SEQUENCE OF LINEAR PROGRAM FOR ROBUST PHASE RETRIEVAL

Seonho Kim and Kiryung Lee

The Ohio State University, USA

ABSTRACT

We consider a robust phase retrieval problem that aims to recover a signal from its absolute measurements corrupted with sparse noise. The least absolute deviation (LAD) provides a robust estimation against outliers. However, the corresponding optimization problem is nonconvex. We propose an "unregularized" iterative convexification approach to LAD through a sequence of linear programs (SLP). We provide a non-asymptotic convergence analysis under the standard Gaussian assumption of the measurement vectors. The SLP algorithm, when suitably initialized, linearly converges to the ground truth at optimal sample complexity up to a numerical constant. Furthermore, SLP empirically outperforms existing methods that provide a comparable performance guarantee.

***Index Terms*—** phase retrieval, outliers, least absolute deviation, linear program, convex optimization

1. INTRODUCTION

Phase retrieval refers to the recovery of signals from magnitude-only measurements. It arises in numerous applications including X-ray crystallography, diffraction and array imaging, and optics [1–4]. We consider the robust phase retrieval from the measurements that are corrupted with sparse noise. For example, such a scenario arises in imaging applications [5].

A suite of methods designed for plain phase retrieval has been adapted to address outliers. For instance, anchored regression [6] and PhaseMax [7] formulate phase retrieval as a linear program when provided with an initial estimate. RobustPhaseMax [8] modifies these methods to offer robust estimation by introducing an auxiliary variable to describe the outliers. In another example, Reshaped Wirtinger Flow (RWF) [9] and Amplitude Flow [10] follow a subgradient descent approach. Median-RWF [11] is a variant of these methods tailored for robust phase retrieval. Specifically, Median-RWF identifies and excludes outliers in each iteration by median-based thresholding on the consistency of the current estimate to the measurements.

A different approach based on the *least absolute deviation* (LAD) has been proposed in a later work [12]. LAD has been a popular approach to regression with outliers. The LAD formulation of phase retrieval from squared magnitude measurements is cast as a nonconvex optimization problem. The prox-linear algorithm implements an iterative convexification of LAD through linearization of the convex measurement model and regularization on updates relative to the previous iterate.

All of these works guarantee the exact recovery by their estimators under comparable conditions on sample complexity and outliers. However, the empirical performance of Median-RWF and the prox-linear algorithm has shown significant improvement over RobustPhaseMax. In this paper, we present an unregularized iterative convexification method for LAD in robust phase retrieval. This method achieves a performance guarantee comparable to previous estimators but yields superior empirical results.

Notations : Boldface lowercase letters denote column vectors. We use $\|\cdot\|_1$ and $\|\cdot\|_2$ to denote the ℓ_1 norm and the Euclidean norm respectively. For brevity, the shorthand notation $[n]$ denotes the set $\{1,\dots,n\}$ for $n \in \mathbb{N}$.

This work was supported in part by NSF CAREER Award CCF-1943201.

2. PROBLEM FORMULATION AND ALGORITHM

We consider phase retrieval from magnitude measurements corrupted with sparse noise. Let $\boldsymbol{x}_\star \in \mathbb{R}^d$ and $\{\boldsymbol{a}_i\}_{i=1}^m \subset \mathbb{R}^d$ denote the unknown ground-truth signal and the known measurement vectors. Then the measurements are written as

$$b_i = \begin{cases} \xi_i & \text{if } i \in I_{\text{out}} \\ |\langle \boldsymbol{a}_i, \boldsymbol{x}_\star \rangle| & \text{if } i \in [m] \setminus I_{\text{out}} \end{cases} \tag{1}$$

where $I_{\text{out}} \subset [m]$ collects the unknown indices of outliers.

