SDP Achieves Exact Minimax Optimality in Phase Synchronization

Chao Gao and Anderson Y. Zhang

***Abstract*— We study the phase synchronization problem with noisy measurements $\boldsymbol{Y} = \boldsymbol{z}^*\boldsymbol{z}^{*\mathrm{H}} + \sigma \boldsymbol{W} \in \mathbb{C}^{n\times n}$, where $\boldsymbol{z}^*$ is an n-dimensional complex unit-modulus vector and $\boldsymbol{W}$ is a complex-valued Gaussian random matrix. It is assumed that each entry $\boldsymbol{Y_{jk}}$ is observed with probability $\boldsymbol{p}$. We prove that an SDP relaxation of the MLE achieves the error bound $(1+o(1))\frac{\sigma^2}{2np}$ under a normalized squared ℓ_2 loss. This result matches the minimax lower bound of the problem, and even the leading constant is sharp. The analysis of the SDP is based on an equivalent non-convex programming whose solution can be characterized as a fixed point of the generalized power iteration lifted to a higher dimensional space. This viewpoint unifies the proofs of the statistical optimality of three different methods: MLE, SDP, and generalized power method. The technique is also applied to the analysis of the SDP for $\mathbb{Z}_2$ synchronization, and we achieve the minimax optimal error $\exp\left(-(1-o(1))\frac{np}{2\sigma^2}\right)$ with a sharp constant in the exponent.**

***Index Terms*— Angular synchronization, minimax risk, SDP, $\mathbb{Z}_2$ synchronization, Burer-Monteiro factorization.**

I. Introduction

CONSIDER the problem of phase synchronization [1] with observations

$$Y_{jk} = z_j^* \bar{z}_k^* + \sigma W_{jk} \in \mathbb{C}, \tag{1}$$

for $1 \le j < k \le n$, where $\bar{z}_k^*$ stands for the complex conjugate of z_k^*. Our goal is to estimate $z_1^*, \cdots, z_n^* \in \mathbb{C}_1 = \{x \in \mathbb{C} : |x| = 1\}$. Since $|z_j^*| = 1$, we can write $z_j^* = e^{i\theta_j^*}$ with some $\theta_j^* \in (0, 2\pi]$ for all $j \in [n]$, and thus Y_{jk} is understood to be a noisy observation of the pairwise difference between two angles θ_j^* and θ_k^*. Following [2]–[5], we consider an additive noise model and we assume that W_{jk} is a standard complex Gaussian variable independently for all $1 \le j < k \le n$.[1]

Manuscript received 31 August 2021; accepted 7 March 2022. Date of publication 18 April 2022; date of current version 13 July 2022. The work of Chao Gao was supported in part by NSF CAREER Award under Grant DMS-1847590 and 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 S. Jalali, Associate Editor for Probability and Statistics.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3167603.

Digital Object Identifier 10.1109/TIT.2022.3167603

[1]For $W_{jk} \sim \mathcal{CN}(0,1)$, we have $\mathrm{Re}(W_{jk}) \sim \mathcal{N}\left(0, \frac{1}{2}\right)$ and $\mathrm{Im}(W_{jk}) \sim \mathcal{N}\left(0, \frac{1}{2}\right)$ independently.

Recently, minimax risk of estimating $z^* \in \mathbb{C}_1^n$ has been studied by [6] under the loss function

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

We note that the minimization of $a \in \mathbb{C}_1^n$ in the definition of (2) is necessary, since a global rotation of the angles $\theta_1^*, \cdots, \theta_n^*$ does not change the distribution of the observations $\{Y_{jk}\}_{1\le j<k\le n}$. It was proved by [6] that the minimax risk of phase synchronization has the following lower bound

$$\inf_{\widehat{z}\in\mathbb{C}_1^n} \sup_{z\in\mathbb{C}_1^n} \mathbb{E}_z \ell(\widehat{z}, z) \ge (1-o(1))\frac{\sigma^2}{2n}, \tag{3}$$

and the maximum likelihood estimator (MLE), defined as a global maximizer of

$$\max_{z\in\mathbb{C}_1^n} z^{\mathrm{H}} Y z, \tag{4}$$

is proved to achieve the error bound $(1+o(1))\frac{\sigma^2}{2n}$, and is therefore asymptotically minimax optimal. However, the optimization problem (4) is a constrained quadratic programming that is generally known to be NP-hard. This motivates researchers to consider a convex relaxation of (4) in the form of semi-definite programming (SDP) [1], [3], [5], [7], [8]. Write $Z = zz^{\mathrm{H}}$. For any $z \in \mathbb{C}_1^n$, Z is a complex positive-semidefinite Hermitian matrix whose diagonal entries are all one. The SDP relaxation of (4) is then defined as

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

A global maximizer of (5), denoted as $\widehat{Z}$, can thus be used as an estimator of the matrix $z^*z^{*\mathrm{H}}$. The tightness of the SDP (5) has been thoroughly investigated in the literature of phase synchronization. When $\sigma^2 = O(n^{1/2})$, it was proved by [3] that the solution to (5) is a rank-one matrix $\widehat{Z} = \widehat{z}\widehat{z}^{\mathrm{H}}$, with $\widehat{z}$ being a global maximizer of (4). This result was recently proved by [5] to hold under a weaker condition $\sigma^2 = O\left(\frac{n}{\log n}\right)$. Given the tightness of the SDP and the minimax optimality of the MLE in [6], we can immediately claim that the SDP (5) is also asymptotically minimax optimal under the condition $\sigma^2 = O\left(\frac{n}{\log n}\right)$. Without the condition $\sigma^2 = O\left(\frac{n}{\log n}\right)$, whether SDP is still statistically optimal remains as an open question in the literature.

In this paper, we study the statistical properties of the SDP (5) directly without the need to establish any connection

between SDP and MLE. This allows us to go beyond the condition $\sigma^2 = O\left(\frac{n}{\log n}\right)$ and we are able to derive sharp statistical error bounds of the SDP (5) as long as $\sigma^2 = o(n)$. According to the minimax lower bound (3), the condition $\sigma^2 = o(n)$ is necessary for any estimator to have an error rate of a nontrivial order. To formally state our main result, we introduce a more general statistical estimation setting that allows the possibility of missing data. Instead of observing Y_{jk} for all $1 \le j < k \le n$, 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 $1 \le j < k \le n$, and we only observe Y_{jk} that follows (1) when $A_{jk} = 1$. The SDP can be extended to this more general setting by replacing all Y_{jk}'s with $A_{jk}Y_{jk}$'s in (5). The formula will be given by (13) in Section II.

Theorem 1: Assume $\frac{np}{\sigma^2} \to \infty$ and $\frac{np}{\log n} \to \infty$. Let $\widehat{Z}$ be a global maximizer of the SDP (13) and $\widehat{z}_j = u_j/|u_j|$ for $j \in [n]$ with $u \in \mathbb{C}^n$ being the leading eigenvector of $\widehat{Z}$. There exists some $\delta = o(1)$ such that

$$\frac{1}{n^2}\|\widehat{Z} - z^* z^{*\mathrm{H}}\|_{\mathrm{F}}^2 \le (1+\delta)\frac{\sigma^2}{np},$$
$$\ell(\widehat{z}, z^*) \le (1+\delta)\frac{\sigma^2}{2np},$$

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

Compared with the minimax lower bound (Theorem 2 in Section II), Theorem 1 shows that SDP leads to both minimax optimal estimations of the matrix $z^* z^{*\mathrm{H}}$ and of the vector z^*. The two error bounds are not just rate-optimal, but the leading constants are sharp as well. We remark that both conditions $\sigma^2 = o(np)$ and $\frac{np}{\log n} \to \infty$ are essential for the results of the above theorem to hold. Since the minimax risk of the problem is of order $\frac{\sigma^2}{np}$, the condition $\sigma^2 = o(np)$, which is equivalent to $\frac{\sigma^2}{np} = o(1)$, guarantees that the minimax risk is of smaller order than the trivial one. The order $O(1)$ is trivial, as it can simply be achieved by random guess. The condition $\frac{np}{\log n} \to \infty$ guarantees that the random graph A is connected with high probability. Our technical analysis would still go through when p is of the same order as $\frac{\log n}{n}$, but in this regime only rate optimality is achieved.

Our analysis of the SDP does not rely on its connection to the MLE, and it is therefore fundamentally different from the approaches considered by [3], [5], [8]. To study the statistical properties of SDP directly, we consider the following iteration procedure,[2]

$$V_j^{(t)} = \frac{\sum_{k\in[n]\setminus\{j\}} \bar{Y}_{jk} V_k^{(t-1)}}{\left\|\sum_{k\in[n]\setminus\{j\}} \bar{Y}_{jk} V_k^{(t-1)}\right\|} \in \mathbb{C}^n,\ j = 1, \cdots, n. \quad (6)$$

Define the matrix $V^{(t)} \in \mathbb{C}^{n\times n}$ with its jth column being $V_j^{(t)}$. The above iteration can be shorthanded as $V^{(t)} = f(V^{(t-1)})$. We use (6) as a non-convex characterization of the SDP (5), because the solution to (5) can always be written as $\widehat{Z} = \widehat{V}^{\mathrm{H}}\widehat{V}$ for some $\widehat{V} \in \mathbb{C}^{n\times n}$ satisfying the fixed-point equation $\widehat{V} = f(\widehat{V})$. Note that the iterative procedure (6) resembles the formula of the generalized power method (GPM) [1], [5], [9],[3]

$$z_j^{(t)} = \frac{\sum_{k\in[n]\setminus\{j\}} Y_{jk} z_k^{(t-1)}}{\left|\sum_{k\in[n]\setminus\{j\}} Y_{jk} z_k^{(t-1)}\right|} \in \mathbb{C}, \quad j = 1, \cdots, n. \quad (7)$$

We can therefore think of (6) as a lift of the GPM (7) into a higher dimensional space. This allows us to analyze the statistical error of SDP from an iterative algorithm perspective, and previous techniques of analyzing general iterative algorithms in [10], [11] can be borrowed for the current purpose. To understand the exact statistical error of SDP, we establish the following convergence result for the iterative procedure (6),

$$\ell(V^{(t)}, z^*) \le \delta\ell(V^{(t-1)}, z^*) + \text{ optimal statistical error, for all } t \ge 1, \quad (8)$$

for some $\delta = o(1)$ with high probability, as long as it is properly initialized. Here, with slight abuse of notation, the loss of $\widehat{V}$ is defined by

$$\ell(\widehat{V}, z^*) = \min_{a\in\mathbb{C}^n:\|a\|^2=1} \frac{1}{n}\sum_{j=1}^n \|\widehat{V}_j - \bar{z}_j^* a\|^2, \quad (9)$$

which is natural given that the matrix $\widehat{Z} = \widehat{V}^{\mathrm{H}}\widehat{V}$ is used to estimate $z^* z^{*\mathrm{H}}$. Since the SDP solution is a fixed point of the iteration (6), the convergence result (8) directly leads to the sharp statistical error bounds in Theorem 1.

Our analysis of SDP through (6) also unifies the understandings of the GPM and the MLE. Given the relation between (6) and (7), the convergence result (8) directly implies

$$\ell(z^{(t)}, z^*) \le \delta\ell(z^{(t-1)}, z^*) + \text{ optimal statistical error, for all } t \ge 1, \quad (10)$$

for some $\delta = o(1)$ with high probability, as long as the GPM is properly initialized. This provides an alternative proof to the minimax optimality of the GPM that has been previously established by [6]. In addition, just as the SDP can be viewed as a fixed point of the iteration (6), the MLE can be viewed as a fixed point of the iteration (7). The minimax optimality of the MLE can also be derived. To summarize, we are able to show the exact minimax optimality of SDP, GPM, and MLE using a single proof based on the iterative procedure (6).

In addition to phase synchronization, we also establish the optimality of the SDP for $\mathbb{Z}_2$ synchronization. In the setting of $\mathbb{Z}_2$ synchronization, one observes $Y_{jk} = z_j^* z_k^* + \sigma W_{jk} \in \mathbb{R}$ for $1 \le j < k \le n$, and the goal is to estimate $z_1^*, \cdots, z_n^* \in \{-1, 1\}$. Assume $W_{jk} \sim \mathcal{N}(0,1)$ and each Y_{jk} is observed with probability p, we show that the SDP for $\mathbb{Z}_2$ synchronization achieves the error

$$\exp\left(-(1-o(1))\frac{np}{2\sigma^2}\right). \quad (11)$$

We also prove a matching lower bound for this problem. Since $\mathbb{Z}_2$ synchronization is a discrete parameter estimation problem, the minimax risk is an exponential function of the

[2]When the denominator (6) is zero, take $V_j^{(t)} = V_j^{(t-1)}$.

