OVER-PARAMETERIZED NETWORK SOLVES PHASE RETRIEVAL EFFECTIVELY

Ji Li† *and Chao Wang*‡

† Department of Mathematics, National University of Singapore, Singapore
‡ Department of Statistics and Data Science, National University of Singapore, Singapore

ABSTRACT

Phase retrieval estimates a complex signal vector from the modulus observation of its linear projections. The problem is the core problem in numerous applications. The phase-less/sign-less observation makes it a nonconvex problem in the gradient flow-based approaches. To enable the convergence to the global solution, an initialization close to the solution is indispensable for these methods. In this paper, we propose using an over-parameterized network to represent the unknown signal to solve the problem. With the help of over-parameterization, a gradient flow-based algorithm enables locating solution with vanishing objective. We show that the introduction of the network in phase retrieval problem benefits the gradient flow method to find the optimal solution. Our proposed method outperforms the existing methods with best recovery performance. Code is publicly accessible at `https://github.com/Chilie/PR_Repara`.

Index Terms— Phase retrieval, Nonconvex optimization, Network-based re-parameterization

1. INTRODUCTION

The problem of phase retrieval arises in a numerous applications, such as astronomy [1, 2], optics [3], ptychography [4, 5], as well as coherent diffraction imaging [6]. The core problem is to estimate the complex signal $\boldsymbol{x} \in \mathbb{C}^n$ from the phase-less linear projection observations $|\boldsymbol{A}\boldsymbol{x}| = \boldsymbol{b}$. Here the modulus is an element-wise operation. These broad applications demand an efficient and effective algorithm for phase retrieval. Note that it is a nonconvex problem notoriously difficult to solve. And there are many algorithms to deal with the problem. These methods can be divided into three categories, we introduce them one by one in the following.

The alternating projection-based methods solve the problem by casting it as a set feasibility problem. These methods are built on the two projections onto the span of the measurement matrix and the object domain. Among these approaches, we mention the pioneer works of Gerchberg and Saxton [7] and Fienup [8, 9]. Other mathematical interpretations and extension of these methods include [10, 11, 12, 13, 14].

Thanks to National Natural Science Foundation of China (Grant No. 11801025) for funding.

The gradient flow-based methods propose solving the nonconvex problem via gradient descent with a good initialization. These methods are based on different loss choices: Such as intensity observation-based least squares [15], Poisson noise maximum-a-priori [16] and modulus observation based least squares [17, 18]. Note that these methods require an initialization generation scheme to produce the initialization, then they can provide a convergence guarantee. However, these methods fail to produce success recovery once the initial point is randomly chosen. The stagnation at the local stationary point is the main difficulty of the gradient flow-based methods. The dependence of initialization limits their practical applications.

The convexification-based methods exploit the Schur relaxation of quadratic problem to lift the original vector domain problem into semidefinite programming in matrix variable domain, which can be solved efficiently by standard convex routines. Two of the main approaches are PhaseLift [19, 20] and PhaseCut [21]. The dimension of the lifted problem makes the phase retrieval intractable.

The above approaches have their issues when applying to practical usage. Though some attempts to circumvent the difficulties have been proposed, they are not very effective and stable to solve the phase retrieval. The lack of phase/sign information of the linear projection is generally compensated by redundant measurements. Actually, the redundant measurement is a prerequisite to ensure the solution uniqueness up to some trivial ambiguities. It has been proven that $m = 2n - 1$ and $m = 4n$ are necessary for uniquely determining the solution to phase retrieval for a general measurement matrix $\boldsymbol{A}$ for real and complex scenarios [22, 23, 24]. For numerical algorithms, the empirical recovery performance of one algorithm with different ratios m/n differs. If competitive recovery is achieved with lower ratio m/n, it will be a huge cost saving of computation and observation time. In this paper, we propose a novel solution to phase retrieval with improved empirical recovery performance in terms of the ratio m/n.