The LAD estimator given the measurements $\{(\boldsymbol{a}_i, b_i)\}_{i=1}^m$ as in (1) minimizes the cost function given by

$$\ell(\boldsymbol{x}) := \frac{1}{m} \sum_{i=1}^m \left| |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle| - b_i \right|. \tag{2}$$

The cost function in (2) is written as $\ell(\boldsymbol{x}) = (h \circ F)(\boldsymbol{x})$, where $h : \mathbb{R}^m \to \mathbb{R}$ and $F : \mathbb{R}^d \to \mathbb{R}^m$ are respectively defined by

$$h(\boldsymbol{z}) = \|\boldsymbol{z}\|_1 = \sum_{i=1}^m |z_i| \tag{3}$$

and

$$F(\boldsymbol{x}) = (|\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle| - b_i)_{i=1}^m. \tag{4}$$

A similar LAD estimator for a smooth measurement model has been studied by Duchi and Ruan [12]. They adopted an iterative algorithm formulated in the spirit of a Gauss-Newton method for convex composite optimization [13]. Inspired by their algorithm, we consider a similar iterative algorithm given by

$$\boldsymbol{x}_{k+1} = \underset{\boldsymbol{x} \in \mathbb{R}^d}{\operatorname{argmin}}\ h(F(\boldsymbol{x}_k) + F'(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)) \qquad (5)$$

where $F'(\boldsymbol{x}_k) \in \mathbb{R}^{m \times d}$ denotes the Clarke's generalized Jacobian matrix at $\boldsymbol{x}_k$ [14]. The algorithm by (5) is different from the approach [12] in the following two aspects: i) F is not a smooth mapping; ii) There is no explicit constraint on the amount of update $\|\boldsymbol{x} - \boldsymbol{x}_k\|_2$. Note that the optimization problem in each update by (5) is convex. Therefore, we consider this approach as iterative convexification of the LAD formulation for robust phase retrieval.

Furthermore, given h and F respectively defined by (3) and (4), the optimization in (5) can be solved by a simple linear program. First, note that (5) is written as

$$\boldsymbol{x}_{k+1} = \underset{\boldsymbol{x} \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{m} \left|\operatorname{sign}(\langle \boldsymbol{a}_i, \boldsymbol{x}_k \rangle)\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle - b_i\right|. \qquad (6)$$

Then, by introducing auxiliary variables $\boldsymbol{t} := [t_1; \dots; t_m] \in \mathbb{R}^m$, (6) becomes equivalent to

$$\begin{aligned} \underset{\boldsymbol{x} \in \mathbb{R}^d, (t_i)_{i=1}^m}{\operatorname{minimize}} \quad & \langle \boldsymbol{t}, \boldsymbol{1}_m \rangle \\ \text{subject to} \quad & t_i \geq \operatorname{sign}(\langle \boldsymbol{a}_i, \boldsymbol{x}_k \rangle)\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle - b_i, \\ & t_i \geq -\operatorname{sign}(\langle \boldsymbol{a}_i, \boldsymbol{x}_k \rangle)\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle + b_i, \quad \forall i \in [m], \end{aligned} \qquad (7)$$

where $\boldsymbol{1}_m = [1, \dots, 1]^\mathsf{T} \in \mathbb{R}^m$. In other words, since solving (5) at each iteration executes a linear program, the iterative algorithm in (5) involves a sequence of linear programs. For this reason, we use the acronym SLP to refer to our method.

2.1. Comparison with the prox-linear method

Duchi and Ruan [12] considered a similar LAD with the cost function $\ell_{\mathrm{quad}} = h \circ F_{\mathrm{quad}}$ where $F_{\mathrm{quad}}(\boldsymbol{x}) = (|\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|^2 - |b_i|^2)_{i=1}^m$. Note that F_{quad} is of class $\mathcal{C}^1$. Their prox-linear algorithm is a variation of the Gauss-Newton method [13] so that $\boldsymbol{x}_{k+1}$ is updated as a minimizer of

$$h(F_{\mathrm{quad}}(\boldsymbol{x}_k) + F'_{\mathrm{quad}}(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)) + \lambda\|\boldsymbol{x} - \boldsymbol{x}_k\|_2^2$$

with respect to $\boldsymbol{x}$, where $F'_{\mathrm{quad}}(\boldsymbol{x}_k) \in \mathbb{R}^{m \times d}$ is the Jacobian matrix at $\boldsymbol{x}_k$ and $\lambda > 0$ is a penalty parameter for the regularizer. The original Gauss-Newton method instead considered a constraint on $\|\boldsymbol{x} - \boldsymbol{x}_k\|_2$. In the prox-linear algorithm, the regularizer ensures that estimates are updated in close proximity to the current one. This prevents overshooting in a descent direction and favors the approximation of the local linearization. As we demonstrate in the following section, suitably initialized SLP even without any such regularizer shows monotone linear convergence to the ground truth.

