

On the Estimation Performance and Convergence Rate of the Generalized Power Method for Phase Synchronization

Huikang Liu* Man-Chung Yue† Anthony Man-Cho So‡

October 28, 2016

Abstract

An estimation problem of fundamental interest is that of phase (or angular) synchronization, in which the goal is to recover a collection of phases (or angles) using noisy measurements of relative phases (or angle offsets). It is known that in the Gaussian noise setting, the maximum likelihood estimator (MLE) has an expected squared ℓ_2 -estimation error that is on the same order as the Cramér-Rao lower bound. Moreover, even though the MLE is an optimal solution to a non-convex quadratic optimization problem, it can be found with high probability using semidefinite programming (SDP), provided that the noise power is not too large. In this paper, we study the estimation and convergence performance of a recently-proposed low-complexity alternative to the SDP-based approach, namely, the generalized power method (GPM). Our contribution is twofold. First, we bound the rate at which the estimation error decreases in each iteration of the GPM and use this bound to show that all iterates—not just the MLE—achieve an estimation error that is on the same order as the Cramér-Rao bound. Our result holds under the least restrictive assumption on the noise power and gives the best provable bound on the estimation error known to date. It also implies that one can terminate the GPM at any iteration and still obtain an estimator that has a theoretical guarantee on its estimation error. Second, we show that under the same assumption on the noise power as that for the SDP-based method, the GPM will converge to the MLE at a linear rate with high probability. This answers a question raised in [3] and shows that the GPM is competitive in terms of both theoretical guarantees and numerical efficiency with the SDP-based method. At the heart of our convergence rate analysis is

*Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. E-mail: hkliu@se.cuhk.edu.hk

†Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. E-mail: mcyue@se.cuhk.edu.hk

‡Department of Systems Engineering and Engineering Management, and, by courtesy, CUHK-BGI Innovation Institute of Trans-omics, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. E-mail: mancho@se.cuhk.edu.hk

a new error bound for the non-convex quadratic optimization formulation of the phase synchronization problem, which could be of independent interest. As a by-product, we give an alternative proof of a result in [3], which asserts that every second-order critical point of the aforementioned non-convex quadratic optimization formulation is globally optimal in a certain noise regime.

1 Introduction

The problem of phase synchronization is concerned with the estimation of a collection of phases¹ based on noisy measurements of the relative phases. Formally, let $z^\star \in \mathbb{T}^n = \{w \in \mathbb{C}^n : |w_1| = \dots = |w_n| = 1\}$ be an unknown phase vector. Given noisy measurements of the form

$$C_{j\ell} = z_j^\star \bar{z}_\ell^\star + \Delta_{j\ell} \quad \text{for } 1 \leq j < \ell \leq n, \quad (1)$$

where $\bar{(\cdot)}$ denotes the complex conjugate and $\Delta_{j\ell} \in \mathbb{C}$ is the noise in the measurement of the relative phase $z_j^\star \bar{z}_\ell^\star$, our goal is to find an estimate $\hat{z} \in \mathbb{T}^n$ of $z^\star \in \mathbb{T}^n$ that best fits those measurements in the least-squares sense. In other words, we are interested in solving the following optimization problem:

$$\hat{z} \in \arg \min_{z \in \mathbb{T}^n} \sum_{1 \leq j < \ell \leq n} |C_{j\ell} - z_j \bar{z}_\ell|^2. \quad (2)$$

Despite its simple description, the phase synchronization problem arises in a number of applications, including clock synchronization in wireless networks [6], signal reconstruction from phaseless measurements [1, 19], and ranking of items based on noisy pairwise comparisons [5]. For further discussions on the applications of phase synchronization, we refer the reader to [2] and the references therein.

Although Problem (2) may seem to involve an objective function that is quartic in the decision variable z , it can actually be reformulated as a complex quadratic optimization problem with unit-modulus constraints. Indeed, by writing the measurements (1) more compactly as $C = (z^\star)(z^\star)^H + \Delta$, where $(\cdot)^H$ denotes the Hermitian transpose and Δ is a Hermitian matrix whose diagonal entries are zero and the above-diagonal entries are given by $\{\Delta_{j\ell} : 1 \leq j < \ell \leq n\}$, and by noting that $|z_j \bar{z}_\ell|^2 = 1$ for $1 \leq j < \ell \leq n$ because $z \in \mathbb{T}^n$, we see that Problem (2) is equivalent to

$$\hat{z} \in \arg \max_{z \in \mathbb{T}^n} \{f(z) = z^H C z\}. \quad (\text{QP})$$

As it turns out, Problem (QP) is NP-hard in general [18]. Over the past two decades or so, many different approaches to tackling Problem (QP) have been proposed. One popular approach is to apply the *semidefinite relaxation* (SDR) technique, which will lead to a polynomial-time algorithm for computing

¹Throughout the paper, the term “phase” refers to a complex number with unit modulus.

a feasible but typically sub-optimal solution to Problem (QP) (see [11] for an overview of the technique). Interestingly, the *approximation accuracy* of the SDR solution, measured by the relative gap between the objective value of the SDR solution and the optimal value of Problem (QP), can be established under various assumptions on C [15, 14]. However, since our goal is to estimate the unknown phase vector z^* , a more relevant measure of the quality of the SDR solution is its *estimation error*, which intuitively can be defined as the distance between the SDR solution and the target phase vector z^* . Unfortunately, the aforementioned approximation accuracy results do not automatically translate into estimation error results. In an attempt to fill this gap, Bandeira et al. [2] considered a Gaussian noise model and studied the estimation error of the SDR solution. Specifically, suppose that the measurement noise takes the form $\Delta = \sigma W$, where W is a Wigner matrix (i.e., a Hermitian random matrix whose diagonal entries are zero and the above-diagonal entries are i.i.d. standard complex normal random variables) and $\sigma^2 > 0$ is the noise power. It is shown in [2] that if $\sigma = O(n^{1/4})$, then with high probability the standard SDR of Problem (QP) has a unique optimal solution that is of rank one; i.e., the SDR is tight. This implies that a global maximizer \hat{z} of Problem (QP), which in this case is also a maximum likelihood estimator (MLE) of the target phase vector of z^* , can be found in polynomial time. Moreover, the expected squared ℓ_2 -estimation error of \hat{z} is bounded above by $O(\sigma^2)$. This matches (up to constants) the Cramér-Rao lower bound developed in [4], which applies to any unbiased estimator of z^* . As an aside, although large instances of the standard SDR of Problem (QP) may be costly to solve using interior-point methods, they can be solved quite efficiently in practice by numerical methods that exploit structure; see, e.g., [22, 20, 21]. However, unlike interior-point methods, which are known to converge in polynomial time, most of these methods do not have convergence rate guarantees.

Besides the aforementioned SDR-based method, one can also employ the *generalized power method* (GPM) [8] (see also [12]) to tackle Problem (QP). When specialized to Problem (QP), the GPM can be viewed as a gradient method on the manifold \mathbb{T}^n and is much easier to implement than the SDR-based method. In a very recent work, Bounal [3] analyzed the convergence behavior of the GPM under the same Gaussian noise model used in [2] and showed that if $\sigma = O(n^{1/6})$, then with high probability the GPM will converge to a global maximizer of Problem (QP) when initialized by the eigenvector method in [13]. This result is significant, since in general the GPM may not even converge to a single point, let alone to a global optimizer of the problem at hand. However, it does not give the rate at which the GPM converges to the global maximizer. Moreover, compared with the result obtained for the SDR approach in [2], we see that the above result holds only in the more restrictive noise regime of $\sigma = O(n^{1/6})$. Although numerical experiments in [2, 3] indicate that both the SDR-based method and the GPM can find a global maximizer of Problem (QP) even when σ is on the order of $n^{1/2}/\text{polylog}(n)$, proving this rigorously remains an elusive task.

Motivated by the preceding discussion, our goal in this paper is to gain a

deeper understanding of the GPM when it is applied to Problem (QP) under the same Gaussian noise model used in [2, 3]. The starting point of our investigation is the following curious facts: Using Proposition 1, which first appears in an earlier version of this paper, Boumal [3] showed that modulo constants, the expected squared ℓ_2 -estimation error of the initial iterate of the GPM, viz. the one produced by the eigenvector method in [13], already matches the Cramér-Rao bound. Moreover, in the noise regime $\sigma = O(n^{1/6})$, we know by the results in [2, 3] that the same is true for the limit point of the sequence of iterates generated by the GPM, as it is a global maximizer of Problem (QP). In view of these facts, it is natural to ask whether the intermediate iterates generated by the GPM also achieve an estimation error that is on the same order as the Cramér-Rao bound, and if so, whether the GPM actually reduces the estimation error in each iteration. Our first contribution is to resolve both of these questions in the affirmative and to bound the rate at which the estimation error decreases in each iteration. Specifically, we show that even at the noise level $\sigma = O(n^{1/2})$, the expected squared ℓ_2 -estimation errors of the iterates do not exceed $(c_1 + c_2\tau^k)\sigma^2$, where $c_1, c_2 > 0$, $\tau \in (0, 1)$ are some explicitly given constants and k is the iteration counter; see the discussion after Corollary 1. An interesting aspect of this result is that it holds regardless of whether the iterates converge or not (recall that the convergence result in [3] holds only for noise level up to $O(n^{1/6})$). Thus, from a statistical estimation viewpoint, one can terminate the GPM at any iteration and still obtain an estimator whose estimation error is on the same order as the Cramér-Rao bound. Moreover, the leading constant in the estimation error becomes smaller as one runs more iterations of the GPM. This explains in part the numerical observation in [3] that the GPM can often return a good estimate of z^* even when the noise level is close to $O(n^{1/2})$. To the best of our knowledge, the bound we obtained on the ℓ_2 -estimation error of any accumulation point generated by the GPM holds under the least restrictive noise level requirement and is the best known to date in the Gaussian noise setting.

