

# Exact Minimax Estimation for Phase Synchronization

Chao Gao<sup>ID</sup> and Anderson Y. Zhang

**Abstract**—We study the phase synchronization problem with measurements  $Y = z^* z^{*H} + \sigma W \in \mathbb{C}^{n \times n}$ , where  $z^*$  is an  $n$ -dimensional complex unit-modulus vector and  $W$  is a complex-valued Gaussian random matrix. It is assumed that each entry  $Y_{jk}$  is observed with probability  $p$ . We prove that the minimax lower bound of estimating  $z^*$  under the squared  $\ell_2$  loss is  $(1 - o(1)) \frac{\sigma^2}{2p}$ . We also show that both generalized power method and maximum likelihood estimator achieve the error bound  $(1 + o(1)) \frac{\sigma^2}{2p}$ . Thus,  $\frac{\sigma^2}{2p}$  is the exact asymptotic minimax error of the problem. Our upper bound analysis involves a precise characterization of the statistical property of the power iteration. The lower bound is derived through an application of van Trees' inequality.

**Index Terms**—Angular synchronization, minimax risk, power method, maximum likelihood estimator.

## I. INTRODUCTION

THE phase synchronization problem [1] is to estimate  $n$  unknown angles  $\theta_1^*, \dots, \theta_n^*$  from noisy measurements of  $(\theta_j^* - \theta_k^*) \bmod 2\pi$ . In this paper, we consider the following additive model [2]–[5]:

$$Y_{jk} = z_j^* z_k^* + \sigma W_{jk} \in \mathbb{C}, \quad (1)$$

for all  $1 \leq j < k \leq n$ , where we use the notation  $\bar{x}$  for the complex conjugate of  $x$ . We assume that each  $z_j^* \in \mathbb{C}_1 = \{x \in \mathbb{C} : |x| = 1\}$  and we can thus write it as  $z_j^* = e^{i\theta_j^*}$ . The additive noise  $W_{jk}$  in (1) is assumed to be i.i.d. standard complex Gaussian.<sup>1</sup> Without the constraint that  $z^* \in \mathbb{C}_1^n$ , (1) is recognized as a standard matrix spiked model. Our goal in this paper is to study minimax optimal estimation of the vector  $z^* \in \mathbb{C}_1^n$  under the loss function

$$\ell(\hat{z}, z^*) = \min_{a \in \mathbb{C}_1} \sum_{j=1}^n |\hat{z}_j a - z_j^*|^2. \quad (2)$$

Manuscript received October 9, 2020; revised August 23, 2021; accepted August 30, 2021. Date of publication September 14, 2021; date of current version November 22, 2021. The work of Chao Gao was supported in part by NSF CAREER Award DMS-1847590 and in part by NSF under Grant CCF-1934931. The work of Anderson Y. Zhang was supported in part by NSF under Grant DMS-2112988. (Corresponding author: Chao Gao.)

Chao Gao is with the Department of Statistics, The University of Chicago, Chicago, IL 60637 USA (e-mail: chaogao@galton.uchicago.edu).

Anderson Y. Zhang is with the Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: ayz@wharton.upenn.edu).

Communicated by M. Lops, Associate Editor for Detection and Estimation. Digital Object Identifier 10.1109/TIT.2021.3112712

<sup>1</sup>For  $W_{jk} \sim \mathcal{CN}(0, 1)$ , we have  $\text{Re}(W_{jk}) \sim \mathcal{N}(0, \frac{1}{2})$  and  $\text{Im}(W_{jk}) \sim \mathcal{N}(0, \frac{1}{2})$  independently.

We remark that the minimization over a global phase in the definition of (2) is necessary. This is because the global phase is not identifiable from the pairwise observations (1).

Various estimation procedures have been considered and studied in the literature. For example, the maximum likelihood estimator (MLE) is defined as a global maximizer of the following constrained optimization problem

$$\max_{z \in \mathbb{C}_1^n} z^H Y z, \quad (3)$$

where  $Y$  is Hermitian with  $Y_{jk} = \bar{Y}_{kj}$  for all  $1 \leq k < j \leq n$  and  $Y_{jj} = 0$  for all  $j \in [n]$ . Note that (3) can be shown to be equivalent to  $\min_{z \in \mathbb{C}_1^n} \sum_{1 \leq j < k \leq n} |Y_{jk} - z_j \bar{z}_k|^2$ . It was shown in [6] that the MLE satisfies  $\ell(\hat{z}, z^*) \leq C\sigma^2$  with high probability for some constant  $C > 0$ . However, the optimization (3) is nonconvex and computationally infeasible in general. To address this problem, generalized power method (GPM) [7]–[9] and semi-definite programming (SDP) have been considered in the literature to approximate the solution of (3). The generalized power method is defined through the iteration,<sup>2</sup>

$$z_j^{(t)} = \frac{\sum_{k \in [n] \setminus \{j\}} Y_{jk} z_k^{(t-1)}}{\left| \sum_{k \in [n] \setminus \{j\}} Y_{jk} z_k^{(t-1)} \right|}. \quad (4)$$

In other words, one repeatedly computes the product  $Y z^{(t-1)}$  and projects this vector to  $\mathbb{C}_1^n$  through entrywise normalization. When the iteration (4) is initialized by the eigenvector method,<sup>3</sup> [4] shows that  $z^{(t)}$  converges to the global maximizer of (3) at a linear rate under the noise level condition  $\sigma^2 = O\left(\frac{n}{\log n}\right)$ . For its statistical performance, [10] shows  $\ell(z^{(t)}, z^*) \leq C\sigma^2$  with high probability for some constant  $C > 0$  eventually. The semidefinite programming is a convex relaxation of (3). It refers to the following optimization problem,

$$\max_{Z = Z^H \in \mathbb{C}^{n \times n}} \text{Tr}(YZ) \quad \text{subject to } \text{diag}(Z) = I_n \text{ and } Z \succeq 0. \quad (5)$$

In general, the solution of (5) is an  $n \times n$  matrix that does not directly leads to an estimator of  $z^*$ . When  $\sigma^2 = O(n^{1/2})$ , it was proved by [6] that the solution to (5) is a rank-one matrix  $\hat{Z} = \hat{z} \hat{z}^H$ , with  $\hat{z}$  being a global maximizer of (3). This result was recently proved by [4] to hold under a weaker

<sup>2</sup>When the denominator of (4) is zero, take  $z_j^{(t)} = z_j^{(t-1)}$ .

<sup>3</sup>The eigenvector method estimates  $z_j^*$  by  $\hat{u}_j$ , where  $\hat{u} \in \mathbb{C}^n$  is the leading eigenvector of  $Y$ .

condition  $\sigma^2 = O\left(\frac{n}{\log n}\right)$ . Given the fact that SDP solves (3), we know that it also achieves the same high-probability error bound  $\ell(\hat{z}, z^*) \leq C\sigma^2$  as that of the MLE under the additional condition  $\sigma^2 = O\left(\frac{n}{\log n}\right)$ .

Despite these estimation procedures studied in the literature, it remains an open problem what the optimal error under the loss (2) is. In this paper, we establish a minimax lower bound for phase synchronization. We show that

$$\inf_{\hat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\hat{z}, z) \geq (1 - \delta) \frac{\sigma^2}{2}, \quad (6)$$

for some  $\delta = o(1)$  under the condition that  $\sigma^2 = o(n)$ . Here, by  $\inf_{\hat{z} \in \mathbb{C}_1^n}$ , we mean the infimum over all measurable functions (of the data) whose range is  $\mathbb{C}_1^n$ . The notation  $\mathbb{E}_z$  is the expectation operator induced by the data generating process (1) indexed by  $z$ . This provides a stronger characterization of the fundamental limits of the phase synchronization problem than the Cramér-Rao lower bound developed in [11], [12], which is only established for unbiased estimators. Instead, the lower bound in (6) holds for both unbiased and biased estimators.

Moreover, in this paper, we prove both the MLE and the GPM achieve the error bound

$$\ell(\hat{z}, z^*) \leq (1 + \delta) \frac{\sigma^2}{2}, \quad (7)$$

for some  $\delta = o(1)$  with high probability under the same condition  $\sigma^2 = o(n)$ . In other words, these two estimators are not only rate-optimal, but are also exactly asymptotically minimax by achieving the correct leading constant in front of the optimal rate. In addition, since we know by the result of [4] that the solution of the SDP is a rank-one matrix  $\hat{z}\hat{z}^H$  with  $\hat{z}$  being the MLE, the SDP also achieves the optimal error bound (7) as a direct consequence, but under a stronger condition  $\sigma^2 = O\left(\frac{n}{\log n}\right)$ .

To formally state our main result, we introduce a more general statistical estimation setting that allows the possibility of missing entries [13], [14]. Instead of observing  $Y_{jk}$  for all  $1 \leq j < k \leq n$ , we assume each  $Y_{jk}$  is observed with probability  $p$ . In other words, consider a random graph  $A_{jk} \sim \text{Bernoulli}(p)$  independently for all  $1 \leq j < k \leq n$ , and we only observe  $Y_{jk}$  that follows (1) when  $A_{jk} = 1$ . Define  $A_{jk} = A_{kj}$  for  $1 \leq k < j \leq n$  and  $A_{jj} = 0$  for  $j \in [n]$ . The full observations can be organized into two Hermitian matrices  $A$  and  $A \circ Y$ , where  $\circ$  denotes the matrix Hadamard product. The MLE and the GPM can be extended by replacing  $Y_{jk}$  in (3) and (4) with  $A_{jk}Y_{jk}$ .

**Theorem 1:** Assume  $\frac{np}{\sigma^2} \rightarrow \infty$  and  $\frac{np}{\log n} \rightarrow \infty$ . Then, there exists some  $\delta = o(1)$  such that

$$\inf_{\hat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\hat{z}, z) \geq (1 - \delta) \frac{\sigma^2}{2p}. \quad (8)$$

Moreover, both MLE and GPM achieve the error bound

$$\ell(\hat{z}, z^*) \leq (1 + \delta) \frac{\sigma^2}{2p}, \quad (9)$$

with probability at least  $1 - n^{-8} - \exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right)$ .

Theorem 1 immediately implies (6) and (7) as a special case of  $p = 1$ . We remark that both conditions  $\frac{np}{\sigma^2} \rightarrow \infty$  and  $\frac{np}{\log n} \rightarrow \infty$  are essential for the results of the above theorem to hold. Since the minimax risk of the problem is  $\frac{\sigma^2}{2p}$ , the condition  $\frac{np}{\sigma^2} \rightarrow \infty$ , which is equivalent to  $\frac{\sigma^2}{2p} = o(n)$ , guarantees that the minimax risk is of smaller order than the trivial one. The order  $n$  is trivial, since  $\ell(z, z^*) \leq 4n$  for any  $z, z^* \in \mathbb{C}_1^n$ . When  $p = 1$ , the necessity of  $\sigma^2 = o(n)$  for a consistent estimation is understood in the literature [4], [15]–[17]. In fact, the MLE and the GPM are no longer optimal when  $\sigma^2 \asymp n$ , and optimal phase synchronization under this regime can be achieved by approximate message passing [9], [16], [18]. The condition  $\frac{np}{\log n} \rightarrow \infty$  guarantees that the random graph  $A$  is connected with high probability. It is known that when  $p \leq c \frac{\log n}{n}$  for some sufficiently small constant  $c > 0$ , the random graph has several disjoint components, which makes the recovery of  $z^*$  up to a global phase impossible.