Due to the use of the quadratic measurement model and the squared ℓ_2 regularizer, each iteration of the prox-linear algorithm corresponds to a quadratic program (QP). In contrast, SLP avoids using the quadratic functions, allowing us to utilize a simple linear program (LP) instead of QP. This results in significant computational savings, as QP has notably higher computational costs compared to LP [15, 16]. To accelerate solving those QPs, they also proposed the proximal operator graph splitting method. However, the resulting algorithm is still more computationally expensive than SLP as it involves the multiplication and eigenvalue decomposition of unstructured dense matrices in each iteration.

3. THEORETICAL RESULTS OF SLP ALGORITHM

[17]

In this section, we present the convergence of SLP with respect to the following outlier model.

Assumption 1: The outliers are supported on arbitrarily fixed set I_{out} with $|I_{\mathrm{out}}| = \eta m$ but their magnitudes $|\xi_i|$ can be adversarial.

Theorem 3.1. *Let $\delta \in (0, 1)$ be fixed. Suppose that Assumption 1 holds with $\eta \in [0, 1/4]$ and $\{\boldsymbol{a}_i\}_{i=1}^m$ are independent copies of $\boldsymbol{a} \sim \mathrm{Normal}(\boldsymbol{0}, \boldsymbol{I}_d)$. Then there exist absolute constants $C, c > 0$ and $\nu_\eta \in (0, 1)$ depending only on η, for which the following statement holds for all $\boldsymbol{x}_\star \in \mathbb{R}^d$ with probability at least $1 - \exp(-cd)$: If an initial estimate $\boldsymbol{x}_0$ obeys $\operatorname{dist}(\boldsymbol{x}_0, \boldsymbol{x}_\star) \leq \sin(2/25)\|\boldsymbol{x}_\star\|_2$ and*

$$m \geq Cd, \qquad (8)$$

then the sequence $(\boldsymbol{x}_k)_{k \in \mathbb{N}}$ by SLP satisfies

$$\operatorname{dist}(\boldsymbol{x}_k, \boldsymbol{x}_\star) \leq \nu_\eta^k \operatorname{dist}(\boldsymbol{x}_0, \boldsymbol{x}_\star), \qquad (9)$$

where $\operatorname{dist}(\boldsymbol{x}, \boldsymbol{x}_\star) = \min_{\alpha \in \{\pm 1\}} \|\boldsymbol{x} - \alpha \boldsymbol{x}_\star\|_2$.

Theorem 3.1 establishes a local linear convergence of SLP, implying that without any regularizer, an estimate by SLP converges to the ground truth by avoiding overshooting while staying within a neighborhood of the ground truth. As shown in (8), the sample complexity m to recover the ground truth, $\boldsymbol{x}_\star$, is linearly dependent on d. This sample complexity is comparable with the results of [8, 11, 12] for exact recovery of ground truth.

While the estimators have comparable results with respect to sample complexity and initialization, there exist subtle differences in the assumptions on outliers and the computational costs among these estimators.

First, we compare the degree of adversary in assumptions on outliers. Hand and Voroninski [8] assumed the highest adversary so that the support and magnitudes of sparse noise can be chosen depending on all measurement vectors in their analysis of RobustPhaseMax. Assumption 1, which coincides with the assumption by Zhang et al. [11] for Median-RWF, considers an arbitrarily fixed support of sparse noise but the magnitudes can be adversarial. Duchi and Ruan [12] used the lowest adversary so that the support of sparse noise is chosen randomly. They considered two distinct models on the dependence of magnitudes of sparse noise on measurement vectors.