[3]When the denominator (7) is zero, take $z_j^{(t)} = z_j^{(t-1)}$.

signal-to-noise ratio, compared with the polynomial function for phase synchronization. Despite being a continuous optimization method, the SDP is able to adapt to the discreteness of the problem. The exponential rate (11) has been previously derived for $p = 1$ by [12]. Our analysis based on the iterative algorithm perspective generalizes their result to more general values of $p \gg \frac{\log n}{n}$.

a) Paper organization: The rest of the paper is organized as follows. In Section II, we establish the statistical optimality of the SDP for phase synchronization. The implications of the SDP analysis on the statistical error bounds of GPM and MLE are discussed in Section III. The analysis of the SDP for $\mathbb{Z}_2$ synchronization is presented in Section IV. Finally, Section V collects all the technical proofs of the paper.

b) Notation: For $d \in \mathbb{N}$, we write $[d] = \{1, \ldots, d\}$. 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, $\mathrm{Re}(x)$ for its real part, $\mathrm{Im}(x)$ for its imaginary part, and $|x|$ for its modulus. For a complex vector $x \in \mathbb{C}^d$, we use $\|x\| = \sqrt{\sum_{j=1}^d |x_j|^2}$ for its norm. For a matrix $B = (B_{jk}) \in \mathbb{C}^{d_1 \times d_2}$, we use $B^{\mathrm{H}} \in \mathbb{C}^{d_2 \times d_1}$ for its conjugate transpose such that $B^{\mathrm{H}} = (\bar{B}_{kj})$. The Frobenius norm and operator norm of B are defined by $\|B\|_{\mathrm{F}} = \sqrt{\sum_{j=1}^{d_1} \sum_{k=1}^{d_2} |B_{jk}|^2}$ and $\|B\|_{\mathrm{op}} = \sup_{u \in \mathbb{C}^{d_1}, v \in \mathbb{C}^{d_2} : \|u\| = \|v\| = 1} u^{\mathrm{H}} B v$. We use $\mathsf{Tr}(B)$ for the trace of a squared matrix B. 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 $a_n \le C b_n$ for some constant $C > 0$ independent of n. We also write $a_n = o(b_n)$ or $\frac{b_n}{a_n} \to \infty$ when $\limsup_n \frac{a_n}{b_n} = 0$.

II. Main Results

A. Problem Settings

Recall that we observe a random graph $A_{jk} \sim \text{Bernoulli}(p)$ independently for all $1 \le j < k \le n$. For each pair (j, k), we observe $Y_{jk} = z_j^* \bar{z}_k^* + \sigma W_{jk}$ with $W_{jk} \sim \mathcal{CN}(0, 1)$ whenever $A_{jk} = 1$. The observations can be organized as an adjacency matrix A and a masked version of the pairwise interactions $A \circ Y$. All the matrices A, W, and Y are Hermitian as we define $A_{jk} = A_{kj}$, $W_{jk} = \bar{W}_{kj}$, and $Y_{jk} = \bar{Y}_{kj}$ for all $1 \le k < j \le n$ and $A_{jj} = W_{jj} = 0$ and $Y_{jj} = 1$ for all $j \in [n]$. Hence we have the matrix representation $Y = z^* z^{*\mathrm{H}} + \sigma W$.

To estimate the vector $z^* \in \mathbb{C}_1^n$, the MLE is defined as a global maximizer of the following optimization problem

$$\max_{z \in \mathbb{C}_1^n} z^{\mathrm{H}} (A \circ Y) z. \tag{12}$$

Since (12) is computationally infeasible, we consider the following convex relaxation of (12) via SDP,

$$\max_{Z = Z^{\mathrm{H}} \in \mathbb{C}^{n \times n}} \mathsf{Tr}((A \circ Y) Z) \text{ subject to } \mathrm{diag}(Z) = I_n \text{ and } Z \succeq 0. \tag{13}$$

The goal of our paper is to establish the statistical optimality of the SDP (13). We first provide a minimax lower bound as the benchmark of the problem.

Theorem 2 (Theorem 4.1 of [6]): Assume $\sigma^2 = o(np)$. Then, we have

$$\inf_{\widehat{Z} \in \mathbb{C}^{n \times n}} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \frac{1}{n^2} \|\widehat{Z} - z z^{\mathrm{H}}\|_{\mathrm{F}}^2 \ge (1 - \delta) \frac{\sigma^2}{np},$$

$$\inf_{\widehat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\widehat{z}, z) \ge (1 - \delta) \frac{\sigma^2}{2np},$$

for some $\delta = o(1)$.

The above theorem has been established by [6] as the minimax lower bound for phase synchronization. In fact, Theorem 4.1 of [6] only states the lower bound result for the loss function $\ell(\widehat{z}, z)$. However, the proof of Theorem 4.1 of [6] actually established the lower bound under the loss $\frac{1}{n^2} \|\widehat{Z} - z z^{\mathrm{H}}\|_{\mathrm{F}}^2$, and the lower bound for $\ell(\widehat{z}, z)$ is proved as a direct consequence in view of the inequality

$$\inf_{\widehat{z} \in \mathbb{C}_1^n} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \ell(\widehat{z}, z) \ge \frac{1}{2} \inf_{\widehat{Z} \in \mathbb{C}^{n \times n}} \sup_{z \in \mathbb{C}_1^n} \mathbb{E}_z \frac{1}{n^2} \|\widehat{Z} - z z^{\mathrm{H}}\|_{\mathrm{F}}^2.$$

Since the solution of the SDP (13) is a matrix, it is natural to study the statistical error under $\frac{1}{n^2} \|\widehat{Z} - z z^{\mathrm{H}}\|_{\mathrm{F}}^2$ in addition to the loss $\ell(\widehat{z}, z)$.

B. A Convergence Lemma

Our analysis of the SDP (13) relies on an equivalent non-convex characterization. Since Z is a positive semi-definite Hermitian matrix, it admits a decomposition

$$Z = V^{\mathrm{H}} V,$$

for some $V \in \mathbb{C}^{n \times n}$. Let V_j be the jth column of V, and we have $Z_{jk} = V_j^{\mathrm{H}} V_k$. In particular, the constraint $\mathrm{diag}(Z) = I_n$ can be written as $Z_{jj} = \|V_j\|^2 = 1$ for all $j \in [n]$. Replacing Z by $V^{\mathrm{H}} V$, the SDP (13) can be equivalently represented as

$$\max_{V \in \mathbb{C}^{n \times n}} \mathsf{Tr}((A \circ Y) V^{\mathrm{H}} V) \text{ subject to } \|V_j\|^2 = 1 \text{ for all } j \in [n]. \tag{14}$$

The formulation (14) is closely related to the Burer-Monteiro problem [13], [14] for the SDP except that here V is still an $n \times n$ matrix without dimension reduction. This non-convex formulation allows us to derive sharp statistical error bounds of the SDP (13).

We analyze (14) through the following iteration procedure,

$$V_j^{(t)} = \begin{cases} \dfrac{\sum_{k \in [n] \setminus \{j\}} A_{jk} \bar{Y}_{jk} V_k^{(t-1)}}{\left\| \sum_{k \in [n] \setminus \{j\}} A_{jk} \bar{Y}_{jk} V_k^{(t-1)} \right\|}, \\ \qquad \text{if } \sum_{k \in [n] \setminus \{j\}} A_{jk} \bar{Y}_{jk} V_k^{(t-1)} \ne 0, \\ V_j^{(t-1)}, \text{ if } \sum_{k \in [n] \setminus \{j\}} A_{jk} \bar{Y}_{jk} V_k^{(t-1)} = 0. \end{cases} \tag{15}$$

Let us shorthand the above formula by

$$V^{(t)} = f(V^{(t-1)}), \tag{16}$$

by introducing a map $f : \mathbb{C}_1^{n \times n} \to \mathbb{C}_1^{n \times n}$ such that the jth column of $f(V^{(t-1)})$ is given by (15). We use the notation $\mathbb{C}_1^{n \times n}$ for the set of $n \times n$ complex matrices whose columns

all have unit norms. The update (16) can be seen as a local approach (or more precisely, a block coordinate ascent approach) [15], [16] to solve (14). To see why this is true, consider the following local optimization problem

$$\max_{V_j\in\mathbb{C}^n:\|V_j\|^2=1}\left[V_j^{\mathrm{H}}\left(\sum_{k\in[n]\backslash\{j\}}A_{jk}\bar{Y}_{jk}V_k^{(t-1)}\right)+\left(\sum_{k\in[n]\backslash\{j\}}A_{jk}\bar{Y}_{jk}V_k^{(t-1)}\right)^{\mathrm{H}}V_j\right]. \quad (17)$$

The objective of (17) collects the terms in the expansion of $\mathsf{Tr}((A\circ Y)V^{\mathrm{H}}V)=\sum_{jk}A_{jk}\bar{Y}_{jk}V_j^{\mathrm{H}}V_k$ that depend on V_j and replaces V_k by $V_k^{(t-1)}$ for all $k\in[n]\backslash\{j\}$. By simple algebra, we can see the solution of (17) is exactly (15).

Let $\widehat{V}$ be a global maximizer of (14). The matrix $\widehat{V}$ must be a fixed point of the map f,

$$\widehat{V}=f(\widehat{V}). \quad (18)$$

To see why (18) holds, we consider the local optimization problem (17) with $V_k^{(t-1)}$ replaced by $\widehat{V}_k$ for all $k\in[n]\backslash\{j\}$. Thus, as long as $\widehat{V}$ maximizes (14), its jth column $\widehat{V}_j$ must maximize this local optimization problem, which then implies the fixed-point equation (18).

Since the SDP solution $\widehat{Z}=\widehat{V}^{\mathrm{H}}\widehat{V}$ is an estimator of the matrix $z^*z^{*\mathrm{H}}$, we can think of $\widehat{V}_j$ as an estimator of $\bar{z}_j^*$ embedded in $\mathbb{C}^n$. Note that

$$z_j^*\bar{z}_k^*=z_j^*a^{\mathrm{H}}a\bar{z}_k^*,$$

for any $a\in\mathbb{C}^n$ such that $\|a\|^2=1$, and thus we can embed each $\bar{z}_j^*$ in $\mathbb{C}^n$ by considering the vector $\bar{z}_j^*a\in\mathbb{C}^n$. This motivates the definition of the loss function $\ell(\widehat{V},z^*)$ given in (9). The following lemma characterizes the evolution of this loss function through the map f.

Lemma 1: Assume $\frac{np}{\sigma^2}>c_1$ and $\frac{np}{\log n}>c_2$ for some sufficiently large constants $c_1,c_2>0$. Then, for any $\gamma\in[0,1/16)$, we have

$$\mathbb{P}\Big(\ell(f(V),z^*)\le\delta_1\ell(V,z^*)+(1+\delta_2)\frac{\sigma^2}{2np}\text{ for all }V\in\mathbb{C}_1^{n\times n}\text{ such that }\ell(V,z^*)\le\gamma\Big)\ge1-(2n)^{-1}-\exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right).$$

where $\delta_1=C_1\sqrt{\frac{\log n+\sigma^2}{np}}$ and $\delta_2=C_2\left(\gamma^2+\frac{\log n+\sigma^2}{np}\right)^{1/4}$ for some constants $C_1,C_2>0$.

The lemma shows that for any $V\in\mathbb{C}_1^{n\times n}$ that has a nontrivial error, the matrix $f(V)$ will have an error that is smaller by a multiplicative factor δ_1 up to an additive term $(1+\delta_2)\frac{\sigma^2}{2np}$. Define $V^*\in\mathbb{C}_1^{n\times n}$ with the jth column given by $V_j^*=\bar{z}_j^*a$ for some $a\in\mathbb{C}^n$ that satisfies $\|a\|^2=1$. We immediately have

$$\ell(f(V^*),z^*)\le(1+\delta_2)\frac{\sigma^2}{2np}.$$

Thus, the additive term $(1+\delta_2)\frac{\sigma^2}{2np}$ can be understood as the oracle statistical error given the knowledge of z^*.

The two conditions $\frac{np}{\sigma^2}\ge c_1$ and $\frac{np}{\log n}\ge c_2$ are essential for the result to hold. While $\frac{np}{\sigma^2}\ge c_1$ makes sure that the statistical error $\frac{\sigma^2}{2np}$ is of a nontrivial order, the condition $\frac{np}{\log n}\ge c_2$ guarantees that the random graph is connected. We can slightly strengthen both conditions to $\frac{np}{\sigma^2}\to\infty$ and $\frac{np}{\log n}\to\infty$ so that both δ_1 and δ_2 are varnishing.

C. Statistical Optimality of SDP

In this section, we show the result of Lemma 1 implies the statistical optimality of the SDP (13). Since the solution of the SDP can be written as $\widehat{Z}=\widehat{V}^{\mathrm{H}}\widehat{V}$ with $\widehat{V}$ satisfying the fixed-point equation (18), we can apply the result of Lemma 1 to $\widehat{V}=f(\widehat{V})$ as long as a crude bound $\ell(\widehat{V},z^*)\le\gamma$ can be proved for some $\gamma<1/16$.

