Geometrical Properties and Accelerated Gradient Solvers of Non-convex Phase Retrieval

Yi Zhou, Huishuai Zhang and Yingbin Liang[1,2]

***Abstract*— We consider recovering a signal $\boldsymbol{x} \in \mathbb{R}^n$ from the magnitudes of Gaussian measurements by minimizing a second order yet non-smooth loss function. By exploiting existing concentration results of the loss function, we show that the non-convex loss function satisfies several quadratic geometrical properties. Based on these geometrical properties, we characterize the linear convergence of the sequence of function graph generated by the gradient flow on minimizing the loss function. Furthermore, we propose an accelerated version of the gradient flow, and establish an in-exact linear convergence of the generated sequence of function graph by exploiting the quadratic geometries of the loss function. Then, we verify the numerical advantages of the proposed algorithms over other state-of-art algorithms.**

I. Introduction

Consider taking a set of linear Gaussian measurements of an underlying signal $\boldsymbol{x} \in \mathbb{R}^n$, and we only observe the magnitudes of the measurements, i.e.,

$$y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|, \quad i = [m]. \tag{1}$$

Here, $\{y_i\}_i$ are the magnitude observations and $\{\boldsymbol{a}_i\}_i$ are measurement vectors with entries generated by i.i.d normal distribution. This observation model naturally arises in the so called phase retrieval applications [1], [2], [3], [4], [5], [6]. There, the physical device (such as photon detectors) can only record the magnitudes of the measurements, and the goal is to recover the underlying signal, up to a global sign (or phase in the complex case), based on the magnitude observations $\{y_i\}_i$ and the measurement vectors $\{\boldsymbol{a}_i\}_i$.

If we observe the exact linear measurements (i.e., with both magnitudes and sign), then a natural approach for recovering the signal is to find the minimum least square loss, i.e., $\min_{\boldsymbol{z}} \sum_{i=1}^m (\boldsymbol{a}_i^T \boldsymbol{z} - y_i)^2$. Inherit from this idea, [3], [7] aim to recover the underlying signal $\boldsymbol{x}$ in eq. (1) (up to a global sign) by considering the phase retrieval problem

$$\textbf{(P)} \quad \min_{\boldsymbol{z} \in \mathbb{R}^n} \ell(\boldsymbol{z}) := \frac{1}{2m} \sum_{i=1}^m \left(|\boldsymbol{a}_i^\top \boldsymbol{z}| - y_i \right)^2. \tag{2}$$

Clearly, the form of **(P)** is similar to that of the least square loss for solving linear systems, except that we are only given the magnitude observations. Although the magnitude operation $|\cdot|$ destroys the convexity and the smoothness of **(P)**, gradient flow (also called Wirtinger flow) with a proper initialization can still solve it efficiently when the number of magnitude observations satisfies $m > \mathcal{O}(n)$ [7].

Another approach to avoid non-smoothness in **(P)** is to raise the order of the magnitude observations, and consider

$$\min_{\boldsymbol{z} \in \mathbb{R}^n} \hat{\ell}(\boldsymbol{z}) := \frac{1}{4m} \sum_{i=1}^m \left(|\boldsymbol{a}_i^\top \boldsymbol{z}|^2 - y_i^2 \right)^2. \tag{3}$$

This problem has recently been considered in [8] and is solved by gradient flow with sample complexity $m > \mathcal{O}(n \log n)$. Albeit the smoothness of the above loss function, the variable is of order four, and hence makes the loss function over curved. Consequently, this formulation requires more magnitude observations and induces higher computational cost compared to the lower order loss in **(P)** [7].

Our goal in this paper is to further explore the geometrical properties of the loss function in **(P)** for solving phase retrieval. Moreover, we want to characterize different convergence behaviors of gradient flow on solving **(P)**, and explore the possibility of acceleration.

A. Our Contributions

In this paper, we consider solving **(P)** for recovering the signal in eq. (1). Specifically, by exploiting existing concentration results on **(P)** in [7], we show that the loss function in **(P)** preserves several quadratic geometries in a local neighborhood around $\pm\boldsymbol{x}$. Based on these geometrical properties, we establish the linear convergence of the sequence of function graph generated by the gradient flow in [7] on solving **(P)**. Moreover, we develop an accelerated version of the gradient flow by adding a momentum term, and establish an in-exact linear convergence of the generated sequence of function graph by exploiting the proposed geometrical properties. Then, we verify the numerical advantages of the proposed algorithms over other state-of-art algorithms.