2. OPTIMIZATION WITH REPARAMETERIZATION

In this section, we recall the intensity measurement-based least-squares formulation of the phase retrieval problem. Then we propose our network-based re-parameterization of the unknown signal $\boldsymbol{x}$. With such over-parameterization, the deep network benefits the gradient descent method to decrease the loss. Compared to the loss objective with respect to the original variable $\boldsymbol{x}$, the stagnation of optimization at the stationary point is mitigated. The success recovery rate is much improved.

2.1. Least-squares formulation

From the modulus measurement $\boldsymbol{b} = |\boldsymbol{A}\boldsymbol{x}|$, where $\boldsymbol{x} \in \mathbb{C}^n$, measurement matrix $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ and observation data $\boldsymbol{b} \in \mathbb{R}^m$, we can formulate the least-squares loss:

$$\min_{\boldsymbol{x}} \quad \mathcal{L}(\boldsymbol{x}) := \frac{1}{4m} \left\| |\boldsymbol{A}\boldsymbol{x}|^2 - \boldsymbol{b}^2 \right\|_2^2 . \tag{1}$$

Note that the objective $\mathcal{L}(\boldsymbol{x})$ is a real-valued function of complex-variable, thus the gradient computation can be derived by the Wirtinger derivative [15].

Assume the observation is noise-free and the solution to the phase retrieval problem is unique up to the trivial ambiguities, once the loss $\mathcal{L}$ is decreased to zero with machine accuracy, a solution is obtained. However, for the nonconvex problem (1), it is hard to optimize the loss towards the global optimal solution. More especially, the landscape of the nonconvex loss is generally complicated, and there may exist several local minima (a.k.a. stationary points with zero gradient and positive local Hessian), then gradient flow-based method can not escape from the local minima. The complicated landscape is the main difficulty of the nonconvex problems, so is the phase retrieval problem.

The existing literature normally start from an initial guess that is close to the global minima, which much increases the possibility of locating the solution. The initialization can be obtained by a certain spectral method [15, 25, 16] for a particular configuration of the problem: The entries of measurement matrix $\boldsymbol{A}$ are distributed as normal distribution, *i.e.*, $a_{ij} \sim \mathcal{N}(0,1)$ or $a_{ij} \sim \mathcal{N}(0,1/2)+1i\mathcal{N}(0,1/2)$ for real and complex cases respectively. Under such configuration, the convergence of gradient descent to global minima is guaranteed with the *measurement complexity* $m \geq c_0 n \log n$ where c_0 is a sufficiently large numerical constant. This configuration is widely assumed for the latter follow-up studies of gradient flow-based methods for phase retrieval. With the help of good initialization, the *computational complexity* is reduced by devising other loss objectives; see *e.g.* [16, 18]. These methods inevitably produce failed recovery once the good initialization is not available. Noted that for the Gaussian matrix $\boldsymbol{A}$, the least-squares loss (1) conveys benign landscape without spurious local minimizers when the *measurement complexity* $m \geq c_1 n \log^3 n$ where c_1 is a sufficiently large numerical constant.

In realistic setting, due to the difficulty of taking a good initialization, gradient flow-based methods are not the primary choices for phase retrieval. Towards the issue raised by the initialization, we aim to seek a gradient flow-based method with the negligible effect of the initialization on the recovery performance for phase retrieval.

2.2. Network-based re-parameterization

Motivated by the expression of the over-parameterized network and its solvability by the gradient flow method, we propose exploiting such advantage of learning over-parameterized network to solve the nonconvex phase retrieval (1). It is counter-intuitive that the nonconvexity of learning over-parameterized network via gradient based methods does not raise the convergence issue. The solution with vanishing objective can be located via gradient descent, which is illustrated and analyzed in the deep learning community; see *e.g.* [26, 27, 28]. In the following, we first introduce the re-parameterized network function, and then we formulate the loss for phase retrieval with the network-based re-parameterization.