Lemma 2: Assume $\frac{np}{\log n}>c$ for some sufficiently large constant $c>0$. Let $\widehat{Z}=\widehat{V}^{\mathrm{H}}\widehat{V}$ be a global maximizer of the SDP (13). Then, there exits some constant $C>0$ such that

$$\ell(\widehat{V},z^*)\le C\sqrt{\frac{\sigma^2+1}{np}},$$

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

Under the condition that $\frac{np}{\sigma^2}$ and $\frac{np}{\log n}$ are sufficiently large, we have $\ell(\widehat{V},z^*)\le\gamma$ for some $\gamma<1/16$. Thus, Lemma 1 and the fact $\widehat{V}=f(\widehat{V})$ imply that

$$\ell(\widehat{V},z^*)\le\delta_1\ell(\widehat{V},z^*)+(1+\delta_2)\frac{\sigma^2}{2np}. \quad (19)$$

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

Theorem 3: Assume $\frac{np}{\sigma^2}>c_1$ and $\frac{np}{\log n}>c_2$ for some sufficiently large constants $c_1,c_2>0$. Let $\widehat{Z}=\widehat{V}^{\mathrm{H}}\widehat{V}$ be a global maximizer of the SDP (13). Then, there exists some $\delta=C\left(\frac{\log n+\sigma^2}{np}\right)^{1/4}$ for some constant $C>0$, such that

$$\begin{aligned}\ell(\widehat{V},z^*)&\le(1+\delta)\frac{\sigma^2}{2np},\\ \frac{1}{n^2}\|\widehat{Z}-z^*z^{*\mathrm{H}}\|_{\mathrm{F}}^2&\le(1+\delta)\frac{\sigma^2}{np},\end{aligned}$$

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

Theorem 3 gives sharp statistical error bounds for both loss functions $\ell(\widehat{V},z^*)$ and $\frac{1}{n^2}\|\widehat{Z}-z^*z^{*\mathrm{H}}\|_{\mathrm{F}}^2$. While the result for $\ell(\widehat{V},z^*)$ is derived from (19), the result for $\frac{1}{n^2}\|\widehat{Z}-z^*z^{*\mathrm{H}}\|_{\mathrm{F}}^2$ is a consequence of the inequality

$$\frac{1}{n^2}\|\widehat{V}^{\mathrm{H}}\widehat{V}-z^*z^{*\mathrm{H}}\|_{\mathrm{F}}^2\le2\ell(\widehat{V},z^*),$$

which is established by Lemma 11 in Section V-A. Compared with the minimax lower bound in Theorem 2, we can conclude that the SDP (13) is minimax optimal for the estimation of the matrix $z^*z^{*\mathrm{H}}$. It not only achieves the optimal rate, but the leading constant is also sharp when $\sigma^2=o(np)$ and $\frac{np}{\log n}\to\infty$. Figure 1 verifies the correctness of the leading constants of the two loss functions. Both loss functions are

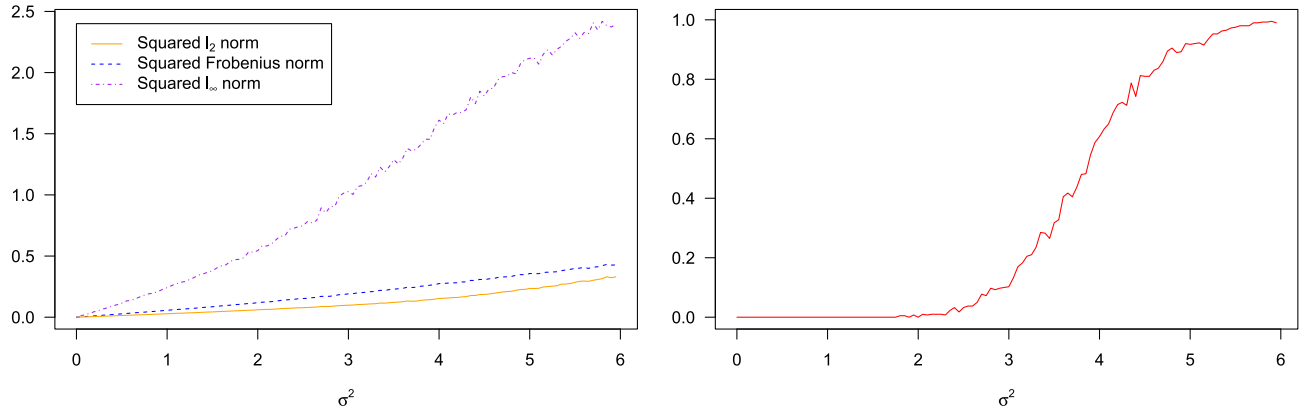

Fig. 1. *Left: average values of* $\ell(\widehat{V}, z^*)$ *(orange solid),* $\frac{1}{n^2}\|\widehat{Z} - z^* z^{*\mathrm{H}}\|_{\mathrm{F}}^2$ *(blue dashed) and* $\max_{1\le j\le n} \|\widehat{V}_j - \bar{z}_j^* a\|^2$ *(purple dash-dotted) with* $n = 100$, $p = 0.2$ *and* σ^2 *varying in* $[0, 6]$ *across* 400 *independent experiments. The vector* $a \in \mathbb{C}^n$ *used in the squared* ℓ_∞ *loss is the minimizer of the right hand side of (9).* Right: probability of the event that the second eigenvalue of $\widehat{Z}$ is nonzero (i.e., $\widehat{Z}$ is not rank-one) for the same experiments.

approximately linear at least when σ^2 is small. When σ^2 and np are of the same order, SDP does not have the optimal constant anymore, and its asymptotics is predicted by a very different technique [14]. We also remark that the ℓ_2 error control does not imply that each individual z_j^* can be accurately recovered. This is also reflected in Figure 1 with the comparison between ℓ_2 and ℓ_∞ loss.

We emphasize that our proof of the optimality of the SDP is based on a direct statistical error analysis, regardless of whether the SDP relaxation is tight or not. It is shown by [5] that the tightness of SDP (when the solution has rank one) requires $\sigma^2 = O\left(\frac{n}{\log n}\right)$ at least when $p = 1$. When $\sigma^2 = o(np)$, it is possible that SDP is not tight but still statistically optimal. This point is also illustrated by Figure 1.

Since the solution of the SDP is a matrix, some post-processing step is required to obtain a vector estimator for z^*. This can easily be done by extracting the leading eigenvector of $\widehat{Z}$. Let $u \in \mathbb{C}^n$ be the leading eigenvector of $\widehat{Z}$, and define $\widehat{z}$ with each entry $\widehat{z}_j = u_j/|u_j|$. If $u_j = 0$ we can take $\widehat{z}_j = 1$. The statistical optimality of $\widehat{z}$ is established by the following result. Recall that for two vectors in $\mathbb{C}_1^n$, the definition of the loss $\ell(\widehat{z}, z^*)$ is given by (2).

Theorem 4: Assume $\frac{np}{\sigma^2} > c_1$ and $\frac{np}{\log n} > c_2$ for some sufficiently large constants $c_1, c_2 > 0$. Let $\widehat{Z} = \widehat{V}^{\mathrm{H}}\widehat{V}$ be a global maximizer of the SDP (13). Then,

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

with probability at least $1 - 2n^{-9} - \exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right)$ for some constant $C > 0$.

Compared with the minimax lower bound in Theorem 2, the SDP (13) is also minimax optimal for the estimation of the vector z^* in phase synchronization. Theorem 3 and Theorem 4 together establish Theorem 1.

III. Implications on Generalized Power Method and MLE

In this section, we show that the analysis of the SDP through Lemma 1 also leads to statistical optimality of the generalized power method (GPM) and the maximum likelihood estimator (MLE). We note that it has already been established by [6] that both GPM and MLE achieve the optimal error bound $(1 + o(1))\frac{\sigma^2}{2np}$ under the loss $\ell(\widehat{z}, z^*)$. By deriving the same results using the analysis of the SDP, we can unify the three proofs and thus form a coherent understanding of the three different methods.

The iteration of GPM of phase synchronization is

$$z_j^{(t)} = \begin{cases} \dfrac{\sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}z_k^{(t-1)}}{\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}z_k^{(t-1)}\right|}, \\ \qquad \text{if } \sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}z_k^{(t-1)} \neq 0, \\ z_j^{(t-1)}, \quad \text{if } \sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}z_k^{(t-1)} = 0. \end{cases} \tag{20}$$

The similarity between (20) and (15) is obvious. To make an explicit connection between the two iteration procedures, we can embed (20) into the space of (15). Let $e_1 \in \mathbb{C}^n$ be the first canonical vector with the first entry 1 and the remaining entries all 0. It is easy to check that as long as $V_j^{(t-1)} = \bar{z}_j^{(t-1)}e_1$ for all $j \in [n]$, we also have $V_j^{(t)} = \bar{z}_j^{(t)}e_1$ for all $j \in [n]$. This is because once the columns $V_1^{(t)}, \cdots, V_n^{(t)}$ lie in the same one-dimensional subspace for some t, the iteration (15) remains in this subspace. Thus, the formula (15) exactly describes the GPM iteration (20). In addition to the connection between (20) and (15), the two loss functions $\ell(V, z^*)$ and $\ell(z, z^*)$ are also equivalent. Under the condition that $V_j = \bar{z}_j e_1$ for all $j \in [n]$, we have

$$\ell(V, z^*) = \ell(z, z^*).$$

Therefore, Lemma 1 directly implies that

$$\ell(g(z), z^*) \le \delta_1 \ell(z, z^*) + (1 + \delta_2)\frac{\sigma^2}{2np}, \tag{21}$$

uniformly over all $z \in \mathbb{C}_1^n$ such that $\ell(z, z^*) < 1/16$ with high probability. The map $g : \mathbb{C}_1^n \to \mathbb{C}_1^n$ is defined so that (20) can be shorthanded by $z^{(t)} = g(z^{(t-1)})$.

From (21), we know that as long as $\ell(z^{(t-1)}, z^*) \le \gamma$ for some $\gamma < 1/16$, the next step of power iteration (20) satisfies

$$\ell(z^{(t)}, z^*) \le \delta_1 \ell(z^{(t-1)}, z^*) + (1 + \delta_2)\frac{\sigma^2}{2np}. \tag{22}$$

The condition $\ell(z^{(t-1)}, z^*) \le \gamma$ then implies $\ell(z^{(t)}, z^*) \le \delta_1\gamma + (1+\delta_2)\frac{\sigma^2}{2np}$. Given that $\frac{\sigma^2}{2np}$ is sufficiently small, we can always choose $\gamma < 1/16$ that satisfies $\frac{\sigma^2}{2np} \le \frac{\gamma}{2}$. Therefore, $\ell(z^{(t)}, z^*) \le \gamma$. Thus, a simple induction argument implies that (22) holds for all $t \ge 1$ as long as $\ell(z^{(0)}, z^*) \le \gamma$. The one-step iteration bound (22) immediately implies the linear convergence

$$\ell(z^{(t)}, z^*) \le \delta_1^t \ell(z^{(0)}, z^*) + \frac{1+\delta_2}{1-\delta_1}\frac{\sigma^2}{2np}, \tag{23}$$

for all $t \ge 1$. It has been shown by [6] that the initial error condition $\ell(z^{(0)}, z^*) \le \gamma < 1/16$ is satisfied by a simple eigenvector method. That is, $z_j^{(0)} = v_j/|v_j|$ with $v \in \mathbb{C}^n$ being the leading eigenvector of the matrix $A \circ Y$. Then, (23) implies $\ell(z^{(t)}, z^*) \le (1+o(1))\frac{\sigma^2}{2np}$ for all $t \ge \log\left(\frac{1}{\sigma^2}\right)$.

The optimality of the MLE can be derived from a similar embedding argument. Let $\widehat{z}$ be a global maximizer of (12). By the definition of $\widehat{z}$, its jth entry must satisfy

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

as long as $\sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}\widehat{z}_k \ne 0$. By letting $\widehat{V} = e_1\widehat{z}^{\mathrm{H}}$, it can be shown that the fixed-point equation $\widehat{V} = f(\widehat{V})$ holds. Given the equivalence of the loss $\ell(\widehat{V}, z^*) = \ell(\widehat{z}, z^*)$, as long as we can show a crude bound $\ell(\widehat{z}, z^*) \le \gamma < 1/16$ for the MLE, the inequality (19) holds and it can be written as

$$\ell(\widehat{z}, z^*) \le \delta_1\ell(\widehat{z}, z^*) + (1+\delta_2)\frac{\sigma^2}{2np},$$

which implies $\ell(\widehat{z}, z^*) \le \frac{1+\delta_2}{1-\delta_1}\frac{\sigma^2}{2np}$ after rearrangement. The crude bound $\ell(\widehat{z}, z^*) \le \gamma < 1/16$ can be easily established for the MLE using the argument in [6] or by a similar argument to the proof of Lemma 2, and thus we obtain the optimal error bound $\ell(\widehat{z}, z^*) \le (1+o(1))\frac{\sigma^2}{2np}$ for the MLE.

IV. SDP for $\mathbb{Z}_2$ Synchronization

In this section, we show our analysis of SDP can also be applied to $\mathbb{Z}_2$ synchronization and leads to a sharp exponential statistical error rate. Suppose we observe a random graph $A_{jk} \sim \text{Bernoulli}(p)$ independently for all $1 \le j < k \le n$. For each pair (j,k), we observe $Y_{jk} = z_j^*z_k^* + \sigma W_{jk}$ with $z_j^*, z_k^* \in \{-1,1\}$ and $W_{jk} \sim \mathcal{N}(0,1)$ whenever $A_{jk} = 1$. In $\mathbb{Z}_2$ synchronization, our goal is to estimate the binary vector $z^* \in \{-1,1\}^n$ from observations $\{A_{jk}\}_{1\le j<k\le n}$ and $\{A_{jk}Y_{jk}\}_{1\le j<k\le n}$. We organize the data into two matrices A and $A \circ Y$. Both the matrices A and Y are symmetric as we define $Y_{jk} = Y_{kj}$ and $A_{jk} = A_{kj}$ for all $1 \le k < j \le n$ and $Y_{jj} = A_{jj} = 0$ for all $j \in [n]$.

With slight abuse of notation, we consider the loss function

$$\ell(\widehat{z}, z) = \min_{a\in\{-1,1\}} \frac{1}{n}\sum_{j=1}^n |\widehat{z}_j - z_j a|^2,$$

for any $\widehat{z}, z \in \{-1,1\}^n$. Since $|\widehat{z}_j - z_j a|^2 = 4\mathbb{I}\{\widehat{z}_j \ne z_j a\}$, the loss $\ell(\widehat{z}, z)$ is also called the misclassification proportion in a clustering problem [11], [17]. We first present the minimax lower bound of $\mathbb{Z}_2$ synchronization under this loss function.

Theorem 5: Assume $\frac{np}{\sigma^2} > c_1$ and $\frac{np}{\log n} > c_2$ for some sufficiently large constants $c_1, c_2 > 0$. Then, we have

$$\inf_{\widehat{Z}\in\mathbb{R}^{n\times n}} \sup_{z\in\{-1,1\}^n} \mathbb{E}_z \frac{1}{n^2}\|\widehat{Z} - zz^{\mathrm{T}}\|_{\mathrm{F}}^2 \ge \exp\left(-(1+\delta)\frac{np}{2\sigma^2}\right),$$

$$\inf_{\widehat{z}\in\{-1,1\}^n} \sup_{z\in\{-1,1\}^n} \mathbb{E}_z \ell(\widehat{z}, z) \ge \exp\left(-(1+\delta)\frac{np}{2\sigma^2}\right),$$

where $\delta = C\sqrt{\frac{\log n + \sigma^2}{np}}$ for some constant $C > 0$.

