Robust Phase Retrieval with Non-Convex Penalties

Reza Mirzaeifard*, Naveen K. D. Venkategowda§, Stefan Werner*
*Dept. of Electronic Systems, Norwegian University of Science and Technology-NTNU, Norway
§Department of Science and Technology, Linköping University, Sweden
E-mails: {reza.mirzaeifard, stefan.werner}@ntnu.no, naveen.venkategowda@liu.se

***Abstract*—This paper proposes an alternating direction method of multiplier (ADMM) based algorithm for solving the sparse robust phase retrieval with non-convex and non-smooth sparse penalties, such as minimax concave penalty (MCP). The accuracy of the robust phase retrieval, which employs an l_1 based estimator to handle outliers, can be improved in a sparse situation by adding a non-convex and non-smooth penalty function, such as MCP, which can provide sparsity with a low bias effect. This problem can be effectively solved using a novel proximal ADMM algorithm, and under mild conditions, the algorithm is shown to converge to a stationary point. Several simulation results are presented to verify the accuracy and efficiency of the proposed approach compared to existing methods.**

***Index Terms*—Robust phase retrieval, MCP, non-convex and non-smooth optimization, sparse learning, ADMM**

I. Introduction

Retrieving phases is a challenging problem in several fields, including X-ray crystallography [1], astronomy [2], quantum tomography [3], optics [4], and microscopy [5]. In these applications, amplitude or intensity can be measured, but not phase, because measuring phase is difficult and expensive. Phase retrieval, therefore, involves reconstructing the signal from phaseless measurements.

Initial phase retrieval efforts focused on non-convex alternating projection algorithms for Fourier transform models, namely the error reduction techniques. Gerchberg and Saxton (GS) [6] and Fienup [7] are well-known error reduction methods. There has been a growing interest in optimization-based approaches for phase retrieval in recent years. PhaseLift [8], PhaseCut [9], and PhaseMax [10], for example, reformulate phase retrieval as a convex problem to take advantage of convex optimization. Alternatively, a non-convex formulation can be obtained by minimizing an intensity-based loss function, such as Wirtinger flow (WF) [11], and its variations [12,13]. Furthermore, robust phase retrieval algorithms have been proposed in recent years to deal with outliers, e.g., [14,15]. Since outliers are inevitable due to sensor anomalies and recording errors, robust phase retrieval is critical in practical applications.

In many phase retrieval applications, such as image processing [16], the signal acquired are sparse in nature. Hence, many phase retrieval methods have been generalized to their sparse versions [17,18]. An approach, for example, is to use a threshold method to provide sparsity solutions for k-sparse signals [19]. However, the value k needs to be known, which is not practical in many cases. As an alternative, the objective function can likewise be penalized by adding an l_1 penalty [20]. However, the l_1 norm has a few drawbacks, including its inability to enforce sufficient sparsity and its tendency to penalize large coefficients excessively [21]. For this reason, we need a sparsity penalty that can distinguish between zero and non-zero coefficients of the model. To that end, one may consider using non-convex and non-smooth penalties, such as minimax concave penalty (MCP) [22] and smoothly clipped absolute deviation (SCAD) [23].

While non-convex and non-smooth penalties may improve estimation accuracy in many problems [24,25], because of their non-convexity and non-smoothness, they complicate optimization. For penalized robust penalized phase retrieval, in particular, proposing an optimization algorithm is more challenging due to its non-convex and non-smooth nature. The alternative direction method of multiplier (ADMM) is a potent optimization algorithm that can handle penalized problems efficiently and effectively in parallel. However, since robust phase retrieval penalized with non-convex and non-smooth penalties does not exhibit Lipschitz differentiability, it is unclear whether the existing ADMM-based non-convex optimization algorithms [26]–[29] can reach a stationary point. Recently, a Moreau envelope-based augmented Lagrangian method (MEAL) [30] has been developed for constraint optimization of non-smooth and non-convex functions that satisfy implicit Lipschitz differentiability or implicit bounded subgradient conditions. These conditions are more general than Lipschitz differentiability conditions and include a group of non-smooth functions as well. However, MEAL and its variations [30] cannot handle penalized robust penalized phase retrieval optimization because it has two non-smooth parts. It is still necessary to develop an ADMM-based algorithm that can handle fully non-smooth and non-convex two-part problems.