Next, we study the convergence behavior of the GPM when it is applied to Problem (QP). Our second contribution is to show that in the Gaussian noise setting, if $\sigma = O(n^{1/4})$ and the GPM is initialized by the eigenvector method, then with high probability the sequence of iterates generated by the GPM will converge linearly to a global maximizer of Problem (QP) (which is an MLE of z^*); see Corollary 2. The significance of this result is twofold. First, compared with the result in [3], the noise level requirement for the convergence of the GPM is relaxed from $O(n^{1/6})$ to $O(n^{1/4})$, thus matching the noise level requirement for the tightness of the SDR-based method. Second, our result answers a question raised in [3] concerning the convergence rate of the GPM and contributes to the growing literature on the design and analysis of fast algorithms for structured non-convex optimization problems (see, e.g., [17] and the references therein for an overview). Key to our analysis is a new *error bound* for Problem (QP), which provides a computable estimate of the distance between any given point on \mathbb{T}^n and the set of second-order critical points (which includes the global maximizers) of Problem (QP); see Propositions 3 and 4. As a by-product, we show that every

second-order critical point of Problem (QP) is still a global maximizer under a slightly less restrictive noise level requirement than [3]; see the discussion after the proof of Proposition 4. We remark that error bounds have long played an important role in the convergence rate analysis of iterative methods; see, e.g., [7, 16, 24, 25, 9, 10, 23] for some recent developments. However, most of the error bounds in the cited works are for convex optimization problems. By contrast, our error bound is developed for the non-convex problem (QP), which could be of independent interest.

We end this section by introducing the notations needed. Let $\mathbf{1}$ denote the vector of all ones and \mathbb{H}^n denote the set of $n \times n$ Hermitian matrices. For a complex vector $v \in \mathbb{C}^n$, let $\text{Diag}(v)$ denote the diagonal matrix whose diagonal elements are given by the entries of v , $|v|$ denote the vector of entry-wise moduli of v , and $\frac{v}{|v|}$ denote the vector of entry-wise normalizations of v ; i.e.,

$$(\text{Diag}(v))_{jj} = v_j, \quad |v|_j = |v_j|, \quad \left(\frac{v}{|v|}\right)_j = \begin{cases} \frac{v_j}{|v_j|} & \text{if } v_j \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

For a complex matrix $M \in \mathbb{C}^{n \times n}$, let $\text{diag}(M)$ denote the vector whose entries are the diagonal elements of M , $\|M\|_{\text{op}}$ denote its operator norm, and $\|M\|_F$ denote its Frobenius norm.

Since the measurements $\{C_{j\ell} : 1 \leq j < \ell \leq n\}$ in (1) are invariant under multiplication of a common phase to the target phase vector z^* , we can only identify z^* up to a global phase. This motivates us to define the ℓ_q -distance (where $q \in [1, \infty]$) between two phase vectors $w, z \in \mathbb{T}^n$ by

$$d_q(w, z) = \min_{\theta \in [0, 2\pi)} \|w - e^{i\theta} z\|_q.$$

2 Preliminaries

In this section, we review the GPM for solving Problem (QP) and collect some basic facts that will be used in our subsequent analysis.

The GPM is an iterative method that was introduced in [8] for maximizing a convex function over a compact set. In each iteration of the GPM, an affine minorant of the objective function at the current iterate is maximized over the feasible set to obtain the next iterate. When specialized to Problem (QP), the maximization performed in each iteration admits a closed-form solution, and the GPM takes the following form:

Algorithm 1 Generalized Power Method for Problem (QP)

```
1: input: objective matrix  $C \in \mathbb{H}^n$ , step size  $\alpha > 0$ , initial point  $z^0 \in \mathbb{T}^n$ 
2: for  $k = 0, 1, \dots$  do
3:   if termination criterion is met then
4:     return  $z^k$ 
5:   else
6:      $w^k \leftarrow (I + \frac{\alpha}{n}C) z^k$ 
7:      $z^{k+1} \leftarrow \frac{w^k}{|w^k|}$ 
8:   end if
9: end for
```

Algorithm 1 can be viewed as a projected gradient method (see lines 6-7), though it is not necessarily a Riemannian gradient method on the manifold \mathbb{T}^n ; see [3, Remark 1]. Due to the non-convexity of Problem (QP), given an arbitrary initial point, Algorithm 1 may not converge to any useful point (if it converges at all). To tackle this issue, Boumal [3] proposed to use the *eigenvector estimator* $v_C \in \mathbb{T}^n$ (cf. [13]) to initialize Algorithm 1. Specifically, let $u \in \mathbb{C}^n$ be a leading eigenvector of $C \in \mathbb{H}^n$ and $a \in \mathbb{C}^n$ be any vector satisfying $a^H u \neq 0$. Then, the vector v_C is defined by

$$(v_C)_j = \begin{cases} \frac{u_j}{|u_j|} & \text{if } u_j \neq 0, \\ \frac{a^H u}{|a^H u|} & \text{otherwise} \end{cases} \quad \text{for } j = 1, \dots, n. \quad (3)$$

As shown in [3], the advantage of initializing Algorithm 1 with $z^0 = v_C$ is twofold. First, the vector v_C is close to the target phase vector z^* in the following sense:

Fact 1. ([3, Lemma 6]) *Let $v_C \in \mathbb{T}^n$ be given by (3). Then, we have*

$$d_2(v_C, z^*) \leq \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}}.$$

Second, under some mild assumptions on the measurement noise Δ and step size α , the iterates generated by Algorithm 1 will converge to a global maximizer of Problem (QP):

Fact 2. ([3, Theorem 3]) *Suppose that (i) the measurement noise Δ satisfies $\|\Delta\|_{\text{op}} = O(n^{2/3})$ and $\|\Delta z^*\|_{\infty} = O(n^{2/3}\sqrt{\log n})$, (ii) the step size α satisfies $\alpha \leq \frac{n}{\|\Delta\|_{\text{op}}}$, and (iii) the initial point z^0 is given by $z^0 = v_C$. Then, the iterates generated by Algorithm 1 will converge to a global maximizer of Problem (QP).*

It should be noted that even allowing for the multiplication of a common phase, a global maximizer \hat{z} of Problem (QP) may not equal to the target phase vector z^* . Thus, an immediate question is whether global maximizers of Problem (QP) are close to z^* . The following result shows that the answer is affirmative:

Fact 3. ([2, Lemma 4.1]) Suppose that $\hat{z} \in \mathbb{C}^n$ satisfies $\|\hat{z}\|_2^2 = n$ and $f(z^*) \leq f(\hat{z})$ (e.g., if \hat{z} is a global maximizer of Problem (QP)). Then, we have

$$d_2(\hat{z}, z^*) = \sqrt{2(n - |\hat{z}^H z^*|)} \leq \frac{4\|\Delta\|_{\text{op}}}{\sqrt{n}}.$$

Lastly, let us record a useful property of Algorithm 1. Recall that $\tilde{z} \in \mathbb{T}^n$ is a *second-order critical point* of Problem (QP) if $w^H S(\tilde{z})w \geq 0$ for all $w \in T_{\tilde{z}}\mathbb{T}^n$, where

$$S(z) = \Re \{ \text{Diag}(\text{diag}(Cz z^H)) \} - C$$

and

$$T_z\mathbb{T}^n = \{w \in \mathbb{C}^n : \Re \{w_i \bar{z}_i\} = 0 \text{ for } i = 1, \dots, n\}$$

is the tangent space to \mathbb{T}^n at $z \in \mathbb{T}^n$; see [2, 3]. By considering the second-order necessary optimality conditions of Problem (QP), it can be shown that every global maximizer of Problem (QP) is a second-order critical point. The following result asserts that a second-order critical point of Problem (QP) is (i) a fixed point of Algorithm 1 and (ii) close to the target phase vector z^* if the measurement noise Δ is not too large.

Fact 4. ([3, Lemmas 7, 14, 15, and 16]) Let $\tilde{z} \in \mathbb{T}^n$ be any second-order critical point of Problem (QP) and $\tilde{C} = \frac{n}{\alpha} (I + \frac{\alpha}{n}C) = C + \frac{n}{\alpha}I$. Then, for any $\alpha > 0$,

$$|(C\tilde{z})_j| = (C\tilde{z})_j \overline{(\tilde{z}_j)} \quad \text{and} \quad |(\tilde{C}\tilde{z})_j| = (\tilde{C}\tilde{z})_j \overline{(\tilde{z}_j)} \quad \text{for } j = 1, \dots, n.$$

Consequently, we have $\tilde{z}^H C\tilde{z} = \|C\tilde{z}\|_1$, $\tilde{z}^H \tilde{C}\tilde{z} = \|\tilde{C}\tilde{z}\|_1$, and

$$\left(\text{Diag}(|\tilde{C}\tilde{z}|) - \tilde{C} \right) \tilde{z} = (\text{Diag}(|C\tilde{z}|) - C) \tilde{z} = \mathbf{0}.$$

Moreover, if $\|\Delta\|_{\text{op}} \leq \frac{n}{16}$, then

$$|(z^*)^H \tilde{z}| \geq n - 4\|\Delta\|_{\text{op}} \quad \text{and} \quad d_2(\tilde{z}, z^*) \leq \sqrt{8\|\Delta\|_{\text{op}}}.$$

To distinguish the different points of interest on \mathbb{T}^n , we shall reserve the notations z^* , \hat{z} , and \tilde{z} to denote the target phase vector, a global maximizer of Problem (QP), and a second-order critical point of Problem (QP), respectively in the sequel.