When $p = 1$, the above result has been proved by [12], but the lower bound result for a general p is unknown in the literature. Compared with Theorem 2, the minimax lower bound for $\mathbb{Z}_2$ synchronization is an exponential function of the signal-to-noise ratio, a consequence of the discreteness of the problem.

To estimate $z^* \in \{-1,1\}^n$, the MLE is defined as the global maximizer of the following optimization problem

$$\max_{z\in\{-1,1\}^n} z^{\mathrm{T}}(A \circ Y)z. \tag{24}$$

Similar to (13), a convex relaxation of (24) leads to the following SDP,

$$\max_{Z=Z^{\mathrm{T}}\in\mathbb{R}^{n\times n}} \mathsf{Tr}((A\circ Y)Z) \text{ subject to } \operatorname{diag}(Z) = I_n \text{ and } Z \succeq 0. \tag{25}$$

The SDP for $\mathbb{Z}_2$ synchronization is almost in the exact form of (13). The only difference between (25) and (13) is that the optimization (25) is over real symmetric matrices and the optimization of (13) is over complex Hermitian matrices.

Our analysis of the SDP (25) for $\mathbb{Z}_2$ synchronization relies on a non-convex characterization that is similar to (14). For any Z that is a positive semi-definite real symmetric matrix, it admits a decomposition $Z = V^{\mathrm{T}}V$ for some $V \in \mathbb{R}^{n\times n}$. By writing the jth column of V as V_j, we can replace the constraint $\operatorname{diag}(Z) = I_n$ by $\|V_j\|^2 = 1$ for all $j \in [n]$. Then, an equivalent non-convex form of the SDP (25) is

$$\max_{V\in\mathbb{R}^{n\times n}} \mathsf{Tr}((A\circ Y)V^{\mathrm{T}}V) \text{ subject to } \|V_j\|^2 = 1 \text{ for all } j \in [n]. \tag{26}$$

We will study the solution of (26) using the following loss function,

$$\ell(\widehat{V}, z) = \min_{a\in\mathbb{R}^n:\|a\|^2=1} \frac{1}{n}\sum_{j=1}^n \|\widehat{V}_j - z_j a\|^2.$$

By the same argument that leads to (18), we know that if $\widehat{V}$ is a global maximizer of (26), it will satisfy the equation $\widehat{V} = f(\widehat{V})$, where $f : \mathbb{R}_1^{n\times n} \to \mathbb{R}_1^{n\times n}$ is a map such that the jth column of $f(\widehat{V})$ is given by

$$[f(\widehat{V})]_j = \begin{cases} \dfrac{\sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}\widehat{V}_k^{(t-1)}}{\left\|\sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}\widehat{V}_k^{(t-1)}\right\|}, \\ \qquad \text{if } \sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}\widehat{V}_k^{(t-1)} \ne 0, \\ \widehat{V}_j^{(t-1)}, \text{ if } \sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}\widehat{V}_k^{(t-1)} = 0. \end{cases}$$

Here, we use the notation $\mathbb{R}_1^{n\times n}$ for set of $n\times n$ real matrices whose columns all have unit norms. For each $j\in[n]$, define the random variable

$$U_j=\frac{\sigma}{(n-1)p}\sum_{k\in[n]\setminus\{j\}} z_k^* A_{jk}W_{jk}.$$

The following lemma characterizes the evolution of the loss $\ell(V,z^*)$ through the map f.

Lemma 3: Assume $\frac{np}{\sigma^2}\geq c_2$ and $\frac{np}{\log n}\geq c_2$ for some sufficiently large constants $c_1,c_2>0$. Then, for any $\gamma\in[0,1/16)$, we have

$$\mathbb{P}\Bigg(\ell(f(V),z^*)\leq\frac{1}{2}\ell(V,z^*)+\frac{4}{n}\sum_{j=1}^n\mathbb{I}\{|U_j|>1-\delta\}$$
$$\text{for all } V\in\mathbb{R}_1^{n\times n} \text{ such that } \ell(V,z^*)\leq\gamma\Bigg)\geq 1-2n^{-9},$$

where $\delta=C\left(\sqrt{\gamma}+\sqrt{\frac{\log n+\sigma^2}{np}}\right)$ for some constant $C>0$.

Lemma 3 immediately implies that for any $\widehat{V}$ that satisfies the fixed-point equation $\widehat{V}=f(\widehat{V})$ and the crude error bound $\ell(\widehat{V},z^*)\leq\gamma<1/16$, we have

$$\ell(\widehat{V},z^*)\leq\frac{8}{n}\sum_{j=1}^n\mathbb{I}\{|U_j|>1-\delta\},\tag{27}$$

with high probability. The property of the random variable $\frac{8}{n}\sum_{j=1}^n\mathbb{I}\{|U_j|>1-\delta\}$ can be easily analyzed, and we present the following lemma.

Lemma 4: Assume $\frac{np}{\sigma^2}\geq c_1$ and $\frac{np}{\log n}\geq c_2$ for some sufficiently large constants $c_1,c_2>0$. Then, for any $\delta\in(0,1)$, we have

$$\frac{8}{n}\sum_{j=1}^n\mathbb{I}\{|U_j|>1-\delta\}\leq\exp\left(-(1-\delta')\frac{np}{2\sigma^2}\right),$$

with probability at least $1-\exp\left(-\sqrt{\frac{np}{\sigma^2}}\right)-n^{-9}$, where $\delta'=C\left(\delta+\sqrt{\frac{\log n}{np}}\right)$ for some constant $C>0$. If we additionally assume $(1-\delta')\frac{np}{2\sigma^2}>\log n$, then

$$\frac{8}{n}\sum_{j=1}^n\mathbb{I}\{|U_j|>1-\delta\}=0,$$

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

We also need a lemma to establish a crude error bound for $\ell(\widehat{V},z^*)$.

Lemma 5: Assume $\frac{np}{\log n}\geq c$ for some constant $c>0$. Let $\widehat{Z}=\widehat{V}^{\mathrm{T}}\widehat{V}$ be a global maximizer of the SDP (25). Then, there exits some constant $C>0$ such that

$$\ell(\widehat{V},z^*)\leq C\sqrt{\frac{\sigma^2+1}{np}},$$

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

The results of Lemma 3, Lemma 4 and Lemma 5 immediately imply the statistical optimality of the SDP (25).

Theorem 6: Assume $\frac{np}{\sigma^2}\geq c_1$ and $\frac{np}{\log n}\geq c_2$ for some sufficiently large constants $c_1,c_2>0$. Let $\widehat{Z}=\widehat{V}^{\mathrm{T}}\widehat{V}$ be a global maximizer of the SDP (25) and $u\in\mathbb{R}^n$ be the leading eigenvector of $\widehat{Z}$. Define $\widehat{z}$ with each entry $\widehat{z}_j=u_j/|u_j|$. If $u_j=0$ we can take $\widehat{z}_j=1$. Then, there exists some $\delta=C\left(\frac{\log n+\sigma^2}{np}\right)^{\frac{1}{4}}$ for some constant $C>0$, such that

$$\begin{aligned}\ell(\widehat{V},z^*) &\leq \exp\left(-(1-\delta)\frac{np}{2\sigma^2}\right),\\ \frac{1}{n^2}\|\widehat{Z}-z^*z^{*\mathrm{T}}\|_{\mathrm{F}}^2 &\leq \exp\left(-(1-\delta)\frac{np}{2\sigma^2}\right),\\ \ell(\widehat{z},z^*) &\leq \exp\left(-(1-\delta)\frac{np}{2\sigma^2}\right),\end{aligned}$$

with probability at least $1-\exp\left(-\sqrt{\frac{np}{\sigma^2}}\right)-2n^{-9}$. Moreover, if we additionally assume $\sigma^2<(1-\epsilon)\frac{np}{2\log n}$ for some arbitrarily small constant $\epsilon>0$, the SDP solution $\widehat{Z}$ is a rank-one matrix that satisfies $\widehat{Z}=z^*z^{*\mathrm{T}}$ with probability at least $1-\exp\left(-\sqrt{\frac{np}{\sigma^2}}\right)-2n^{-9}$.

While the first conclusion of the theorem is a direct consequence of Lemma 3, the second conclusion can be derived from the inequality

$$\frac{1}{n^2}\|\widehat{V}^{\mathrm{T}}\widehat{V}-z^*z^{*\mathrm{T}}\|_{\mathrm{F}}^2\leq 2\ell(\widehat{V},z^*),$$

which is established by Lemma 11 in Section V-A. The result for the loss $\ell(\widehat{z},z^*)$ is resulted from a matrix perturbation bound [18].

Theorem 6 has established the minimax optimality of the SDP (25) for $\mathbb{Z}_2$ synchronization in view of the matching lower bound results in Theorem 5. The special case $p=1$ recovers the results of [12]. Moreover, under the condition $\sigma^2<(1-\epsilon)\frac{np}{2\log n}$, we show that the SDP solution $\widehat{Z}$ is exactly rank-one and therefore rounding through the leading eigenvector is not needed. This result generalizes the exact recovery threshold of $\mathbb{Z}_2$ synchronization when $p=1$ [3], [19], [20]. The phenomenon that SDP can achieve exact recovery has also been revealed in community detection under stochastic block models [21]–[26].

We shall compare Theorem 6 to Theorem 3 and Theorem 4. Though the two SDPs (25) and (13) have the same type of constraints, the difference of the domain implies two types of convergence rates $\exp\left(-(1-o(1))\frac{np}{2\sigma^2}\right)$ and $(1+o(1))\frac{\sigma^2}{2np}$. It is quite surprising that the SDP (25), a continuous optimization problem, is able to achieve an exponential rate, which is typical for a discrete problem. The adaptation of the SDP (25) to the discrete structure is a consequence of the fact that both (25) and (26) are optimization problems over $\mathbb{R}^{n\times n}$. We make this effect explicit by bounding the statistical error by the random variable $\frac{8}{n}\sum_{j=1}^n\mathbb{I}\{|U_j|>1-\delta\}$ in Lemma 3.

To close this section, we briefly discuss the implications of Lemma 3 on the MLE (24) and the generalized power method defined by the iteration procedure

$$z_j^{(t)}=\begin{cases}\frac{\sum_{k\in[n]\setminus\{j\}}A_{jk}Y_{jk}z_k^{(t-1)}}{\left|\sum_{k\in[n]\setminus\{j\}}A_{jk}Y_{jk}z_k^{(t-1)}\right|},\\ \qquad\text{if } \sum_{k\in[n]\setminus\{j\}}A_{jk}Y_{jk}z_k^{(t-1)}\neq 0,\\ z_j^{(t-1)},\ \text{if } \sum_{k\in[n]\setminus\{j\}}A_{jk}Y_{jk}z_k^{(t-1)}=0.\end{cases}\tag{28}$$

We note that the iteration (28) is real-valued so that we always have $z_j^{(t)}\in\{-1,1\}$, which makes it different from (20). The statistical optimality of the generalized power method (28)

has been established by [11] for $\mathbb{Z}_2$ synchronization when $p = 1$. Following the same argument in Section III, we can embed both MLE and GPM into $\mathbb{R}_1^{n\times n}$, and thus Lemma 3 also implies that both MLE and GPM achieve the optimal rate $\exp\left(-(1-o(1))\frac{np}{2\sigma^2}\right)$ for a general p as well. Just as what we have for phase synchronization, the analyses of MLE, GPM, and SDP for $\mathbb{Z}_2$ synchronization are all based on Lemma 3, and thus we have unified the three different methods from an iterative algorithm perspective.

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 lemmas of the SDP analyses, Lemma 1 and Lemma 3, are proved in Section V-B and Section V-C, respectively. We then prove the main results including Theorem 3, Theorem 4 and Theorem 6 in Section V-D. Theorem 5 is proved in Section V-E. Finally, the proofs of Lemma 2, Lemma 5 and Lemma 4 are given in Section V-F.

A. Some Auxiliary Lemmas

Lemma 6: Assume $\frac{np}{\log n} \to \infty$. Then, there exists a constant $C > 0$, such that

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

and

$$\|A - \mathbb{E}A\|_{\mathrm{op}} \le C\sqrt{np},$$

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 [27]. □

The following result is essentially Corollary 3.11 of [28]. The specific form that we need is from Lemma 5.2 in [6].

Lemma 7 (Corollary 3.11 of [28]): Assume $\frac{np}{\log n} \to \infty$. Then, there exists a constant $C > 0$, such that

$$\|A \circ W\|_{\mathrm{op}} \le C\sqrt{np},$$

with probability at least $1 - n^{-10}$. The result holds for both complex W in Section II and real W in Section IV.

Lemma 8 (Lemma 5.3 of [6]): Assume $\frac{np}{\log n} > c$ for some sufficiently large constant $c > 0$. Consider independent random variables $X_{jk} \sim \mathcal{N}(0,1)$ for $1 \le j < k \le n$. Assume $X_{kj} = X_{jk}$ for $1 \le j < k \le n$. Then, we have

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

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

Lemma 9 (Lemma 13 of [29]): Consider independent random variables $X_j \sim \mathcal{N}(0,1)$ and $E_j \sim \text{Bernoulli}(p)$. Then,

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

for any $t > 0$.