In this paper, we propose an ADMM based algorithm for solving robust phase retrieval penalized with non-convex and non-smooth sparse penalties, e.g., MCP. Since the robust phase retrieval objective function and MCP are non-convex and non-smooth, demonstrating the convergence of ADMM is a challenging issue. We are able to demonstrate the convergence of our ADMM method by using an innovative technique based on Moreau envelope functions and the presence of a implicit Lipschitz subgradient property for the robust phase retrieval objective function. Additionally, we validate our theoretical claims through numerical simulations. Our results show that

This work was supported in part by the Research Council of Norway.

the proposed algorithm outperforms state-of-the-art methods.

II. Problem Formulation

In real phase retrieval, the aim is to recover a signal $\mathbf{x}$ based on the amplitude of its linear measurements:

$$y_i = |\langle \mathbf{a}_i^{\mathrm{T}}, \mathbf{x}\rangle|^2, \quad \forall i \in \{1, \cdots, N\} \tag{1}$$

where $\mathbf{a}_i \in \mathbb{R}^M$ are observations, $y_i \in \mathbb{R}$ are the measured intensities, and $\mathbf{x} \in \mathbb{R}^M$ is the target signal we wish to recover. In practice, however, real-world applications exhibit noise, which may corrupt measurements and lead to incorrect solutions. In order to deal with such outliers, a robust formulation can be used as follows:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{i=1}^{N} l_i(\mathbf{x}), \quad \text{with} \quad l_i(\mathbf{x}) = |y_i - |\langle \mathbf{a}_i^{\mathrm{T}}, \mathbf{x}\rangle|^2|. \tag{2}$$

In order to improve the quality of inference, the robust phase retrieval loss function can be penalized in a suitable manner based on *a priori* information about the model coefficients. Once the penalty $P_{\lambda,\zeta}(\mathbf{x})$ is incorporated, the optimization problem takes the form:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{N} \sum_{i=1}^{N} l_i(\mathbf{x}) + \omega P_{\lambda,\zeta}(\mathbf{x}), \tag{3}$$

where ω is a regularization parameter. By using an auxiliary variable $\mathbf{z}$, (3) can be rewritten as:

$$\begin{aligned} \hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad & \frac{1}{N} \sum_{i=1}^{N} l_i(\mathbf{x}_i) + \omega P_{\lambda,\zeta}(\mathbf{z}). \\ \text{subject to} \quad & \mathbf{z} = \mathbf{x}_i, \ \ \forall i \in \{1, \cdots, N\} \end{aligned} \tag{4}$$

l_1 norm is popular as a sparse penalty function, but it leads to estimation bias. To overcome this limitation, in this paper, we investigate using an MCP as $P_{\lambda,\zeta}(\mathbf{w}) = \sum_{p=1}^{P} g_{\lambda,\zeta}(w_p)$ [23], which is a non-convex and non-smooth function, to provide sparsity in the estimated signal. The definition of MCP is given by [23]:

$$g_{\lambda,\zeta}^{\mathrm{MCP}}(w_p) = \begin{cases} \lambda|w_p| - \frac{w_p^2}{2\zeta}, & |w_p| \leq \zeta\lambda \\ \frac{\zeta\lambda^2}{2}, & |w_p| > \zeta\lambda \end{cases} \quad \text{for } \zeta \geq 1. \tag{5}$$

It is noteworthy that MCP is weakly convex with $\rho \geq \frac{1}{\zeta}$, meaning that there is a function convex function $h(x)$ such that $h(x) = g_{\lambda,\zeta}^{\mathrm{MCP}}(w_p) + \frac{\rho}{2}\|w_p\|_2^2$ [24]. Additionally, each $l_i(.)$ is also weakly convex with $\rho = 2\|\mathbf{a}_i\|_2^2$ [15].

In the next section, we propose an ADMM based algorithm for solving the optimization problem (4).