We utilize the full connection network $G(\boldsymbol{\theta};\boldsymbol{z}_0)$ to represent the unknown $\boldsymbol{x} \in \mathbb{C}^n$, where $\boldsymbol{z}_0 \in \mathbb{C}^n$ is the fixed input of the network and $\boldsymbol{\theta}$ is the parameter weight associated to the network. Suppose there are L layers in G, then G can be expressed as

$$G(\boldsymbol{\theta};\boldsymbol{z}_0) = \boldsymbol{z}_0+\boldsymbol{W}_L(\sigma(\boldsymbol{W}_{L-1}(\cdots\sigma(\boldsymbol{W}_0\boldsymbol{z}_0+\boldsymbol{b}_0))+\boldsymbol{b}_{L-1}))+\boldsymbol{b}_L,$$

where σ is the activation function and the parameter wight $\boldsymbol{\theta}$ comprises of $\{\boldsymbol{W}_i, \boldsymbol{b}_i\}_{i=0}^{L}$. Here we adopt the skip connection structure in G for its better performance to avoid the stagnation at sub-optimal point.

With the re-parameterization, we can formulate the loss objective for phase retrieval

$$\min_{\boldsymbol{\theta}} \quad \frac{1}{4m} \left\| |\boldsymbol{A}G(\boldsymbol{\theta};\boldsymbol{z}_0)|^2 - \boldsymbol{b}^2 \right\|_2^2 , \tag{2}$$

where the optimization variable becomes the weight $\boldsymbol{\theta}$ of network G, and the solution is given by $\boldsymbol{x} = G(\boldsymbol{\theta};\boldsymbol{z}_0)$.

Compared to the dimension of the unknown $\boldsymbol{x}$ in original formulation (1), the dimension of unknown $\boldsymbol{\theta}$ of our re-parameterized loss (2) is much larger than $2n$. However, such over-parameterization actually facilitates the optimization of (2). Note that there may exist several points $\boldsymbol{\theta}$ corresponding to the solution $\boldsymbol{x}$, our aim for (2) is to decrease the loss towards zero. It is observed that achieving (approximate) zero-loss for (2) is always ensured for the particular configuration with Gaussian measurement matrix $\boldsymbol{A}$. It means that we can always find a solution $G(\boldsymbol{\theta};\boldsymbol{z}_0)$ such that loss in (1) approaching zero. For gradient methods for (1) without benign landscape, it always stagnate at local minimizers.

3. EXPERIMENTS

In this section, we study the performance of our proposed network-based re-parameterization approach by extensive computer simulations. We evaluate and compare the performance of seven methods. They include two alternating projection-based methods: Douglas-Rachford (DR, a.k.a. HIO with $\beta = 1$) [8], Graph Projection Splitting (GPS) [10]; four gradient flow-based methods: WirtFlow [15], TWF [16], TAF [29] and RAT [18], and the proposed method: phase retrieval via network-based re-parameterization, denoted by PR_NR. We do not include the convexification-based method for its intractability for the high dimensional problems. For the compared methods, we set their parameters as default or as suggested in their papers. Note that for four compared gradient flow methods, we follow the initialization generating routines in their implementations. For the remaining methods and PR_NR, the same random initialization is used.

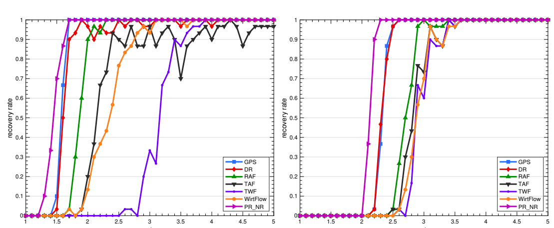
To quantitatively compare the empirical recovery performance, we compute the relative error of the recovery $\boldsymbol{x}$ and the truth signal $\boldsymbol{x}^*$:

$$\text{rel. error} := \min_{|c|=1} \frac{\|c\boldsymbol{x} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|},$$

where the minimum is attained when $c = \boldsymbol{x}^*\boldsymbol{x}/|\boldsymbol{x}^*\boldsymbol{x}|$. Within these implementations, three stopping criteria are exploited: The loss objective drops below $1e-10$, the relative error of the recovery and the truth signal is below $1e-3$ when the true signal is accessible, and the algorithm achieves the maximum iteration number 5000.