Lemma 10: The following three statements hold:

1) For any $x, y \in \mathbb{C}^n$ such that $\|y\| = 1$ and $\mathrm{Re}(y^{\mathrm{H}}x) > 0$, we have

$$\left\| \frac{x}{\|x\|} - y \right\|^2 \le \frac{\|(I_n - yy^{\mathrm{H}})x\|^2 + |\mathrm{Im}(y^{\mathrm{H}}x)|^2}{|\mathrm{Re}(y^{\mathrm{H}}x)|^2}.$$

2) For any $x, y \in \mathbb{R}^n$ such that $\|y\| = 1$ and $y^{\mathrm{T}}x > 0$, we have

$$\left\| \frac{x}{\|x\|} - y \right\|^2 \le \frac{\|(I_n - yy^{\mathrm{T}})x\|^2}{|y^{\mathrm{T}}x|^2}.$$

3) For any $x \in \mathbb{C}$ such that $\mathrm{Re}(x) > 0$, we have

$$\left| \frac{x}{|x|} - 1 \right|^2 \le \frac{|\mathrm{Im}(x)|^2}{|\mathrm{Re}(x)|^2}.$$

Proof: It is easy to see that the last two statements are special cases of the first one. Thus, we only need to prove the first statement. Note that

$$\begin{aligned} \left\| \frac{x}{\|x\|} - y \right\|^2 &= \left\| \frac{(I_n - yy^{\mathrm{H}})x + (y^{\mathrm{H}}x)y}{\|x\|} - y \right\|^2 \\ &= \frac{\|(I_n - yy^{\mathrm{H}})x\|^2 + |y^{\mathrm{H}}x - \|x\||^2}{\|x\|^2} \\ &= \frac{b^2 + \left(a - \sqrt{a^2+b^2}\right)^2}{a^2+b^2}, \end{aligned}$$

where a, b are defined as $a = \mathrm{Re}(y^{\mathrm{H}}x) > 0$ and $b = \sqrt{|\mathrm{Im}(y^{\mathrm{H}}x)|^2 + \|(I_n - yy^{\mathrm{H}})x\|^2}$. Since

$$\frac{b^2 + \left(a - \sqrt{a^2+b^2}\right)^2}{a^2+b^2} = \frac{2b^2}{a^2+b^2+a\sqrt{a^2+b^2}} \le \frac{b^2}{a^2},$$

the proof is complete. □

Lemma 11: For any $V = (V_1, \cdots, V_n) \in \mathbb{C}^{n\times n}$ and any $z \in \mathbb{C}_1^n$ such that $\|V_j\| = 1$ for all $j \in [n]$, we have

$$n^{-2}\|V^{\mathrm{H}}V - zz^{\mathrm{H}}\|_{\mathrm{F}}^2 \le 2\ell(V, z).$$

For any $V = (V_1, \cdots, V_n) \in \mathbb{R}^{n\times n}$ and any $z \in \{-1, 1\}^n$ such that $\|V_j\| = 1$ for all $j \in [n]$, we have

$$n^{-2}\|V^{\mathrm{T}}V - zz^{\mathrm{T}}\|_{\mathrm{F}}^2 \le 2\ell(V, z).$$

Proof: We only prove the complex version of the inequality. The real version follows the same argument. By definition, we have

$$\begin{aligned} \ell(V,z) = \Bigg(2 - \max_{a\in\mathbb{C}^n:\|a\|^2=1} \Bigg(a^{\mathrm{H}} \left(\frac{1}{n}\sum_{j=1}^{n} z_jV_j \right) \\ + \left(\frac{1}{n}\sum_{j=1}^{n} z_jV_j \right)^{\mathrm{H}} a \Bigg) \Bigg) = 2\left(1 - \left\| \frac{1}{n}\sum_{j=1}^{n} z_jV_j \right\| \right). \end{aligned}$$

Then,

$$\begin{aligned} & n^{-2}\|V^{\mathrm{H}}V - zz^{\mathrm{H}}\|_{\mathrm{F}}^2 \\ = \;& \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}|V_j^{\mathrm{H}}V_l - z_j\bar{z}_l|^2 \\ \le \;& \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}\left(2 - V_j^{\mathrm{H}}V_l\bar{z}_j z_l - V_l^{\mathrm{H}}V_j z_j\bar{z}_l\right) \\ = \;& 2\left(1 - \left\|\frac{1}{n}\sum_{j=1}^{n} z_j V_j\right\|^2\right). \end{aligned}$$

Therefore, $n^{-2}\|V^{\mathrm{H}}V - zz^{\mathrm{H}}\|_{\mathrm{F}}^2 \le \ell(V,z)\left(2 - \frac{1}{2}\ell(V,z)\right) \le 2\ell(V,z)$, and the proof is complete. □

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 $\ell(f(V), z^*)$, 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 6, Lemma 7, and Lemma 8, we know that

$$\min_{j\in[n]}\sum_{k\in[n]\setminus\{j\}} A_{jk} \ge (n-1)p - C\sqrt{np\log n}, \quad (29)$$

$$\max_{j\in[n]}\sum_{k\in[n]\setminus\{j\}} A_{jk} \le (n-1)p + C\sqrt{np\log n}, \quad (30)$$

$$\|A - \mathbb{E}A\|_{\mathrm{op}} \le C\sqrt{np}, \quad (31)$$

$$\|A\circ W\|_{\mathrm{op}} \le C\sqrt{np}, \quad (32)$$

$$\sum_{j=1}^{n}\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\mathrm{Im}(\bar{W}_{jk}\bar{z}_k^* z_j^*)\right|^2 \le \frac{n^2p}{2}\left(1 + C\sqrt{\frac{\log n}{n}}\right), \quad (33)$$

$$\sum_{j=1}^{n}\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\mathrm{Re}(\bar{W}_{jk}\bar{z}_k^* z_j^*)\right|^2 \le \frac{n^2p}{2}\left(1 + C\sqrt{\frac{\log n}{n}}\right), \quad (34)$$

all hold with probability at least $1 - n^{-9}$ for some constant $C > 0$. To establish (33)-(34), note that $\sqrt{2}\mathrm{Im}(\bar{W}_{jk}\bar{z}_k^* z_j^*), \sqrt{2}\mathrm{Re}(\bar{W}_{jk}\bar{z}_k^* z_j^*)$ are all independently standard normally distributed for $1 \le j < k \le n$. We also have $-\mathrm{Im}(\bar{W}_{jk}\bar{z}_k^* z_j^*) = \mathrm{Im}(W_{jk}z_k^*\bar{z}_j^*) = \mathrm{Im}(\bar{W}_{kj}\bar{z}_j^* z_k^*)$ and $\mathrm{Re}(\bar{W}_{jk}\bar{z}_k^* z_j^*) = \mathrm{Re}(W_{jk}z_k^*\bar{z}_j^*) = \mathrm{Re}(\bar{W}_{kj}\bar{z}_j^* z_k^*)$ for any $1 \le j < k \le n$.

In addition to (29)-(34), we need another high-probability inequality. For a sufficiently small ρ such that $\frac{\rho^2 np}{\sigma^2}$ is sufficiently large, we want to upper bound the random variable $\sum_{j=1}^{n}\mathbb{I}\left\{\frac{2\sigma}{np}\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}\bar{z}_k^*\right| > \rho\right\}$. The existence of such ρ is guaranteed by the condition $\frac{np}{\sigma^2}$ is sufficiently large, and the specific choice will be given later. We first bound its expectation by Lemma 9,

$$\begin{aligned} & \sum_{j=1}^{n}\mathbb{P}\left\{\frac{2\sigma}{np}\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}\bar{z}_k^*\right| > \rho\right\} \\ \le \;& \sum_{j=1}^{n}\mathbb{P}\left\{\frac{2\sigma}{np}\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\mathrm{Re}(\bar{W}_{jk}\bar{z}_k^*)\right| > \frac{\rho}{2}\right\} \\ & + \sum_{j=1}^{n}\mathbb{P}\left\{\frac{2\sigma}{np}\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\mathrm{Im}(\bar{W}_{jk}\bar{z}_k^*)\right| > \frac{\rho}{2}\right\} \\ \le \;& 4n\exp\left(-\frac{\rho^2 np}{256\sigma^2}\right) + 4n\exp\left(-\frac{\rho np}{8\sigma}\right). \end{aligned}$$

By Markov inequality, we have

$$\begin{aligned} & \sum_{j=1}^{n}\mathbb{I}\left\{\frac{2\sigma}{np}\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}\bar{z}_k^*\right| > \rho\right\} \\ & \le \frac{4\sigma^2}{\rho^2 p}\exp\left(-\frac{1}{16}\sqrt{\frac{\rho^2 np}{\sigma^2}}\right), \end{aligned} \quad (35)$$

with probability at least

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

Finally, we conclude that the events (29)-(35) 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 $V \in \mathbb{C}_1^{n\times n}$ such that $\ell(V, z^*) \le \gamma$, we can define $\widetilde{V} \in \mathbb{C}^{n\times n}$ such that

$$\widetilde{V}_j = \frac{\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{Y}_{jk}V_k}{\sum_{k\in[n]\setminus\{j\}} A_{jk}},$$

for each $j \in [n]$. Denote $\widehat{V} = f(V)$ then $\widehat{V}_j = \widetilde{V}_j/\|\widetilde{V}_j\|$ for each coordinate such that $\widetilde{V}_j \ne 0$.

The condition $\ell(V, z^*) \le \gamma$ implies there exists some $b \in \mathbb{C}^n$ such that $\|b\| = 1$ and $n\ell(V, z^*) = \|V - bz^{*\mathrm{H}}\|_{\mathrm{F}}^2 \le \gamma n$. By direct calculation, we can write

$$\begin{aligned} z_j^*\widetilde{V}_j = b + & \frac{\sum_{k\in[n]\setminus\{j\}} A_{jk}z_k^*(V_k - \bar{z}_k^* b)}{\sum_{k\in[n]\setminus\{j\}} A_{jk}} \\ & + \frac{\sigma z_j^* b\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}\bar{z}_k^*}{\sum_{k\in[n]\setminus\{j\}} A_{jk}} \\ & + \frac{\sigma z_j^*\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}(V_k - \bar{z}_k^* b)}{\sum_{k\in[n]\setminus\{j\}} A_{jk}} \end{aligned}$$

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

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

$$
\begin{aligned}
z_j^* a^{\mathrm{H}} \widetilde{V}_j &= \|a_0\| - \frac{1}{n-1} z_j^* a^{\mathrm{H}}(V_j - \bar{z}_j^* b) \\
&\quad + a^{\mathrm{H}} F_j + a^{\mathrm{H}} b G_j + a^{\mathrm{H}} H_j, \qquad (36) \\
\|(I_n - aa^{\mathrm{H}})\widetilde{V}_j\| &\le \frac{1}{n-1}\|V_j - \bar{z}_j^* b\| + \|F_j\| \\
&\quad + \|(I_n - aa^{\mathrm{H}})b\|\,|G_j| + \|H_j\|, \qquad (37)
\end{aligned}
$$

where

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

By Lemma 10, we have the bound

$$
\|\widehat{V}_j - \bar{z}_j^* a\|^2 \le \frac{\|(I_n - aa^{\mathrm{H}})\widetilde{V}_j\|^2 + |\mathrm{Im}(z_j^* a^{\mathrm{H}}\widetilde{V}_j)|^2}{|\mathrm{Re}(z_j^* a^{\mathrm{H}}\widetilde{V}_j)|^2}, \qquad (38)
$$

whenever $\mathrm{Re}(z_j^* a^{\mathrm{H}}\widetilde{V}_j) > 0$ holds. Since

$$
\begin{aligned}
\|a_0 - b\| &= \left\| \frac{1}{n-1}\sum_{k=1}^{n} z_k^*(V_k - \bar{z}_k^* b) \right\| \\
&\le \frac{1}{n-1}\sqrt{n}\|V - bz^{*\mathrm{H}}\|_{\mathrm{F}} \\
&\le \frac{n}{n-1}\sqrt{\gamma} \le 2\sqrt{\gamma},
\end{aligned}
$$

we have $\|a - b\| \le 2\|a_0 - b\| \le 4\sqrt{\gamma}$. Therefore,

$$
\begin{aligned}
\|a_0\| &\ge \|b\| - \|a_0 - b\| \ge 1 - 2\sqrt{\gamma}, \qquad (39) \\
|a^{\mathrm{H}} b - 1| &= |(a-b)^{\mathrm{H}} b| \le \|a - b\| \le 4\sqrt{\gamma}, \qquad (40) \\
\|(I_n - aa^{\mathrm{H}})b\| &\le \|a - b\| + |a^{\mathrm{H}} b - 1| \le 8\sqrt{\gamma}. \qquad (41)
\end{aligned}
$$

We also have

$$
\left| \frac{1}{n-1} z_j^* a^{\mathrm{H}}(V_j - \bar{z}_j^* b) \right| \le \frac{1}{n-1}\|V_j - \bar{z}_j^* b\| \le \frac{\sqrt{\gamma n}}{n-1}. \qquad (42)
$$

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

$$
\mathrm{Re}\left(z_j^* a^{\mathrm{H}} \widetilde{V}_j \right) \ge 1 - 2\sqrt{\gamma} - \frac{\sqrt{\gamma n}}{n-1} - 3\rho \ge 1 - 3(\sqrt{\gamma} + \rho) > 0, \qquad (43)
$$

where we have used (36), (39), and (42), and the last inequality is due to the assumption that $\gamma < 1/16$ and ρ is sufficiently small. Hence, the event $\{\widetilde{V}_j = 0\}$ is included in the event $\{\|F_j\| \vee |G_j| \vee \|H_j\| > \rho\}$.