III. Proposed Method

For solving (4) with an ADMM algorithm, we write the associated augmented Lagrangian function as follows:

$$\begin{aligned} \mathcal{L}_{\rho_u}(\mathbf{X}, \mathbf{z}, \mathbf{U}) = & \frac{1}{N} \sum_{i=1}^{N} l_i(\mathbf{x}_i) + \omega P_{\lambda,\zeta}(\mathbf{z}) \\ & + \sum_{i=1}^{N} \left(\mathbf{u}_i^{\mathrm{T}}(\mathbf{z} - \mathbf{x}_i) + \frac{\rho_u}{2}\|\mathbf{x}_i - \mathbf{z}\|_2^2 \right), \end{aligned} \tag{6}$$

where $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_N]$ is dual variable, $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$, and ρ_u is the dual penalty parameter. According to our understanding, existing ADMM-based approaches [26,28] cannot guarantee convergence to a stationary point in the absence of Lipschitz differentiability and convexity of the objective function. In the following, we derive a proximal ADMM that guarantees convergence in the above mentioned settings.

By defining an auxiliary variable $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_N]$, the proximal augmented Lagrangian can be defined as:

$$\Psi_{\rho_u,\rho_h}(\mathbf{X}, \mathbf{z}, \mathbf{U}, \mathbf{H}) = \mathcal{L}_{\rho_u}(\mathbf{X}, \mathbf{z}, \mathbf{U}) + \frac{\rho_h}{2}||\mathbf{X} - \mathbf{H}||_2^2, \tag{7}$$

where ρ_h is a penalty parameter. The proximal term in (7) assists in obtaining a convergence result using the Moreau envelope function. The $(m+1)$th iteration of our proposed proximal ADMM algorithm can be expressed as:

$$\mathbf{z}^{(m+1)} = \arg\min_{\mathbf{z}} \Psi_{\rho_u,\rho_h}(\mathbf{X}^{(m)}, \mathbf{z}, \mathbf{U}^{(m)}, \mathbf{H}^{(m)}), \tag{8a}$$

$$\mathbf{X}^{(m+1)} = \arg\min_{\mathbf{X}} \Psi_{\rho_u,\rho_h}(\mathbf{X}, \mathbf{z}^{(m+1)}, \mathbf{U}^{(m)}, \mathbf{H}^{(m)}), \tag{8b}$$

$$\mathbf{H}^{(m+1)} = \mathbf{H}^{(m)} - \eta(\mathbf{H}^{(m)} - \mathbf{X}^{(m+1)}), \tag{8c}$$

$$\mathbf{U}^{(m+1)} = \mathbf{U}^{(m)} + \rho_u(\mathbf{z}^{(m+1)}\mathbf{1}_N^{\mathrm{T}} - \mathbf{X}^{(m+1)}), \tag{8d}$$

where $\eta \in (0, 2)$ is a step size.

More precisely, the $\mathbf{z}$-step update in $(m+1)$th can be written as:

$$\mathbf{z}^{(m+1)} = \mathrm{Prox}_{p_{\lambda,\zeta}} \left(\frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i^{(m)} - \frac{\mathbf{u}_i^{(m)}}{\rho_u}); \frac{\omega}{N\rho_u} \right), \tag{9}$$

where $\mathrm{Prox}_h(\mathbf{w}; \kappa) = \arg\min_{\mathbf{x}} \left\{ h(\mathbf{x}) + \frac{1}{2\kappa}\|\mathbf{x} - \mathbf{w}\|_2^2 \right\}$. Moreover, for each $\mathbf{x}_i$, separate updates can be performed as follows:

$$\mathbf{x}_i^{(m+1)} = \mathrm{Prox}_{l_i} \left(\frac{\rho_u \mathbf{z} + \rho_h \mathbf{h}_i + \mathbf{u}_i}{\rho_u + \rho_h}; \frac{1}{N(\rho_u + \rho_h)} \right). \tag{10}$$

It can be shown that by considering $\mathbf{w}_1 = \mathbf{w} - \frac{2\kappa \mathbf{a}_i^{\mathrm{T}}\mathbf{w}}{2\kappa\|\mathbf{a}_i\|_2^2 + 1}\mathbf{a}_i$, $\mathbf{w}_2 = \mathbf{w} - \frac{2\kappa \mathbf{a}_i^{\mathrm{T}}\mathbf{w}}{2\kappa\|\mathbf{a}_i\|_2^2 - 1}\mathbf{a}_i$, $\mathbf{w}_3 = \mathbf{w} - \frac{\mathbf{a}_i^{\mathrm{T}}\mathbf{w} + \sqrt{y_i}}{\|\mathbf{a}_i\|_2^2}\mathbf{a}_i$, and $\mathbf{w}_4 = \mathbf{w} - \frac{\mathbf{a}_i^{\mathrm{T}}\mathbf{w} - \sqrt{y_i}}{\|\mathbf{a}_i\|_2^2}\mathbf{a}_i$, for $0 < \kappa \leq \frac{1}{2\|\mathbf{a}_i\|_2^2}$ we have:

$$\mathrm{Prox}_{l_i}(\mathbf{w}; \kappa) = \begin{cases} \mathbf{w}_1, & (\frac{\mathbf{a}_i^{\mathrm{T}}\mathbf{w}}{2\kappa\|\mathbf{a}_i\|_2^2 + 1})^2 \geq y_i \\ \mathbf{w}_2, & (\frac{\mathbf{a}_i^{\mathrm{T}}\mathbf{w}}{2\kappa\|\mathbf{a}_i\|_2^2 - 1})^2 \leq y_i \\ \mathbf{w}_3, & -\sqrt{y_i} < \frac{\mathbf{a}_i^{\mathrm{T}}\mathbf{w} + \sqrt{y_i}}{2\kappa\|\mathbf{a}_i\|_2^2} < \sqrt{y_i} \\ \mathbf{w}_4. & -\sqrt{y_i} < \frac{\mathbf{a}_i^{\mathrm{T}}\mathbf{w} - \sqrt{y_i}}{2\kappa\|\mathbf{a}_i\|_2^2} < \sqrt{y_i} \end{cases} \tag{11}$$

Finally, we can adapt the stopping criteria in [31] to our algorithm. The proposed algorithm for solving the sparse robust phase retrieval is summarized in Algorithm 1.

Algorithm 1: Proximal ADMM (PADM) for sparse robust phase retrieval

Initialize $\mathbf{z}^{(0)}$, $\mathbf{H}^{(0)}$, and $\mathbf{U}^{(0)}$ to zero vectors, $\eta \in (0, 2)$, and $\{\mathbf{x}_i^{(0)}\}_{i=1}^N$ as in [11];
repeat
  Update $\mathbf{z}^{(m+1)}$ by (9);
  for $i = 1, \ldots, N$ **do**
    Update $\mathbf{x}_i^{(m+1)}$ by (10)-(11);
  end
  Update $\mathbf{H}^{(m+1)}$ by (8c);
  Update $\mathbf{U}^{(m+1)}$ by (8d);
until *the convergence criterion is met*;

IV. Convergence Proof

In this section, a convergence analysis of the Algorithm 1 is presented. We construct our convergence proof based on the following assumption.

Assumption 1. *Each of $l_i(\cdot)$ for $0 < \kappa < \frac{1}{2\|\mathbf{a}_i\|_2^2}$ satisfies the implicit Lipschitz subgradient property, which means that $\forall \mathbf{w} \in Prox_{l_i}^{-1}(\mathbf{u}, \kappa), \forall \mathbf{w}' \in Prox_{l_i}^{-1}(\mathbf{v}, \kappa)$, there exists an $L_f > 0$ such that:*

$$\|\nabla \mathcal{M}_{l_i}(\mathbf{w}, \kappa) - \nabla \mathcal{M}_{l_i}(\mathbf{w}', \kappa)\| \leq L_f \|\mathbf{u} - \mathbf{v}\|, \quad (12)$$

in which $\mathcal{M}_h(\mathbf{w}; \kappa) = \min_{\mathbf{x}} \left\{ h(\mathbf{x}) + \frac{1}{2\kappa}\|\mathbf{x} - \mathbf{w}\|_2^2 \right\}$.

By considering that the the Moreau envelope function $\mathcal{M}_h(\mathbf{w}; \kappa)$ is smooth, the Assumption 1 can be satisfied if each $\text{Prox}_{l_i}(\mathbf{w}, \kappa)$ is locally invertible. When $M = 1$, it is easy to demonstrate that the assumption 1 holds. Based on Assumption 1 one can derive the following lemma.