B. Related Work

In the early age, error-reduction methods are proposed to solve the phase retrieval problem empirically [2], [3]. Recently, many modern methods are proposed for phase retrieval with performance guarantee, for instance, convex relaxations via phase lifting [4], [5], [6] and via phase cut [9], alternating minimization algorithm in [10], first-order gradient-like algorithms in [8], [11], [12], [13], second-order trust-region algorithm in [14], and stochastic Kaczmarz algorithm in [15].

Geometrical properties of the phase retrieval problem has been discussed in [8], [11], [12], [14] for smooth and higher

[1] Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA. {yzhou35, hzhang23, yliang06}@syr.edu.

[2] This work is supported in part by the grants AFOSR FA9550-16-1-0077 and NSF ECCS 16-09916.

order losses, and in [7] for the non-smooth and lower order loss in **(P)**. Specifically, [8] establishes the geometrical regularity condition of the loss in eq. (3), and [14] further establishes its Lipschitz gradient property and restricted strongly convex property. The geometrical regularity condition is also established in [11], [12] for a smooth loss function fitting the Poisson observation model, and in [7] for **(P)**.

C. Organization and Notations

The rest of this paper is organized as follows. In Section II we present the geometrical properties of **(P)**. Section III characterizes the performance guarantee of gradient flow and its accelerated version based on the proposed geometrical properties. Section IV compares the proposed algorithms with other competitive algorithms numerically. Finally, Section V concludes the paper.

Throughout the paper, $\boldsymbol{x}$ is the underlying signal to recover, $\{\boldsymbol{a}_i\}_i$ are Gaussian vectors and $\{y_i\}_i$ are the magnitude observations. We denote the vector $\boldsymbol{z}e^{\psi(\boldsymbol{z})}$ simply as $\boldsymbol{z}$, where $\psi(\boldsymbol{z})$ is the phase term defined by

$$\psi(\boldsymbol{z}) = \begin{cases} 0, & \text{if} \|\boldsymbol{z}-\boldsymbol{x}\| \le \|\boldsymbol{z}+\boldsymbol{x}\|, \\ \pi, & \text{otherwise.} \end{cases}$$

By this convention, $\|\boldsymbol{z}-\boldsymbol{x}\|$ coincides the minimal Euclidean distance between $\boldsymbol{z}$ and $\boldsymbol{x}$ up to a global sign difference, and the induced l_2 ball centered at $\boldsymbol{x}$ with radius ϵ is denoted as $\mathcal{B}(\boldsymbol{x},\epsilon)$. The binary indicator function is denoted as $\mathbf{1}_{\{\cdot\}}$.

II. Geometrical Properties of Phase Retrieval

Recall the gradient operator introduced in [7] for **(P)**, i.e.,

$$\nabla \ell(\boldsymbol{z}) \overset{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{a}_i^\top \boldsymbol{z} - y_i \cdot \operatorname{sgn}(\boldsymbol{a}_i^\top \boldsymbol{z})\right)\boldsymbol{a}_i, \tag{4}$$

where $\operatorname{sgn}(\cdot)$ is the sign function for nonzero arguments, and we adopt the convention that $\operatorname{sgn}(0)=0$. Based on the notion of gradient, it is shown in [7] that **(P)** satisfies the local geometrical regularity condition, i.e., for all $\boldsymbol{z}\in\mathcal{B}(\boldsymbol{x},c\|\boldsymbol{x}\|)$ and some $\gamma,\lambda>0$

$$\langle \nabla\ell(\boldsymbol{z}), \boldsymbol{z}-\boldsymbol{x}\rangle \ge \frac{\gamma}{2}\|\nabla\ell(\boldsymbol{z})\|^2 + \frac{\lambda}{2}\|\boldsymbol{z}-\boldsymbol{x}\|^2. \tag{5}$$

The above regularity condition guarantees that $\pm\boldsymbol{x}$ are the unique local minimizers. Moreover, it ensures a linear convergence of the distance residual $\|\boldsymbol{z}-\boldsymbol{x}\|$ generated by the gradient flow $\boldsymbol{z}\leftarrow\boldsymbol{z}-\gamma\nabla\ell(\boldsymbol{z})$ [7]. The proof of the regularity condition is depend on several concentration results when the number of magnitude observations satisfies $m>\mathcal{O}(n)$ [1]. These concentration results provide a convenient tool to explore other geometrical properties of **(P)**.