By (38), we obtain the bound

$$
\begin{aligned}
&\|\widehat{V}_j - \bar{z}_j^* a\|^2 \\
&\le \frac{\|(I_n - aa^{\mathrm{H}})\widetilde{V}_j\|^2 + |\mathrm{Im}(z_j^* a^{\mathrm{H}}\widetilde{V}_j)|^2}{|\mathrm{Re}(z_j^* a^{\mathrm{H}}\widetilde{V}_j)|^2} \\
&\quad \times \mathbb{I}\{\|F_j\| \vee |G_j| \vee \|H_j\| \le \rho\} \\
&\quad + 4\mathbb{I}\{\|F_j\| \vee |G_j| \vee \|H_j\| > \rho\} \\
&\le \frac{\left(\frac{1}{n-1}\|V_j - \bar{z}_j^* b\| + \|F_j\| + 8\sqrt{\gamma}|G_j| + \|H_j\| \right)^2}{(1 - 3(\sqrt{\gamma} + \rho))^2} \\
&\quad + \frac{\left(\frac{1}{n-1}\|V_j - \bar{z}_j^* b\| + \|F_j\| + |\mathrm{Im}(a^{\mathrm{H}} b G_j)| + \|H_j\| \right)^2}{(1 - 3(\sqrt{\gamma} + \rho))^2} \\
&\quad + 4\mathbb{I}\{\|F_j\| > \rho\} + 4\mathbb{I}\{|G_j| > \rho\} + 4\mathbb{I}\{\|H_j\| > \rho\} \\
&\le \frac{(1+\eta)|\mathrm{Im}(a^{\mathrm{H}} b G_j)|^2 + 256\gamma|G_j|^2}{(1 - 3(\sqrt{\gamma} + \rho))^2} \\
&\quad + \frac{(7 + 4\eta^{-1})\left(\frac{1}{(n-1)^2}\|V_j - \bar{z}_j^* b\|^2 + \|F_j\|^2 + \|H_j\|^2 \right)}{(1 - 3(\sqrt{\gamma} + \rho))^2} \\
&\quad + 4\mathbb{I}\{\|F_j\| > \rho\} + 4\mathbb{I}\{|G_j| > \rho\} + 4\mathbb{I}\{\|H_j\| > \rho\},
\end{aligned}
$$

for some η to be specified later. The last inequality above is due to Jensen's inequality.

c) Step 3: Analysis of each error term: Next, we will analyze the error terms F_j, H_j and G_j separately. By triangle inequality, (29) and (30), we have

$$
\begin{aligned}
&\|F_j\| \\
&\le \left\| \frac{\sum_{k\in[n]\setminus\{j\}} (A_{jk} - p) z_k^*(V_k - \bar{z}_k^* b)}{\sum_{k\in[n]\setminus\{j\}} A_{jk}} \right\| \\
&+ \left\| p \sum_{k\in[n]\setminus\{j\}} z_k^*(V_k - \bar{z}_k^* b) \right\| \left| \frac{1}{\sum_{k\in[n]\setminus\{j\}} A_{jk}} - \frac{1}{(n-1)p} \right| \\
&\le \frac{2}{np} \left\| \sum_{k\in[n]\setminus\{j\}} (A_{jk} - p) z_k^*(V_k - \bar{z}_k^* b) \right\| \\
&+ p\sqrt{n}\|V - bz^{*\mathrm{H}}\|_{\mathrm{F}} \frac{2\left| \sum_{k\in[n]\setminus\{j\}} (A_{jk} - p) \right|}{n^2 p^2} \\
&\le \frac{2}{np} \left\| \sum_{k\in[n]\setminus\{j\}} (A_{jk} - p) z_k^*(V_k - \bar{z}_k^* b) \right\| \\
&+ C_1 \frac{\sqrt{p \log n}}{np}\|V - bz^{*\mathrm{H}}\|_{\mathrm{F}}.
\end{aligned}
$$

Using (31), we have

$$\begin{aligned}\sum_{j=1}^n \|F_j\|^2 &\le \frac{8}{n^2p^2}\sum_{j=1}^n \left\| \sum_{k\in[n]\setminus\{j\}} (A_{jk}-p) z_k^*(V_k - \bar{z}_k^* b)\right\| \\ &\quad + 2C_1^2 \frac{\log n}{np}\|V - bz^{*\mathrm{H}}\|_{\mathrm{F}}^2 \\ &\le \frac{8}{n^2p^2}\|A-\mathbb{E}A\|_{\mathrm{op}}^2 \|V-bz^{*\mathrm{H}}\|_{\mathrm{F}}^2 \\ &\quad + 2C_1^2\frac{\log n}{np}\|V-bz^{*\mathrm{H}}\|_{\mathrm{F}}^2 \\ &\le C_2 \frac{\log n}{np}\|V-bz^{*\mathrm{H}}\|_{\mathrm{F}}^2. \end{aligned} \tag{44}$$

The above bound also implies

$$\sum_{j=1}^n \mathbb{I}\{\|F_j\| > \rho\} \le \rho^{-2}\sum_{j=1}^n \|F_j\|^2 \le \frac{C_2}{\rho^2}\frac{\log n}{np}\|V-bz^{*\mathrm{H}}\|_{\mathrm{F}}^2.$$

Similarly, we can also bound the error terms that depend on H_j. By (29) and (32), we have

$$\begin{aligned}\sum_{j=1}^n \|H_j\|^2 &\le \frac{2\sigma^2}{n^2p^2}\sum_{j=1}^n \left\| \sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}(V_k - \bar{z}_k^* b)\right\|^2 \\ &= \frac{2\sigma^2}{n^2p^2}\|(V-bz^{*\mathrm{H}})(A\circ W)^{\mathrm{H}}\|_{\mathrm{F}}^2 \\ &\le \frac{2\sigma^2}{n^2p^2}\|A\circ W\|_{\mathrm{op}}^2\|V-bz^{*\mathrm{H}}\|_{\mathrm{F}}^2 \\ &\le C_3\frac{\sigma^2}{np}\|V-bz^{*\mathrm{H}}\|_{\mathrm{F}}^2, \end{aligned} \tag{45}$$

and thus

$$\sum_{j=1}^n \mathbb{I}\{\|H_j\| > \rho\} \le \rho^{-2}\sum_{j=1}^n \|H_j\|^2 \le \frac{C_3}{\rho^2}\frac{\sigma^2}{np}\|V-bz^{*\mathrm{H}}\|_{\mathrm{F}}^2.$$

For the contribution of G_j, we use (29) and (35), and have

$$\begin{aligned}&\sum_{j=1}^n \mathbb{I}\{|G_j| > \rho\} \\ &\le \sum_{j=1}^n \mathbb{I}\left\{\frac{2\sigma}{np}\left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}\bar{z}_k^*\right| > \rho\right\} \\ &\le \frac{4\sigma^2}{\rho^2 p}\exp\left(-\frac{1}{16}\sqrt{\frac{\rho^2 np}{\sigma^2}}\right). \end{aligned} \tag{46}$$

Next, we study the main error term $|\mathrm{Im}(a^{\mathrm{H}}bG_j)|^2$. By (29), we have

$$\begin{aligned}&\sum_{j=1}^n |\mathrm{Im}(a^{\mathrm{H}}bG_j)|^2 \\ &\le \left(1+C_4\sqrt{\frac{\log n}{np}}\right)^2 \frac{\sigma^2}{n^2p^2}\sum_{j=1}^n \\ &\quad \left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\mathrm{Im}(\bar{W}_{jk}z_j^*\bar{z}_k^* a^{\mathrm{H}}b)\right|^2 \\ &\le (1+\eta)\left(1+C_4\sqrt{\frac{\log n}{np}}\right)^2 \frac{\sigma^2}{n^2p^2}\sum_{j=1}^n \\ &\quad \left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\mathrm{Im}(\bar{W}_{jk}z_j^*\bar{z}_k^*)\right|^2 \\ &\quad + (1+\eta^{-1})\left(1+C_4\sqrt{\frac{\log n}{np}}\right)^2 \frac{\sigma^2}{n^2p^2}\sum_{j=1}^n \\ &\quad \left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\mathrm{Re}(\bar{W}_{jk}z_j^*\bar{z}_k^*)\right|^2 |\mathrm{Im}(a^{\mathrm{H}}b)|^2. \end{aligned}$$

By (40), we have

$$|\mathrm{Im}(a^{\mathrm{H}}b)| = |\mathrm{Im}(a^{\mathrm{H}}b-1)| \le |a^{\mathrm{H}}b-1| \le 4\sqrt{\gamma}.$$

Together with (33) and (34), we have

$$\sum_{j=1}^n |\mathrm{Im}(a^{\mathrm{H}}bG_j)|^2 \le \left(1+C_5\left(\eta+\eta^{-1}\gamma+\sqrt{\frac{\log n}{np}}\right)\right)\frac{\sigma^2}{2p}. \tag{47}$$

We also have

$$\sum_{j=1}^n |G_j|^2 \le \frac{2\sigma^2}{n^2p^2}\sum_{j=1}^n \left|\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}z_j^*\bar{z}_k^*\right|^2 \le C_6\frac{\sigma^2}{p}, \tag{48}$$

by (33) and (34).

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

$$\begin{aligned}&n\ell(\widehat{V}, z^*) \\ &\le \sum_{j=1}^n \|\widehat{V}_j - \bar{z}_j^* a\|^2 \\ &\le \left(1+C_7\left(\rho+\eta+\sqrt{\gamma}+\eta^{-1}\gamma+\sqrt{\frac{\log n}{np}}\right)\right)\frac{\sigma^2}{2p} \\ &\quad + \frac{16\sigma^2}{\rho^2 p}\exp\left(-\frac{1}{16}\sqrt{\frac{\rho^2np}{\sigma^2}}\right) \\ &\quad + C_7\left(\eta^{-1}+\rho^{-2}\right)\frac{\log n+\sigma^2}{np} n\ell(V, z^*). \end{aligned}$$

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

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

Therefore, we have

$$\ell(\widehat{V}, z^*) \le \left(1 + C_8\left(\gamma^2 + \frac{\log n + \sigma^2}{np}\right)^{1/4}\right)\frac{\sigma^2}{2np} + C_8\sqrt{\frac{\log n + \sigma^2}{np}}\ell(V, z^*).$$

Since the above inequality is derived from the conditions (29)-(35) and $\ell(V, z^*) \le \gamma$, it holds uniformly over all $V \in \mathbb{C}_1^{n\times n}$ such that $\ell(V, z^*) \le \gamma$ 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 3

Similar to the proof of Lemma 1, we organize the proof of Lemma 3 into four steps.

a) Step 1: Some high-probability events: We already know that (29), (30) and (31) hold with probability at least $1 - n^{-9}$. We also have

$$\|A \circ W\|_{\mathrm{op}} \le C\sqrt{np}, \tag{49}$$

with probability at least $1 - n^{-9}$ by Lemma 7. Note that the matrix W in (49) is real-valued, compared with the complex version of the bound (32). Another high-probability event we need is for the random variable $\sum_{j=1}^n \left|\sum_{k\in[n]\setminus\{j\}} A_{jk}W_{jk}z_j^*z_k^*\right|^2$. By Lemma 8 and with a similar analysis that leads to (34), we can conclude that with probability at least $1 - n^{-9}$,

$$\sum_{j=1}^n \left|\sum_{k\in[n]\setminus\{j\}} A_{jk}W_{jk}z_j^*z_k^*\right|^2 \le n^2p\left(1 + C\sqrt{\frac{\log n}{n}}\right). \tag{50}$$

In the end, we conclude that the events (29), (30), (31), (49) and (50) hold simultaneously with probability at least $1 - 2n^{-9}$.

b) Step 2: Error decomposition: For any $V \in \mathbb{R}_1^{n\times n}$ such that $\ell(V, z^*) \le \gamma$, we can write $\widehat{V} = f(V)$ with each column $\widehat{V}_j = \widetilde{V}_j/\|\widetilde{V}_j\|$, where

$$\widetilde{V}_j = \frac{\sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}V_k}{\sum_{k\in[n]\setminus\{j\}} A_{jk}}.$$

The condition $\ell(V, z^*) \le \gamma$ implies there exists some $b \in \mathbb{R}^n$ such that $\|b\| = 1$ and $n\ell(V, z^*) = \|V - bz^{*\mathrm{T}}\|_{\mathrm{F}}^2 \le \gamma n$. By direct calculation, we can write

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

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

$$\begin{aligned} z_j^*a^{\mathrm{T}}\widetilde{V}_j &= \|a_0\| - \frac{1}{n-1}z_j^*a^{\mathrm{T}}(V_j - z_j^*b) + a^{\mathrm{T}}F_j \\ &\quad + a^{\mathrm{T}}bG_j + a^{\mathrm{T}}H_j, \\ \|(I_n - aa^{\mathrm{T}})\widetilde{V}_j\| &\le \frac{1}{n-1}\|V_j - z_j^*b\| + \|F_j\| \\ &\quad + \|(I_n - aa^{\mathrm{T}})b\||G_j| + \|H_j\|, \end{aligned}$$

where

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

By Lemma 10, we have the bound

$$\|\widehat{V}_j - z_j^*a\|^2 \le \frac{\|(I_n - aa^{\mathrm{T}})\widetilde{V}_j\|^2}{|z_j^*a^{\mathrm{T}}\widetilde{V}_j|^2}, \tag{51}$$

whenever $z_j^*a^{\mathrm{T}}\widetilde{V}_j > 0$ holds. Since

$$\begin{aligned} \|a_0 - b\| &= \left\|\frac{1}{n-1}\sum_{k=1}^n z_k^*(V_k - z_k^*b)\right\| \\ &\le \frac{1}{n-1}\sqrt{n}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}} \le \frac{2}{\sqrt{n}}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}}, \end{aligned}$$

we have $\|a - b\| \le 2\|a_0 - b\| \le \frac{4}{\sqrt{n}}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}}$. Therefore,

$$\|a_0\| \ge \|b\| - \|a_0 - b\| \ge 1 - \frac{2}{\sqrt{n}}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}}, \tag{52}$$

$$|a^{\mathrm{T}}b - 1| = |(a - b)^{\mathrm{T}}b| \le \|a - b\| \le \frac{4}{\sqrt{n}}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}}, \tag{53}$$

$$\|(I_n - aa^{\mathrm{T}})b\| \le \|a - b\| + |a^{\mathrm{T}}b - 1| \le \frac{8}{\sqrt{n}}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}}. \tag{54}$$

We also have

$$\begin{aligned} \left|\frac{1}{n-1}z_j^*a^{\mathrm{T}}(V_j - z_j^*b)\right| &\le \frac{1}{n-1}\|V_j - z_j^*b\| \\ &\le \frac{1}{n-1}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}}. \end{aligned} \tag{55}$$

Therefore, as long as $\|F_j\| \vee \|H_j\| \le \rho$ and $|G_j| \le 1 - 4\rho$, we have

$$\begin{aligned} z_j^*a^{\mathrm{T}}\widetilde{V}_j &\ge 1 - \frac{2}{\sqrt{n}}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}} - \frac{1}{n-1}\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}} \\ &\quad - \|F_j\| - |G_j| - \|H_j\| \\ &\ge 1 - 3\sqrt{\gamma} - 2\rho - (1 - 4\rho) \\ &\ge \rho, \end{aligned}$$

where we have used (52) and (55), and we set ρ to satisfy $\rho \geq 3\sqrt{\gamma}$. The specific choice of ρ will be given later. Hence, the event $\{\widetilde{V}_j = 0\}$ is included in the event $\{\|F_j\| \vee \|H_j\| > \rho$ or $|G_j| > 1 - 4\rho\}$.