Lemma 1. *For all $m \geq 1$, the following inequality is held*:

$$\rho_u^{-1}\|\mathbf{u}_i^{(m+1)} - \mathbf{u}_i^{(m)}\|_2^2 \leq \rho_u^{-1}\big((L_f + \rho_h)^2 \|\mathbf{x}_i^{(m+1)} - \mathbf{x}_i^{(m)}\| + \rho_h^2 \|\mathbf{h}_i^{(m)} - \mathbf{h}_i^{(m-1)}\|_2^2\big), \quad \forall i \in \{1, \cdots, N\}. \quad (13)$$

Proof. This can be demonstrated by adapting [30, Lemma 4]. □

The convergence proof can be illustrated by ensuring that all conditions in [30, Lemma 3] are met. There is only one increasing step in each iteration of the algorithm, which is the update of the dual update step. By considering Assumption 1, as one can see in Lemma 1, the amount of increase in the dual-update step is bound by the primal and auxiliary variables. Due to this, we can tune the parameters of the proximal augmented Lagrangian to ensure that the *sufficient decrease condition* in [30, Lemma 3] is met. Furthermore, the subgradient of the proximal augmented Lagrangian based on each of its inputs can be easily shown to be bound in every iteration, which is sufficient for validating the *bounded subgradient condition* of [30, Lemma 3]. Finally, by knowing that the robust phase retrieval function is a coercive function and adopting the [30, proposition 3], which shows that under certain conditions the proximal augmented Lagrangian is lower bounded in each step, and with the knowledge that the proximal augmented Lagrangian is a continuous function based on its inputs, we can show that the *continuity condition* of [30, Lemma 3] holds.

V. Simulation Result

This section compares the proposed proximal ADMM algorithm (PADM) for MCP penalized robust phase retrieval with sparse PhaseliftOff (SPhaseliftOff) [18], a well-known algorithm for sparse phase retrieval, and robust phase retrieval with the subgradient method (Sub) [32] in three scenarios. The relative error equal to $\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}$ is used as a performance measure, and all results were derived from averaging over 100 independent trials.

The first scenario compares the accuracy of algorithms when there are outliers in the measurements. For generating measurements, the following settings were chosen:

$$y_i = |\langle \mathbf{a}_i^{\mathrm{T}}, \mathbf{x} \rangle|^2 + \epsilon_i, \quad \forall i \in \{1, \cdots, N\} \quad (14)$$

where $\mathbf{a}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \mathbf{I})$ and $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I}) \odot \mathbf{B}_s$, in which $\odot$ represents the Hadamard product and $\mathbf{B}_s \in \mathbb{R}^{M \times M}$ is a random diagonal matrix with $s = \frac{3}{10}$ of the diagonal elements are non-zero and equal to one. In addition, $\{\epsilon_i\}_{i=1}^N$ are components of the noise sampled i.i.d from a mixture of exponential distributions in the form of $\sum_{i=1}^2 c_i \lambda_i e^{-\lambda_i v}$, for $v >= 0$, in which $c_1 = 0.9$, $c_2 = 0.1$, and $\lambda_2 = \frac{\lambda_1}{10}$. In terms of SNR ($\gamma$), λ_1 can be derived as $\lambda_1 = \sqrt{\frac{N \times 21.8 \times 10^{\frac{\gamma}{10}}}{\sum_{i=1}^N |\mathbf{a}_i^{\mathrm{T}}\mathbf{x}|^2}}$. In this setting, 10% of the measurement can be assumed as outliers. Simulation results were obtained for $(N, M) = (150, 25)$ and varying SNR values from -10 dB to 40 dB, in steps of 5 dB. Fig. 1 illustrates that PADM with MCP outperforms other methods in terms of relative error.

The second scenario compares the algorithms' performance under a noisy phase schema. The following settings were used to generate measurements:

$$y_i = |\langle \mathbf{a}_i^{\mathrm{T}}, \mathbf{x} + \boldsymbol{\epsilon}_i \rangle|^2, \quad \forall i \in \{1, \cdots, N\}, \quad (15)$$

where $\mathbf{a}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \mathbf{I})$, and $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I}) \odot \mathbf{B}_s$ with $s = \frac{3}{10}$. Also, each component of noise $\boldsymbol{\epsilon}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I})$, where $\sigma = \sqrt{\frac{\sum_{i=1}^N |\mathbf{a}_i^{\mathrm{T}}\mathbf{x}|^2}{N \times M^2 \times 10^{\frac{\gamma}{10}}}}$. Simulation results were derived for $(N, M) = (150, 25)$ and varying SNR values from 0 dB to 50 dB, in steps of 5 dB. In this scenario, PADM with MCP performs better than other methods, as shown in Fig. 2.