More importantly, the faction of outliers is not clearly specified in the analysis of RobustPhaseMax and Median-RWF. They only presented that there exists a numerical constant so that if the fraction of outliers is below the threshold, then the exact recovery is achieved with high probability. In contrast, the analysis of the prox-linear algorithm [12] and SLP (Theorem 3.1) demonstrated that these methods can tolerate the fraction of outliers up to $1/4$. As shown in the numerical results in Section 4, RobustPhaseMax showed inferior empirical performance. It provides exact recovery with high probability for only a small fraction of outliers η. In contrast, other methods, such as Median-RWF, prox-linear algorithm, and SLP, improve the tolerance level for outliers.

Moreover, the theoretical results can also be compared in terms of their computational costs. RobustPhaseMax is solved using a single linear program, making it the most computationally efficient method. However, this comes at the expense of its inferior empirical performance. The per-iteration cost of Median-RWF is lower than those of both the prox-linear algorithm and SLP. Yet, these latter methods offer better empirical performance. While the analysis of the prox-linear algorithm indicates faster quadratic convergence compared to the linear convergence of SLP, the per-iteration cost of SLP is significantly lower. As a result, we were not able to obtain a direct comparison of computational cost in theory. Empirically, however, SLP converges faster than the prox-linear algorithm and does so with reduced computational costs, as shown in Section 4.

Finally, all convergence analysis by Theorem 3.1 and previous work [8, 11, 12] require an initialization within a neighborhood of the ground truth. The size of the basin of convergence was determined as an explicit numerical constant in [8] and up to a numerical constant in Theorem 3.1 and [11, 12]. Various initialization techniques for robust phase retrieval have been developed with performance guarantees [11, 12]. These initialization methods apply to all of the considered robust estimators (RobustPhaseMax, Median-RWF, prox-linear algorithm, and SLP) so that they provide the desired accuracy without amplifying sample complexity in the subsequent estimator.

4. NUMERICAL RESULTS

This section studies the empirical performance of SLP relative to its theoretical analysis in Theorem 3.1 and to the competing methods for robust phase retrieval. In all experiments, we generated observations and signals as follows. The measurement vectors are generated so that $\{\boldsymbol{a}_i\}_{i=1}^m \overset{i.i.d.}{\sim} \mathrm{Normal}(\boldsymbol{0}, \boldsymbol{I}_d)$, which has been employed in Theorem 3.1 and all analogous theoretical analyses of the other methods. The ground truth signal is generated as $\boldsymbol{x}_\star \sim \mathrm{Normal}(\boldsymbol{0}, \boldsymbol{I}_d)$ independently from the measurement vectors. The outlier indices are randomly selected. We consider various random and deterministic scenarios for the outlier magnitudes. This setting is slightly different from the uniform performance guarantee in Theorem 3.1. Recall that all considered methods require an initial estimate. For this purpose, we adopt the spectral method by Zhang et al. [11].

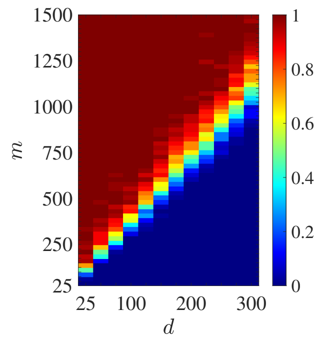
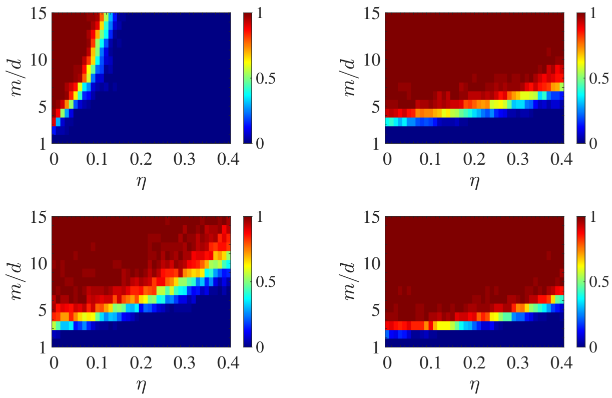


Fig. 1: Phase transition of success rate by SLP per the number of measurements m and the dimension d. The outlier measurements are randomly generated with respect to the Cauchy distribution. The fraction of outliers is fixed to $\eta = 0.25$.