3 Estimation Performance of the GPM

Facts 1 and 3 show that both the eigenvector estimator v_C and global maximizers of Problem (QP) are close to the target phase vector z^* . In this section, we show that the same is true for all intermediate iterates of Algorithm 1. In fact, we establish a stronger result: We show that the ℓ_2 - and ℓ_∞ -estimation errors of the iterates decrease in each iteration of Algorithm 1 and provide explicit bounds on the rates of decrease.

To begin, let us introduce our first result, which concerns the ℓ_2 -estimation errors of the iterates:

Theorem 1. Suppose that (i) the measurement noise Δ satisfies $\|\Delta\|_{\text{op}} \leq \frac{n}{16}$, (ii) the step size α satisfies $\alpha \geq 2$, and (iii) the initial point z^0 is given by $z^0 = v_C$. Then, the sequence of iterates $\{z^k\}_{k \geq 0}$ generated by Algorithm 1 satisfies

$$d_2(z^{k+1}, z^*) \leq \mu^{k+1} \cdot d_2(z^0, z^*) + \frac{\nu}{1-\mu} \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}} \leq \left(\mu^{k+1} + \frac{\nu}{1-\mu} \right) \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}}$$

for $k = 0, 1, \dots$, where

$$\mu = \frac{16(\alpha\|\Delta\|_{\text{op}} + n)}{(7\alpha + 8)n} < 1, \quad \nu = \frac{2\alpha}{7\alpha + 8}. \quad (4)$$

Theorem 1 has two noteworthy features. First, it does not assume that Algorithm 1 converges. Second, it provides a bound on the ℓ_2 -estimation error of each iterate generated by Algorithm 1. As such, one can terminate Algorithm 1 at any iteration and still has a guarantee on the quality of the estimator.

To further illustrate the usefulness of Theorem 1, recall from Facts 1 and 3 that the ℓ_2 -estimation errors of the initial point v_C and the global maximizers of Problem (QP) are bounded above by $\frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}}$ and $\frac{4\|\Delta\|_{\text{op}}}{\sqrt{n}}$, respectively. Now, if we take $\alpha = 4$ in Algorithm 1, then under the assumptions of Theorem 1, we have $\mu \leq \frac{5}{9}$ and $\nu = \frac{2}{9}$. This implies that any accumulation point z^∞ generated by Algorithm 1 satisfies

$$d_2(z^\infty, z^*) \leq \frac{4\|\Delta\|_{\text{op}}}{\sqrt{n}},$$

which matches the bound on the ℓ_2 -estimation error of any global maximizer of Problem (QP). Furthermore, if we let $\alpha \rightarrow \infty$, which can be interpreted as using the update $z^{k+1} \leftarrow \frac{Cz^k}{|Cz^k|}$ in line 7 of Algorithm 1, then

$$d_2(z^{k+1}, z^*) \leq \left(\left(\frac{1}{7} \right)^{k+1} + \frac{1}{3} \right) \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}}$$

for $k = 0, 1, \dots$. In this case, our bound is even better than that in Fact 3 when k is sufficiently large.

Next, we present our result on the ℓ_∞ -estimation errors of the iterates:

Theorem 2. Under the same assumptions as Theorem 1, the sequence of iterates $\{z^k\}_{k \geq 0}$ generated by Algorithm 1 satisfies

$$d_\infty(z^{k+1}, z^*) \leq \gamma^{k+1} \cdot d_\infty(z^0, z^*) + \frac{\zeta \cdot \mu^k}{1-\gamma/\mu} + \frac{\omega}{1-\gamma}$$

for $k = 0, 1, 2, \dots$, where

$$\gamma = \frac{16}{7\alpha + 8} < 1, \quad \zeta = \frac{128\alpha\|\Delta\|_{\text{op}}^2}{(7\alpha + 8)n^{3/2}}, \quad \omega = \frac{16\alpha}{7\alpha + 8} \left(\frac{\nu}{1-\mu} \frac{8\|\Delta\|_{\text{op}}^2}{n^{3/2}} + \frac{\|\Delta z^*\|_\infty}{n} \right),$$

and μ, ν are given in (4), so that $\gamma/\mu < 1$.

To prove Theorems 1 and 2, we need the following technical results:

Proposition 1. *For any $w \in \mathbb{C}^n$, $z \in \mathbb{T}^n$, and $q \in [1, \infty]$, we have*

$$\left\| \frac{w}{|w|} - z \right\|_q \leq 2 \|w - z\|_q.$$

Proof. Without loss of generality, we may assume that $z = \mathbf{1}$. By definition of $\frac{w}{|w|}$, it suffices to show that for $j = 1, \dots, n$,

$$\left| \left(\frac{w}{|w|} \right)_j - 1 \right| \leq 2|w_j - 1|.$$

The above inequality holds trivially if $w_j = 0$. Hence, we may focus on the case where $w_j \neq 0$. We claim that

$$|e^{i\phi} - 1| \leq 2|re^{i\phi} - 1| \quad \text{for any } \phi \in [0, 2\pi) \text{ and } r \geq 0.$$

To prove this, observe that $|re^{i\phi} - 1|^2 = r^2 - 2r \cos \phi + 1$. Thus, we have

$$\arg \min_{r \geq 0} |re^{i\phi} - 1|^2 = \begin{cases} 0 & \text{if } \phi \in [\frac{\pi}{2}, \frac{3\pi}{2}], \\ \cos \phi & \text{if } \phi \in [0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi), \end{cases}$$

from which it follows that

$$\min_{r \geq 0} |re^{i\phi} - 1|^2 = \begin{cases} 1 & \text{if } \phi \in [\frac{\pi}{2}, \frac{3\pi}{2}], \\ \sin^2 \phi & \text{if } \phi \in [0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi). \end{cases} \quad (5)$$

Now, for $\phi \in [\frac{\pi}{2}, \frac{3\pi}{2}]$, by the triangle inequality and (5), we have

$$|e^{i\phi} - 1| \leq 2 \leq 2|re^{i\phi} - 1| \quad \text{for any } r \geq 0.$$

On the other hand, for $\phi \in [0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi)$, we use the half-angle formula and (5) to get

$$|e^{i\phi} - 1| = \sqrt{2(1 - \cos \phi)} = 2 \left| \sin \frac{\phi}{2} \right| \leq 2 |\sin \phi| \leq 2|re^{i\phi} - 1| \quad \text{for any } r \geq 0.$$

Combining the above two cases, the proof is completed. \square

Proposition 2. *Let $\{z^k\}_{k \geq 0}$ be the sequence of iterates generated by Algorithm 1 with $\alpha > 0$. For $q \in [1, \infty]$ and $k = 0, 1, \dots$, define*

$$\begin{aligned} \theta_k &= \arg \min_{\theta \in [0, 2\pi)} \|z^k - e^{i\theta} z^*\|_q, \\ \epsilon^k &= e^{-i\theta_k} (z^k - e^{i\theta_k} z^*), \\ \beta_k &= 1 + \alpha + \frac{\alpha}{n} (z^*)^H (\epsilon^k). \end{aligned}$$

Then, for any $r \in \mathbb{C}$ and $k \in \{0, 1, \dots\}$, we have

$$d_q(z^{k+1}, z^*) \leq 2\|rg^k - z^*\|_q,$$

where

$$g^k = \beta_k z^* + \left(I + \frac{\alpha}{n} \Delta\right) \epsilon^k + \frac{\alpha}{n} \Delta z^*.$$

Proof. Consider a fixed $k \in \{0, 1, \dots\}$. By definition, we have

$$\begin{aligned} w^k &= \left(I + \frac{\alpha}{n} C\right) z^k \\ &= e^{i\theta_k} \left(I + \frac{\alpha}{n} ((z^*) (z^*)^H + \Delta)\right) (z^* + \epsilon^k) \\ &= \left[\left(1 + \alpha + \frac{\alpha}{n} (z^*)^H (\epsilon^k)\right) z^* + \left(I + \frac{\alpha}{n} \Delta\right) \epsilon^k + \frac{\alpha}{n} \Delta z^*\right] e^{i\theta_k} \\ &= g^k e^{i\theta_k}. \end{aligned}$$

Since $z^{k+1} = \frac{w^k}{|w^k|}$, it follows from Proposition 1 that for any $r \in \mathbb{C} \setminus \{0\}$,

$$d_q(z^{k+1}, z^*) \leq \left\| \frac{g^k}{|g^k|} - z^* \right\|_q = \left\| \frac{rg^k}{|rg^k|} - z^* \right\|_q \leq 2\|rg^k - z^*\|_q.$$

Since the above inequality holds for all $r \in \mathbb{C} \setminus \{0\}$, by taking $r \rightarrow 0$, we see that it holds for $r = 0$ as well. \square

We are now ready to prove Theorems 1 and 2.