Under a noiseless schema, the final scenario compares the convergence rate of the algorithms. Measurements were generated with the following settings:

$$y_i = |\langle \mathbf{a}_i^{\mathrm{T}}, \mathbf{x} \rangle|^2, \quad \forall i \in \{1, \cdots, N\}, \quad (16)$$

where $\mathbf{a}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \mathbf{I})$, and $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I}) \odot \mathbf{B}_s$ with $s = \frac{1}{10}$. Simulation results were obtained for $(N, M) = (60, 10)$. In Fig. 3, PADM with MCP exhibits a faster convergence rate than Sub. Additionally, SPhaseLiftOff suffers from an inner loop and each iteration solves an ADMM algorithm, which in total is slower than our algorithm.

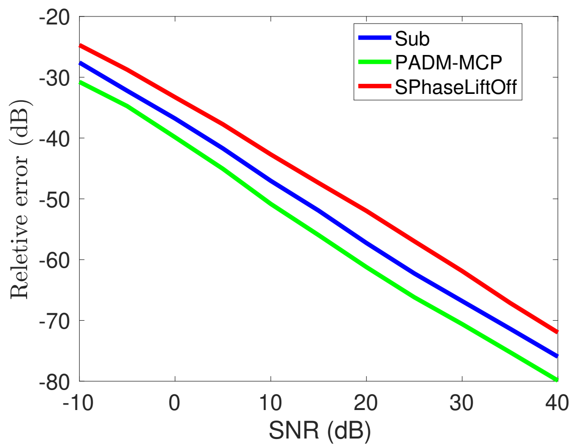

Fig. 1. Relative error vs. SNR

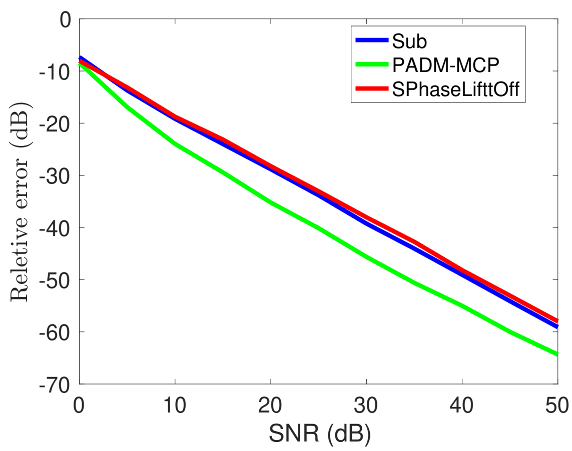

Fig. 2. Relative error vs. SNR

VI. Conclusion

This paper introduced a new proximal variant of ADMM called PADM for sparse robust phase retrieval problems penalized with non-convex penalties. In the presence of an implicit Lipschitz subgradient property for the robust phase retrieval function, our analysis demonstrated that the proposed algorithm can converge to a stationary point. According to the simulation results, the proposed algorithm with MCP penalty outperforms state-of-the-art methods in terms of accuracy and convergence speed.

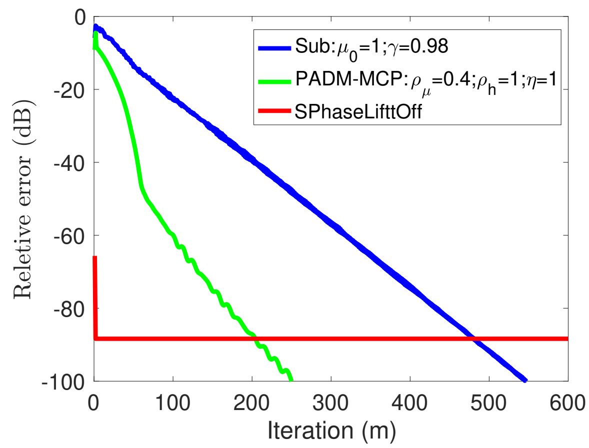

Fig. 3. Relative error vs. iteration

References

[1] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature*, vol. 400, no. 6742, pp. 342–344, July 1999.

[2] C. Fienup and J. Dainty, *Phase retrieval and image reconstruction for astronomy*, 1987, vol. 231.

[3] M. Kech and M. Wolf, "From quantum tomography to phase retrieval and back," in *IEEE International Conference on Sampling Theory and Applications*, 2015, pp. 173–177.

[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 Process. Mag.*, vol. 32, no. 3, pp. 87–109, Apr. 2015.