A natural way to understand the geometry of **(P)** is to analyze its first order expansion. In fact, by careful calculation we can verify that

[1]Order-wise optimal for identifying an n dimensional linear system.

Lemma 1. *Consider the loss function in **(P)**. Then for all $\boldsymbol{z}_1,\boldsymbol{z}_2\in\mathbb{R}^n$, the first order expansion of $\ell(\boldsymbol{z}_1)$ at point $\boldsymbol{z}_2$ satisfies*

$$\ell(\boldsymbol{z}_1)-\ell(\boldsymbol{z}_2)-\langle \boldsymbol{z}_1-\boldsymbol{z}_2, \nabla\ell(\boldsymbol{z}_2)\rangle = \frac{1}{m}\sum_{i=1}^{m}|\boldsymbol{a}_i^\top(\boldsymbol{z}_2-\boldsymbol{z}_1)|^2 - 4y_i|\boldsymbol{a}_i^\top\boldsymbol{z}_1|\mathbf{1}_{\{(\boldsymbol{a}_i^\top\boldsymbol{z}_1)(\boldsymbol{a}_i^\top\boldsymbol{z}_2)<0\}}.$$

The two terms on the right hand side of the above equality corresponds to the residual of the first order approximation. Clearly, the first term $\frac{1}{m}\sum_{i=1}^{m}|\boldsymbol{a}_i^\top(\boldsymbol{z}_2-\boldsymbol{z}_1)|^2$ takes a least square form, thus contributing a quadratic geometry to the loss function. On the other hand, the second term $4y_i|\boldsymbol{a}_i^\top\boldsymbol{z}_1|\mathbf{1}_{\{(\boldsymbol{a}_i^\top\boldsymbol{z}_1)(\boldsymbol{a}_i^\top\boldsymbol{z}_2)<0\}}$ is induced by the magnitude operation in **(P)**, and brings both non-convexity and non-smoothness. Fortunately, existing concentration results on **(P)** provides a convenient tool to further control the residual in Lemma 1, and lead to the following characterization of geometrical properties of **(P)**.

Lemma 2. *Let $m>\mathcal{O}(n)$. Then for all $\boldsymbol{z}_1,\boldsymbol{z}_2\in\mathbb{R}^n$ and all $\boldsymbol{z}\in\mathcal{B}(\boldsymbol{x},\epsilon\|\boldsymbol{x}\|)$, **(P)** satisfies*

$$\ell(\boldsymbol{z}_1)-\ell(\boldsymbol{z}_2)-\langle\boldsymbol{z}_1-\boldsymbol{z}_2,\nabla\ell(\boldsymbol{z}_2)\rangle \le \frac{L}{2}\|\boldsymbol{z}_2-\boldsymbol{z}_1\|^2, \tag{6}$$

$$\ell(\boldsymbol{x})-\ell(\boldsymbol{z})-\langle\boldsymbol{x}-\boldsymbol{z},\nabla\ell(\boldsymbol{z})\rangle \ge \frac{\mu}{2}\|\boldsymbol{z}-\boldsymbol{x}\|^2, \tag{7}$$

$$\ell(\boldsymbol{z})-\ell(\boldsymbol{x}) \le \frac{L}{2}\|\nabla\ell(\boldsymbol{z})\|^2, \tag{8}$$

$$\ell(\boldsymbol{z})-\ell(\boldsymbol{x}) \ge \frac{\mu}{2}\|\boldsymbol{z}-\boldsymbol{x}\|^2 \tag{9}$$

with probability at least $1-\exp(-\mathcal{O}(m))$, where L,μ are positive numerical constants satisfying $\mu L<1$.