Proof of Theorem 1. We prove by induction that for $k = 0, 1, \dots$, the following inequalities hold:

$$\|\epsilon^k\|_2 \leq \frac{\sqrt{n}}{2}, \quad (6)$$

$$d_2(z^{k+1}, z^*) \leq \mu \cdot d_2(z^k, z^*) + \nu \cdot \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}}. \quad (7)$$

Indeed, by the definition of ϵ^0 , Fact 1, and the assumption that $\|\Delta\|_{\text{op}} \leq \frac{n}{16}$, we have $\|\epsilon^0\|_2 = d_2(z^0, z^*) \leq \frac{\sqrt{n}}{2}$. This implies that

$$\begin{aligned} |\beta_0| &\geq \left| 1 + \alpha + \frac{\alpha}{n} \Re((z^*)^H (\epsilon^0)) \right| \\ &= \left| 1 + \alpha + \frac{\alpha}{2n} (\|z^* + \epsilon^0\|_2^2 - \|z^*\|_2^2 - \|\epsilon^0\|_2^2) \right| \\ &= \left| 1 + \alpha + \frac{\alpha}{2n} (\|z^0\|_2^2 - \|z^*\|_2^2 - \|\epsilon^0\|_2^2) \right| \\ &\geq 1 + \frac{7\alpha}{8}, \end{aligned} \quad (8)$$

where the last inequality follows from the fact that $\|z^0\|_2^2 = \|z^*\|_2^2 = n$ and $\|\epsilon^0\|_2 \leq \frac{\sqrt{n}}{2}$. Hence, by taking $r = \beta_0^{-1}$ (which is well-defined) and $q = 2$ in Proposition 2 and using (8), we have

$$\begin{aligned}
d_2(z^1, z^*) &\leq 2 \left\| \beta_0^{-1} \left(I + \frac{\alpha}{n} \Delta \right) \epsilon^0 + \beta_0^{-1} \frac{\alpha}{n} \Delta z^* \right\|_2 \\
&\leq 2 |\beta_0^{-1}| \cdot \left[\left\| \left(I + \frac{\alpha}{n} \Delta \right) \epsilon^0 \right\|_2 + \frac{\alpha}{n} \|\Delta z^*\|_2 \right] \\
&\leq \frac{16}{7\alpha + 8} \left[\left(1 + \frac{\alpha}{n} \|\Delta\|_{\text{op}} \right) \|\epsilon^0\|_2 + \frac{\alpha}{\sqrt{n}} \|\Delta\|_{\text{op}} \right] \\
&= \mu \cdot d_2(z^0, z^*) + \nu \cdot \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}}. \tag{9}
\end{aligned}$$

Now, suppose that (6) and (7) hold for some $k \geq 0$. By the inductive hypothesis and the assumption that $\|\Delta\|_{\text{op}} \leq \frac{n}{16}$ and $\alpha \geq 2$, we have

$$\begin{aligned}
\|\epsilon^{k+1}\|_2 &= d_2(z^{k+1}, z^*) \leq \mu \cdot d_2(z^k, z^*) + \nu \cdot \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}} \\
&= \mu \|\epsilon^k\|_2 + \nu \cdot \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}} \\
&\leq \frac{8(\alpha\|\Delta\|_{\text{op}} + n)}{(7\alpha + 8)\sqrt{n}} + \frac{16\alpha\|\Delta\|_{\text{op}}}{(7\alpha + 8)\sqrt{n}} \\
&\leq \frac{\sqrt{n}}{2}.
\end{aligned}$$

Using the same argument as the derivation of the inequality (8), we have $|\beta_{k+1}| \geq 1 + \frac{7\alpha}{8}$. Hence, following the same derivation as the inequality (9), we obtain $d_2(z^{k+2}, z^*) \leq \mu \cdot d_2(z^{k+1}, z^*) + \nu \cdot \frac{8\|\Delta\|_{\text{op}}}{\sqrt{n}}$. This completes the inductive step.

To complete the proof of Theorem 1, it remains to unroll (7) and use Fact 1. \square

Proof of Theorem 2. From the proof of Theorem 1, we have $\|\epsilon^k\|_2 \leq \frac{\sqrt{n}}{2}$ and $|\beta_k^{-1}| \leq \frac{8}{7\alpha+8}$ for $k = 0, 1, \dots$. By taking $r = \beta_k^{-1}$ and $q = \infty$ in Proposition 2

and using Theorem 1, we compute

$$\begin{aligned}
& d_\infty(z^{k+1}, z^*) \\
& \leq 2\|\beta_k^{-1}g^k - z^*\|_\infty \\
& \leq 2|\beta_k^{-1}| \cdot \left[\|\epsilon^k\|_\infty + \frac{\alpha}{n} (\|\Delta\epsilon^k\|_\infty + \|\Delta z^*\|_\infty) \right] \\
& \leq \frac{16}{7\alpha + 8} \left[d_\infty(z^k, z^*) + \frac{\alpha}{n} (\|\Delta\epsilon^k\|_2 + \|\Delta z^*\|_\infty) \right] \\
& \leq \frac{16}{7\alpha + 8} \left[d_\infty(z^k, z^*) + \frac{\alpha}{n} (\|\Delta\|_{\text{op}} \cdot \|\epsilon^k\|_2 + \|\Delta z^*\|_\infty) \right] \\
& \leq \frac{16}{7\alpha + 8} \left[d_\infty(z^k, z^*) + \frac{\alpha}{n} \left(\left(\mu^k + \frac{\nu}{1-\mu} \right) \frac{8\|\Delta\|_{\text{op}}^2}{\sqrt{n}} + \|\Delta z^*\|_\infty \right) \right] \\
& = \gamma \cdot d_\infty(z^k, z^*) + \zeta \cdot \mu^k + \omega. \tag{10}
\end{aligned}$$

Since $\alpha \geq 2$, we have $\gamma \in (0, 1)$ and

$$\frac{\gamma}{\mu} = \frac{16}{(7\alpha + 8)} \frac{(7\alpha + 8)n}{16(\alpha\|\Delta\|_{\text{op}} + n)} = \frac{n}{\alpha\|\Delta\|_{\text{op}} + n} \in (0, 1).$$

It follows from (10) that

$$\begin{aligned}
d_\infty(z^{k+1}, z^*) & \leq \gamma \cdot d_\infty(z^k, z^*) + \zeta \cdot \mu^k + \omega \\
& \leq \gamma^{k+1} \cdot d_\infty(z^0, z^*) + \zeta \sum_{j=0}^k \gamma^j \mu^{k-j} + \omega \sum_{j=0}^k \gamma^j \\
& \leq \gamma^{k+1} \cdot d_\infty(z^0, z^*) + \frac{\zeta \cdot \mu^k}{1 - \gamma/\mu} + \frac{\omega}{1 - \gamma}.
\end{aligned}$$

This completes the proof. \square

By specializing the above results to the Gaussian noise setting, we obtain the following corollary:

Corollary 1. *Suppose that the measurement noise Δ takes the form $\Delta = \sigma W$, where $\sigma^2 > 0$ is the noise power satisfying $\sigma \in \left(0, \frac{\sqrt{n}}{48}\right]$ and $W \in \mathbb{H}^n$ is a Wigner matrix. Suppose further that the step size α satisfies $\alpha \geq 2$ and the initial point z^0 is given by $z^0 = v_C$. Then, with probability at least $1 - 2n^{-5/4} - 2e^{-n/2}$, the sequence of iterates $\{z^k\}_{k \geq 0}$ generated by Algorithm 1 satisfies*

$$\begin{aligned}
d_2(z^{k+1}, z^*) & \leq \left(\frac{\alpha + 16}{7\alpha + 8} \right)^{k+1} d_2(z^0, z^*) + \frac{24\alpha}{3\alpha - 4} \sigma, \\
d_\infty(z^{k+1}, z^*) & \leq \left(\frac{16}{7\alpha + 8} \right)^{k+1} d_\infty(z^0, z^*) + \left(\frac{\alpha + 16}{7\alpha + 8} \right)^{k+1} \frac{\sqrt{n}}{2} \\
& \quad + \frac{48\alpha}{7\alpha + 8} \left(\sqrt{\log n} + \frac{24\alpha}{3\alpha - 4} \sigma \right) \frac{\sigma}{\sqrt{n}}
\end{aligned}$$

for $k = 0, 1, \dots$

Proof. By [2, Proposition 3.3], we have $\|W\|_{\text{op}} \leq 3\sqrt{n}$ and $\|Wz^*\|_{\infty} \leq 3\sqrt{n \log n}$ with probability at least $1 - 2n^{-5/4} - 2e^{-n/2}$. The result then follows by combining these estimates with the bounds in Theorems 1 and 2. \square

Note that by Fact 1 and [2, Proposition 3.3], we have $d_2(z^0, z^*) \leq 24\sigma$ with high probability. Hence, for $\alpha > 4$ and k sufficiently large, the bound on the ℓ_2 -estimation error will be strictly less than 12σ , which is better than that obtained from Fact 3 for any maximum likelihood estimator (which is a global maximizer of Problem (QP)) of the target phase vector z^* . Furthermore, for $k = 0, 1, \dots$, since $d_2(z^k, z^*)^2 \leq 2n$, we have

$$\begin{aligned} & \mathbb{E} [d_2(z^{k+1}, z^*)^2] \\ & \leq 1152 \left(1 - 2n^{-5/4} - 2e^{-n/2}\right) \left[\left(\frac{\alpha + 16}{7\alpha + 8}\right)^{2(k+1)} + \left(\frac{\alpha}{3\alpha - 4}\right)^2 \right] \sigma^2 \\ & \quad + \left(2n^{-5/4} + 2e^{-n/2}\right) (2n) \\ & \leq (c_1 + c_2\tau^k)\sigma^2 \end{aligned}$$

for some constants $c_1, c_2 > 0$ and $\tau \in (0, 1)$. This shows that the expected squared ℓ_2 -estimation errors of the iterates generated by Algorithm 1 are all on the order of σ , which matches the Cramér-Rao bound developed in [4]. It is worth noting that the above conclusions hold even when the noise level is $\sigma = O(n^{1/2})$, which is the least restrictive among similar results in the literature; cf. [2, 3]. Our result explains in part the excellent numerical estimation performance of the GPM observed in [3] even when the noise level is close to $O(n^{1/2})$.

