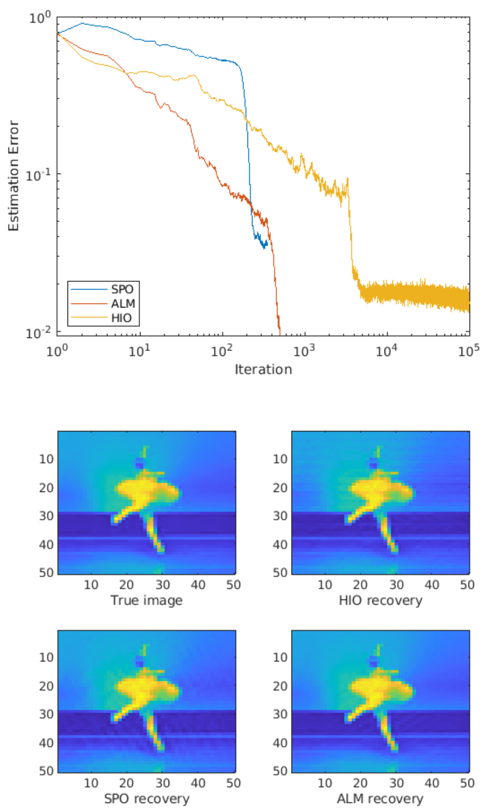

Figure 2. An easy case: evolution of the estimation error (top); and recovered images (bottom).

We also compared the three PR methods in terms of convergence performance. Here, we present the trajectories of the estimation error versus iteration count, as well as the recovered images by the three methods, for an easy case and a hard case, depicted in Fig. 2 and Fig. 3, respectively. For the easy one, both SPO and our ALM algorithm converged to a reasonable solution after several hundred iterations, but HIO took up to 10^5 iterations to find a solution of lower quality. Further, our ALM was faster than S in this case. For the hard case, all methods struggled for a long time to find an acceptable solution. In particular, our ALM converged in less than 10^3 iterations, while SPO took about 10^4 iterations, and HIO was again very slow. Clearly, our ALM is much faster than the other two methods here. Moreover, SPO encountered many plateaus, which can cause premature stopping if the stopping criterion is not carefully set. In a nutshell, our ALM has consistent merits over HIO and SPO in terms of convergence speed as well as recovery performance on both the easy and hard cases.

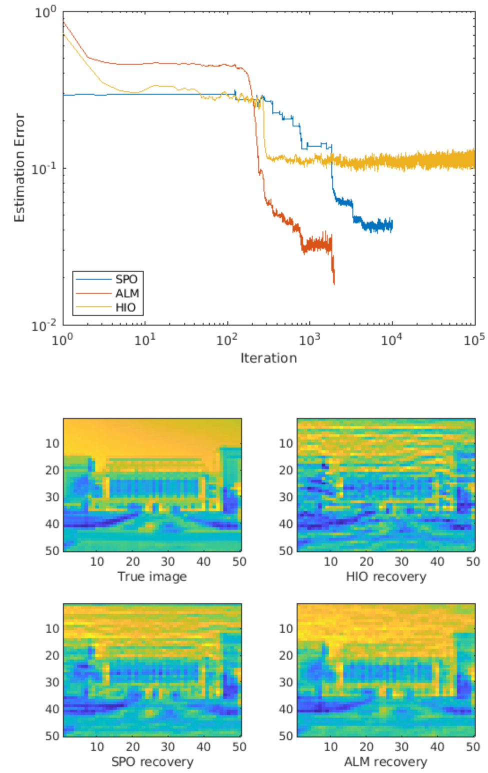

Figure 3. A hard case: evolution of the estimation error (top); and recovered images (bottom).

4. Discussion

In this paper, we developed a new constrained nonconvex optimization formulation for Fourier phase retrieval, and proposed an efficient second-order algorithm based on the augmented Lagrangian method. Preliminary tests using natural images showcase the merits of the proposed method relative to several popular Fourier phase retrieval methods.

Acknowledgments

GW was supported by NSF grant 1901134. We thank Dr. Stefano Marchesini for his insightful feedback and suggestions on this work, and for sharing his codes for the paper Marchesini (2007) with us. We also acknowledge the Minnesota Supercomputing Institute at the University of Minnesota for providing resources that contributed to the research results reported in this paper.