By (51), we obtain the bound

$$\begin{aligned}
&\|\widehat{V}_j - z_j^* a\|^2 \\
&\leq \frac{\|(I_n - aa^{\mathrm{T}})\widetilde{V}_j\|^2}{|z_j^* a^{\mathrm{T}}\widetilde{V}_j|^2}\mathbb{I}\{\|F_j\| \vee \|H_j\| \leq \rho, |G_j| \leq 1 - 4\rho\} \\
&+ 4\mathbb{I}\{\|F_j\| \vee \|H_j\| > \rho \text{ or } |G_j| > 1 - 4\rho\} \\
&\leq \frac{1}{\rho^2}\bigg(\frac{1}{n-1}\|V_j - z_j^* b\| + \|F_j\| \\
&\quad + \|(I_n - aa^{\mathrm{T}})b\||G_j| + \|H_j\|\bigg)^2 \\
&+ 4\mathbb{I}\{\|F_j\| > \rho\} + 4\mathbb{I}\{|G_j| > 1 - 4\rho\} + 4\mathbb{I}\{\|H_j\| > \rho\} \\
&\leq \frac{4\|V_j - z_j^* b\|^2}{\rho^2(n-1)^2} + \frac{4\|F_j\|^2}{\rho^2} + \frac{4\|(I_n - aa^{\mathrm{T}})b\|^2|G_j|^2}{\rho^2} \\
&+ \frac{4\|H_j\|^2}{\rho^2} + \frac{4\|F_j\|^2}{\rho^2} + \frac{4\|H_j\|^2}{\rho^2} + 4\mathbb{I}\{|G_j| > 1 - 4\rho\} \\
&\leq \frac{4\|V_j - z_j^* b\|^2}{\rho^2(n-1)^2} + \frac{8\|F_j\|^2}{\rho^2} + \frac{256\|V - bz^{*\mathrm{T}}\|_{\mathrm{F}}^2|G_j|^2}{n\rho^2} \\
&+ \frac{8\|H_j\|^2}{\rho^2} + 4\mathbb{I}\{|G_j| > 1 - 4\rho\}.
\end{aligned}$$

We have used (54), Jensen's inequality and Markov's inequality in the above derivation.

c) Step 3: Analysis of each error term: Next, we will analyze the error terms F_j, H_j and G_j separately. Following the same analysis that leads to (44), (45) and (48), we have

$$\begin{aligned}
\sum_{j=1}^n \|F_j\|^2 &\leq C_1\frac{\log n}{np}\|V - bz^{*\mathrm{H}}\|_{\mathrm{F}}^2, \\
\sum_{j=1}^n \|H_j\|^2 &\leq C_2\frac{\sigma^2}{np}\|V - bz^{*\mathrm{H}}\|_{\mathrm{F}}^2, \\
\sum_{j=1}^n |G_j|^2 &\leq C_3\frac{\sigma^2}{p}.
\end{aligned}$$

Note that the above three bounds are based on the events (29), (30), (31), (49) and (50).

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

$$\begin{aligned}
&n\ell(\widehat{V}, z^*) \\
&\leq \sum_{j=1}^n \|\widehat{V}_j - z_j^* a\|^2 \\
&\leq \left(\frac{4}{\rho^2(n-1)^2} + \frac{8C_1\log n + (8C_2 + 256C_3)\sigma^2}{\rho^2 np}\right) n\ell(V, z^*) \\
&\quad + 4\sum_{j=1}^n \mathbb{I}\{|G_j| > 1 - 4\rho\}.
\end{aligned}$$

Set

$$\rho^2 = \left(C_4\frac{\log n + \sigma^2}{np}\right) \vee (3\gamma),$$

for some sufficiently large constant C_4 such that $\frac{4}{\rho^2(n-1)^2} + \frac{8C_1\log n + (8C_2 + 256C_3)\sigma^2}{\rho^2 np} \leq \frac{1}{2}$. Then, we have

$$\begin{aligned}
&\ell(\widehat{V}, z^*) \\
&\leq \frac{1}{2}\ell(V, z^*) + \frac{4}{n}\sum_{j=1}^n \mathbb{I}\{|G_j| > 1 - 4\rho\} \\
&\leq \frac{1}{2}\ell(V, z^*) + \frac{4}{n}\sum_{j=1}^n \mathbb{I}\bigg\{\frac{\sigma}{(n-1)p}\bigg|\sum_{k\in[n]\setminus\{j\}} z_k^* A_{jk} W_{jk}\bigg| > \\
&\quad \left(1 - C_5\sqrt{\frac{\log n + \sigma^2}{np}} - \sqrt{3\gamma}\right)\bigg\},
\end{aligned}$$

where the last inequality is by (29) and (30). Note that $\frac{\log n + \sigma^2}{np}$ is sufficiently small and $\gamma < 1/16$, we have $\delta < 1$ with $\delta = C_5\sqrt{\frac{\log n + \sigma^2}{np}} + \sqrt{3\gamma}$. Since the above about is derived from the conditions (29), (30), (31), (49) and (50) and $\ell(V, z^*) \leq \gamma$, it holds uniformly over all $V \in \mathbb{R}_1^{n\times n}$ such that $\ell(V, z^*) \leq \gamma$ with probability at least $1 - 2n^{-9}$. The proof is complete.

D. Proofs of Theorem 3, Theorem 4, and Theorem 6

Proof of Theorem 3: We obtain (19) as a consequence of Lemma 1 and Lemma 2, which immediately implies the first conclusion. The second conclusion is a consequence of Lemma 11. □

Proof of Theorem 4: By Theorem 3, we have $\|\widehat{V} - bz^{*\mathrm{H}}\|_{\mathrm{F}}^2 \leq \frac{\sigma^2}{p}$ with high probability for some $b \in \mathbb{C}^n$ such that $\|b\| = 1$. Since $\widehat{V} = f(\widehat{V})$, we can follow the same analysis in the proof of Lemma 1 and obtain the bound

$$\|\widehat{V} - az^{*\mathrm{H}}\|_{\mathrm{F}}^2 \leq (1 + \delta)\frac{\sigma^2}{2p}, \tag{56}$$

with high probability, where $\delta = C\left(\frac{\log n + \sigma^2}{np}\right)^{1/4}$ and $a = a_0/\|a_0\|$ with $a_0 = b + \frac{1}{n-1}\sum_{k=1}^n z_k^*(\widehat{V}_k - \bar{z}_k^* b)$. Let $\widetilde{z} = \widehat{V}^{\mathrm{H}}\widehat{a}$ where $\widehat{a}$ is the leading left singular vector of $\widehat{V}$. By the definition of $\widehat{z}$, we can write $\widehat{z}_j = \widetilde{z}_j/|\widetilde{z}_j|$ for all $j \in [n]$ with non-zero $\widetilde{z}_j$.

By (56) and Wedin's sin-theta theorem [30], we have

$$\|\widehat{a} - ha\|^2 \leq \frac{\|\widehat{V} - az^{*\mathrm{H}}\|_{\mathrm{F}}^2}{n} \leq \frac{\sigma^2}{np}, \tag{57}$$

form some $h \in \mathbb{C}_1$. Define $d_0 = a^{\mathrm{H}}\widehat{a}$ and $d = d_0/|d_0|$. With $\widetilde{z}_j = \widehat{V}_j^{\mathrm{H}}\widehat{a}$, we have

$$\widetilde{z}_j\bar{z}_j^*\bar{d} = |d_0| + h\bar{d}\bar{z}_j^*(\widehat{V}_j - \bar{z}_j^* a)^{\mathrm{H}}a + \bar{d}(\widehat{V}_j - \bar{z}_j^* a)^{\mathrm{H}}(\widehat{a} - ha)\bar{z}_j^*. \tag{58}$$

By Lemma 10, we have the bound

$$|\widehat{z}_j - dz_j^*|^2 = \left|\frac{\widetilde{z}_j\bar{z}_j^*\bar{d}}{|\widetilde{z}_j\bar{z}_j^*\bar{d}|} - 1\right|^2 \leq \frac{|\mathrm{Im}(\widetilde{z}_j\bar{z}_j^*\bar{d})|^2}{|\mathrm{Re}(\widetilde{z}_j\bar{z}_j^*\bar{d})|^2}, \tag{59}$$

as long as $\mathrm{Re}(\widetilde{z}_j\bar{z}_j^*\bar{d}) > 0$. By (57), we have $|d_0| \geq \mathrm{Re}(h\widehat{a}^{\mathrm{H}}a) \geq 1 - \frac{\sigma^2}{2np}$ and $|\bar{d}(\widehat{V}_j - \bar{z}_j^*a)^{\mathrm{H}}(\widehat{a} - ha)\bar{z}_j^*| \leq \sqrt{\frac{\sigma^2}{np}}\|\widehat{V}_j - \bar{z}_j^*a\|$. Moreover,

$$|h\bar{d}\bar{z}_j^*(\widehat{V}_j - \bar{z}_j^*a)^{\mathrm{H}}a| \leq |(\widehat{V}_j - \bar{z}_j^*a)^{\mathrm{H}}a| = |\widehat{V}_j^{\mathrm{H}}a\bar{z}_j^* - 1| = |a^{\mathrm{H}}\widehat{V}_jz_j^* - 1|.$$

Therefore,

$$\mathrm{Re}(\widetilde{z}_j\bar{z}_j^*\bar{d}) \geq 1 - \frac{\sigma^2}{2np} - |a^{\mathrm{H}}\widehat{V}_jz_j^* - 1| - \sqrt{\frac{\sigma^2}{np}}\|\widehat{V}_j - \bar{z}_j^*a\|. \quad (60)$$

In the following, we are going to establish a lower bound for (60) using some similar analysis as in the proof of Lemma 1. Since $\widehat{V} = f(\widehat{V})$, we can write $\widehat{V}_j = \widetilde{V}_j/\|\widetilde{V}_j\|$ for all non-zero $\widetilde{V}_j$, where

$$\widetilde{V}_j = \frac{\sum_{k\in[n]\setminus\{j\}} A_{jk}Y_{jk}\widehat{V}_k}{\sum_{k\in[n]\setminus\{j\}} A_{jk}}.$$

Similar to the decomposition (36), we can write

$$a^{\mathrm{H}}\widetilde{V}_jz_j^* = \|a_0\| - \frac{1}{n-1}z_j^*a^{\mathrm{H}}(\widehat{V}_j - \bar{z}_j^*b) + a^{\mathrm{H}}F_j + a^{\mathrm{H}}bG_j + a^{\mathrm{H}}H_j,$$

where

$$F_j = \frac{\sum_{k\in[n]\setminus\{j\}} A_{jk}z_k^*(\widehat{V}_k - \bar{z}_k^*b)}{\sum_{k\in[n]\setminus\{j\}} A_{jk}} - \frac{1}{n-1}\sum_{k=1}^n z_k^*(\widehat{V}_k - \bar{z}_k^*b),$$
$$G_j = \frac{\sigma z_j^*\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}\bar{z}_k^*}{\sum_{k\in[n]\setminus\{j\}} A_{jk}},$$
$$H_j = \frac{\sigma z_j^*\sum_{k\in[n]\setminus\{j\}} A_{jk}\bar{W}_{jk}(\widehat{V}_k - \bar{z}_k^*b)}{\sum_{k\in[n]\setminus\{j\}} A_{jk}}.$$

By the same argument that leads to (43) with $\gamma = \frac{\sigma^2}{np}$, for any $\rho > 0$, we know that as long as $\|F_j\| \vee |G_j| \vee \|H_j\| \leq \rho$, we have

$$|\mathrm{Re}(a^{\mathrm{H}}\widetilde{V}_jz_j^*) - 1| \leq 3\rho + 3\sqrt{\frac{\sigma^2}{np}}. \quad (61)$$