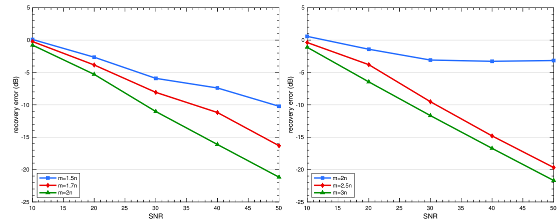
For our proposed PR_NR, input $\boldsymbol{z}_0 \in \mathbb{C}^n$ is a fixed uniformly random vector, and we adopt 4-layer full connection network with skip connection and each layer contains 100 nodes except for the input nodes and output nodes. The ReLU activation function is used. For the gradient method to optimizing (2), we use the Adam optimizer with default parameters in the implementation of PyTorch package.

In the following experiments, we first compare the recovery performance under the Gaussian matrix measurement configuration for both real and complex cases. Then we compare the recovery performance for non-zero mean Gaussian matrix measurement and general coded diffraction pattern measurements.



(a) Real Gaussian (b) Complex Gaussian

Fig. 1. Empirical phase transition for real and complex Gaussian phase retrieval.



(a) Real Gaussian (b) Complex Gaussian

Fig. 2. Recovery performance from noisy measurements for real and complex Gaussian phase retrieval.

3.1. Phase transition

For Gaussian phase retrieval, the benchmark problem is to compare the phase transition of each algorithm. The *phase transition* is the critical point of the ratio of the number of measurements m to the length of the signal n. When m/n is above a certain constant, the unique solution can be located by the solving algorithm for any instance of phase retrieval problem. When m/n is below the constant, there exists at least an instance problem, of which solution can not be located by the solving algorithm. To numerically specify the phase transition, for a fixed signal with length $n = 400$, we generate an array measurements with different measurement complexity $m = \{1n, 1.1n, \ldots, 5n\}$. Once we fix n and m, we generate and solve 30 problems to average the number of successful recoveries. If the relative error goes below $1e-3$, the reconstruction is successful.

We compute the *recovery rate*, *i.e.*, the percentage of the successful recoveries within the 30 instances for each pair (n, m), and plot the curves of recovery rate in Figure 1 (a) and (b) for real and complex problem respectively. It shows that our proposed PR_NR is the best performer among the compared methods with the sharpest phase transition curves. Noted that, although our method is also a gradient flow-based method, compared to other gradient flow-based methods, it is not sensitive to the initialization. Our method can successfully solve phase retrieval from $m = 1.7n$ and $m = 2.3n$ measurements for real and complex cases respectively. To the best of our knowledge, this is the best performer in terms of the phase transition. It is very close to the phase transition in asymptotic setting ($m, n \to \infty$) in [30].

3.2. Robust to noise

We consider the recovery performance when measurements are corrupted by Gaussian noise to validate the robustness to noise. To specify the effect of the noise level, we generate noisy real and complex problems with noise of different SNR levels $10 + 5i$ ($i = 0, 1, \ldots, 8$). We set measurement complexity m to $1.5n, 1.7n, 2n$ and $2n, 2.5m, 3n$ for real and complex problems respectively. The relative recovery error (in dB) versus noise SNR level is plotted in Figure 2 (a) and

Table 1. Averaged Loss and relative error results of our reparameterization approach from 30 instances of the problem with different number of masks. The loss and relative error are separated by the slash symbol.