Since  $\ell(\hat{z}, z) \leq 4n$  and  $\exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right) = o\left(\frac{\sigma^2}{np}\right)$ , the MLE and the GPM also satisfy the in-expectation bound

$$\sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\hat{z}, z) \leq (1 + \delta) \frac{\sigma^2}{2p} + 4n^{-7},$$

for some  $\delta = o(1)$ . We remark that the high-probability exponent  $-8$  of Theorem 1 can actually be replaced by  $-C$  for an arbitrarily large constant  $C > 0$ . This implies the  $4n^{-7}$  term in the above display can also be improved to  $4n^{-(C-1)}$  for any constant  $C > 0$ . If we further assume that  $\sigma^2/p \geq n^{-c}$  for some constant  $c > 0$ , we will have

$$(1 - \delta) \frac{\sigma^2}{2p} \leq \inf_{\hat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\hat{z}, z) \leq (1 + \delta) \frac{\sigma^2}{2p},$$

for some  $\delta = o(1)$ . Thus,  $\frac{\sigma^2}{2p}$  is the exact asymptotic minimax risk for phase synchronization.

Our analysis of the upper bounds relies on a precise statistical characterization of the power iteration map  $f : \mathbb{C}_1^n \rightarrow \mathbb{C}_1^n$ . Let  $f$  be the map that characterizes the iteration of the GPM. That is,  $z^{(t)} = f(z^{(t-1)})$ . We show that as long as  $z \in \mathbb{C}_1^n$  satisfies  $\ell(z, z^*) = o(n)$ , the vector  $f(z)$  must satisfy

$$\ell(f(z), z^*) \leq \delta \ell(z, z^*) + (1 + \delta) \frac{\sigma^2}{2p}, \quad (10)$$

for some  $\delta = o(1)$  with high probability. To be more precise, we prove that the inequality (10) holds uniformly overall  $z \in \mathbb{C}_1^n$  such that  $\ell(z, z^*) \leq \gamma n$  for some  $\gamma = o(1)$  with high probability. The bound (10) immediately leads to the optimality of the GPM. This direct analysis of the power iteration is very different from what has been done in the literature. In the literature, the statistical error bound of the GPM is derived through its convergence to the MLE [4], [7], but that requires a stronger condition  $\sigma^2 = O\left(\frac{n}{\log n}\right)$  at least when  $p = 1$ . In contrast, our analysis of GPM is not based on its relation to the MLE. On the opposite, we analyze the MLE based on its relation to the GPM. The optimality of the MLE can also be derived from (10). This is by showing that MLE is a fixed point of the map  $f$ . That is,  $\hat{z} = f(\hat{z})$ , and

therefore with  $z = \hat{z}$ , we can rearrange (10) into the bound  $\ell(\hat{z}, z) \leq \frac{1+\delta}{1-\delta} \frac{\sigma^2}{2p}$ .

To derive the lower bound result, we show that it is sufficient to analyze the Bayes risk of a subproblem of estimating the relative angle  $z_j \bar{z}_k$  for each  $j \neq k$ ,

$$\inf_{\hat{T}} \int \int \pi(z_j) \pi(z_k) \mathbb{E}_z |\hat{T} - z_j \bar{z}_k|^2 dz_j dz_k, \quad (11)$$

where  $\inf_{\hat{T}}$  means the infimum over all measurable functions of the data, and  $\pi(\cdot)$  is a density function on  $\mathbb{C}_1$ . In other words, the difficulty of estimating a vector in  $\mathbb{C}_1^n$  is determined by the average difficulty of estimating  $z_j \bar{z}_k$  given the knowledge of  $(z_l)_{l \in [n] \setminus \{j,k\}}$  for all  $j \neq k$ . To lower bound (11), we apply a multivariate van Trees' inequality [19] that relates the Bayes risk (11) to the Fisher information of the phase synchronization model.

*a) Paper Organization:* The rest of the paper is organized as follows. In Section II, we establish a key lemma that implies the critical inequality (10). The implications of the key lemma on the statistical error bounds of GPM and MLE are discussed in Section III, which establishes (9) for Theorem 1. The minimax lower bound of phase synchronization is proved in Section IV, which establishes (8) for Theorem 1. Finally, Section V collects the remaining technical proofs of the paper.

*b) Notation:* For  $d \in \mathbb{N}$ , we write  $[d] = \{1, \dots, d\}$ . Given  $a, b \in \mathbb{R}$ , we write  $a \vee b = \max(a, b)$  and  $a \wedge b = \min(a, b)$ . For a set  $S$ , we use  $\mathbb{I}\{S\}$  and  $|S|$  to denote its indicator function and cardinality respectively. For a complex number  $x \in \mathbb{C}$ , we use  $\bar{x}$  for its complex conjugate and  $|x|$  for its modulus. For a matrix  $B = (B_{jk}) \in \mathbb{C}^{d_1 \times d_2}$ , we use  $B^H \in \mathbb{C}^{d_2 \times d_1}$  for its conjugate transpose such that  $B^H = (\bar{B}_{kj})$ . The Frobenius norm and operator norm of  $B$  are defined by  $\|B\|_F = \sqrt{\sum_{j=1}^{d_1} \sum_{k=1}^{d_2} |B_{jk}|^2}$  and  $\|B\|_{\text{op}} = \sup_{u \in \mathbb{C}^{d_1}, v \in \mathbb{C}^{d_2}: \|u\| = \|v\| = 1} u^H B v$ . For  $U, V \in \mathbb{C}^{d_1 \times d_2}$ ,  $U \circ V \in \mathbb{R}^{d_1 \times d_2}$  is the Hadamard product  $U \circ V = (U_{jk} V_{jk})$ . The notation  $\mathbb{P}$  and  $\mathbb{E}$  are generic probability and expectation operators whose distribution is determined from the context. For two positive sequences  $\{a_n\}$  and  $\{b_n\}$ ,  $a_n \lesssim b_n$  or  $a_n = O(b_n)$  means  $a_n \leq C b_n$  for some constant  $C > 0$  independent of  $n$ . We also write  $a_n = o(b_n)$  or  $\frac{a_n}{b_n} \rightarrow 0$  when  $\limsup_n \frac{a_n}{b_n} = 0$ .

## II. A KEY LEMMA

Recall that we observe a random graph  $A_{jk} \sim \text{Bernoulli}(p)$  independently for all  $1 \leq j < k \leq n$ . Whenever  $A_{jk} = 1$ , we also observe  $Y_{jk} = z_j^* \bar{z}_k^* + \sigma W_{jk}$  with  $W_{jk} \sim \mathcal{CN}(0, 1)$ . In summary, the observations contain an adjacency matrix  $A$  and a masked version of pairwise interactions  $A \circ Y$ , which are both Hermitian as we define  $Y_{jk} = \bar{Y}_{kj}$  and  $A_{jk} = A_{kj}$  for all  $1 \leq k < j \leq n$  and  $Y_{jj} = A_{jj} = 0$  for all  $j \in [n]$ .

In this section, we establish a lemma that shows the contraction of the loss function through the generalized power method. Note that  $\mathbb{E}(A \circ Y) = p z^* z^{*H} - p I_n$  with its leading eigenvector proportional to  $z^*$ . We can thus estimate  $z^*/\sqrt{n}$  by computing the leading eigenvector of  $A \circ Y$ . Algorithmically, this leads to the power iteration that repeatedly computes  $(A \circ Y)z$  and then applies  $\ell_2$  normalization. In the setting of

phase synchronization, since we know  $z^* \in \mathbb{C}_1^n$ , the generalized power method then replaces the  $\ell_2$  normalization by a normalization applied to each coordinate. This GPM algorithm can be explicitly written as

$$z_j^{(t)} = \begin{cases} \frac{\sum_{k \in [n] \setminus \{j\}} A_{jk} Y_{jk} z_k^{(t-1)}}{\left| \sum_{k \in [n] \setminus \{j\}} A_{jk} Y_{jk} z_k^{(t-1)} \right|}, & \text{if } \sum_{k \in [n] \setminus \{j\}} A_{jk} Y_{jk} z_k^{(t-1)} \neq 0, \\ z_j^{(t-1)}, & \text{if } \sum_{k \in [n] \setminus \{j\}} A_{jk} Y_{jk} z_k^{(t-1)} = 0. \end{cases} \quad (12)$$

Let us shorthand the above formula as

$$z^{(t)} = f(z^{(t-1)}), \quad (13)$$

by introducing a map  $f: \mathbb{C}_1^n \rightarrow \mathbb{C}_1^n$  such that the  $j$ th entry of  $f(z^{(t-1)})$  is given by (12). The following lemma characterizes the evolution of the loss function (2) through the map  $f$ .

*Lemma 1:* There exist constants  $c_1, c_2, C_1, C_2 > 0$ , such that if  $\frac{np}{\sigma^2} \geq c_1$  and  $\frac{np}{\log n} \geq c_2$ , then, for any  $\gamma \in [0, 1/16]$ , we will have

$$\mathbb{P}\left(\ell(f(z), z^*) \leq \delta_1 \ell(z, z^*) + (1 + \delta_2) \frac{\sigma^2}{2p} \text{ for all } z \in \mathbb{C}_1^n \text{ such that } \ell(z, z^*) \leq \gamma n\right) \geq 1 - n^{-9} - \exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right),$$

where  $\delta_1 = C_1 \sqrt{\frac{\log n + \sigma^2}{np}}$  and  $\delta_2 = C_2 \left(\gamma^2 + \frac{\log n + \sigma^2}{np}\right)^{1/4}$ .

*Remark 1:* The result of Lemma 1 is established for the map  $f$  defined through (12), where we set  $z_j^{(t)} = z_j^{(t-1)}$  if the denominator is 0. We remark that  $z_j^{(t)} = z_j^{(t-1)}$  is not important and it can be replaced by any number in  $\mathbb{C}_1$  without changing the result of Lemma 1.

The lemma shows that for any  $z \in \mathbb{C}_1^n$  that has a nontrivial error, the vector  $f(z)$  will have an error that is smaller by a multiplicative factor  $\delta_1$  up to an additive term  $(1 + \delta_2) \frac{\sigma^2}{2p}$ . Plugging  $z = z^*$  into the inequality reveals that

$$\ell(f(z^*), z^*) \leq (1 + \delta_2) \frac{\sigma^2}{2p}.$$

Thus, the additive term  $(1 + \delta_2) \frac{\sigma^2}{2p}$  can be understood as the oracle error given the knowledge of  $z^*$ . Indeed, we show in Theorem 5 that  $\frac{\sigma^2}{2p}$  is the minimax lower bound for phase synchronization under the loss function (2).

The two conditions  $\frac{np}{\sigma^2} \geq c_1$  and  $\frac{np}{\log n} \geq c_2$  are essentially necessary for the result to hold. While  $\frac{np}{\sigma^2} \geq c_1$  makes sure that the error  $\frac{\sigma^2}{2p}$  is sufficiently small compared with the trivial  $n$ , the condition  $\frac{np}{\log n} \geq c_2$  guarantees that the random graph is connected. We can also slightly strengthen the conditions to  $\frac{np}{\sigma^2} \rightarrow \infty$  and  $\frac{np}{\log n} \rightarrow \infty$  so that both  $\delta_1$  and  $\delta_2$  are vanishing.

Lemma 1 implies that as long as  $\ell(z^{(t-1)}, z^*) \leq \gamma n$  for some  $\gamma < 1/16$ , the next step of power iteration (12) satisfies

$$\ell(z^{(t)}, z^*) \leq \delta_1 \ell(z^{(t-1)}, z^*) + (1 + \delta_2) \frac{\sigma^2}{2p}. \quad (14)$$

The condition  $\ell(z^{(t-1)}, z^*) \leq \gamma n$  then implies  $\ell(z^{(t)}, z^*) \leq \delta_1 \gamma n + (1 + \delta_2) \frac{\sigma^2}{2p}$ . Given that  $\frac{np}{\sigma^2}$  is sufficiently large (for a sufficiently large  $c_1$ ), we can always choose  $\gamma$  not too small so that  $\delta_1 \gamma n + (1 + \delta_2) \frac{\sigma^2}{2p} \leq \gamma n$ . Therefore,  $\ell(z^{(t)}, z^*) \leq \gamma n$ .