Let us go through the above geometrical properties item by item. The first property is well known as the global Lipschitz gradient condition. It implies that **(P)**, although being non-convex, is globally upper bounded by a quadratic function. The rest three properties characterize the local geometries around $\pm\boldsymbol{x}$ (i.e., the local minimizers). Specifically, the second property is known as the restricted strong convexity [16], [17]. This is because it takes the form of strong convexity, but is restricted between the *fixed* point $\boldsymbol{x}$ and another point around it. Thus, it is a weaker notion of strong convexity. The third property upper bounds the residual of the function value by the square of the gradient. It is well known as the Polyak-Łojasiewicz inequality date back to 1960s [18], [19], [20]. As we show in the next section, this property guarantees the linear convergence of the function value residual generated by the gradient flow. The last property further lower bounds the residual of the function value by a quadratic function. It implies that the function grows at least quadratically around the local minimizers, and is usually referred to as quadratic growth property [21], [22], [23]. Conveniently, the quadratic growth property can translate any upper bound of the function value residual into the corresponding upper bound of the distance residual. We note that all the above properties imply a quadratic-like geometry, and are also shared by the least square loss globally. Importantly, none of them necessarily implies (local) convexity.

To summarize, the loss function in **(P)** satisfies the above quadratic geometries (locally around $\pm\boldsymbol{x}$). Thus, with the spectral initialization in Proposition 1, we can exploit these quadratic geometrical properties in designing algorithms for solving **(P)**.

III. Gradient Flows for Solving Phase Retrieval

In general, geometrical properties are the key factors in designing algorithms with performance guarantee for solving optimization problems. In specific, we discuss the gradient flow and its accelerated version for solving **(P)**, and characterize their performance based on the proposed geometrical properties on phase retrieval.

In [7], a gradient flow algorithm (referred to as reshaped Wirtinger Flow) is proposed for solving **(P)** with the spectral initialization in Proposition 1 (See appendix). The details are given in Algorithm 1.

Algorithm 1 Reshaped Wirtinger Flow

Initialization: $\boldsymbol{z}^{(0)}$ via the spectral method;
Gradient loop: for $t = 0 : T-1$ do

$$\boldsymbol{z}^{(t+1)} = \boldsymbol{z}^{(t)} - \eta\nabla\ell(\boldsymbol{z}^{(t)}). \tag{10}$$

Output $\boldsymbol{z}^{(T)}$.

Based on the regularity condition in eq. (5), [7] establishes the linear convergence of the distance residual $\|\boldsymbol{z}^{(t)} - \boldsymbol{x}\|$ generated by the above gradient flow with $\eta = \gamma$. Now, with the Lipschitz gradient property in eq. (6) and the Polyak-Łojasiewicz property in eq. (8), we can also characterize the function value residual $\ell(\boldsymbol{z}^{(t)}) - \ell(\boldsymbol{x})$ generated by the gradient flow. The statement is made formal below.

Theorem 1. *Let $m > \mathcal{O}(n)$ so that eq.* (6) *and eq.* (8) *hold with probability at least $1-\exp(-\mathcal{O}(m))$. Then the sequence $\{\ell(\boldsymbol{z}^{(t)})\}_t$ generated by Algorithm 1 with $\eta = \mu$ satisfies for all $t = [T]$,*

$$\ell(\boldsymbol{z}^{(t)}) \le \left(1 - \frac{\mu}{L}\right)^t \ell(\boldsymbol{z}^{(0)}). \tag{11}$$

Thus, combining with the result in [7], the sequence of the whole function graph [2] $(\boldsymbol{z}^{(t)}, \ell(\boldsymbol{z}^{(t)}))_t$ generated by gradient flow converges linearly to the function graph $(\boldsymbol{x}, \ell(\boldsymbol{x}))$ at the optimal point.

Typically, quadratic-like convex functions, especially strongly convex functions, can be minimized more efficiently by gradient flow with momentum, i.e., Nesterov's acceleration scheme [24]. In addition to the usual gradient step, the scheme adds another linear extrapolation step for acceleration. The details of the accelerated scheme is presented in Algorithm 2, where $\eta > 0$ is the step size and $\beta > 0$ is the momentum of acceleration. Specially, the gradient flow in Algorithm 1 corresponds to zero momentum, i.e., $\beta = 0$.

Naturally, the proposed quadratic geometrical properties in Lemma 2 implies a possibility of applying the acceleration scheme on the gradient flow. However, the non-convexity of **(P)** makes the convergence guarantee established by Nesterov for the accelerated scheme no longer applicable. Hence, we need to take the non-convexity into consideration in the analysis of the accelerated scheme, and characterize its effect on the overall convergence behavior. For this purpose, we introduce the following measure of non-convexity.