4 Convergence Rate of the GPM

Although the results in the previous section show that Algorithm 1 generates increasingly accurate (in the ℓ_2 and ℓ_{∞} sense) estimators of the target phase vector z^* , they do not shed any light on its convergence behavior. On the other hand, recall from Fact 2 that the sequence of iterates generated by Algorithm 1 will converge to a global maximizer of Problem (QP) under suitable assumptions on the measurement noise Δ and step size α . However, it does not give the rate of convergence. In this section, we prove that under weaker assumptions than those of Fact 2, both the sequence of iterates and the associated sequence of objective values generated by Algorithm 1 will converge *linearly* to a global maximizer and the optimal value of Problem (QP), respectively. Specifically, we have the following result:

Theorem 3. *Suppose that (i) the measurement noise Δ satisfies $\|\Delta\|_{\text{op}} \leq \frac{n^{3/4}}{312}$ and $\|\Delta z^*\|_{\infty} \leq \frac{n}{24}$, (ii) the step size α satisfies $\alpha \in \left[4, \frac{n}{\|\Delta\|_{\text{op}}}\right)$, and (iii) the*

initial point z^0 is given by $z^0 = v_C$. Then, the sequence of iterates $\{z^k\}_{k \geq 0}$ generated by Algorithm 1 satisfies

$$\begin{aligned} f(\hat{z}) - f(z^k) &\leq (f(\hat{z}) - f(z^0)) \lambda^k, \\ d_2(z^k, \hat{z}) &\leq a (f(\hat{z}) - f(z^0))^{1/2} \lambda^{k/2} \end{aligned}$$

for $k = 0, 1, \dots$, where $a > 0, \lambda \in (0, 1)$ are quantities that depend only on n and α , and \hat{z} is any global maximizer of Problem (QP).

Theorem 3 improves upon Fact 2 in two aspects. First, Theorem 3 holds under a less restrictive requirement on the measurement noise Δ . Specifically, it requires that $\|\Delta\|_{\text{op}} = O(n^{3/4})$ and $\|\Delta z^*\|_{\infty} = O(n)$, while Fact 2 requires that $\|\Delta\|_{\text{op}} = O(n^{2/3})$ and $\|\Delta z^*\|_{\infty} = O(n^{2/3} \sqrt{\log n})$. Second, Theorem 3 is more quantitative than Fact 2 in the sense that it also gives the rate at which Algorithm 1 converges. Consequently, we resolve an open question raised in [3].

The proof of Theorem 3 consists of two main parts. The first, which is the more challenging part, is to establish the following *error bound* for Problem (QP). Such a bound provides a computable estimate of the distance between any point in a neighborhood of z^* and the set of global maximizers of Problem (QP), which could be of independent interest.

Proposition 3. *Let $\Sigma : \mathbb{T}^n \rightarrow \mathbb{H}^n$ and $\rho : \mathbb{T}^n \rightarrow \mathbb{R}_+$ be defined as*

$$\Sigma(z) = \text{Diag}(|\tilde{C}z|) - \tilde{C}, \quad \rho(z) = \|\Sigma(z)z\|_2 = \left\| \left(\text{Diag}(|\tilde{C}z|) - \tilde{C} \right) z \right\|_2,$$

where $\tilde{C} = C + \frac{n}{\alpha}I$. Under the assumptions of Theorem 3, for any point $z \in \mathbb{T}^n$ satisfying $d_2(z, z^*) \leq \frac{\sqrt{n}}{2}$ and any global maximizer $\hat{z} \in \mathbb{T}^n$ of Problem (QP), we have

$$d_2(z, \hat{z}) \leq \frac{8}{n} \rho(z).$$

Before we prove Proposition 3, several remarks are in order. First, recall from Fact 3 that every global maximizer \hat{z} of Problem (QP) satisfies $d_2(\hat{z}, z^*) \leq \frac{\sqrt{n}}{2}$ whenever $\|\Delta\|_{\text{op}} \leq \frac{n}{8}$. Together with Proposition 3, this shows that up to a global phase, Problem (QP) has a unique global maximizer. Second, the proof of Theorem 1 reveals that the sequence of iterates $\{z^k\}_{k \geq 0}$ generated by Algorithm 1 satisfies $d_2(z^k, z^*) \leq \frac{\sqrt{n}}{2}$ for $k = 0, 1, \dots$ whenever $\|\Delta\|_{\text{op}} \leq \frac{n}{16}$. Thus, the error bound in Proposition 3 applies to the entire sequence $\{z^k\}_{k \geq 0}$. Third, since every global maximizer \hat{z} of Problem (QP) is a second-order critical point, we have $\rho(\hat{z}) = 0$ by Fact 4. Proposition 3 shows that the converse is also true. Hence, we can view ρ as a surrogate measure of optimality and use it to keep track of Algorithm 1's progress.

Proof of Proposition 3. Since

$$\rho(z) = \|\Sigma(z)z\|_2 \geq \|\Sigma(\hat{z})z\|_2 - \|(\Sigma(z) - \Sigma(\hat{z}))z\|_2, \quad (11)$$

it suffices to establish an upper bound on $\|(\Sigma(z) - \Sigma(\hat{z}))z\|_2$ and a lower bound on $\|\Sigma(\hat{z})z\|_2$. Towards that end, recall that

$$\tilde{C} = C + \frac{n}{\alpha}I = (z^*)(z^*)^H + \Delta + \frac{n}{\alpha}I$$

and let

$$\hat{\theta} = \arg \min_{\theta \in [0, 2\pi)} \|z - e^{i\theta} \hat{z}\|_2, \quad \hat{\theta}^* = \arg \min_{\theta \in [0, 2\pi)} \|\hat{z} - e^{i\theta} z^*\|_2.$$

First, we bound

$$\begin{aligned} & \|(\Sigma(z) - \Sigma(\hat{z}))z\|_2 \\ &= \left\| \left(\text{Diag}(|\tilde{C}z|) - \text{Diag}(|\tilde{C}\hat{z}|) \right) z \right\|_2 \\ &= \left(\sum_{j=1}^n \left| (|\tilde{C}z|)_j - (|\tilde{C}\hat{z}|)_j \right| |z_j|^2 \right)^{1/2} \\ &= \left\| |\tilde{C}e^{-i\hat{\theta}}z| - |\tilde{C}\hat{z}| \right\|_2 \\ &\leq \|\tilde{C}(e^{-i\hat{\theta}}z - \hat{z})\|_2 \\ &\leq \|(z^*)(z^*)^H(e^{-i\hat{\theta}}z - \hat{z})\|_2 + \|\Delta(e^{-i\hat{\theta}}z - \hat{z})\|_2 + \frac{n}{\alpha}\|e^{-i\hat{\theta}}z - \hat{z}\|_2 \\ &\leq \sqrt{n} \cdot |(z^*)^H(e^{-i\hat{\theta}}z - \hat{z})| + \left(\|\Delta\|_{\text{op}} + \frac{n}{\alpha} \right) d_2(z, \hat{z}). \end{aligned} \quad (12)$$

By definition of $\hat{\theta}$, we have $\hat{z}^H(e^{-i\hat{\theta}}z) = |\hat{z}^H z|$, which implies that

$$\|e^{-i\hat{\theta}}z - \hat{z}\|_2^2 = 2(n - |\hat{z}^H z|). \quad (13)$$

This, together with Fact 3, yields

$$\begin{aligned} |(z^*)^H(e^{-i\hat{\theta}}z - \hat{z})| &\leq |(z^* - e^{-i\hat{\theta}^*} \hat{z})^H(e^{-i\hat{\theta}}z - \hat{z})| + |(e^{-i\hat{\theta}^*} \hat{z})^H(e^{-i\hat{\theta}}z - \hat{z})| \\ &\leq \|z^* - e^{-i\hat{\theta}^*} \hat{z}\|_2 \cdot \|e^{-i\hat{\theta}}z - \hat{z}\|_2 + |\hat{z}^H z| - n \\ &\leq \frac{4\|\Delta\|_{\text{op}}}{\sqrt{n}} \cdot d_2(z, \hat{z}) + \frac{1}{2}d_2(z, \hat{z})^2. \end{aligned} \quad (14)$$

Upon substituting (14) into (12), we obtain

$$\|(\Sigma(z) - \Sigma(\hat{z}))z\|_2 \leq \left(5\|\Delta\|_{\text{op}} + \frac{n}{\alpha} \right) d_2(z, \hat{z}) + \frac{\sqrt{n}}{2}d_2(z, \hat{z})^2. \quad (15)$$

Next, let $\hat{u} = (I - \frac{1}{n}\hat{z}\hat{z}^H)(e^{-i\hat{\theta}}z - \hat{z})$ be the projection of $e^{-i\hat{\theta}}z - \hat{z}$ onto the orthogonal complement of $\text{span}(\hat{z})$. Hence, we have $\hat{u}^H \hat{z} = 0$ and

$$\|\hat{u}\|_2 \geq \left\| e^{-i\hat{\theta}}z - \hat{z} \right\|_2 - \left\| \frac{1}{n}\hat{z}\hat{z}^H(e^{-i\hat{\theta}}z - \hat{z}) \right\|_2 = d_2(z, \hat{z}) - \frac{1}{2\sqrt{n}}d_2(z, \hat{z})^2, \quad (16)$$

where the last equality follows from (13). Moreover, by definition of $\hat{\theta}^*$, we have $(z^*)^H(e^{-i\hat{\theta}^*}\hat{z}) = |(z^*)^H\hat{z}|$ and

$$|(z^*)^H\hat{z}| = n - \frac{1}{2}\|\hat{z} - e^{i\hat{\theta}^*}z^*\|_2^2. \quad (17)$$