Thus, a simple induction argument implies that (14) holds for all  $t \geq 1$  as long as  $\ell(z^{(0)}, z^*) \leq \gamma n$ . The one-step iteration bound (14) immediately implies the linear convergence

$$\ell(z^{(t)}, z^*) \leq \delta_1^t \ell(z^{(0)}, z^*) + \frac{1 + \delta_2 \sigma^2}{1 - \delta_1 2p}, \quad (15)$$

for all  $t \geq 1$ .

### III. OPTIMALITY OF GENERALIZED POWER METHOD AND MLE

In this section, we show that both the GPM and the MLE achieve the optimal error  $\frac{\sigma^2}{2p}$  by using the conclusion of Lemma 1. Theorem 2 and Theorem 3 together establish the upper bounds (9) for Theorem 1.

#### A. Generalized Power Method

The result of Lemma 1 implies that (15) holds for all  $t \geq 1$ , as long as  $\ell(z^{(0)}, z^*) \leq \gamma n$  for some  $\gamma < 1/16$ . Since  $z^*/\sqrt{n}$  is the leading eigenvector of  $\mathbb{E}(A \circ Y)$ , we can compute the leading eigenvector of  $A \circ Y$  as the initialization of the power method. Let  $\hat{u} \in \mathbb{C}^n$  be the leading eigenvector of  $A \circ Y$ , and we define  $z^{(0)}$  according to

$$z_j^{(0)} = \begin{cases} \frac{\hat{u}_j}{|\hat{u}_j|}, & \hat{u}_j \neq 0, \\ 1, & \hat{u}_j = 0. \end{cases} \quad (16)$$

It is easy to check that  $z^{(0)} \in \mathbb{C}_1^n$ , and thus can be applied by the power iteration (12). This leads to the GPM for phase synchronization, which is formally presented as Algorithm 1.

---

#### Algorithm 1 Generalized Power Method for Phase Synchronization

---

**Input:** The data  $A \circ Y$  and the number of iterations  $t_{\max}$ .

**Output:** The estimator  $\hat{z} = z^{t_{\max}}$ .

```

1 Obtain  $z^{(0)}$  from the leading eigenvector of  $A \circ Y$ 
  by (16);
2 for  $t = 1, \dots, t_{\max}$  do
     $z^{(t)} = f(z^{(t-1)})$ ,
    where  $f(\cdot)$  is defined in (12).
end
```

---

The error bound of  $z^{(0)}$  is given by the following lemma.

**Lemma 2:** There exist constants  $c, C > 0$  such that if  $\frac{np}{\log n} > c$ , then we will have

$$\ell(z^{(0)}, z^*) \leq C \frac{\sigma^2 + 1}{p},$$

with probability at least  $1 - n^{-9}$ .

Under the condition that  $\frac{np}{\sigma^2}$  and  $\frac{np}{\log n}$  are sufficiently large, the error rate of Lemma 2 satisfies  $\ell(z^{(0)}, z^*) \leq \gamma n$  holds for some  $\gamma < 1/16$ . By Lemma 1 and its implication (15), we directly obtain the following result.

**Theorem 2:** There exist constants  $c_1, c_2, C > 0$  such that if  $\frac{np}{\sigma^2} \geq c_1$  and  $\frac{np}{\log n} \geq c_2$ , then the GPM (i.e., Algorithm 1) will satisfy

$$\ell(z^{(t)}, z^*) \leq \left(1 + C \left(\frac{\log n + \sigma^2}{np}\right)^{1/4}\right) \frac{\sigma^2}{2p},$$

for all  $t \geq \log\left(\frac{1}{\sigma^2}\right)$  with probability at least  $1 - 2n^{-9} - \exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right)$ .

We remark that the number of iterations  $\log\left(\frac{1}{\sigma^2}\right)$  required by the theorem can be improved in some special cases. For example, when  $p = 1$ , the error bound of Lemma 2 can be improved to  $\ell(z^{(0)}, z^*) \leq C\sigma^2$  by a matrix perturbation analysis [7]. Then, (14) implies that  $\ell(z^{(1)}, z^*) \leq \left(1 + C \left(\frac{\log n + \sigma^2}{np}\right)^{1/4}\right) \frac{\sigma^2}{2}$ . In other words, when the graph is fully connected, a one-step refinement of power iteration is sufficient to achieve the optimal error.

#### B. Maximum Likelihood Estimator

Next, we discuss how the result of Lemma 1 also implies the optimality of the MLE. According to the data generating process, the MLE is given by

$$\hat{z} = \operatorname{argmin}_{z \in \mathbb{C}_1^n} \sum_{1 \leq j < k \leq n} A_{jk} |Y_{jk} - z_j \bar{z}_k|^2, \quad (17)$$

which is equivalent to  $\operatorname{argmax}_{z \in \mathbb{C}_1^n} z^H (A \circ Y) z$ . By the definition of  $\hat{z}$ , its  $j$ th entry must satisfy

$$\begin{aligned} \hat{z}_j &= \operatorname{argmin}_{z_j \in \mathbb{C}_1} \sum_{k \in [n] \setminus \{j\}} A_{jk} |Y_{jk} - z_j \bar{\hat{z}}_k|^2 \\ &= \frac{\sum_{k \in [n] \setminus \{j\}} A_{jk} Y_{jk} \hat{z}_k}{\left| \sum_{k \in [n] \setminus \{j\}} A_{jk} Y_{jk} \hat{z}_k \right|}. \end{aligned}$$

as long as  $\sum_{k \in [n] \setminus \{j\}} A_{jk} Y_{jk} \hat{z}_k \neq 0$ . In other words, we have

$$\hat{z} = f(\hat{z}),$$

and the MLE is a fixed point of the power iteration. As long as we can establish a crude bound  $\ell(\hat{z}, z^*) \leq \gamma n$  for some  $\gamma < 1/16$ , Lemma 1 automatically leads to the optimal error of the MLE.

**Lemma 3:** There exist constants  $c, C > 0$  such that if  $\frac{np}{\log n} > c$ , then the MLE will have error bound

$$\ell(\hat{z}, z^*) \leq C \frac{\sigma^2 + 1}{p},$$

with probability at least  $1 - n^{-9}$ .

Again, under the condition that  $\frac{np}{\sigma^2}$  and  $\frac{np}{\log n}$  are sufficiently large, we have  $\ell(\hat{z}, z^*) \leq \gamma n$  for some  $\gamma < 1/16$ . Lemma 1 implies that

$$\ell(\hat{z}, z^*) \leq \delta_1 \ell(\hat{z}, z^*) + (1 + \delta_2) \frac{\sigma^2}{2p}.$$

After rearrangement, we obtain the bound  $\ell(\hat{z}, z^*) \leq \frac{1 + \delta_2}{1 - \delta_1} \frac{\sigma^2}{p}$ . The result is summarized into the following theorem.

**Theorem 3:** There exist constants  $c_1, c_2, C > 0$  such that if  $\frac{np}{\sigma^2} \geq c_1$  and  $\frac{np}{\log n} \geq c_2$ , then the MLE (17) will satisfy

$$\ell(\hat{z}, z^*) \leq \left(1 + C \left(\frac{\log n + \sigma^2}{np}\right)^{1/4}\right) \frac{\sigma^2}{2p},$$

with probability at least  $1 - 2n^{-9} - \exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right)$ .

To close this section, we briefly discuss the implication of Lemma 1 on semi-definite programming (SDP) when the



graph is fully connected. In other words, we only consider  $p = 1$ . This is by leveraging a recent result on the connection between SDP and MLE by [4]. Recall the definition of SDP in (5). The following result is a special case of Theorem 5 of [4].

**Theorem 4 ([4]):** Assume  $\sigma^2 \leq c \frac{n}{\log n}$  for some sufficiently large constant  $c > 0$ . Then, with probability at least  $1 - n^{-1}$ , the SDP (5) admits a unique solution  $\hat{z}\hat{z}^H$ , where  $\hat{z}$  is a global optimum of (17) with all  $A_{jk} = 1$ .

This equivalence of SDP and MLE immediately implies the following result.

**Corollary 1:** There exist constants  $c, C > 0$ , such that if  $\sigma^2 \leq c \frac{n}{\log n}$  and  $p = 1$ , then the unique solution of the SDP can be written as  $\hat{z}\hat{z}^H$ , where  $\hat{z}$  satisfies

$$\ell(\hat{z}, z^*) \leq \left(1 + C \left(\frac{\log n + \sigma^2}{np}\right)^{1/4}\right) \frac{\sigma^2}{2},$$

with probability at least  $1 - 2n^{-1} - \exp\left(-\left(\frac{n}{\sigma^2}\right)^{1/4}\right)$ .

Corollary 1 requires a stronger condition  $\sigma^2 \leq c \frac{n}{\log n}$  than what is needed by Theorem 2 and Theorem 3 when  $p = 1$ . This condition is needed for the equivalence between SDP and MLE, which is established via an  $\ell_\infty$  norm argument in [4] and is thus very unlikely to be weakened. Whether the equivalence between SDP and MLE continues to hold when  $p < 1$  is a less clear issue in the literature. In general, the solution to the SDP (5) is not necessarily a rank-one matrix. Obtaining an estimator in  $\mathbb{C}_1^n$  requires a post-processing step such as the rounding method suggested by [16], the consequence of which is unclear to us.

#### IV. MINIMAX LOWER BOUND

We study the minimax lower bound of phase synchronization in this section. The minimax risk is given by

$$\begin{aligned} & \inf_{\hat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\hat{z}, z) \\ &= \inf_{\hat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \left[ \min_{a \in \mathbb{C}_1} \sum_{j=1}^n |\hat{z}_j a - z_j|^2 \right], \end{aligned} \quad (18)$$

where  $\mathbb{E}_z$  is the expectation of  $(A \circ Y, A)$  under the model  $A_{jk} Y_{jk} = A_{jk} z_j \bar{z}_k + \sigma A_{jk} W_{jk}$  and  $A_{jk} \sim \text{Bernoulli}(p)$ . The main difficulty of analyzing (18) is that the loss function  $\ell(\hat{z}, z)$  is not separable in  $j \in [n]$  because of the identity  $\ell(\hat{z}, z) = 2(n - |z^H z^*|)$ . This difficulty can be tackled with the following inequality

$$\ell(z, z^*) \geq \frac{1}{2n} \|\hat{z}\hat{z}^H - z z^H\|_F^2, \quad (19)$$

which is proved in Lemma 9. Since the right hand side of (19) can be written as  $\sum_{1 \leq j \neq k \leq n} |\hat{z}_j \bar{z}_k - z_j \bar{z}_k|^2$ , we can lower bound (18) by

$$\begin{aligned} & \inf_{\hat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\hat{z}, z) \\ & \geq \frac{1}{2n} \inf_{\hat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \|\hat{z}\hat{z}^H - z z^H\|_F^2 \\ & \geq \frac{1}{2n} \inf_{\hat{z}} \sum_{1 \leq j \neq k \leq n} \left( \int \prod_{l=1}^n \pi(z_l) \right) \mathbb{E}_z |\hat{z}_j \bar{z}_k - z_j \bar{z}_k|^2 dz \end{aligned}$$

$$\begin{aligned} & \geq \frac{1}{2n} \sum_{1 \leq j \neq k \leq n} \mathbb{E}_{z_{-(j,k)} \sim \pi} \inf_{\hat{T}} \\ & \int \int \pi(z_j) \pi(z_k) \mathbb{E}_z |\hat{T} - z_j \bar{z}_k|^2 dz_j dz_k, \end{aligned} \quad (20)$$

where  $\pi$  is some density function supported on  $\mathbb{C}_1$  to be specified later, and the notation  $\mathbb{E}_{z_{-(j,k)}}$  means expectation over  $z$  except for its  $j$ th and  $k$ th entries with respect to the distribution  $\pi$ . In the above display, the second inequality is due to the fact that we can lower bound the supremum over  $z$  by an average over  $z$ , i.e.,  $\sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \|\hat{z}\hat{z}^H - z z^H\|_F^2 \geq \int \left( \prod_{l=1}^n \pi(z_l) \right) \mathbb{E}_z \|\hat{z}\hat{z}^H - z z^H\|_F^2 dz$ . From (20), it is sufficient to lower bound

$$\inf_{\hat{T}} \int \int \pi(z_j) \pi(z_k) \mathbb{E}_z |\hat{T} - z_j \bar{z}_k|^2 dz_j dz_k$$

for all  $j \neq k$ . This corresponds to the problem of estimating  $z_j \bar{z}_k$  with other entries of  $z$  assumed to be known.