[2]The graph of a function $f : \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x} \in \mathbb{R}^n$ is defined as $(\boldsymbol{x}, f(\boldsymbol{x})) \in \mathbb{R}^{n+1}$

Algorithm 2 Accelerated Reshaped Wirtinger Flow

Initialization: $\boldsymbol{z}^{(0)} = \boldsymbol{y}^{(0)}$ via the spectral method;
Gradient loop: for $t = 0 : T-1$ do

$$\boldsymbol{z}^{(t+1)} = \boldsymbol{y}^{(t)} - \eta\nabla\ell(\boldsymbol{y}^{(t)}), \tag{12}$$

$$\boldsymbol{y}^{(t+1)} = \boldsymbol{z}^{(t+1)} + \beta(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}). \tag{13}$$

Output $\boldsymbol{z}^{(T)}$.

Definition 1. *For any $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^n$, the corresponding non-convexity $e_{\boldsymbol{z}_1\boldsymbol{z}_2} \in \mathbb{R}$ of the function in **(P)** is defined as*

$$e_{\boldsymbol{z}_1\boldsymbol{z}_2} := -\left[\ell(\boldsymbol{z}_1) - \ell(\boldsymbol{z}_2) - \langle \boldsymbol{z}_1 - \boldsymbol{z}_2, \nabla\ell(\boldsymbol{z}_2)\rangle\right]. \tag{14}$$

That is, the positivity of $e_{\boldsymbol{z}_1\boldsymbol{z}_2}$ measures the violation of convexity at the points $\boldsymbol{z}_1, \boldsymbol{z}_2$, and convexity corresponds to the special case $e_{\boldsymbol{z}_1\boldsymbol{z}_2} \le 0$. Based on the measure of non-convexity and the geometric properties proposed in Lemma 2, we can now characterize the performance of Algorithm 2 for solving **(P)**.

Theorem 2. *Let $m > \mathcal{O}(n)$ so that eq.* (6) *and eq.* (7) *hold with probability at least $1-\exp(-\mathcal{O}(m))$. Set $\eta = \frac{1}{L}, \beta = \frac{1-\sqrt{\frac{\mu}{L}}}{1+\sqrt{\frac{\mu}{L}}}$. Then the sequences $\{\boldsymbol{z}^{(t)}, \boldsymbol{y}^{(t)}\}_t$ generated by Algorithm 2 for solving **(P)** satisfy for all $t = [T]$*

$$\ell(\boldsymbol{z}^{(t)}) \le \left(1 - \sqrt{\frac{\mu}{L}}\right)^t \left[\ell(\boldsymbol{z}^{(0)}) + \frac{\mu}{2}\|\boldsymbol{z}^{(0)} - \boldsymbol{x}\|^2\right] + \epsilon^{(t)}, \tag{15}$$

where

$$\epsilon^{(t)} = \sum_{l=0}^{t-1}\left(1 - \sqrt{\frac{\mu}{L}}\right)^{t-l}\left[e_{\boldsymbol{z}^{(l)}\boldsymbol{y}^{(l)}} - \frac{\sqrt{\mu L}}{2}\|\boldsymbol{z}^{(l)} - \boldsymbol{y}^{(l)}\|^2\right]. \tag{16}$$

The proof is an extension of the original one in [24] to general non-convex cases. Clearly, the characterization on the right hand side of eq. (15) consists of two parts. The first part, compared to eq. (11), attains an accelerated linear convergence rate by a factor of $\sqrt{\frac{\mu}{L}}$. On the other hand, the second part $\epsilon^{(t)}$ characterizes the inexactness induced by the non-convexity $\{e_{\boldsymbol{z}^{(t)}\boldsymbol{y}^{(t)}}\}_t$ along the iterate history. It is in general hard to characterize $e_{\boldsymbol{z}^{(t)}\boldsymbol{y}^{(t)}}$ explicitly due to the coupling with the iterates generated by the accelerated gradient flow. However, eq. (16) implies that $\epsilon^{(t)}$ can be ignored if $e_{\boldsymbol{z}^{(t)}\boldsymbol{y}^{(t)}}$ is majored by $\frac{\sqrt{\mu L}}{2}\|\boldsymbol{z}^{(t)} - \boldsymbol{y}^{(t)}\|^2$, or it vanishes linearly if $e_{\boldsymbol{z}^{(t)}\boldsymbol{y}^{(t)}}$ also vanishes linearly with the factor $(1 - \sqrt{\frac{\mu}{L}})$. In fact, our numerical experiments in Section IV indicate that with a proper momentum β, the practical performance of the accelerated gradient flow

is much faster than the original one. We also note that the characterization of functional residual in eq. (15) can be translated into the distance residual by the quadratic growth property in eq. (9).