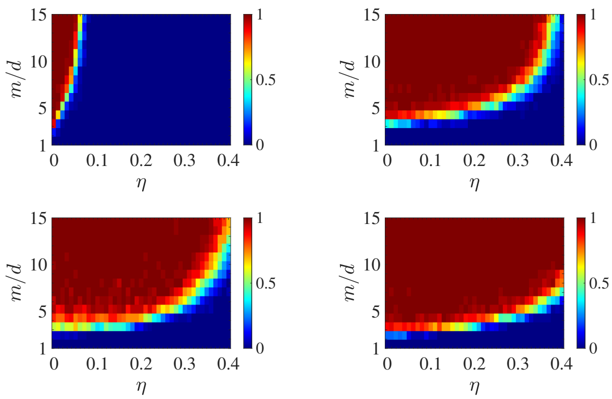
Figure 1 shows the phase transition of success and failure by SLP through Monte Carlo simulations. The success is defined by the criterion $\mathrm{dist}(\widehat{\boldsymbol{x}}, \boldsymbol{x}_\star) \le 10^{-5}$ with $\widehat{\boldsymbol{x}}$ denoting an estimate. The empirical success rate out of 50 trials is displayed. The transition occurs at the boundary where the number of measurements is proportional to the ambient dimension. This empirical result corroborates our theoretical finding in Theorem 3.1.

Next, we compare the empirical performance of SLP to RobustPhaseMax, Median-RWF, and the prox-linear algorithm. Figure 2 displays the phase transition of these methods for a range of the outlier fraction η in three scenarios of the outlier magnitudes. As η increases, m increases for the spectral method to provide a good initialization for the successful recovery of estimators. The first scenario by Zhang et al. [11] draws ξ_i from the uniform distribution on $(-d\|\boldsymbol{x}_\star\|_2^2/2, d\|\boldsymbol{x}_\star\|_2^2/2)$. In the second scenario, ξ_i is drawn from a Cauchy distribution with median 0 and mean-absolute-deviation 1. The third scenario sets ξ_i to 0. The last two scenarios have been considered by Duchi and Ruan [12]. Similar to Figure 1, the success rate out of 50 trials is plotted. The ambient dimension is set to $d = 100$.

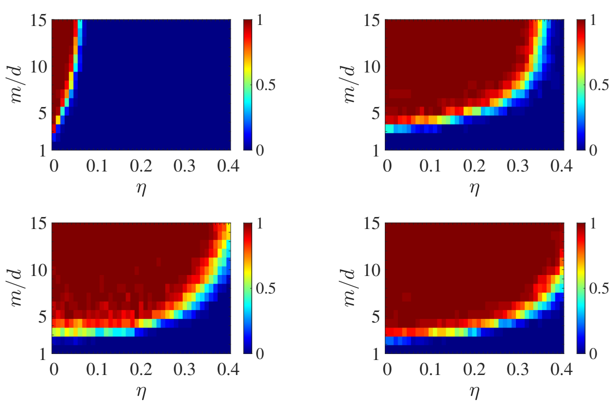
RobustPhaseMax, while providing the strongest theoretical performance guarantee, shows the worst empirical performance in the comparison. There is no consistent dominance between Median-RWF and the prox-linear algorithm.



(a) uniform distribution



(b) Cauchy distribution



(c) zero

Fig. 2: Phase transition of success rate per the measurement ratio m/d and the fraction of outliers η for various outlier magnitude models. For each outlier model: the top-left shows RobustPhaseMax, the top-right depicts Median-RWF, the bottom-left presents prox-linear method, and the bottom-right illustrates SLP.

Median-RWF outperforms the prox-linear in the first scenario, but the other way around in the other scenarios. In contrast, SLP consistently outperforms all the other methods in the three considered scenarios with a significantly lower threshold for the phase transition.