By symmetry, we consider the problem with  $j = 1$  and  $k = 2$ . Given the knowledge of  $z_3, \dots, z_n$ , we can decompose the likelihood function as

$$\begin{aligned} & p(A \circ Y, A) \\ &= p(A) p(A_{12} Y_{12} | A) \left( \prod_{j=3}^n p(A_{1j} Y_{1j} | A) p(A_{2j} Y_{2j} | A) \right) \\ & \times \prod_{3 \leq j < k \leq n} p(A_{jk} Y_{jk} | A). \end{aligned}$$

Since  $p(A) \prod_{3 \leq j < k \leq n} p(A_{jk} Y_{jk} | A)$  is independent of  $z_1 \bar{z}_2$ , the sufficient statistics for estimating  $z_1 \bar{z}_2$  is  $A_{12} Y_{12}, A_{13} Y_{13}, \dots, A_{1n} Y_{1n}$  and  $A_{23} Y_{23}, \dots, A_{2n} Y_{2n}$ . Now we restrict  $z_j \sim \pi$  by requiring  $z_j = a + \sqrt{1 - a^2} i$  with  $a \sim f$  for some density  $f$  supported on the unit interval  $[0, 1]$ . Then, we can represent  $z_1$  and  $z_2$  by

$$z_1 = a + \sqrt{1 - a^2} i \quad \text{and} \quad z_2 = c + \sqrt{1 - c^2} i.$$

Define

$$\begin{aligned} T(a, c) &= ac + \sqrt{1 - a^2} \sqrt{1 - c^2}, \\ S(a, c) &= \sqrt{1 - a^2} c - a \sqrt{1 - c^2}. \end{aligned}$$

With this notation, we have

$$z_1 \bar{z}_2 = T(a, c) + S(a, c) i.$$

The likelihood  $p(A_{12} Y_{12} | A)$  corresponds to the distribution

$$\mathcal{N} \left( A_{12} \begin{pmatrix} T(a, c) \\ S(a, c) \end{pmatrix}, \frac{1}{2} A_{12} \sigma^2 I_2 \right). \quad (21)$$

For  $j = 3, \dots, n$ , we have  $Y_{1j} = z_1 \bar{z}_j + \sigma W_{1j}$ . Since  $z_j$  is known for  $j = 3, \dots, n$ , observing  $Y_{1j}$  is equivalent to observing  $Y_{1j} z_j = z_1 + \sigma W_{1j} z_j$ , and we still have  $W_{1j} z_j \sim \mathcal{CN}(0, 1)$  by the property of complex Gaussian distribution. This argument implies that the likelihood  $\prod_{j=3}^n p(A_{1j} Y_{1j} | A)$  is proportional to the density function of

$$\mathcal{N} \left( \sum_{j=3}^n A_{1j} \begin{pmatrix} a \\ \sqrt{1 - a^2} \end{pmatrix}, \frac{\sigma^2 \sum_{j=3}^n A_{1j}}{2} I_2 \right). \quad (22)$$

Similarly,  $\prod_{j=3}^n p(A_{2j}Y_{2j}|A)$  is proportional to the density function of

$$\mathcal{N}\left(\sum_{j=3}^n A_{2j} \left(\sqrt{\frac{c}{1-c^2}}\right), \frac{\sigma^2 \sum_{j=3}^n A_{2j}}{2} I_2\right). \quad (23)$$

Therefore, we have the identity

$$\begin{aligned} & \inf_{\hat{T}} \int \int \pi(z_1) \pi(z_2) \mathbb{E}_z |\hat{T} - z_1 \bar{z}_2|^2 dz_1 dz_2 \\ &= \inf_{\hat{T}, \hat{S}} \int \int f(a) f(c) \mathbb{E}_{(a,c)} \\ & \quad \left[ (\hat{T} - T(a, c))^2 + (\hat{S} - S(a, c))^2 \right] dadc, \end{aligned} \quad (24)$$

where the expectation  $\mathbb{E}_{(a,c)}$  is under the product of the conditional distributions (21), (22) and (23) with marginal given by  $A_{jk} \sim \text{Bernoulli}(p)$ . The quantity (24) can be lower bounded by van Trees' inequality [19], and the result is given by the following lemma.

**Lemma 4:** There exist constants  $c, C > 0$  such that as long as  $\frac{np}{\sigma^2} \geq c$ , there exists some nice density function  $f$  supported on the unit interval such that

$$\begin{aligned} & \inf_{\hat{T}, \hat{S}} \int \int f(a) f(c) \mathbb{E}_{(a,c)} \left[ (\hat{T} - T(a, c))^2 + (\hat{S} - S(a, c))^2 \right] \\ & \quad dadc \geq \left(1 - C \left(\frac{\sigma^2}{np} + \frac{1}{n}\right)\right) \frac{\sigma^2}{np}. \end{aligned}$$

Plugging the result of Lemma 4 into (20), we have immediately the following theorem, which establishes the lower bound (8) for Theorem 1.

**Theorem 5:** There exist constants  $c, C > 0$  such that as long as  $\frac{np}{\sigma^2} \geq c$ , then we will have

$$\inf_{\hat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\hat{z}, z) \geq \left(1 - C \left(\frac{\sigma^2}{np} + \frac{1}{n}\right)\right) \frac{\sigma^2}{2p}.$$

## V. PROOFS

This section presents the proofs of all technical results in the paper. We first list some auxiliary lemmas in Section V-A. The key lemma that leads to various upper bound results (Lemma 1) is proved in Section V-B. Then, we prove Lemma 4 in Section V-C. Finally, the proofs of Lemma 2 and Lemma 3 are given in Section V-D.

### A. Some Auxiliary Lemmas

**Lemma 5:** There exist constants  $c, C > 0$  such that if  $\frac{np}{\log n} > c$ , then we will have

$$\max_{j \in [n]} \left( \sum_{k \in [n] \setminus \{j\}} (A_{jk} - p) \right)^2 \leq Cnp \log n,$$

and

$$\|A - \mathbb{E}A\|_{\text{op}} \leq C\sqrt{np},$$

with probability at least  $1 - n^{-10}$ .

*Proof:* The first result is a direct application of union bound and Bernstein's inequality. The second result is Theorem 5.2 of [20].  $\square$

**Lemma 6:** There exist constants  $c, C > 0$  such that if  $\frac{np}{\log n} > c$ , then we will have

$$\|A \circ W\|_{\text{op}} \leq C\sqrt{np},$$

with probability at least  $1 - n^{-10}$ .

*Proof:* We use  $\mathbb{P}_A$  for the conditional probability  $\mathbb{P}(\cdot|A)$ . Define the event

$$\mathcal{A} = \left\{ \max_{j \in [n]} \sum_{k \in [n] \setminus \{j\}} A_{jk} \leq 2np \right\}.$$

Under the assumption  $\frac{np}{\log n} > c$ , we have  $\mathbb{P}(\mathcal{A}^c) \leq n^{-11}$  by Bernstein's inequality and a union bound argument. By Corollary 3.11 of [21], we have

$$\sup_{A \in \mathcal{A}} \mathbb{P}_A (\|A \circ \text{Re}(W)\|_{\text{op}} > C_1 \sqrt{np} + t) \leq e^{-t^2/2},$$

for some constant  $C_1 > 0$ . This implies that  $\sup_{A \in \mathcal{A}} \mathbb{P}_A (\|A \circ \text{Re}(W)\|_{\text{op}} > C_2 \sqrt{np}) \leq n^{-11}$  for some constant  $C_2 > 0$ . Thus, we have

$$\begin{aligned} & \mathbb{P} (\|A \circ \text{Re}(W)\|_{\text{op}} > C_2 \sqrt{np}) \leq \mathbb{P}(\mathcal{A}^c) \\ & \quad + \sup_{A \in \mathcal{A}} \mathbb{P}_A (\|A \circ \text{Re}(W)\|_{\text{op}} > C_2 \sqrt{np}) \leq 2n^{-11}. \end{aligned}$$

The same high probability bound also holds for  $\|A \circ \text{Im}(W)\|_{\text{op}}$ . Finally, the desired conclusion is implied by  $\|A \circ W\|_{\text{op}} \lesssim \|A \circ \text{Im}(W)\|_{\text{op}} + \|A \circ \text{Re}(W)\|_{\text{op}}$ .  $\square$

**Lemma 7:** Consider independent random variables  $X_{jk} \sim \mathcal{N}(0, 1)$  for  $1 \leq j < k \leq n$ . Assume  $X_{kj} = X_{jk}$  for  $1 \leq j < k \leq n$ . There exist constants  $c, C > 0$  such that if  $\frac{np}{\log n} > c$ , then we will have

$$\sum_{j=1}^n \left( \sum_{k \in [n] \setminus \{j\}} A_{jk} X_{jk} \right)^2 \leq n(n-1)p + C\sqrt{n^3 p^2 \log n},$$

with probability at least  $1 - n^{-10}$ . The same result holds if  $X_{kj} = -X_{jk}$  is assumed instead for  $1 \leq j < k \leq n$ .

*Proof:* We use  $\mathbb{P}_A$  for the conditional probability  $\mathbb{P}(\cdot|A)$ . We only prove the case where  $X_{kj} = X_{jk}$  for  $1 \leq j < k \leq n$ , as the same proof holds for the other case  $X_{kj} = -X_{jk}$  for  $1 \leq j < k \leq n$ .

Define the event

$$\begin{aligned} \mathcal{A} = & \left\{ \max_{j \in [n]} \sum_{k \in [n] \setminus \{j\}} A_{jk} \leq 2np, \right. \\ & \left. \sum_{j=1}^n \sum_{k \in [n] \setminus \{j\}} A_{jk} \leq n(n-1)p + 5\sqrt{n^2 p \log n} \right\}. \end{aligned}$$

Under the assumption  $\frac{np}{\log n} > c$ , we have  $\mathbb{P}(\mathcal{A}^c) \leq n^{-11}$  by Bernstein's inequality and a union bound argument. Define

$$g(X) = \sqrt{\sum_{j=1}^n \left( \sum_{k \in [n] \setminus \{j\}} A_{jk} X_{jk} \right)^2}.$$

Then, for any  $A \in \mathcal{A}$ , for any two symmetric matrices  $X', X'' \in \mathbb{R}^{n \times n}$ , we have

$$\begin{aligned} & |g(X'') - g(X')| \\ & \leq \sqrt{\sum_{j=1}^n \left( \sum_{k \in [n] \setminus \{j\}} A_{jk} (X''_{jk} - X'_{jk}) \right)^2} \\ & \leq \sqrt{\sum_{j=1}^n \left( \sum_{k \in [n] \setminus \{j\}} A_{jk} \sum_{k \in [n] \setminus \{j\}} (X''_{jk} - X'_{jk})^2 \right)} \\ & \leq \sqrt{2np} \sqrt{\sum_{j=1}^n \sum_{k \in [n] \setminus \{j\}} (X''_{jk} - X'_{jk})^2} \\ & = 2\sqrt{np} \sqrt{\sum_{j=1}^n \sum_{k > j} (X''_{jk} - X'_{jk})^2} \end{aligned}$$

where the first inequality is by the triangle inequality and the second inequality is by the Cauchy-Schwarz inequality. As a consequence,  $g(X)$ , as a function of the upper triangular matrix of  $X$ , has a Lipschitz constant bounded by  $2\sqrt{np}$ . By a standard Gaussian concentration inequality for Lipschitz functions [22], we have

$$\begin{aligned} & \mathbb{P}(|g(X) - \mathbb{E}(g(X)|A)| > t\sqrt{np}) \\ & \leq \mathbb{P}(\mathcal{A}^c) + \sup_{A \in \mathcal{A}} \mathbb{P}_A(|g(X) - \mathbb{E}(g(X)|A)| > t\sqrt{np}) \\ & \leq n^{-11} + 2\exp(-C_1 t^2), \end{aligned}$$

for some constant  $C_1 > 0$ . Therefore, by choosing  $t = C_2\sqrt{\log n}$  for some constant  $C_2 > 0$ , we have

$$g(X) \leq \mathbb{E}(g(X)|A) + C_2\sqrt{np \log n},$$

with probability at least  $1 - 2n^{-11}$ . Additionally, we have

$$\begin{aligned} \mathbb{E}(g(X)|A) & \leq \sqrt{\sum_{j=1}^n \mathbb{E} \left( \sum_{k \in [n] \setminus \{j\}} A_{jk} X_{jk} \right)^2} \\ & = \sqrt{\sum_{j=1}^n \sum_{k \in [n] \setminus \{j\}} A_{jk}}. \end{aligned}$$