Cases	Mask	1	2	3	4
Real	Uniform	2.2e-06/4.8e-01	8.6e-13/4.5e-06	8.6e-13/2.4e-06	8.4e-13/2.2e-06
	Bipolar	9.2e-06/4.9e-01	8.8e-13/2.2e-06	8.5e-13/1.2e-06	8.4e-13/1.1e-06
Complex	Uniform	2.2e-07/6.3e-01	6.1e-06/5.1e-01	7.9e-13/4.5e-06	8.3e-13/2.0e-06
	Bipolar	9.0e-09/6.4e-01	1.9e-05/5.0e-01	7.4e-13/1.6e-06	8.4e-13/9.4e-07

(b). It is shown that when the measurement complexity is above the phase transition critical points, the recovery is stable, and the recovery error is linearly correlated to the noise.

3.3. Non-zero Gaussian measurement configuration

It is observed that the compared four gradient flow-based methods can not perform well for non-zero Gaussian measurement matrix $\boldsymbol{A}$, let alone other general matrices. Thus we also conduct simulations with measurement from Gaussian $\boldsymbol{A}$ with mean 2. We follow above experimental setting for phase transition and noisy measurements. Our proposed method still stably solves the phase retrieval problem, and share the similar phase transition curves for real cases, and it solves the problems when $m \geq 2.6n$ for the complex case.

3.4. Other measurement matrix configuration

We also test our deep network reparameterization approach for non-Gaussian phase retrieval. We follow the Coded Diffraction Pattern (CDP) proposed in [15], where the forward matrix $\boldsymbol{A}$ is the stacked Fourier matrix with random masks. More specially, $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{F}\,\mathrm{diag}(\boldsymbol{m}_1) \\ \cdots \\ \boldsymbol{F}\,\mathrm{diag}(\boldsymbol{m}_L) \end{bmatrix} \in \mathbb{C}^{nL\times n}$ where L is the number of masks. We consider two types of masks $\boldsymbol{m}_i$: *uniform* and *bipolar*. The elements of uniform mask are distributed uniformly on the unit ball in $\mathbb{C}$, while the elements of the bipolar mask are randomly chosen from $\{1, -1\}$ with equal probability. We list the averaged loss and the relative error results in Table 1. It is shown that, it is easy to make the loss vanishing using proposed deep network reparameterization approach. When $L \geq 2$ and 3, the phase retrial with the CDP measurements can be solved efficiently for real and complex-valued signals respectively.

4. CONCLUSION

In this paper, we propose a network-based re-parameterization approach for phase retrieval. For gradient flow-based optimization, the over-parameterization facilitates the minimization towards global optimal solution such that the objective is vanishing. Compared to the gradient flow with respect to the original variables, our proposed method addresses the issue of stagnation at local minima. In the comparison to the existing solvers for phase retrieval, our proposed initialization-independent method outperforms them in terms of recovery phase transition and robustness to noise.

5. REFERENCES

[1] Robert A Gonsalves, "Perspectives on phase rertrieval and phase diversity in astronomy," in *Adaptive Optics Systems IV*, Enrico Marchetti, Laird M. Close, and Jean-Pierre Véran, Eds. International Society for Optics and Photonics, jul 2014, pp. 91482–91482, SPIE-Intl Soc Optical Eng.

[2] C Fienup and J Dainty, "Phase retrieval and image reconstruction for astronomy," in *Image Recovery: Theory and Application*, Henry Stark, Ed. Academic Press, 1987.

[3] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, may 2015.

[4] Jianliang Qian, Chao Yang, A Schirotzek, F Maia, and S Marchesini, "Efficient algorithms for ptychographic phase retrieval," in *Contemporary Mathematics*. American Mathematical Society (AMS), 2014.

[5] Zaiwen Wen, Chao Yang, Xin Liu, and Stefano Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," *Inverse Problems*, vol. 28, no. 11, pp. 115010, 2012.

[6] Emmanuel J Candès, Xiaodong Li, and Mahdi Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, vol. 39, no. 2, pp. 277–299, oct 2015.