IV. Numerical Experiments

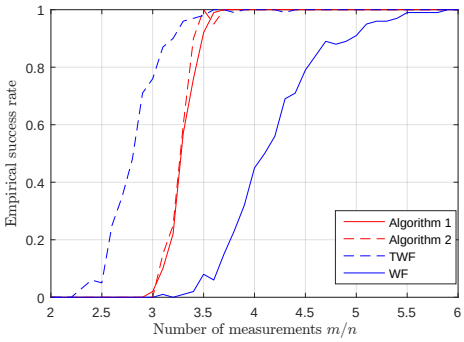
In this section, we empirically demonstrate the efficiency of Algorithm 2 by comparing with Algorithm 1 ,the Wirtinger flow (WF) in [8] for solving eq. (3), and the Truncated Wirtinger flow (TWF) in [11] for solving the Poisson loss. All the algorithms are run with parameter settings suggested in the corresponding papers. The experiments are carried out on a desktop with Intel Core i7 3.4GHz CPU and 12GB RAM.

We first compare the convergence behavior of these methods for solving corresponding loss functions in both the real case and complex case. The underlying signal and measurements are generated by normal distribution with $n = 1000, m = 8n$. All algorithms are initialized by the same spectral method in appendix. The momentum of Algorithm 2 is set to be $\beta = 0.4$, which is found to perform well empirically. In Table I, we list the number of iterations and time cost for those algorithms to achieve the relative error of 10^{-14} averaged over 10 trials. Clearly, Algorithm 2 with acceleration takes much less number of iterations and runs much faster than other methods.

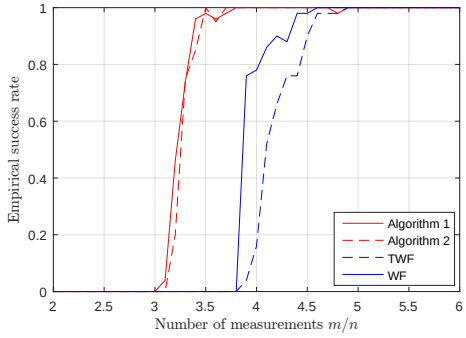
We next compare the empirical successful recovery rate versus the number of measurements for real and complex cases, respectively. The signal dimension is set to be $n = 1000$, and the ratio m/n ranges from 2 to 6. For each m/n, we run 100 trials and count the number of success. For each trial, it is declared to be successful if the output relative error satisfies $\|\boldsymbol{z}^{(T)} - \boldsymbol{x}\|/\|\boldsymbol{x}\| \leq 10^{-5}$ after $T = 1000$ iterations. Figure 1 plots the success rate with respect to m/n for all algorithms. It can be seen that the relationships between success rate and sample complexity of Algorithm 1 and Algorithm 2 are similar in both real and complex cases. This further implies that Algorithm 2 requires the same level of sample complexity as Algorithm 1. Also, they outperforms both TWF and WF for the more practical complex case, although TWF performs better for the real case. Moreover, Algorithm 1 and Algorithm 2 exhibit *shaper* phase transition than TWF and WF.

V. Conclusion

In this paper, we explore in deep the geometries of the phase retrieval problem. We show that the non-convex and non-smooth loss function has quadratic geometries around the local minimizers. These geometrical properties guarantees a linear convergence of the whole function graph generated by the gradient flow. Moreover, they allow us to further accelerate the gradient flow via momentum, and provide an in-exact convergence guarantee that characterizes the effect of non-convexity. We believe that this acceleration scheme can be applied to other gradient based methods as well for solving the phase retrieval problem.



(a) Real case



(b) Complex case

Fig. 1. Comparison of sample complexity among different algorithms.

Appendix I
Initialization via Spectral Method

For all the algorithms discussed in this paper, the initialization is set via the spectral method in [7]. That is, we set $\boldsymbol{z}^{(0)} = \lambda_0 \tilde{\boldsymbol{z}}$ with $\lambda_0 = \frac{mn}{\sum_{i=1}^m \|\boldsymbol{a}_i\|_1} \cdot \left(\frac{1}{m}\sum_{i=1}^m y_i\right)$, and $\tilde{\boldsymbol{z}}$ being the leading eigenvector of

$$\boldsymbol{Y} := \frac{1}{m}\sum_{i=1}^m y_i \boldsymbol{a}_i \boldsymbol{a}_i^\top \mathbf{1}_{\{\lambda_0 < y_i < 5\lambda_0\}}. \tag{17}$$

Then, the following result follows from [7].