By the definition of  $\mathcal{A}$ , we have  $\mathbb{E}(g(X)|A) \leq \sqrt{n(n-1)p} + 5\sqrt{n^2 p \log n}$  on  $\mathcal{A}$ . Hence,

$$g(X) \leq \sqrt{n(n-1)p} + 5\sqrt{n^2 p \log n} + C_2\sqrt{np \log n},$$

with probability at least  $1 - n^{-10}$ , and the desired result is implied by squaring both sides of the above inequality.  $\square$

**Lemma 8 (Lemma 13 of [23]):** Consider independent random variables  $X_j \sim \mathcal{N}(0, 1)$  and  $E_j \sim \text{Bernoulli}(p)$ . Then,

$$\mathbb{P} \left( \left| \sum_{j=1}^n X_j E_j / p \right| > t \right) \leq 2 \exp \left( - \min \left( \frac{pt^2}{16n}, \frac{pt}{2} \right) \right),$$

for any  $t > 0$ .

**Lemma 9:** For any  $z, z^* \in \mathbb{C}_1^n$ , we have

$$n\ell(z, z^*) \leq \|zz^H - z^*z^{*H}\|_F^2 \leq 2n\ell(z, z^*).$$

*Proof:* The definition of  $\ell(z, z^*)$  implies

$$\ell(z, z^*) = 2n - \sup_{|a|=1} (z^H z^* a + z^{*H} z \bar{a}) = 2(n - |z^H z^*|).$$

By direct calculation,

$$\|zz^H - z^*z^{*H}\|_F^2 = 2(n^2 - |z^H z^*|^2).$$

We thus obtain the relation

$$\begin{aligned} \|zz^H - z^*z^{*H}\|_F^2 &= \ell(z, z^*) (n + |z^H z^*|) \\ &= \ell(z, z^*) \left( 2n - \frac{\ell(z, z^*)}{2} \right), \end{aligned}$$

which immediately implies the conclusion.  $\square$

**Lemma 10:** For any  $x \in \mathbb{C}$  such that  $\text{Re}(x) > 0$ ,  $\left| \frac{x}{|x|} - 1 \right| \leq \left| \frac{\text{Im}(x)}{\text{Re}(x)} \right|$ .

*Proof:* Let  $x = a + bi$ , and then

$$\begin{aligned} \left| \frac{x}{|x|} - 1 \right| &= \frac{|x - |x||}{|x|} \\ &= \frac{|a + bi - \sqrt{a^2 + b^2}|}{|x|} \\ &= \frac{\sqrt{(a - \sqrt{a^2 + b^2})^2 + b^2}}{\sqrt{a^2 + b^2}} \\ &= \sqrt{\frac{2b^2}{a^2 + b^2 + a\sqrt{a^2 + b^2}}} \\ &\leq \left| \frac{b}{a} \right| = \left| \frac{\text{Im}(x)}{\text{Re}(x)} \right|. \end{aligned}$$

The proof is complete.  $\square$

**Lemma 11:** For any  $x \in \mathbb{C} \setminus \{0\}$  and any  $y \in \mathbb{C}_1^n$ ,  $\left| \frac{x}{|x|} - y \right| \leq 2|x - y|$ .

*Proof:* We have  $\left| \frac{x}{|x|} - y \right| \leq \left| \frac{x}{|x|} - x \right| + |x - y| = |x| - 1 + |x - y| = |x| - |y| + |x - y| \leq 2|x - y|$ .  $\square$

## B. Proof of Lemma 1

We organize the proof into four steps. We first list a few high-probability events in Step 1. These events are assumed to be true in later steps. Step 2 provides an error decomposition of  $\ell(f(z), z^*)$ , and then each error term in the decomposition will be analyzed and bounded in Step 3. Finally, we combine the bounds and derive the desired result in Step 4.

*a) Step 1. Some high-probability events:* By Lemma 5, Lemma 6 and Lemma 7, we know that

$$\min_{j \in [n]} \sum_{k \in [n] \setminus \{j\}} A_{jk} \geq (n-1)p - C\sqrt{np \log n}, \quad (25)$$

$$\max_{j \in [n]} \sum_{k \in [n] \setminus \{j\}} A_{jk} \leq (n-1)p + C\sqrt{np \log n}, \quad (26)$$

$$\|A - \mathbb{E}A\|_{\text{op}} \leq C\sqrt{np}, \quad (27)$$

$$\|A \circ W\|_{\text{op}} \leq C\sqrt{np}, \quad (28)$$

$$\begin{aligned} & \sum_{j=1}^n \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} \text{Im}(W_{jk} z_k^* \bar{z}_j^*) \right|^2 \\ & \leq \frac{n^2 p}{2} \left( 1 + C\sqrt{\frac{\log n}{n}} \right), \end{aligned} \quad (29)$$

$$\sum_{j=1}^n \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} \operatorname{Re}(W_{jk} z_k^* \bar{z}_j^*) \right|^2 \quad (30)$$

$$\leq \frac{n^2 p}{2} \left( 1 + C \sqrt{\frac{\log n}{n}} \right),$$

all hold with probability at least  $1 - n^{-9}$  for some constant  $C > 0$ . To establish (29)-(30), note that  $\sqrt{2}\operatorname{Im}(W_{jk} z_k^* \bar{z}_j^*)$ ,  $\sqrt{2}\operatorname{Re}(W_{jk} z_k^* \bar{z}_j^*)$  are all independently standard normally distributed for  $1 \leq j < k \leq n$ . We also have  $-\operatorname{Im}(W_{jk} z_k^* \bar{z}_j^*) = \operatorname{Im}(\bar{W}_{jk} \bar{z}_k^* z_j^*) = \operatorname{Im}(W_{kj} z_j^* \bar{z}_k^*)$  and  $\operatorname{Re}(W_{jk} z_k^* \bar{z}_j^*) = \operatorname{Re}(\bar{W}_{jk} \bar{z}_k^* z_j^*) = \operatorname{Re}(W_{kj} z_j^* \bar{z}_k^*)$  for any  $1 \leq j < k \leq n$ .

In addition to (25)-(30), we need another high-probability inequality. For a sufficiently small  $\rho$  such that  $\frac{\rho^2 np}{\sigma^2}$  is sufficiently large, we want to upper bound the random variable  $\sum_{j=1}^n \mathbb{I} \left\{ \frac{2\sigma}{np} \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} z_k^* \right| > \rho \right\}$ . The existence of such  $\rho$  is guaranteed by the condition that  $\frac{np}{\sigma^2}$  is sufficiently large, and the specific choice will be given later. We first bound its expectation by Lemma 8,

$$\begin{aligned} & \sum_{j=1}^n \mathbb{P} \left\{ \frac{2\sigma}{np} \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} z_k^* \right| > \rho \right\} \\ & \leq \sum_{j=1}^n \mathbb{P} \left\{ \frac{2\sigma}{np} \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} \operatorname{Re}(W_{jk} z_k^*) \right| > \frac{\rho}{2} \right\} \\ & \quad + \sum_{j=1}^n \mathbb{P} \left\{ \frac{2\sigma}{np} \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} \operatorname{Im}(W_{jk} z_k^*) \right| > \frac{\rho}{2} \right\} \\ & \leq 4n \exp \left( -\frac{\rho^2 np}{256\sigma^2} \right) + 4n \exp \left( -\frac{\rho np}{8\sigma} \right). \end{aligned}$$

By Markov inequality, we have

$$\begin{aligned} & \sum_{j=1}^n \mathbb{I} \left\{ \frac{2\sigma}{np} \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} z_k^* \right| > \rho \right\} \\ & \leq \frac{4\sigma^2}{\rho^2 p} \exp \left( -\frac{1}{16} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right), \end{aligned} \quad (31)$$

with probability at least

$$\begin{aligned} & 1 - \frac{\rho^2 pn}{\sigma^2} \left( \exp \left( -\frac{\rho^2 np}{256\sigma^2} + \frac{1}{16} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right) \right. \\ & \quad \left. + \exp \left( -\frac{\rho np}{8\sigma} + \frac{1}{16} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right) \right) \\ & \geq 1 - \frac{2\rho^2 pn}{\sigma^2} \exp \left( -\frac{1}{16} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right) \\ & \geq 1 - \exp \left( -\frac{1}{32} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right). \end{aligned}$$

Finally, we conclude that the events (25)-(31) hold simultaneously with probability at least  $1 - n^{-9} - \exp \left( -\frac{1}{32} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right)$ .

*b) Step 2. Error decomposition:* For any  $z \in \mathbb{C}_1^n$  such that  $\ell(z, z^*) \leq \gamma n$ , we can define  $\tilde{z} \in \mathbb{C}^n$  such that

$$\tilde{z}_j = \frac{\sum_{k \in [n] \setminus \{j\}} A_{jk} Y_{jk} z_k}{\sum_{k \in [n] \setminus \{j\}} A_{jk}}$$

for each  $j \in [n]$ . Denote  $\hat{z} = f(z)$  then  $\hat{z}_j = \tilde{z}_j / |\tilde{z}_j|$  for each coordinate such that  $\tilde{z}_j \neq 0$ .