Hence,

$$\begin{aligned} \hat{u}^H\Sigma(\hat{z})\hat{u} &= \hat{u}^H\left(\text{Diag}(|\tilde{C}\hat{z}|) - \tilde{C}\right)\hat{u} \\ &= \hat{u}^H\left(\text{Diag}(|C\hat{z}|) - C\right)\hat{u} \end{aligned} \quad (18)$$

$$\begin{aligned} &\geq \left(\sum_{j=1}^n |(C\hat{z})_j| \cdot |\hat{u}_j|^2\right) - |(z^*)^H\hat{u}|^2 - \hat{u}^H\Delta\hat{u} \\ &\geq \left(|(z^*)^H\hat{z}| - \|\Delta\hat{z}\|_\infty\right)\|\hat{u}\|_2^2 - \left|\hat{u}^H(z^* - e^{-i\hat{\theta}^*}\hat{z})\right|^2 - \|\Delta\|_{\text{op}}\|\hat{u}\|_2^2 \\ &\geq \left(n - \|\Delta\hat{z}\|_\infty - \|\Delta\|_{\text{op}} - \frac{3}{2}\|z^* - e^{-i\hat{\theta}^*}\hat{z}\|_2^2\right)\|\hat{u}\|_2^2 \end{aligned} \quad (19)$$

$$\geq \left(n - \|\Delta\hat{z}\|_\infty - \|\Delta\|_{\text{op}} - \frac{24\|\Delta\|_{\text{op}}^2}{n}\right)\|\hat{u}\|_2^2, \quad (20)$$

where (18) follows from Fact 4 and the fact that \hat{z} is a second-order critical point of Problem (QP), (19) is due to (17), and (20) follows from Fact 3. Since $\Sigma(\hat{z})\hat{z} = \mathbf{0}$ by Fact 4, we obtain from (16) and (20) that

$$\begin{aligned} \|\Sigma(\hat{z})z\|_2 &= \|\Sigma(\hat{z})\hat{u}\|_2 \\ &\geq \left(n - \|\Delta\hat{z}\|_\infty - \|\Delta\|_{\text{op}} - \frac{24\|\Delta\|_{\text{op}}^2}{n}\right)\left(d_2(z, \hat{z}) - \frac{1}{2\sqrt{n}}d_2(z, \hat{z})^2\right). \end{aligned} \quad (21)$$

Now, by Fact 3 and the assumption that $d_2(z, z^*) \leq \frac{\sqrt{n}}{2}$, we have

$$d_2(z, \hat{z}) \leq d_2(z, z^*) + d_2(\hat{z}, z^*) \leq \frac{\sqrt{n}}{2} + \frac{4\|\Delta\|_{\text{op}}}{\sqrt{n}}.$$

This implies that

$$d_2(z, \hat{z})^2 \leq \left(\frac{\sqrt{n}}{2} + \frac{4\|\Delta\|_{\text{op}}}{\sqrt{n}}\right)d_2(z, \hat{z}). \quad (22)$$

Moreover,

$$\begin{aligned} \|\Delta\hat{z}\|_\infty &\leq \|\Delta z^*\|_\infty + \|\Delta(e^{-i\hat{\theta}^*}\hat{z} - z^*)\|_\infty \\ &\leq \|\Delta z^*\|_\infty + \|\Delta\|_{\text{op}} \cdot d_2(\hat{z}, z^*) \\ &\leq \|\Delta z^*\|_\infty + \frac{4\|\Delta\|_{\text{op}}^2}{\sqrt{n}}. \end{aligned} \quad (23)$$

It follows from (11), (15), (21), (22), and (23) that

$$\begin{aligned}
\rho(z) &\geq \left[\left(\frac{1}{2} - \frac{1}{\alpha} \right) n - \frac{3\|\Delta\hat{z}\|_\infty}{4} - \frac{39\|\Delta\|_{\text{op}}}{4} - \frac{16\|\Delta\|_{\text{op}}^2}{n} \right] d_2(z, \hat{z}) \\
&\geq \left[\frac{n}{4} - \frac{3\|\Delta z^*\|_\infty}{4} - \frac{39\|\Delta\|_{\text{op}}}{4} - \frac{3\|\Delta\|_{\text{op}}^2}{\sqrt{n}} - \frac{16\|\Delta\|_{\text{op}}^2}{n} \right] d_2(z, \hat{z}) \\
&\geq \frac{n}{8} d_2(z, \hat{z})
\end{aligned}$$

whenever $\|\Delta\|_{\text{op}} \leq \frac{n^{3/4}}{312}$, $\|\Delta z^*\|_\infty \leq \frac{n}{24}$, and $\alpha \geq 4$. This completes the proof. \square

We note that under a slightly more restrictive noise setting, one can establish an error bound similar to that in Proposition 3 to estimate the distance between any point in a neighborhood of z^* and the set of *second-order critical points* of Problem (QP). Specifically, we have the following result:

Proposition 4. *Let $\Sigma : \mathbb{T}^n \rightarrow \mathbb{H}^n$ and $\rho : \mathbb{T}^n \rightarrow \mathbb{R}_+$ be as in Proposition 3. Suppose that (i) the measurement noise Δ satisfies $\|\Delta\|_{\text{op}} \leq \frac{n^{2/3}}{32768}$ and $\|\Delta z^*\|_\infty \leq \frac{n}{24}$, and (ii) the parameter α satisfies $\alpha \geq 4$. Then, for any point $z \in \mathbb{T}^n$ satisfying $d_2(z, z^*) \leq \frac{\sqrt{n}}{2}$ and any second-order critical point $\tilde{z} \in \mathbb{T}^n$ of Problem (QP), we have*

$$d_2(z, \tilde{z}) \leq \frac{8}{n} \rho(z).$$

Proof. The proof is essentially the same as that of Proposition 3. To save space, let us just highlight the key steps. Similar to (11), we have

$$\rho(z) = \|\Sigma(z)z\|_2 \geq \|\Sigma(\tilde{z})z\|_2 - \|(\Sigma(z) - \Sigma(\tilde{z}))z\|_2. \quad (24)$$

Define

$$\tilde{\theta} = \arg \min_{\theta \in [0, 2\pi)} \|z - e^{i\theta} \tilde{z}\|_2, \quad \tilde{\theta}^* = \arg \min_{\theta \in [0, 2\pi)} \|\tilde{z} - e^{i\theta} z^*\|_2.$$

Then, since $d_2(\tilde{z}, z^*) = \|z^* - e^{-i\tilde{\theta}^*} \tilde{z}\|_2 \leq \sqrt{8\|\Delta\|_{\text{op}}}$ by Fact 4, we have

$$\begin{aligned}
&\|(\Sigma(z) - \Sigma(\tilde{z}))z\|_2 \\
&\leq \sqrt{n} \cdot |(z^*)^H (e^{-i\tilde{\theta}} z - \tilde{z})| + \left(\|\Delta\|_{\text{op}} + \frac{n}{\alpha} \right) d_2(z, \tilde{z}) \\
&\leq \sqrt{n} \left(\|z^* - e^{-i\tilde{\theta}^*} \tilde{z}\|_2 \cdot \|e^{-i\tilde{\theta}} z - \tilde{z}\|_2 + \left| |\tilde{z}^H z| - n \right| \right) \\
&\quad + \left(\|\Delta\|_{\text{op}} + \frac{n}{\alpha} \right) d_2(z, \tilde{z}) \\
&\leq \left(\sqrt{8n\|\Delta\|_{\text{op}}} + \|\Delta\|_{\text{op}} + \frac{n}{\alpha} \right) d_2(z, \tilde{z}) + \frac{\sqrt{n}}{2} d_2(z, \tilde{z})^2. \quad (25)
\end{aligned}$$

Now, let $\tilde{u} = (I - \frac{1}{n}\tilde{z}\tilde{z}^H)(e^{-i\tilde{\theta}}z - \tilde{z})$ be the projection of $e^{-i\tilde{\theta}}z - \tilde{z}$ onto the orthogonal complement of $\text{span}(\tilde{z})$. Then, similar to the derivation of (21), we have

$$\begin{aligned} \|\Sigma(\tilde{z})z\|_2 &= \|\Sigma(\tilde{z})\tilde{u}\|_2 \\ &\geq (|(z^*)^H\tilde{z}| - \|\Delta\tilde{z}\|_\infty)\|\tilde{u}\|_2^2 - \left|\tilde{u}^H(z^* - e^{-i\tilde{\theta}^*}\tilde{z})\right|^2 - \|\Delta\|_{\text{op}}\|\tilde{u}\|_2^2 \\ &\geq (n - \|\Delta\tilde{z}\|_\infty - 13\|\Delta\|_{\text{op}})\left(d_2(z, \tilde{z}) - \frac{1}{2\sqrt{n}}d_2(z, \tilde{z})^2\right). \end{aligned} \quad (26)$$

Moreover, following the derivations of (22) and (23), we have

$$d_2(z, \tilde{z}) \leq \frac{\sqrt{n}}{2} + \sqrt{8\|\Delta\|_{\text{op}}}, \quad (27)$$

$$\|\Delta\tilde{z}\|_\infty \leq \|\Delta z^*\|_\infty + \sqrt{8}\|\Delta\|_{\text{op}}^{3/2}. \quad (28)$$

Upon putting together (24)–(28), we obtain

$$\begin{aligned} \rho(z) &\geq \left[\frac{n}{4} - \frac{3\|\Delta z^*\|_\infty}{4} - 4\sqrt{2n\|\Delta\|_{\text{op}}} - \frac{43\|\Delta\|_{\text{op}}}{4} - \frac{3\sqrt{2}\|\Delta\|_{\text{op}}^{3/2}}{2}\right]d_2(z, \tilde{z}) \\ &\geq \frac{n}{8}d_2(z, \tilde{z}) \end{aligned}$$

whenever $\|\Delta\|_{\text{op}} \leq \frac{n^{2/3}}{32768}$, $\|\Delta z^*\|_\infty \leq \frac{n}{24}$, and $\alpha \geq 4$. This completes the proof. \square

Recall that a global maximizer of Problem (QP) is a second-order critical point. Now, under the assumptions of Proposition 4, we know that every second-order critical point \tilde{z} of Problem (QP) satisfies $d_2(\tilde{z}, z^*) \leq \sqrt{8\|\Delta\|_{\text{op}}} \leq \frac{\sqrt{n}}{2}$; see Fact 4. Thus, Proposition 4 shows that every second-order critical point of Problem (QP) is also a global maximizer, which is unique up to a global phase. This gives an alternative proof of [3, Theorem 4] with a less restrictive requirement on $\|\Delta z^*\|_\infty$ ($\|\Delta z^*\|_\infty = O(n)$ in Proposition 4 vs. $\|\Delta z^*\|_\infty = O(n^{2/3}\sqrt{\log n})$ in [3, Theorem 4]). It remains an open question to determine whether the conclusion of Proposition 4 still holds under the same noise requirement as Proposition 3.