Proposition 1 (Proposition 1, [7]). *Let $m > C(\delta, \epsilon)n$, where C is a positive number only affected by $\delta > 0$ and $\epsilon > 0$. Then, the spectral initialization satisfies*

$$\boldsymbol{z}^{(0)} \in \mathcal{B}(\boldsymbol{x}, \delta\|\boldsymbol{x}\|)$$

with probability at least $1 - \exp(-c'm\epsilon^2)$, where c' is some positive constant.

References

[1] 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," *Signal Processing Magazine, IEEE*, vol. 32, no. 3, pp. 87–109, 2015.

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

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

[4] A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," *Inverse Problems*, vol. 27, no. 1, 2011.

TABLE I

COMPARISON OF ITERATION COUNT AND TIME COST AMONG ALGORITHMS ($n = 1000, m = 8n$)

	Algorithms	Algorithm 1	Algorithm 2	truncated-WF	WF
real case	iterations	72	**50**	182	319.2
	time cost(s)	0.477	**0.331**	1.232	2.104
complex case	iterations	272.7	**148.2**	486.7	915.4
	time cost(s)	6.956	**3.732**	12.815	23.306

[5] E. J. Candès, T. Strohmer, and V. 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, 2013.

[6] D. Gross, F. Krahmer, and R. Kueng, "Improved recovery guarantees for phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, 2015.

[7] H. Zhang and Y. Liang, "Reshaped wirtinger flow for solving quadratic systems of equations," in *Advances in Neural Information Processing Systems*, 2016.

[8] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[9] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.

[10] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[11] Y. Chen and E. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Advances in Neural Information Processing Systems 28*, 2015, pp. 739–747.

[12] H. Zhang, Y. Chi, and Y. Liang, "Provable non-convex phase retrieval with outliers: Median truncated wirtinger flow," *arXiv preprint arXiv:1603.03805*, 2016.

[13] T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow," *arXiv preprint arXiv:1506.03382*, 2015.

[14] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," *arXiv preprint arXiv:1602.06664*, 2016.

[15] K. Wei, "Solving systems of phaseless equations via kaczmarz methods: a proof of concept study," *Inverse Problems*, vol. 31, no. 12, p. 125008, 2015.

[16] H. Zhang and L. Cheng, "Restricted strong convexity and its applications to convergence analysis of gradient-type methods in convex optimization," *Optimization Letters*, vol. 9, no. 5, pp. 961–979, 2015.

[17] H. Zhang, "The restricted strong convexity revisited: analysis of equivalence to error bound and quadratic growth," *Optimization Letters*, pp. 1–17, 2016.

[18] B. T. Polyak, "Gradient methods for minimizing functionals," *Zh. Vychisl. Mat. Mat. Fiz.*, vol. 3, p. 643653, 1963.

[19] S. Łojasiewicz, "A topological property of real analytic subsets," *Coll. du CNRS, Les equations aux derivees partielles*, p. 8789, 1963.

[20] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition," *ArXiv*, Aug 2016. [Online]. Available: http://adsabs.harvard.edu/abs/2016arXiv160804636K

[21] P. Gong and J. Ye, "Linear convergence of variance-reduced projected stochastic gradient without strong convexity," *Arxiv*, vol. abs/1406.1102, 2014. [Online]. Available: http://arxiv.org/abs/1406.1102

[22] J. Liu and S. J. Wright, "Asynchronous stochastic coordinate descent: Parallelism and convergence properties," *SIAM Journal on Optimization*, vol. 25, no. 1, pp. 351–376, 2015.

[23] I. Necoara, Y. Nesterov, and F. Glineur, "Linear convergence of first order methods for non-strongly convex optimization," *ArXiv*, 2015. [Online]. Available: http://adsabs.harvard.edu/abs/2015arXiv150406298N

[24] Y. Nesterov, *Introductory lectures on convex optimization : a basic course*, ser. Applied optimization. Kluwer Academic Publisher., 2004.