The condition  $\ell(z, z^*) \leq \gamma n$  implies there exists some  $b \in \mathbb{C}_1$  such that  $\|z - z^* b\|^2 \leq \gamma n$ . By direct calculation, we can write

$$\begin{aligned} \tilde{z}_j \bar{z}_j^* \bar{b} &= \frac{\sum_{k \in [n] \setminus \{j\}} A_{jk} z_k^* \bar{z}_k^* z_k}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \bar{z}_j^* \bar{b} \\ &+ \frac{\sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} z_k}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \bar{z}_j^* \bar{b} \\ &= 1 + \frac{\sum_{k \in [n] \setminus \{j\}} A_{jk} \bar{z}_k^* \bar{b} (z_k - z_k^* b)}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \\ &+ \frac{\sigma \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} (z_k - z_k^* b) \bar{z}_j^* \bar{b}}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \\ &+ \frac{\sigma \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} z_k^* \bar{z}_j^*}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \end{aligned}$$

which is equal to

$$\begin{aligned} & 1 + \frac{1}{n-1} \sum_{k=1}^n \bar{z}_k^* \bar{b} (z_k - z_k^* b) - \frac{1}{n-1} \bar{z}_j^* \bar{b} (z_j - z_j^* b) \\ & + \left( \frac{\sum_{k \in [n] \setminus \{j\}} A_{jk} \bar{z}_k^* \bar{b} (z_k - z_k^* b)}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \right. \\ & \quad \left. - \frac{1}{n-1} \sum_{k \in [n] \setminus \{j\}} \bar{z}_k^* \bar{b} (z_k - z_k^* b) \right) \\ & + \frac{\sigma \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} (z_k - z_k^* b) \bar{z}_j^* \bar{b}}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \\ & + \frac{\sigma \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} z_k^* \bar{z}_j^*}{\sum_{k \in [n] \setminus \{j\}} A_{jk}}. \end{aligned}$$

Now we define  $a_0 = 1 + \frac{1}{n-1} \sum_{k=1}^n \bar{z}_k^* \bar{b} (z_k - z_k^* b)$  and  $a = a_0 / |a_0|$ , and we have

$$\tilde{z}_j \bar{z}_j^* \bar{a} \bar{b} = |a_0| - \frac{1}{n-1} \bar{z}_j^* \bar{a} \bar{b} (z_j - z_j^* b) + F_j + G_j + H_j,$$

where

$$\begin{aligned} F_j &= \frac{\sum_{k \in [n] \setminus \{j\}} A_{jk} \bar{z}_k^* \bar{a} \bar{b} (z_k - z_k^* b)}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \\ &\quad - \frac{1}{n-1} \sum_{k \in [n] \setminus \{j\}} \bar{z}_k^* \bar{a} \bar{b} (z_k - z_k^* b), \\ G_j &= \frac{\sigma \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} (z_k - z_k^* b) \bar{z}_j^* \bar{a} \bar{b}}{\sum_{k \in [n] \setminus \{j\}} A_{jk}}, \\ H_j &= \frac{\sigma \bar{a} \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} z_k^* \bar{z}_j^*}{\sum_{k \in [n] \setminus \{j\}} A_{jk}}. \end{aligned}$$



By Lemma 10, we have the bound

$$|\hat{z}_j - z_j^* ab|^2 = |\hat{z}_j \bar{z}_j^* \bar{a} \bar{b} - 1|^2 \leq \left| \frac{\text{Im}(\hat{z}_j \bar{z}_j^* \bar{a} \bar{b})}{\text{Re}(\hat{z}_j \bar{z}_j^* \bar{a} \bar{b})} \right|^2, \quad (32)$$

whenever  $\text{Re}(\hat{z}_j \bar{z}_j^* \bar{a} \bar{b}) > 0$  holds. By Cauchy-Schwartz inequality, we have

$$|a_0| \geq 1 - \frac{1}{n-1} \|z^*\| \|z - bz^*\| \geq 1 - 2\sqrt{\gamma}, \quad (33)$$

under the condition  $\|z - z^* b\|^2 \leq \gamma n$ . We also have

$$\frac{1}{n-1} |\bar{z}_j^* \bar{a} \bar{b} (z_j - z_j^* b)| \leq \frac{1}{n-1} |z_j - z_j^* b| \leq \frac{\sqrt{\gamma n}}{n-1}.$$

Therefore, as long as  $|F_j| \vee |G_j| \vee |H_j| \leq \rho$ , we have

$$\text{Re}(\hat{z}_j \bar{z}_j^* \bar{a} \bar{b}) \geq 1 - 3(\sqrt{\gamma} + \rho). \quad (34)$$

Under the assumption that  $\gamma < 1/16$  and with a proper choice of  $\rho$  that will be given later, we have  $3(\sqrt{\gamma} + \rho) < 1$  and then  $\text{Re}(\hat{z}_j \bar{z}_j^* \bar{a} \bar{b})$  is positive. Hence, the event  $\{\hat{z}_j = 0\}$  is included in the event  $\{|F_j| \vee |G_j| \vee |H_j| > \rho\}$ .

By (32), we obtain the bound

$$\begin{aligned} |\hat{z}_j - z_j^* ab|^2 &\leq \left| \frac{\text{Im}(\hat{z}_j \bar{z}_j^* \bar{a} \bar{b})}{\text{Re}(\hat{z}_j \bar{z}_j^* \bar{a} \bar{b})} \right|^2 \mathbb{I}\{|F_j| \vee |G_j| \vee |H_j| \leq \rho\} \\ &\quad + 4\mathbb{I}\{|F_j| \vee |G_j| \vee |H_j| > \rho\} \\ &\leq \frac{1}{(1 - 3(\sqrt{\gamma} + \rho))^2} \left| \text{Im} \left( -\frac{1}{n-1} \bar{z}_j^* \bar{a} \bar{b} (z_j - z_j^* b) \right) \right. \\ &\quad \left. + \text{Im}(F_j) + \text{Im}(G_j) + \text{Im}(H_j) \right|^2 \\ &\quad + 4\mathbb{I}\{|F_j| > \rho\} + 4\mathbb{I}\{|G_j| > \rho\} + 4\mathbb{I}\{|H_j| > \rho\}, \end{aligned}$$

which can be further upper bounded by

$$\begin{aligned} &\leq \frac{1+\eta}{(1 - 3(\sqrt{\gamma} + \rho))^2} |\text{Im}(H_j)|^2 \\ &\quad + \frac{3(1+\eta^{-1})}{(1 - 3(\sqrt{\gamma} + \rho))^2} |\text{Im}(F_j)|^2 \\ &\quad + \frac{3(1+\eta^{-1})}{(1 - 3(\sqrt{\gamma} + \rho))^2} \left| \text{Im} \left( -\frac{1}{n-1} \bar{z}_j^* \bar{a} \bar{b} (z_j - z_j^* b) \right) \right|^2 \\ &\quad + \frac{3(1+\eta^{-1})}{(1 - 3(\sqrt{\gamma} + \rho))^2} |\text{Im}(G_j)|^2 + 4\mathbb{I}\{|F_j| > \rho\} \\ &\quad + 4\mathbb{I}\{|G_j| > \rho\} + 4\mathbb{I}\{|H_j| > \rho\}, \end{aligned}$$

for some  $\eta$  which will be specified later. Here the last inequality is due to the fact that  $(x_1 + x_2)^2 = x_1^2 + x_2^2 + 2(\eta^{1/2}x_1)(\eta^{-1/2}x_2) \leq (1+\eta)x_1^2 + (1+\eta^{-1})x_2^2$  for any  $x_1, x_2 \in \mathbb{R}$ .

c) *Step 3. Analysis of each error term:* Next, we will analyze the error terms  $F_j$ ,  $G_j$  and  $H_j$  separately. By triangle inequality, (25) and (26), we have

$$\begin{aligned} |F_j| &\leq \frac{\left| \sum_{k \in [n] \setminus \{j\}} (A_{jk} - p) \bar{z}_k^* \bar{a} \bar{b} (z_k - z_k^* b) \right|}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} \\ &\quad + \left| p \sum_{k \in [n] \setminus \{j\}} \bar{z}_k^* \bar{a} \bar{b} (z_k - z_k^* b) \right| \left| \frac{1}{\sum_{k \in [n] \setminus \{j\}} A_{jk}} - \frac{1}{(n-1)p} \right| \end{aligned}$$

$$\begin{aligned} &\leq \frac{2}{np} \left| \sum_{k \in [n] \setminus \{j\}} (A_{jk} - p) \bar{z}_k^* (z_k - z_k^* b) \right| \\ &\quad + p\sqrt{n} \|z - bz^*\| \frac{2 \left| \sum_{k \in [n] \setminus \{j\}} (A_{jk} - p) \right|}{n^2 p^2} \\ &\leq \frac{2}{np} \left| \sum_{k \in [n] \setminus \{j\}} (A_{jk} - p) \bar{z}_k^* (z_k - z_k^* b) \right| \\ &\quad + C_1 \frac{\sqrt{p \log n}}{np} \|z - bz^*\|. \end{aligned}$$

Using (27), we have

$$\begin{aligned} \sum_{j=1}^n |F_j|^2 &\leq \frac{8}{n^2 p^2} \sum_{j=1}^n \left| \sum_{k \in [n] \setminus \{j\}} (A_{jk} - p) \bar{z}_k^* (z_k - z_k^* b) \right|^2 \\ &\quad + C_1^2 \frac{\log n}{np} \|z - bz^*\|^2 \\ &\leq \frac{8}{n^2 p^2} \|A - \mathbb{E}A\|_{\text{op}}^2 \|z - bz^*\|^2 + C_1^2 \frac{\log n}{np} \|z - bz^*\|^2 \\ &\leq C_2 \frac{\log n}{np} \ell(z, z^*). \end{aligned}$$

The above bound also implies

$$\sum_{j=1}^n \mathbb{I}\{|F_j| > \rho\} \leq \rho^{-2} \sum_{j=1}^n |F_j|^2 \leq \frac{C_2 \log n}{\rho^2 np} \ell(z, z^*).$$

Similarly, we can also bound the error terms that depend on  $G_j$ . By (25) and (28), we have

$$\begin{aligned} \sum_{j=1}^n |G_j|^2 &\leq \frac{2\sigma^2}{n^2 p^2} \sum_{j=1}^n \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} (z_k - z_k^* b) \right|^2 \\ &\leq \frac{2\sigma^2}{n^2 p^2} \|A \circ W\|_{\text{op}}^2 \|z - bz^*\|^2 \\ &\leq C_3 \frac{\sigma^2}{np} \ell(z, z^*), \end{aligned}$$

and thus

$$\sum_{j=1}^n \mathbb{I}\{|G_j| > \rho\} \leq \rho^{-2} \sum_{j=1}^n |G_j|^2 \leq \frac{C_3 \sigma^2}{\rho^2 np} \ell(z, z^*).$$

For the contribution of  $H_j$ , we use (25) and (31), and have

$$\begin{aligned} &\sum_{j=1}^n \mathbb{I}\{|H_j| > \rho\} \\ &\leq \sum_{j=1}^n \mathbb{I} \left\{ \frac{2\sigma}{np} \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} W_{jk} z_k^* \right| > \rho \right\} \\ &\leq \frac{4\sigma^2}{\rho^2 p} \exp \left( -\frac{1}{16} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right). \end{aligned}$$

Next, we study the main error term  $|\text{Im}(H_j)|^2$ . By (25), we have

$$\begin{aligned} \sum_{j=1}^n |\text{Im}(H_j)|^2 &\leq \left( 1 + C_4 \sqrt{\frac{\log n}{np}} \right)^2 \frac{\sigma^2}{n^2 p^2} \\ &\quad \times \sum_{j=1}^n \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} \text{Im}(W_{jk} z_k^* \bar{z}_j^* \bar{a}) \right|^2, \end{aligned}$$

which can be further upper bounded by

$$\begin{aligned} &\leq (1 + \eta) \left( 1 + C_4 \sqrt{\frac{\log n}{np}} \right)^2 \frac{\sigma^2}{n^2 p^2} \\ &\quad \times \sum_{j=1}^n \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} \operatorname{Im}(W_{jk} z_k^* \bar{z}_j^*) \right|^2 \\ &+ (1 + \eta^{-1}) \left( 1 + C_4 \sqrt{\frac{\log n}{np}} \right)^2 \frac{\sigma^2}{n^2 p^2} \\ &\quad \times \sum_{j=1}^n \left| \sum_{k \in [n] \setminus \{j\}} A_{jk} \operatorname{Re}(W_{jk} z_k^* \bar{z}_j^*) \right|^2 |\operatorname{Im}(\bar{a})|^2. \end{aligned}$$

By (33), we have

$$\begin{aligned} |\operatorname{Im}(\bar{a})| &= \frac{|\operatorname{Im}(a_0)|}{|a_0|} \leq \frac{\frac{1}{n-1} \|z^*\| \|z - bz^*\|}{1 - 2\sqrt{\gamma}} \\ &\leq \frac{\frac{n}{n-1} \sqrt{\gamma}}{1 - 2\sqrt{\gamma}} \leq 20\sqrt{\gamma}, \end{aligned}$$

where the last inequality is due to the assumption that  $\gamma < 1/16$ . Together with (29) and (30), we have

$$\sum_{j=1}^n |\operatorname{Im}(H_j)|^2 \leq \left( 1 + C_5 \left( \eta + \eta^{-1}\gamma + \sqrt{\frac{\log n}{np}} \right) \right) \frac{\sigma^2}{2p}.$$