[7] Ralph W Gerchberg and W Owen Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237, 1972.

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

[9] James R Fienup, "Phase retrieval algorithms: A personal tour [invited]," *Applied Optics*, vol. 52, no. 1, pp. 45–56, dec 2012.

[10] Ji Li and Hongkai Zhao, "Solving phase retrieval via graph projection splitting," *Inverse Problems*, vol. 36, no. 5, pp. 055003, 2020.

[11] D Russell Luke, "Relaxed averaged alternating reflections for diffraction imaging," *Inverse Problems*, vol. 21, no. 1, pp. 37–50, nov 2004.

[12] D Russell Luke, Heinz H Bauschke, and Patrick L Combettes, "Hybrid projection–reflection method for phase retrieval," *Journal of the Optical Society of America A*, vol. 20, no. 6, pp. 1025–1034, 2003.

[13] Heinz H Bauschke, Patrick L Combettes, and D Russell Luke, "Phase retrieval, error reduction algorithm, and fienup variants: A view from convex optimization," *Journal of the Optical Society of America A*, vol. 19, no. 7, pp. 1334–1345, 2002.

[14] Yan Shuo Tan and Roman Vershynin, "Phase retrieval via randomized kaczmarz: Theoretical guarantees," 2017.

[15] Emmanuel J Candès, Xiaodong Li, and Mahdi Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, apr 2015.

[16] Yuxin Chen and Emmanuel J Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Communications on pure and applied mathematics*, vol. 70, no. 5, pp. 822–883, 2017.

[17] Gang Wang, Georgios B Giannakis, and Yonina 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.

[18] Gang Wang, Georgios B Giannakis, Yousef Saad, and Jie Chen, "Solving most systems of random quadratic equations," *Advances in Neural Information Processing Systems*, vol. 2017, pp. 1868–1878, 2017.

[19] Emmanuel J Candès, Yonina C Eldar, Thomas Strohmer, and Vladislav Voroninski, "Phase retrieval via matrix completion," *SIAM Review*, vol. 57, no. 2, pp. 225–251, jan 2015.

[20] Emmanuel J Candès, Thomas Strohmer, and Vladislav Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, nov 2012.

[21] Irène Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1–2, pp. 47–81, dec 2015.

[22] Radu Balan, Pete Casazza, and Dan Edidin, "On signal reconstruction without phase," *Applied and Computational Harmonic Analysis*, vol. 20, no. 3, pp. 345–356, may 2006.

[23] Radu Balan, "Reconstruction of signals from magnitudes of redundant representations: The complex case," *Foundations of Computational Mathematics*, pp. 1–45, apr 2015.

[24] Radu Balan, "The fisher information matrix and the CRLB in a non-Awgn model for the phase retrieval problem," in *2015 International Conference on Sampling Theory and Applications (SampTA)*. may 2015, Institute of Electrical & Electronics Engineers (IEEE).

[25] Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi, "Phase retrieval using alternating minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4814–4826, sep 2015.

[26] Samet Oymak and Mahdi Soltanolkotabi, "Overparameterized nonlinear learning: Gradient descent takes the shortest path?," in *International Conference on Machine Learning*. PMLR, 2019, pp. 4951–4960.

[27] Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song, "A convergence theory for deep learning via over-parameterization," in *International Conference on Machine Learning*. PMLR, 2019, pp. 242–252.

[28] Simon Du, Jason Lee, Haochuan Li, Liwei Wang, and Xiyu Zhai, "Gradient descent finds global minima of deep neural networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 1675–1685.

[29] Gang Wang, Georgios B. Giannakis, Jie Chen, and Mehmet Akçakaya, "Sparta: Sparse phase retrieval via truncated amplitude flow," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 3974–3978.

[30] Junjie Ma, Ji Xu, and Arian Maleki, "Optimization-based amp for phase retrieval: The impact of initialization and ℓ_2 regularization," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3600–3629, 2019.