Now, let us proceed to the second part of the proof of Theorem 3. Our goal is to prove the following proposition, which elucidates the key properties of Algorithm 1:

Proposition 5. *Under the assumptions of Theorem 3, the sequence of iterates $\{z^k\}_{k \geq 0}$ generated by Algorithm 1 satisfies the following for $k = 0, 1, \dots$, where $a_0, a_1, a_2 > 0$ are quantities that depend only on n and α , and \hat{z} is any global maximizer of Problem (QP):*

- (a) (Sufficient Ascent) $f(z^{k+1}) - f(z^k) \geq a_0 \cdot \|z^{k+1} - z^k\|_2^2$.
- (b) (Cost-to-Go Estimate) $f(\hat{z}) - f(z^k) \leq a_1 \cdot d_2(z^k, \hat{z})^2$.

(c) (Safeguard) $\rho(z^k) \leq a_2 \cdot \|z^{k+1} - z^k\|_2$.

Proof. We begin by proving (a). Recalling that $\tilde{C} = C + \frac{n}{\alpha}I$, we have

$$\begin{aligned} & f(z^{k+1}) - f(z^k) \\ &= (z^{k+1} - z^k)^H \tilde{C} (z^{k+1} - z^k) - 2(z^k)^H \tilde{C} (z^k) + 2\Re\{(z^{k+1})^H \tilde{C} (z^k)\}. \end{aligned}$$

We claim that $\Re\{(z^{k+1})^H \tilde{C} (z^k)\} \geq (z^k)^H \tilde{C} (z^k)$. This follows from the fact that

$$(z^{k+1})^H \tilde{C} (z^k) = \left(\frac{\tilde{C} z^k}{|\tilde{C} z^k|} \right)^H \tilde{C} (z^k)$$

is a real number and

$$z^{k+1} \in \arg \max_{z \in \mathbb{T}^n} \Re\{z^H \tilde{C} z\}.$$

Hence, by the assumption on α , we have

$$f(z^{k+1}) - f(z^k) \geq (z^{k+1} - z^k)^H \tilde{C} (z^{k+1} - z^k) \geq a_0 \cdot \|z^{k+1} - z^k\|_2^2$$

with $a_0 = \lambda_{\min}(\Delta + \frac{n}{\alpha}I) > 0$.

Next, we prove (b). Let $\hat{\theta}_k = \arg \min_{\theta \in [0, 2\pi)} \|z^k - e^{i\theta} \hat{z}\|_2$. Then, we have

$$\begin{aligned} f(\hat{z}) - f(z^k) &= \hat{z}^H \tilde{C} \hat{z} - (z^k)^H \tilde{C} (z^k) \\ &= \|\tilde{C} \hat{z}\|_1 - (z^k)^H \tilde{C} (z^k) \end{aligned} \quad (29)$$

$$\begin{aligned} &= (z^k)^H \left(\text{Diag}(|\tilde{C} \hat{z}|) - \tilde{C} \right) (z^k) \\ &= (e^{-i\hat{\theta}_k} z^k - \hat{z})^H \left(\text{Diag}(|\tilde{C} \hat{z}|) - \tilde{C} \right) (e^{-i\hat{\theta}_k} z^k - \hat{z}) \quad (30) \\ &\leq \left(\|\tilde{C}\|_{\text{op}} + \|\tilde{C} \hat{z}\|_{\infty} \right) d_2(z^k, \hat{z})^2, \end{aligned}$$

where both (29) and (30) follow from Fact 4. Now, observe that

$$\begin{aligned} & \|\tilde{C}\|_{\text{op}} + \|\tilde{C} \hat{z}\|_{\infty} \\ &\leq \|C\|_{\text{op}} + \|C \hat{z}\|_{\infty} + \frac{2n}{\alpha} \\ &\leq \|(z^*) (z^*)^H\|_{\text{op}} + \|\Delta\|_{\text{op}} + \|(z^*) (z^*)^H \hat{z}\|_{\infty} + \|\Delta \hat{z}\|_{\infty} + \frac{2n}{\alpha} \\ &\leq n + \|\Delta\|_{\text{op}} + |(z^*)^H \hat{z}| + \|\Delta z^*\|_{\infty} + \frac{4\|\Delta\|_{\text{op}}^2}{\sqrt{n}} + \frac{2n}{\alpha} \end{aligned} \quad (31)$$

$$\leq 2n + \|\Delta\|_{\text{op}} + \|\Delta z^*\|_{\infty} + \frac{4\|\Delta\|_{\text{op}}^2}{\sqrt{n}} + \frac{2n}{\alpha} \quad (32)$$

$$< 3n, \quad (33)$$

where (31) follows from (23) and the fact that $\|(z^*)(z^*)^H\|_{\text{op}} = n$, (32) follows from (17) and Fact 3, and (33) is due to the assumptions on α , $\|\Delta\|_{\text{op}}$, and $\|\Delta z^*\|_{\infty}$. Hence, we conclude that

$$f(\hat{z}) - f(z^k) \leq a_1 \cdot d_2(z^k, \hat{z})^2$$

for some $a_1 \in (0, 3n)$.

Lastly, we prove (c). By definition of z^{k+1} , we have

$$\text{Diag}(|\tilde{C}z^k|)(z^{k+1} - z^k) = \left(\tilde{C} - \text{Diag}(|\tilde{C}z^k|)\right) z^k.$$

It follows that

$$\rho(z^k) = \left\| \text{Diag}(|\tilde{C}z^k|)(z^{k+1} - z^k) \right\|_2 \leq \|\tilde{C}z^k\|_{\infty} \|z^{k+1} - z^k\|_2.$$

Now, recall from the proof of Theorem 1 that $d_2(z^k, z^*) \leq \frac{\sqrt{n}}{2}$ for $k = 0, 1, \dots$. Upon letting $\theta_k^* = \arg \min_{\theta \in [0, 2\pi)} \|z^k - e^{i\theta} z^*\|_2$, we obtain

$$\begin{aligned} \|\tilde{C}z^k\|_{\infty} &\leq \|(z^*)(z^*)^H z^k\|_{\infty} + \|\Delta z^k\|_{\infty} + \frac{n}{\alpha} \\ &\leq |(z^*)^H z^k| + \|\Delta(e^{-i\theta_k^*} z^k - z^*)\|_{\infty} + \|\Delta z^*\|_{\infty} + \frac{n}{\alpha} \\ &\leq n + \|\Delta\|_{\text{op}} \cdot d_2(z^k, z^*) + \|\Delta z^*\|_{\infty} + \frac{n}{\alpha} \\ &< \frac{3n^{5/4}}{2}, \end{aligned}$$

where the last inequality is due to the assumptions on α , $\|\Delta\|_{\text{op}}$, and $\|\Delta z^*\|_{\infty}$. It follows that

$$\rho(z^k) \leq a_2 \cdot \|z^{k+1} - z^k\|_2$$

for some $a_2 \in \left(0, \frac{3n^{5/4}}{2}\right)$. \square

Armed with Propositions 3 and 5, we are now ready to prove Theorem 3.

Proof of Theorem 3. By Propositions 3 and 5, we have

$$\begin{aligned} f(\hat{z}) - f(z^{k+1}) &= (f(\hat{z}) - f(z^k)) - (f(z^{k+1}) - f(z^k)) \\ &\leq a_1 \cdot d_2(z^k, \hat{z})^2 - (f(z^{k+1}) - f(z^k)) \\ &\leq \frac{64a_1}{n^2} \rho(z^k)^2 - (f(z^{k+1}) - f(z^k)) \\ &\leq \frac{64a_1 a_2^2}{n^2} \|z^{k+1} - z^k\|_2^2 - (f(z^{k+1}) - f(z^k)) \\ &\leq \left(\frac{64a_1 a_2^2}{a_0 n^2} - 1\right) (f(z^{k+1}) - f(\hat{z}) + f(\hat{z}) - f(z^k)). \end{aligned}$$

Since $f(\hat{z}) \geq f(z^k)$ for $k = 0, 1, \dots$, we may assume without loss of generality that $a' = \frac{64a_1a_2^2}{a_0n^2} > 1$. It then follows that

$$f(\hat{z}) - f(z^{k+1}) \leq \frac{a' - 1}{a'} (f(\hat{z}) - f(z^k)),$$

which yields

$$f(\hat{z}) - f(z^k) \leq (f(\hat{z}) - f(z^0)) \lambda^k$$

with $\lambda = \frac{a' - 1}{a'} \in (0, 1)$. Furthermore, we have

$$\begin{aligned} d_2(z^k, \hat{z})^2 &\leq \frac{64}{n^2} \rho(z^k)^2 \\ &\leq \frac{64a_2^2}{n^2} \|z^{k+1} - z^k\|_2^2 \\ &\leq \frac{64a_2^2}{a_0n^2} (f(z^{k+1}) - f(z^k)) \\ &\leq \frac{64a_2^2}{a_0n^2} (f(\hat{z}) - f(z^k)) \\ &\leq \frac{64a_2^2}{a_0n^2} (f(\hat{z}) - f(z^0)) \lambda^k, \end{aligned}$$

which implies that

$$d_2(z^k, \hat{z}) \leq a (f(\hat{z}) - f(z^0))^{1/2} \lambda^{k/2}$$

with $a = \sqrt{\frac{64a_2^2}{a_0n^2}}$. This completes the proof. \square

Again, we can specialize Theorem 3 to the Gaussian noise setting. This leads to the following corollary, which can be proven by combining Theorem 3 with the probabilistic estimates in [2, Proposition 3.3]; cf. Corollary 1:

Corollary 2. *Suppose that the measurement noise Δ takes the form $\Delta = \sigma W$, where $\sigma^2 > 0$ is the noise power satisfying $\sigma \in \left(0, \frac{n^{1/4}}{936}\right]$ and $W \in \mathbb{H}^n$ is a Wigner matrix. Suppose further that the step size α satisfies $\alpha \in [4, 312n^{1/4})$ and the initial point z^0 is given by $z^0 = v_C$. Then, with probability at least $1 - 2n^{-5/4} - 2e^{-n/2}$, the sequence of iterates $\{z^k\}_{k \geq 0}$ generated by Algorithm 1 satisfies*

$$\begin{aligned} f(\hat{z}) - f(z^k) &\leq (f(\hat{z}) - f(z^0)) \lambda^k, \\ d_2(z^k, \hat{z}) &\leq a (f(\hat{z}) - f(z^0))^{1/2} \lambda^{k/2} \end{aligned}$$

for $k = 0, 1, \dots$, where $a > 0, \lambda \in (0, 1)$ are quantities that depend only on n and α , and \hat{z} is any global maximizer of Problem (QP).