The last error term we need to analyze is  $\left| \operatorname{Im} \left( -\frac{1}{n-1} \bar{z}_j^* \bar{a} b (z_j - z_j^* b) \right) \right|^2$ . It can simply be bounded by  $\frac{1}{(n-1)^2} |z_j - z_j^* b|^2$ .

d) *Step 4. Combining the bounds:* Plugging all the individual error bounds obtained in Step 3 into the error decomposition in Step 2, we obtain

$$\begin{aligned} &\ell(\hat{z}, z^*) \\ &\leq \sum_{j=1}^n |\hat{z}_j - z_j^* a b|^2 \\ &\leq \left( 1 + C_6 \left( \rho + \sqrt{\gamma} + \eta + \eta^{-1}\gamma + \sqrt{\frac{\log n}{np}} \right) \right) \frac{\sigma^2}{2p} \\ &\quad + \frac{16\sigma^2}{\rho^2 p} \exp \left( -\frac{1}{16} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right) \\ &\quad + C_6 (\eta^{-1} + \rho^{-2}) \frac{\log n + \sigma^2}{np} \ell(z, z^*). \end{aligned}$$

We set

$$\eta = \sqrt{\gamma + \frac{\log n + \sigma^2}{np}} \quad \text{and} \quad \rho^2 = \sqrt{32} \sqrt{\frac{\log n + \sigma^2}{np}}.$$

Then, since  $\frac{\rho^2 np}{\sigma^2}$  is sufficiently large, we have

$$\frac{16\sigma^2}{\rho^2 p} \exp \left( -\frac{1}{16} \sqrt{\frac{\rho^2 np}{\sigma^2}} \right) \leq \frac{\sigma^2}{\rho^2 p} \left( \frac{\sigma^2}{\rho^2 np} \right)^2 \leq \frac{\sigma^2}{p} \sqrt{\frac{\sigma^2}{np}}.$$

Therefore, we have

$$\begin{aligned} \ell(\hat{z}, z^*) &\leq \left( 1 + C_7 \left( \gamma^2 + \frac{\log n + \sigma^2}{np} \right)^{1/4} \right) \frac{\sigma^2}{2p} \\ &\quad + C_7 \sqrt{\frac{\log n + \sigma^2}{np}} \ell(z, z^*). \end{aligned}$$

Since the above inequality is derived from the condition (25)-(31) and  $\ell(z, z^*) \leq \gamma n$ . It holds uniformly overall  $z \in \mathbb{C}_n^1$  such that  $\ell(z, z^*) \leq \gamma n$  with probability at least  $1 - n^{-9} - \exp \left( -\left(\frac{np}{\sigma^2}\right)^{1/4} \right)$ . The proof is complete.

### C. Proof of Lemma 4

We use the version of van Trees' inequality presented as Equation (11) of the paper [19]. We have

$$\begin{aligned} &\inf_{\hat{T}, \hat{S}} \int \int f(a) f(c) \mathbb{E}_{(a,c)} \\ &\quad \left[ (\hat{T} - T(a, c))^2 + (\hat{S} - S(a, c))^2 \right] da dc \\ &\geq \int \int \operatorname{Tr}(J(a, c)) f(a) f(c) da dc - I(f), \end{aligned} \quad (35)$$

where  $I(f)$  is the information of the prior which will be elaborated later. The matrix  $J(a, c)$  is given by  $J(a, c) = AB^{-1}A$ , where

$$\begin{aligned} A &= \frac{\partial(T, S)}{\partial(a, c)} \\ &= \begin{pmatrix} c - \sqrt{1-c^2} \frac{a}{\sqrt{1-a^2}} & a - \sqrt{1-a^2} \frac{c}{\sqrt{1-c^2}} \\ -\frac{ac}{\sqrt{1-a^2}} - \sqrt{1-c^2} & \sqrt{1-a^2} + \frac{ac}{\sqrt{1-c^2}} \end{pmatrix}, \end{aligned}$$

and  $B$  is the Fisher information matrix. Due to the independence between  $A_{12}Y_{12}$  and  $\{A_{1j}Y_{1j}\}_{j \geq 3} \cup \{A_{2j}Y_{2j}\}_{j \geq 3}$ , we have  $B = B_1 + B_2$  with

$$B_1 = \frac{2p}{\sigma^2} \begin{pmatrix} \left| \frac{\partial T}{\partial a} \right|^2 + \left| \frac{\partial S}{\partial a} \right|^2 & \frac{\partial T}{\partial a} \frac{\partial T}{\partial c} \\ \frac{\partial T}{\partial a} \frac{\partial T}{\partial c} & \left| \frac{\partial T}{\partial c} \right|^2 + \left| \frac{\partial S}{\partial c} \right|^2 \end{pmatrix},$$

and

$$B_2 = \frac{2(n-2)p}{\sigma^2} \begin{pmatrix} \frac{1}{1-a^2} & 0 \\ 0 & \frac{1}{1-c^2} \end{pmatrix},$$

In the following we show how to derive  $[B_2]_{11}$ . Let  $(\xi, \zeta)^T$  be a bivariate normal random vector distributed according to (22). Thus, the part of the likelihood function that involves  $a$  is proportional to  $l(\xi, \zeta, \{A_{1j}\}_{j \geq 3}; a)$  which is defined as

$$\begin{aligned} &l(\xi, \zeta, \{A_{1j}\}_{j \geq 3}; a) \\ &= \exp \left( -\frac{1}{\sigma^2 \sum_{j \geq 3} A_{1j}} \left( \left( \xi - a \sum_{j \geq 3} A_{1j} \right)^2 \right. \right. \\ &\quad \left. \left. + \left( \zeta - \sqrt{1-a^2} \sum_{j \geq 3} A_{1j} \right)^2 \right) \right). \end{aligned}$$

Then

$$\begin{aligned} [B_2]_{11} &= -\mathbb{E} \frac{\partial^2 \log l(\xi, \zeta, \{A_{1j}\}_{j \geq 3}; a)}{\partial a^2} \\ &= \mathbb{E} \frac{2\zeta}{\sigma^2 (1-a^2)^{3/2}} = \frac{2(n-2)p}{(1-a^2)\sigma^2}. \end{aligned}$$

The rest of the entries of  $B_1$  and  $B_2$  can be calculated analogously and are omitted here.

We then have

$$\begin{aligned} &\operatorname{Tr}(J(a, c)) \\ &\geq \operatorname{Tr}(AB_2^{-1}A^T) - |\operatorname{Tr}(A((B_1 + B_2)^{-1} - B_2^{-1})A^T)| \\ &\geq \|B_2^{-1/2}A^T\|_F^2 \left( 1 - \|B_2^{1/2}(B_1 + B_2)^{-1}B_2^{1/2} - I_2\|_F \right). \end{aligned}$$

By direct calculation,  $\|B_2^{-1/2}A^\top\|_F^2 = \frac{2\sigma^2}{(n-2)p}$ . Therefore,

$$\begin{aligned} & \int \int \text{Tr}(J(a, c))\pi(a)\pi(c)dadc \\ & \geq \frac{\sigma^2}{p(n-2)} \left( 1 - \int \int \|B_2^{1/2}(B_1 + B_2)^{-1}B_2^{1/2} - I_2\|_F \right. \\ & \quad \left. f(a)f(c)dadc \right). \end{aligned}$$

Now we set  $f$  to be a density function of a distribution supported on the interval  $[0.4, 0.6]$ . We set

$$f(\theta) \propto \exp\left(-\frac{1}{1-10|\theta-1/2|^2}\right) \mathbb{I}\{0.4 < \theta < 0.6\}.$$

This makes  $f$  satisfy the regularity condition required by [19] so that van Trees' inequality (35) holds. By the choice of  $f$ , we know that both  $a$  and  $c$  are bounded away from 0 and 1. As a result, we have  $\|B_2\|_{\text{op}} \lesssim \frac{np}{\sigma^2}$ ,  $\|B_2^{-1}\|_{\text{op}} \lesssim \frac{\sigma^2}{np}$  and  $\|(B_1 + B_2)^{-1}\|_{\text{op}} \lesssim \frac{\sigma^2}{np}$ . Thus,

$$\begin{aligned} & \|B_2^{1/2}(B_1 + B_2)^{-1}B_2^{1/2} - I_2\|_F \\ & \leq \|B_2\|_{\text{op}}\|(B_1 + B_2)^{-1} - B_2^{-1}\|_F \\ & \leq \|B_2\|_{\text{op}}\|B_2^{-1}\|_{\text{op}}\|(B_1 + B_2)^{-1}\|_{\text{op}}\|B_1\|_F \\ & \lesssim n^{-1}. \end{aligned}$$

Hence, we have

$$\int \int \text{Tr}(J(a, c))f(a)f(c)dadc \geq \frac{\sigma^2}{(n-2)p} (1 - C_1 n^{-1}).$$

Finally, we need to provide an upper bound for  $I(f)$ . The definition of  $I(f)$  is given by

$$\begin{aligned} I(f) &= \int_{[0.4, 0.6]^2} \frac{1}{f(\theta_1)f(\theta_2)} \sum_{j,k,l \in \{1,2\}} \\ & \left( \frac{\partial}{\partial \theta_k} K_{jk}(\theta) f(\theta_1) f(\theta_2) \right) \left( \frac{\partial}{\partial \theta_l} K_{jl}(\theta) f(\theta_1) f(\theta_2) \right) d\theta, \end{aligned}$$

where  $\theta = (\theta_1, \theta_2) = (a, c)$  and  $K(\theta) = AB^{-1}$ , which is a  $2 \times 2$  matrix depending on the value of  $\theta = (a, c)$ . By the regularity conditions satisfied by the choice of  $f$ , we have

$$I(f) \leq C_2 \max_{\theta \in [0.4, 0.6]^2} \max_{j,k \in \{1,2\}} \left| \frac{\partial}{\partial \theta_k} K_{jk}(\theta) \right|^2.$$

The above bound can be explicitly calculated via the definitions of  $A$ ,  $B_1$  and  $B_2$ , but we omit the tedious details. Intuitively, the contribution of  $B_1$  is negligible compared with that of  $B_2$ , and the contribution of  $B_2$  is of order  $np/\sigma^2$ . Thus,  $I(f) \leq C_3 \left(\frac{\sigma^2}{np}\right)^2$ . This leads to the lower bound

$$\begin{aligned} & \inf_{\hat{T}} \int \int \pi(z_j)\pi(z_k)\mathbb{E}_z[\hat{T} - z_j\bar{z}_k]^2 dz_j dz_k \\ & \geq \frac{\sigma^2}{np} \left( 1 - C \left( \frac{1}{n} + \frac{\sigma^2}{np} \right) \right). \end{aligned}$$

The desired result is obtained by plugging the above lower bound to (20). This completes the proof.