[5] J. Miao, T. Ishikawa, B. Johnson, E. H. Anderson, B. Lai, and K. O. Hodgson, "High resolution 3d x-ray diffraction microscopy," *Physical review letters*, vol. 89, no. 8, p. 088303, Aug. 2002.

[6] R. W. Gerchberg and W. O. Saxton, "Practical algorithm for determination of phase from image and diffraction plane pictures," *OPTIK*, vol. 35, no. 2, pp. 237–246, Nov. 1972.

[7] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.

[8] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, Mar. 2013.

[9] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Math. Program.*, vol. 149, no. 1, pp. 47–81, Dec. 2015.

[10] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2675–2689, Apr. 2018.

[11] E. J. Candes, 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.

[12] G. Wang and G. Giannakis, "Solving random systems of quadratic equations via truncated generalized gradient flow," *Adv. Neural Inf. Process. Syst.*, vol. 29, Dec. 2016.

[13] G. Wang, G. B. Giannakis, and J. Chen, "Scalable solvers of random quadratic equations via stochastic truncated amplitude flow," *IEEE Trans.Signal Process.*, vol. 65, no. 8, pp. 1961–1974, Apr. 2017.

[14] C. Qian, X. Fu, N. D. Sidiropoulos, L. Huang, and J. Xie, "Inexact alternating optimization for phase retrieval in the presence of outliers," *IEEE Trans.Signal Process.*, vol. 65, no. 22, pp. 6069–6082, Nov. 2017.

[15] 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, Sep. 2019.

[16] Y. Shechtman, Y. C. Eldar, A. Szameit, and M. Segev, "Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing," *Opt. exp.*, vol. 19, no. 16, pp. 14 807–14 822, 2011.

[17] T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow," *Ann. Stat.*, vol. 44, no. 5, pp. 2221–2251, Oct. 2016.

[18] Y. Xia and Z. Xu, "Sparse phase retrieval via phaseliftoff," *IEEE Trans. Signal Process.*, vol. 69, pp. 2129–2143, Mar. 2021.

[19] G. Wang, L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen, "Sparse phase retrieval via truncated amplitude flow," *IEEE Trans.Signal Process.*, vol. 66, no. 2, pp. 479–491, Jan. 2017.

[20] H. Ohlsson, A. Yang, R. Dong, and S. Sastry, "CPRL–an extension of

compressive sensing to the phase retrieval problem," *Adv. Neural Inf. Process. Syst.*, vol. 25, Dec. 2012.

[21] A. Rakotomamonjy, R. Flamary, and G. Gasso, "DC proximal Newton for nonconvex optimization problems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 636–647, Mar. 2015.

[22] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Annals stat.*, vol. 38, no. 2, pp. 894–942, Apr. 2010.

[23] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.

[24] R. Mirzaeifard, N. K. Venkategowda, V. C. Gogineni, and S. Werner, "ADMM for sparse-penalized quantile regression with non-convex penalties," in *30th European Signal Processing Conference*. IEEE, 2022, pp. 2046–2050.

[25] R. Mirzaeifard, V. C. Gogineni, N. K. Venkategowda, and S. Werner, "Dynamic graph topology learning with non-convex penalties," in *30th European Signal Processing Conference*. IEEE, 2022, pp. 682–686.

[26] Y. Wang, W. Yin, and J. Zeng, "Global convergence of admm in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, Jan. 2019.

[27] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, Jan. 2016.

[28] M. Yashtini, "Convergence and rate analysis of a proximal linearized ADMM for nonconvex nonsmooth optimization," *J. Glob. Optim.*, pp. 1–27, May 2022.

[29] A. Themelis and P. Patrinos, "Douglas–Rachford splitting and ADMM for nonconvex optimization: Tight convergence results," *SIAM J. Optim*, vol. 30, no. 1, pp. 149–181, Jan. 2020.

[30] J. Zeng, W. Yin, and D.-X. Zhou, "Moreau envelope augmented lagrangian method for nonconvex optimization with linear constraints," *J. Sci. Comput.*, vol. 91, no. 2, pp. 1–36, Apr. 2022.

[31] S. Boyd, N. Parikh, and E. Chu, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Now Publishers Inc, 2011.

[32] D. Davis, D. Drusvyatskiy, K. J. MacPhee, and C. Paquette, "Subgradient methods for sharp weakly convex functions," *J. Optim. Theory Appl.*, vol. 179, no. 3, pp. 962–982, Sep. 2018.