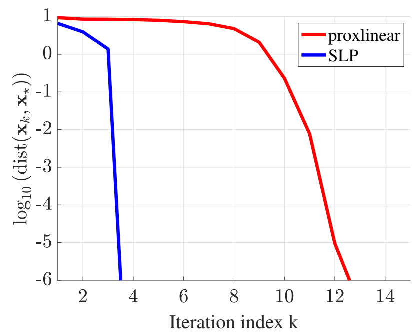


Fig. 3: Convergence of SLP and the prox-linear algorithm in the iteration count.

In the last experiment, we compare the convergence speed of SLP and the prox-linear algorithm. Here, we set $m = 1,500, d = 200$, and $\eta = 0.3$ under the Cauchy distribution scenario in Figure 2. Figure 3 illustrates how $\log_{10} \mathrm{dist}(\boldsymbol{x}_k, \boldsymbol{x}_\star)$ decays over the iteration index k. The median over 10 trials is plotted. In their theoretical analyses, the prox-linear algorithm converges faster at a quadratic rate than the linear convergence of SLP in Theorem 3.1. However, as shown in Figure 3, SLP empirically converges significantly faster than the prox-linear algorithm in the iteration count. Furthermore, due to a lower per-iteration cost of SLP than that of the prox-linear algorithm, the gap in the empirical convergence of the two methods will be more significant when it is measured in the flop count.

5. CONCLUSION

The least absolute deviation (LAD) has been a popular statistical method for regression in the presence of outliers. We consider a sequence of linear programs (SLP) to solve a LAD formulation of robust phase retrieval with the magnitude-only measurement model. As an unregularized Gauss-Newton type algorithm, SLP provides a significantly lower per-iteration cost than a similar approach known as the prox-linear algorithm [12].

We established a local convergence analysis of SPL under the standard Gaussian measurement when the support of sparse noise is arbitrarily fixed but magnitudes can be adversarial. A suitably initialized SPL converges linearly to the ground truth when the number of measurements m is proportional to the signal length d and the outlier fraction is up to $1/4$. The performance guarantee holds with high probability for all signals. This theoretical result is comparable to existing prior art in the literature. Furthermore, the numerical results show that SPL outperforms the existing guaranteed methods for various outlier models.

6. REFERENCES

[1] A. Walther, "The question of phase retrieval in optics," *Optica Acta: International Journal of Optics*, vol. 10, no. 1, pp. 41–49, 1963.

[2] O. Bunk, A. Diaz, F. Pfeiffer, C. David, B. Schmitt, D. K. Satapathy, and J. F. Van Der Veen, "Diffractive imaging for periodic samples: retrieving one-dimensional concentration profiles across microfluidic channels," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 63, no. 4, pp. 306–314, 2007.

[3] A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," *Inverse Problems*, vol. 27, no. 1, p. 015005, 2010.

[4] 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 Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.

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

[6] S. Bahmani and J. Romberg, "Phase retrieval meets statistical learning theory: A flexible convex relaxation," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 252–260.

[7] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2675–2689, 2018.

[8] P. Hand and V. Voroninski, "Corruption robust phase retrieval via linear programming," *arXiv preprint arXiv:1612.03547*, 2016.

[9] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, "A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms," *Journal of Machine Learning Research*, 2017.

[10] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 773–794, 2017.

[11] H. Zhang, Y. Chi, and Y. Liang, "Median-truncated nonconvex approach for phase retrieval with outliers," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7287–7310, 2018.

[12] J. C. Duchi and F. Ruan, "Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval," *Information and Inference: A Journal of the IMA*, vol. 8, no. 3, pp. 471–529, 2019.

[13] J. V. Burke and M. C. Ferris, "A Gauss–Newton method for convex composite optimization," *Mathematical Programming*, vol. 71, no. 2, pp. 179–194, 1995.

[14] F. Clarke, *Optimization and Nonsmooth Analysis*, ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1990.

[15] J. van den Brand, "A deterministic linear program solver in current matrix multiplication time," in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2020, pp. 259–278.

[16] E. John and E. A. Yıldırım, "Implementation of warm-start strategies in interior-point methods for linear programming in fixed dimension," *Computational Optimization and Applications*, vol. 41, no. 2, pp. 151–183, 2008.

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