#### D. Proofs of Lemma 2 and Lemma 3

The two lemmas can be proved via the same argument, and thus we present the proofs together. For any  $z \in \mathbb{C}^n$  such that  $\|z\|^2 = n$ , we have

$$\|p^{-1}A \circ Y - zz^H\|_F^2 = \|p^{-1}A \circ Y\|_F^2 + n^2 - 2p^{-1}z^H(A \circ Y)z.$$

Therefore,  $\min_{\|z\|^2=n} \|p^{-1}A \circ Y - zz^H\|_F^2$  is equivalent to  $\max_{\|z\|^2=n} z^H(A \circ Y)z$ . We can thus write  $z^{(0)}$  as  $z_j^{(0)} = \frac{\hat{z}_j}{|\hat{z}_j|}$ , where

$$\hat{z} = \underset{\|z\|^2=n}{\text{argmin}} \|p^{-1}A \circ Y - zz^H\|_F^2. \quad (36)$$

On the other hand, for any  $z \in \mathbb{C}_1^n$ ,

$$\begin{aligned} & \sum_{1 \leq j < k \leq n} A_{jk} |Y_{jk} - z_j \bar{z}_k|^2 \\ &= \sum_{1 \leq j < k \leq n} A_{jk} (|Y_{jk}|^2 + 1) \\ & \quad - \sum_{1 \leq j < k \leq n} A_{jk} Y_{jk} (z_j \bar{z}_k + \bar{z}_j z_k) \\ &= \sum_{1 \leq j < k \leq n} A_{jk} (|Y_{jk}|^2 + 1) - z^H(A \circ Y)z \\ &= \sum_{1 \leq j < k \leq n} A_{jk} (|Y_{jk}|^2 + 1) \\ & \quad - \frac{p}{2} (\|p^{-1}A \circ Y\|_F^2 + n^2 - \|p^{-1}A \circ Y - zz^H\|_F^2). \end{aligned}$$

Therefore,  $\min_{z \in \mathbb{C}_1^n} \sum_{1 \leq j < k \leq n} A_{jk} |Y_{jk} - z_j \bar{z}_k|^2$  is equivalent to  $\min_{z \in \mathbb{C}_1^n} \|p^{-1}A \circ Y - zz^H\|_F^2$ , and the MLE can be equivalently written as

$$\hat{z} = \underset{z \in \mathbb{C}_1^n}{\text{argmin}} \|p^{-1}A \circ Y - zz^H\|_F^2. \quad (37)$$

Now we analyze (36) and (37) simultaneously. For  $\hat{z}$  that is either (36) or (37), we have

$$\|p^{-1}A \circ Y - \hat{z}\hat{z}^H\|_F^2 \leq \|p^{-1}A \circ Y - z^* z^{*H}\|_F^2.$$

Rearranging the above inequality, we have

$$\begin{aligned} & \|\hat{z}\hat{z}^H - z^* z^{*H}\|_F^2 \\ & \leq 2 |\text{Tr}((\hat{z}\hat{z}^H - z^* z^{*H})(p^{-1}A \circ Y - z^* z^{*H}))|, \end{aligned}$$

which implies

$$\begin{aligned} & \|\hat{z}\hat{z}^H - z^* z^{*H}\|_F \\ & \leq 2 \left| \text{Tr} \left( \left( \frac{\hat{z}\hat{z}^H - z^* z^{*H}}{\|\hat{z}\hat{z}^H - z^* z^{*H}\|_F} \right) (p^{-1}A \circ Y - z^* z^{*H}) \right) \right|. \end{aligned} \quad (38)$$

Since  $\frac{\hat{z}\hat{z}^H - z^* z^{*H}}{\|\hat{z}\hat{z}^H - z^* z^{*H}\|_F}$  is a Hermitian matrix of rank at most two, it has the spectral decomposition

$$\frac{\hat{z}\hat{z}^H - z^* z^{*H}}{\|\hat{z}\hat{z}^H - z^* z^{*H}\|_F} = \lambda_1 uu^H + \lambda_2 vv^H,$$

where  $u, v$  are complex unit vectors orthogonal to each other and  $\lambda_1, \lambda_2$  are real and satisfy  $\lambda_1^2 + \lambda_2^2 = 1$ . Then, we bound the right hand side of (38) by

$$\begin{aligned} & 2|\lambda_1| |u^H(p^{-1}A \circ Y - z^* z^{*H})u| \\ & \quad + 2|\lambda_2| |v^H(p^{-1}A \circ Y - z^* z^{*H})v| \\ & \leq 2(|\lambda_1| + |\lambda_2|) \|p^{-1}A \circ Y - z^* z^{*H}\|_{\text{op}} \\ & \leq 2\sqrt{2} \|p^{-1}A \circ Y - z^* z^{*H}\|_{\text{op}}. \end{aligned}$$

Therefore,

$$\begin{aligned} & \|\widehat{z}\widehat{z}^H - z^*z^{*H}\|_F \\ & \leq 4\sqrt{2}\|p^{-1}A \circ Y - z^*z^{*H}\|_{\text{op}} \\ & \leq \frac{1}{p}\|(A - \mathbb{E}A) \circ z^*z^{*H}\|_{\text{op}} + \frac{\sigma}{p}\|A \circ W\|_{\text{op}}. \end{aligned}$$

By Lemma 5,

$$\begin{aligned} & \|(A - \mathbb{E}A) \circ z^*z^{*H}\|_{\text{op}} \\ & = \sup_{\|u\|=1} \left| \sum_{1 \leq j \neq k \leq n} (A_{jk} - p) z_j^* \bar{z}_k^* u_j \bar{u}_k \right| \\ & \leq \|A - \mathbb{E}A\|_{\text{op}} \\ & \leq C_1 \sqrt{np}, \end{aligned}$$

with probability at least  $1 - n^{-10}$ . By Lemma 6,  $\|A \circ W\|_{\text{op}} \leq C_2 \sqrt{np}$  with probability at least  $1 - n^{-10}$ . Thus,

$$\|\widehat{z}\widehat{z}^H - z^*z^{*H}\|_F^2 \leq C_3 \frac{n(\sigma^2 + 1)}{p},$$

with probability at least  $1 - 2n^{-10}$ . For the MLE  $\widehat{z}$ , it satisfies  $\widehat{z} \in \mathbb{C}_1^n$ , and thus we can use Lemma 9 and obtain the bound

$$\ell(\widehat{z}, z^*) \leq \frac{1}{n} \|\widehat{z}\widehat{z}^H - z^*z^{*H}\|_F^2 \leq C_3 \frac{\sigma^2 + 1}{p},$$

with probability at least  $1 - 2n^{-10}$ . For  $z^{(0)}$ , its definition implies that

$$z_j^{(0)} \bar{z}_k^{(0)} = \frac{\widehat{z}_j \widehat{z}_k}{|\widehat{z}_j \widehat{z}_k|},$$

and therefore, we have

$$|z_j^{(0)} \bar{z}_k^{(0)} - z_j^* \bar{z}_k^*| \leq 2|\widehat{z}_j \widehat{z}_k - z_j^* \bar{z}_k^*|,$$

by Lemma 11. Use this inequality, and we have

$$\|z^{(0)} z^{(0)H} - z^* z^{*H}\|_F \leq 2\|\widehat{z}\widehat{z}^H - z^* z^{*H}\|_F,$$

and thus  $\|z^{(0)} z^{(0)H} - z^* z^{*H}\|_F^2 \leq C_4 \frac{n(\sigma^2 + 1)}{p}$  with probability at least  $1 - 2n^{-10}$ . By Lemma 9, we have

$$\ell(z^{(0)}, z^*) \leq \frac{1}{n} \|z^{(0)} z^{(0)H} - z^* z^{*H}\|_F^2 \leq C_4 \frac{\sigma^2 + 1}{p},$$

with probability at least  $1 - 2n^{-10}$ . The proof is complete.

#### ACKNOWLEDGMENT

The authors would like to thank the associate editor and the anonymous reviewers, whose comments helped improve the presentation of the article.

#### REFERENCES

- [1] A. Singer, "Angular synchronization by eigenvectors and semidefinite programming," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 1, pp. 20–36, Jan. 2011.
- [2] E. Abbe, L. Massoulie, A. Montanari, A. Sly, and N. Srivastava, "Group synchronization on grids," 2017, *arXiv:1706.08561*. [Online]. Available: <http://arxiv.org/abs/1706.08561>
- [3] T. Gao and Z. Zhao, "Multi-frequency phase synchronization," in *Proc. Int. Conf. Mach. Learn.*, vol. 97, 2020, pp. 2132–2141.
- [4] Y. Zhong and N. Boumal, "Near-optimal bounds for phase synchronization," *SIAM J. Optim.*, vol. 28, no. 2, pp. 989–1016, 2018.
- [5] S. Ling, "Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods," 2020, *arXiv:2008.05341*. [Online]. Available: <http://arxiv.org/abs/2008.05341>
- [6] A. S. Bandeira, N. Boumal, and A. Singer, "Tightness of the maximum likelihood semidefinite relaxation for angular synchronization," *Math. Program.*, vol. 163, nos. 1–2, pp. 145–167, May 2017.
- [7] N. Boumal, "Nonconvex phase synchronization," *SIAM J. Optim.*, vol. 26, no. 4, pp. 2355–2377, 2016.
- [8] F. Filbir, F. Krahmer, and O. Melnyk, "On recovery guarantees for angular synchronization," 2020, *arXiv:2005.02032*. [Online]. Available: <http://arxiv.org/abs/2005.02032>
- [9] A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, "Message-passing algorithms for synchronization problems over compact groups," *Commun. Pure Appl. Math.*, vol. 71, no. 11, pp. 2275–2322, Nov. 2018.
- [10] H. Liu, M.-C. Yue, and A. M.-C. So, "On the estimation performance and convergence rate of the generalized power method for phase synchronization," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2426–2446, 2017.
- [11] N. Boumal, "On intrinsic Cramér–Rao bounds for Riemannian submanifolds and quotient manifolds," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1809–1821, Apr. 2013.
- [12] N. Boumal, A. Singer, P.-A. Absil, and V. D. Blondel, "Cramér–Rao bounds for synchronization of rotations," *Inf. Inference*, vol. 3, no. 1, pp. 1–39, Mar. 2014.
- [13] E. Romanov and M. Gavish, "The noise-sensitivity phase transition in spectral group synchronization over compact groups," *Appl. Comput. Harmon. Anal.*, vol. 49, no. 3, pp. 935–970, Nov. 2020.
- [14] N. Boumal and X. Cheng, "Concentration of the Kirchhoff index for Erdos–Rényi graphs," *Syst. Control Lett.*, vol. 74, pp. 74–80, Dec. 2014.
- [15] A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, "Optimality and sub-optimality of PCA for spiked random matrices and synchronization," 2016, *arXiv:1609.05573*. [Online]. Available: <http://arxiv.org/abs/1609.05573>
- [16] A. Javanmard, A. Montanari, and F. Ricci-Tersenghi, "Phase transitions in semidefinite relaxations," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 16, pp. E2218–E2223, Apr. 2016.
- [17] M. Lelarge and L. Miolane, "Fundamental limits of symmetric low-rank matrix estimation," *Probab. Theory Rel. Fields*, vol. 173, pp. 859–929, Apr. 2018.
- [18] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," *Ann. Statist.*, vol. 49, no. 1, pp. 321–345, Feb. 2021.
- [19] R. D. Gill and B. Y. Levit, "Applications of the van Trees inequality: A Bayesian Cramér–Rao bound," *Bernoulli*, vol. 1, nos. 1–2, pp. 59–79, 1995.
- [20] J. Lei and A. Rinaldo, "Consistency of spectral clustering in stochastic block models," *Ann. Statist.*, vol. 43, no. 1, pp. 215–237, 2015.
- [21] A. S. Bandeira and R. van Handel, "Sharp nonasymptotic bounds on the norm of random matrices with independent entries," *Ann. Probab.*, vol. 44, no. 4, pp. 2479–2506, 2016.
- [22] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces," *Publications Mathématiques l'Institut Hautes Etudes Scientifiques*, vol. 81, no. 1, pp. 73–205, 1995.
- [23] C. Gao, Y. Lu, Z. Ma, and H. H. Zhou, "Optimal estimation and completion of matrices with biclustering structures," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 5602–5630, 2016.

**Chao Gao** received the B.S. degree in mathematics from Zhejiang University, Hangzhou, China, in 2010, and the M.A. and Ph.D. degrees in statistics from Yale University in 2011 and 2016, respectively. He has been an Assistant Professor of statistics with The University of Chicago since 2016. He was a recipient of the NSF CAREER Award in 2019. He received the 2021 Tweedie New Researcher Award from the Institute of Mathematical Statistics (IMS).

**Anderson Y. Zhang** received the B.S. degree in statistics from Zhejiang University, Hangzhou, China, in 2013, and the Ph.D. degree in statistics and data science from Yale University in 2018. He was a William H. Kruskal Instructor with the Department of Statistics, The University of Chicago, in 2018. He has been an Assistant Professor of statistics and data science with the University of Pennsylvania since 2019.