References

Asl, A. and Overton, M. L. Behavior of limited memory bfgs when applied to nonsmooth functions and their nesterov smoothings. *arXiv:2006.11336*, 2020.

Barnett, A., Epstein, C. L., Greengard, L., and Magland, J. Geometry of the phase retrieval problem. *arXiv:1808.10747*, 2018.

Bauschke, H. H., Combettes, P. L., and Luke, D. R. Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization. *Journal of the Optical Society of America A*, 19(7):1334–1345, 2002.

Bendory, T., Beinert, R., and Eldar, Y. C. Fourier phase retrieval: Uniqueness and algorithms. In *Compressed Sensing and its Applications*, pp. 55–91. Springer, 2017.

Clarke, F. H. *Optimization and Nonsmooth Analysis*. Society for Industrial and Applied Mathematics, 1990.

Combettes, P. L. and Pesquet, J.-C. Proximal splitting methods in signal processing. In *Springer Optimization and Its Applications*, pp. 185–212. Springer New York, 2011. doi: 10.1007/978-1-4419-9569-8_10.

Curtis, F. E., Mitchell, T., and Overton, M. L. A BFGS-SQP method for nonsmooth, nonconvex, constrained optimization and its evaluation using relative minimization profiles. *Optimization Methods and Software*, 32(1):148–181, July 2016.

Elser, V., Rankenburg, I., and Thibault, P. Searching with iterated maps. *Proceedings of the National Academy of Sciences*, 104 (2):418–423, Jan. 2007.

Fannjiang, A. and Strohmer, T. The numerics of phase retrieval. *arXiv:2004.05788*, 2020.

Fannjiang, A. and Zhang, Z. Fixed point analysis of Douglas–Rachford splitting for ptychography and phase retrieval. *SIAM Journal on Imaging Sciences*, 13(2):609–650, Jan. 2020.

Fienup, J. R. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769, 1982.

Gabay, D. Chapter IX applications of the method of multipliers to variational inequalities. In *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, pp. 299–331. Elsevier, 1983.

Gerchberg, R. W. and Saxton, W. O. A practical algorithm for the determination of the phase from image and diffraction phase pictures. *Optik*, (237), 1972.

Hayes, M. The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30 (2):140–154, 1982.

Lee, J. D., Sun, Y., and Saunders, M. A. Proximal Newton-type methods for minimizing composite functions. *SIAM Journal on Optimization*, 24(3):1420–1443, Jan. 2014.

Levin, E. and Bendory, T. A note on Douglas-Rachford, gradients, and phase retrieval. *arXiv:1911.13179*, 2019.

Lewis, A. S. and Overton, M. L. Nonsmooth optimization via quasi-Newton methods. *Mathematical Programming*, 141(1-2): 135–163, Feb, 2012.

Li, X., Sun, D., and Toh, K.-C. A highly efficient semismooth newton augmented Lagrangian method for solving Lasso problems. *SIAM Journal on Optimization*, 28(1):433–458, Jan. 2018.

Lindstorm, S. B. and Sims, B. Survey: Sixty years of Douglas-Rachford. *Journal of the Australian Mathematical Society*, pp. 1–38, Feb. 2020.

Luke, D. R. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21(1):37, 2004.

Luke, D. R. Proximal methods for image processing. In *Topics in Applied Physics*, pp. 165–202. Springer International Publishing, 2020. doi: 10.1007/978-3-030-34413-9_6.

Marchesini, S. A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments*, 78, 2006.

Marchesini, S. Phase retrieval and saddle point optimization. *Journal of the Optical Society of America A*, 24:3289–3296, 2007.

Themelis, A. and Patrinos, P. Douglas–Rachford splitting and ADMM for Nonconvex Optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, Jan. 2020.

Tripathi, A., Leyffer, S., Munson, T., and Wild, S. M. Visualizing and improving the robustness of phase retrieval algorithms. *Procedia Computer Science*, 51:815–824, 2015.

Wen, Z., Yang, C., Liu, X., and Marchesini, S. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems*, 28(11):115010, 2012.

Yang, L., Sun, D., and Toh, K.-C. SDPNAL+: A majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Mathematical Programming Computation*, 7(3):331–366, May 2015.