Corollary 2 shows that in the Gaussian noise setting, Algorithm 1 will converge to a global maximizer of Problem (QP) at a linear rate with high probability for noise level up to $\sigma = O(n^{1/4})$. This matches the noise level requirement for the tightness of the SDR-based method established in [2, Theorem 2.1]. As the GPM typically has lower complexity than the SDR-based method in tackling Problem (QP), we see that the former is competitive with the latter in terms of both theoretical guarantees and numerical efficiency.

5 Conclusion

In this paper, we conducted a comprehensive analysis of the estimation and convergence performance of the GPM for tackling the phase synchronization problem. First, under the assumption that the measurement noise Δ satisfies $\|\Delta\|_{\text{op}} = O(n)$, we established bounds on the rates of decrease in the ℓ_2 - and ℓ_∞ -estimation errors of the iterates generated by the GPM. As a corollary, we showed that in the Gaussian noise setting (i.e., $\Delta = \sigma W$, where $\sigma > 0$ is the noise level and W is a Wigner matrix), the expected squared ℓ_2 -estimation errors of the iterates are decreasing and all are on the same order as that of the MLE even when the noise level is $\sigma = O(n^{1/2})$. The above result holds regardless of whether the iterates converge or not and yields the best provable bound on the estimation error of any accumulation point generated by the GPM under the least restrictive noise requirement currently known. Second, we showed that when the measurement noise Δ and target phase vector z^* satisfy $\|\Delta\|_{\text{op}} = O(n^{3/4})$ and $\|\Delta z^*\|_\infty = O(n)$, the GPM will converge linearly to a global maximizer of Problem (QP). This not only resolves an open question in [3] concerning the convergence *rate* of the GPM but also improves upon the noise requirement $\|\Delta\|_{\text{op}} = O(n^{2/3})$ and $\|\Delta z^*\|_\infty = O(n^{2/3}\sqrt{\log n})$ that is imposed in [3] to establish just the convergence of the GPM. Our result implies that in the Gaussian noise setting, the GPM will converge linearly to a global maximizer of Problem (QP) in the noise regime $\sigma = O(n^{1/4})$. This is the same regime for which the computationally heavier SDR-based method in [2] is provably tight. To establish our convergence rate result, we developed a new error bound for the non-convex problem (QP). As a by-product, we showed that every second-order critical point of Problem (QP) is globally optimal if $\|\Delta\|_{\text{op}} = O(n^{2/3})$ and $\|\Delta z^*\|_\infty = O(n)$. This slightly improves upon the corresponding result in [3]. An interesting future direction would be to extend the GPM and the machinery developed in this paper to design and analyze first-order methods for other (non-convex) quadratic optimization problems.

Acknowledgement

We thank Nicolas Boumal for his helpful comments on an earlier version of our manuscript.

References

- [1] Boris Alexeev, Afonso S. Bandeira, Matthew Fickus, and Dustin G. Mixon. Phase Retrieval with Polarization. *SIAM Journal on Imaging Sciences*, 7(1):35–66, 2014.
- [2] Afonso S. Bandeira, Nicolas Boumal, and Amit Singer. Tightness of the Maximum Likelihood Semidefinite Relaxation for Angular Synchronization. Accepted for publication in *Mathematical Programming, Series A*, 2016.
- [3] Nicolas Boumal. Nonconvex Phase Synchronization. Manuscript, available at <http://arxiv.org/abs/1601.06114>, 2016.
- [4] Nicolas Boumal, Amit Singer, P.-A. Absil, and Vincent D. Blondel. Cramér–Rao Bounds for Synchronization of Rotations. *Information and Inference: A Journal of the IMA*, 3(1):1–39, 2014.
- [5] Mihai Cucuringu. Sync–Rank: Robust Ranking, Constrained Ranking and Rank Aggregation via Eigenvector and SDP Synchronization. *IEEE Transactions on Network Science and Engineering*, 3(1):58–79, 2016.
- [6] Arvind Giridhar and P. R. Kumar. Distributed Clock Synchronization over Wireless Networks: Algorithms and Analysis. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 4915–4920, 2006.
- [7] Ke Hou, Zirui Zhou, Anthony Man-Cho So, and Zhi-Quan Luo. On the Linear Convergence of the Proximal Gradient Method for Trace Norm Regularization. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: Proceedings of the 2013 Conference*, pages 710–718, 2013.
- [8] Michel Journée, Yurii Nesterov, Peter Richtárik, and Rodolphe Sepulchre. Generalized Power Method for Sparse Principal Component Analysis. *Journal of Machine Learning Research*, 11(Feb.):517–553, 2010.
- [9] Guoyin Li and Ting Kei Pong. Calculus of the Exponent of Kurdyka–Lojasiewicz Inequality and Its Applications to Linear Convergence of First–Order Methods. Manuscript, available at <http://arxiv.org/abs/1602.02915>, 2016.
- [10] Huikang Liu, Weijie Wu, and Anthony Man-Cho So. Quadratic Optimization with Orthogonality Constraints: Explicit Lojasiewicz Exponent and Linear Convergence of Line–Search Methods. In *Proceedings of the 33rd International Conference on Machine Learning (ICML 2016)*, pages 1158–1167, 2016.
- [11] Zhi-Quan Luo, Wing-Kin Ma, Anthony Man-Cho So, Yinyu Ye, and Shuzhong Zhang. Semidefinite Relaxation of Quadratic Optimization Problems. *IEEE Signal Processing Magazine*, 27(3):20–34, 2010.

- [12] Ronny Luss and Marc Teboulle. Conditional Gradient Algorithms for Rank-One Matrix Approximations with a Sparsity Constraint. *SIAM Review*, 55(1):65–98, 2013.
- [13] A. Singer. Angular Synchronization by Eigenvectors and Semidefinite Programming. *Applied and Computational Harmonic Analysis*, 30(1):20–36, 2011.
- [14] Anthony Man-Cho So. Probabilistic Analysis of the Semidefinite Relaxation Detector in Digital Communications. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2010)*, pages 698–711, 2010.
- [15] Anthony Man-Cho So, Jiawei Zhang, and Yinyu Ye. On Approximating Complex Quadratic Optimization Problems via Semidefinite Programming Relaxations. *Mathematical Programming, Series B*, 110(1):93–110, 2007.
- [16] Anthony Man-Cho So and Zirui Zhou. Non-Asymptotic Convergence Analysis of Inexact Gradient Methods for Machine Learning Without Strong Convexity. Manuscript, available at http://www.se.cuhk.edu.hk/~manchos/papers/inexact_GM_conv.pdf, 2015.
- [17] Ju Sun, Qing Qu, and John Wright. When Are Nonconvex Problems Not Scary? Manuscript, available at <http://arxiv.org/abs/1510.06096>, 2015.
- [18] Onur Toker and Hitay Özbay. On the Complexity of Purely Complex μ Computation and Related Problems in Multidimensional Systems. *IEEE Transactions on Automatic Control*, 43(3):409–414, 1998.
- [19] Aditya Viswanathan and Mark Iwen. Fast Angular Synchronization for Phase Retrieval via Incomplete Information. In *Proceedings of SPIE*, volume 9597, pages 959718–1–959718–8, 2015.
- [20] Hoi-To Wai, Wing-Kin Ma, and Anthony Man-Cho So. Cheap Semidefinite Relaxation MIMO Detection Using Row-by-Row Block Coordinate Descent. In *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, pages 3256–3259, 2011.
- [21] Zaiwen Wen, Donald Goldfarb, and Katya Scheinberg. Block Coordinate Descent Methods for Semidefinite Programming. In Miguel F. Anjos and Jean B. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166 of *International Series in Operations Research and Management Science*, pages 533–564. Springer Science+Business Media, LLC, New York, 2012.

- [22] Zaiwen Wen, Donald Goldfarb, and Wotao Yin. Alternating Direction Augmented Lagrangian Methods for Semidefinite Programming. *Mathematical Programming Computation*, 2(3):203–230, 2010.
- [23] Hui Zhang. New Analysis of Linear Convergence of Gradient-Type Methods via Unifying Error Bound Conditions. Manuscript, available at <http://arxiv.org/abs/1606.00269>, 2016.
- [24] Zirui Zhou and Anthony Man-Cho So. A Unified Approach to Error Bounds for Structured Convex Optimization Problems. Manuscript, available at <http://arxiv.org/abs/1512.03518>, 2015.
- [25] Zirui Zhou, Qi Zhang, and Anthony Man-Cho So. $\ell_{1,p}$ -Norm Regularization: Error Bounds and Convergence Rate Analysis of First-Order Methods. In *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*, pages 1501–1510, 2015.