Moreover,

$$\begin{aligned}
&|\mathrm{Im}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)| \\
&\leq \frac{1}{n-1}\|\widehat{V}_j - \bar{z}_j^*b\| + \|F_j\| + |\mathrm{Im}(a^{\mathrm{H}}bG_j)|^2 + \|H_j\| \\
&\leq \frac{1}{n-1}\sqrt{\frac{\sigma^2}{p}} + \|F_j\| + |\mathrm{Im}(a^{\mathrm{H}}bG_j)|^2 + \|H_j\| \\
&\leq \frac{1}{n-1}\sqrt{\frac{\sigma^2}{p}} + 3\rho. \quad (62)
\end{aligned}$$

By a similar bound to (37), we also have

$$\|(I_n - aa^{\mathrm{H}})\widetilde{V}_j\| \leq \frac{1}{n-1}\|\widehat{V}_j - \bar{z}_j^*b\| + \|F_j\| + |G_j| + \|H_j\| \leq \frac{1}{n-1}\sqrt{\frac{\sigma^2}{p}} + 3\rho.$$

With the decomposition $\|\widetilde{V}_j\|^2 = \|(I_n - aa^{\mathrm{H}})\widetilde{V}_j\|^2 + |\mathrm{Im}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)|^2 + |\mathrm{Re}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)|^2$, we have

$$\begin{aligned}
&\left|\|\widetilde{V}_j\|^2 - 1\right| \\
&\leq \|(I_n - aa^{\mathrm{H}})\widetilde{V}_j\|^2 + |\mathrm{Im}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)|^2 + \left||\mathrm{Re}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)|^2 - 1\right| \\
&\leq 4\rho + 4\sqrt{\frac{\sigma^2}{np}}. \quad (63)
\end{aligned}$$

Let ρ be a sufficiently small with explicit expression to be given later. Together with the assumption that $\frac{\sigma^2}{np}$ is also sufficiently small, both (61) and (63) can be upper bounded by $1/2$, which implies $\mathrm{Re}(a^{\mathrm{H}}\widetilde{V}_jz_j^*) > 1/2$ and $\widetilde{V}_j \neq 0$ respectively. Then $\widehat{V}_j = \widetilde{V}_j/\|\widetilde{V}_j\|$ leads to the bound

$$\begin{aligned}
&|a^{\mathrm{H}}\widehat{V}_jz_j^* - 1| \\
&\leq \left|\frac{\mathrm{Re}(a^{\mathrm{H}}\widetilde{V}_jz_j^*) - \|\widetilde{V}_j\|}{\|\widetilde{V}_j\|}\right| + \frac{|\mathrm{Im}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)|}{\|\widetilde{V}_j\|} \\
&\leq \left|\frac{\mathrm{Re}(a^{\mathrm{H}}\widetilde{V}_jz_j^*) - 1}{\|\widetilde{V}_j\|}\right| + \left|\frac{1 - \|\widetilde{V}_j\|}{\|\widetilde{V}_j\|}\right| + \frac{|\mathrm{Im}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)|}{\|\widetilde{V}_j\|} \\
&\leq C_1\left(\rho + \sqrt{\frac{\sigma^2}{np}}\right),
\end{aligned}$$

where we use (61)-(63). By Lemma 10, we have the bound

$$\|\widehat{V}_j - \bar{z}_j^*a\|^2 \leq \frac{\|(I_n - aa^{\mathrm{H}})\widetilde{V}_j\|^2 + |\mathrm{Im}(z_j^*a^{\mathrm{H}}\widetilde{V}_j)|^2}{|\mathrm{Re}(z_j^*a^{\mathrm{H}}\widetilde{V}_j)|^2} \leq C_2\left(\rho^2 + \frac{\sigma^2}{n^2p}\right). \quad (64)$$

Plugging the above two bounds into (60), we have

$$\mathrm{Re}(\widetilde{z}_j\bar{z}_j^*\bar{d}) \geq 1 - C_3\left(\rho + \sqrt{\frac{\sigma^2}{np}}\right),$$

which is positive since ρ and $\frac{\sigma^2}{np}$ are sufficiently small. Therefore, we have $\mathrm{Re}(\widetilde{z}_j\bar{z}_j^*\bar{d}) > 0$ and the bound (59) holds when $\|F_j\| \vee |G_j| \vee \|H_j\| \leq \rho$. Also this implies the event $\{\widetilde{z}_j = 0\}$ is included in the event $\{\|F_j\| \vee |G_j| \vee \|H_j\| > \rho\}$. As a consequence, we have

$$\begin{aligned}
&|\widehat{z}_j - dz_j^*|^2 \\
&\leq \frac{|\mathrm{Im}(\widetilde{z}_j\bar{z}_j^*\bar{d})|^2}{\left(1 - C_3\left(\rho + \sqrt{\frac{\sigma^2}{np}}\right)\right)^2}\mathbb{I}\{\|F_j\| \vee |G_j| \vee \|H_j\| \leq \rho\} \\
&\quad + 4\mathbb{I}\{\|F_j\| \vee |G_j| \vee \|H_j\| > \rho\}.
\end{aligned}$$

Now we need to bound $|\mathrm{Im}(\widetilde{z}_j\bar{z}_j^*\bar{d})|$ according to the expansion (58). We have

$$\begin{aligned}
&|\mathrm{Im}(\widetilde{z}_j\bar{z}_j^*\bar{d})| \\
&\leq \left|\mathrm{Im}(\bar{z}_j^*(\widehat{V}_j - \bar{z}_j^*a)^{\mathrm{H}}a)\right| + |\mathrm{Im}(h\bar{d})|\left|\mathrm{Re}(\bar{z}_j^*(\widehat{V}_j - \bar{z}_j^*a)^{\mathrm{H}}a)\right| \\
&\quad + \|\widehat{V}_j - \bar{z}_j^*a\|\|\widehat{a} - ha\|. \quad (65)
\end{aligned}$$

By (57) and (64), the third term in the bound (65) can be further bounded by $C_2\sqrt{\frac{\sigma^2}{np}}\left(\rho + \sqrt{\frac{\sigma^2}{n^2p}}\right)$. To bound the

second term on the right hand side of (65), we have $|\mathrm{Im}(h\bar{d})| \le |\mathrm{Im}(ha^{\mathrm{H}}\widehat{a})| \le \sqrt{1-|\mathrm{Re}(ha^{\mathrm{H}}a)|^2} \le \sqrt{\frac{\sigma^2}{np}}$ by (57). Together with (61), we obtain the bound $3\sqrt{\frac{\sigma^2}{np}}\left(\rho+\sqrt{\frac{\sigma^2}{np}}\right)$. By (63), we can bound the first term in the bound (65) by

$$\left|\mathrm{Im}(\bar{z}_j^*(\widehat{V}_j-\bar{z}_j^*a)^{\mathrm{H}}a)\right| = \left|\mathrm{Im}(\widehat{V}_j^{\mathrm{H}}a\bar{z}_j^*)\right| = \left|\mathrm{Im}(a^{\mathrm{H}}\widehat{V}_jz_j^*)\right| \le \frac{\left|\mathrm{Im}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)\right|}{1-4\left(\rho+\sqrt{\frac{\sigma^2}{np}}\right)}.$$

Then, we have

$$\begin{aligned}
&|\widehat{z}_j - dz_j^*|^2 \\
&\le \left(1+C_4\left(\rho+\sqrt{\frac{\sigma^2}{np}}\right)\right)\left|\mathrm{Im}(a^{\mathrm{H}}\widetilde{V}_jz_j^*)\right|^2 \\
&+ C_5\frac{\sigma^2}{np}\left(\rho^2+\frac{\sigma^2}{np}\right) + 4\mathbb{I}\{\|F_j\|\vee|G_j|\vee\|H_j\|>\rho\} \\
&\le \left(1+C_6\left(\rho+\sqrt{\frac{\sigma^2}{np}}+\eta\right)\right)|\mathrm{Im}(a^{\mathrm{H}}bG_j)|^2 \\
&+ C_7\eta^{-1}\left(\frac{\sigma^2}{n^2p}+\|F_j\|^2+\|H_j\|^2\right) \\
&+ C_5\frac{\sigma^2}{np}\left(\rho^2+\frac{\sigma^2}{np}\right)+4\rho^{-2}\left(\|F_j\|^2+\|H_j\|^2\right)+4\mathbb{I}\{|G_j|>\rho\},
\end{aligned}$$

for some η to be specified later, where the last inequality is by (62). Summing over $j\in[n]$, we obtain

$$\begin{aligned}
n\ell(\widehat{z},z^*) &\le \sum_{j=1}^n|\widehat{z}_j-dz_j^*|^2 \\
&\le \left(1+C_6\left(\rho+\sqrt{\frac{\sigma^2}{np}}+\eta\right)\right)\sum_{j=1}^n|\mathrm{Im}(a^{\mathrm{H}}bG_j)|^2 \\
&+ C_7\eta^{-1}\frac{\sigma^2}{np}+C_5\frac{\sigma^2}{p}\left(\rho^2+\frac{\sigma^2}{np}\right) \\
&+ (C_7\eta^{-1}+4\rho^{-2})\sum_{j=1}^n\left(\|F_j\|^2+\|H_j\|^2\right) \\
&+ 4\sum_{j=1}^n\mathbb{I}\{|G_j|>\rho\}.
\end{aligned}$$

By the same argument that leads to the bound (44)-(47) (with $\gamma=\frac{\sigma^2}{np}$ in (47)), we have

$$\begin{aligned}
&\sum_{j=1}^n(\|F_j\|^2+\|H_j\|^2) \le C'\frac{\log n}{np}\|\widehat{V}-bz^{*\mathrm{H}}\|_{\mathrm{F}}^2 \le C'\frac{\log n}{np}\frac{\sigma^2}{p}, \\
&\sum_{j=1}^n\mathbb{I}\{|G_j|>\rho\} \le \frac{4\sigma^2}{\rho^2p}\exp\left(-\frac{1}{16}\sqrt{\frac{\rho^2np}{\sigma^2}}\right), \\
&\sum_{j=1}^n|\mathrm{Im}(a^{\mathrm{H}}bG_j)|^2 \le \left(1+C''\left(\eta+\eta^{-1}\frac{\sigma^2}{np}+\sqrt{\frac{\log n}{np}}\right)\right)\frac{\sigma^2}{2p}.
\end{aligned}$$

Take $\eta=\rho^2=\sqrt{\frac{\log n+\sigma^2}{np}}$, and we have some constant $C'''>0$ such that

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

Note that the above bound is derived from conditions (29)-(34), and thus the result holds with high probability. □

Proof of Theorem 6: The first conclusion is an immediate consequence of Lemma 3, Lemma 4 and Lemma 5. By Lemma 11, we also obtain the second conclusion. For the last conclusion, we have $|\widehat{z}_j-z_j^*| \le 2|\sqrt{n}u_j-z_j^*|$ and $|\widehat{z}_j+z_j^*| \le 2|\sqrt{n}u_j+z_j^*|$ by the definition of $\widehat{z}_j$. Then,

$$\begin{aligned}
\ell(\widehat{z},z^*) &\le 4\left(\|u-z^*/\sqrt{n}\|^2\wedge\|u+z^*/\sqrt{n}\|^2\right) \\
&\le \frac{16}{n^2}\|\widehat{Z}-z^*z^{*\mathrm{T}}\|_{\mathrm{F}}^2,
\end{aligned}$$

by Davis-Kahan theorem [18]. Thus, we can derive the third conclusion from the second one. Finally, when $\sigma^2<(1-\epsilon)\frac{np}{2\log n}$, we know from (27) and Lemma 4 that $\ell(\widehat{V},z^*)=0$. Lemma 11 implies that $\|\widehat{Z}-z^*z^{*\mathrm{T}}\|_{\mathrm{F}}^2=0$ and thus $\widehat{Z}=z^*z^{*\mathrm{T}}$ is a rank-one matrix. □

E. Proof of Theorem 5

Since $\ell(\widehat{z},z)=2\left(1-\frac{1}{n}|\widehat{z}^{\mathrm{T}}z|\right)$ and $n^{-2}\|\widehat{z}\widehat{z}^{\mathrm{T}}-zz^{\mathrm{T}}\|_{\mathrm{F}}^2 = 2\left(1-\frac{1}{n^2}|\widehat{z}^{\mathrm{T}}z|^2\right)$, we have

$$n^{-2}\|\widehat{z}\widehat{z}^{\mathrm{T}}-zz^{\mathrm{T}}\|_{\mathrm{F}}^2 = \ell(\widehat{z},z)\left(1+\frac{1}{n}|\widehat{z}^{\mathrm{T}}z|\right) \le 2\ell(\widehat{z},z),$$

and thus

$$\begin{aligned}
&\inf_{\widehat{z}\in\{-1,1\}^n}\sup_{z\in\{-1,1\}^n}\mathbb{E}_z\ell(\widehat{z},z) \\
\ge\ &\inf_{\widehat{z}\in\{-1,1\}^n}\sup_{z\in\{-1,1\}^n}\mathbb{E}_z\frac{1}{2n^2}\|\widehat{z}\widehat{z}^{\mathrm{T}}-zz^{\mathrm{T}}\|_{\mathrm{F}}^2 \\
\ge\ &\frac{1}{2}\inf_{\widehat{Z}\in\mathbb{R}^{n\times n}}\sup_{z\in\{-1,1\}^n}\mathbb{E}_z\frac{1}{n^2}\|\widehat{Z}-zz^{\mathrm{T}}\|_{\mathrm{F}}^2.
\end{aligned}$$

It suffices to prove a lower bound for the loss $\|\widehat{Z}-zz^{\mathrm{T}}\|_{\mathrm{F}}^2$. We lower bound the minimax risk by a Bayes risk

$$\begin{aligned}
&\inf_{\widehat{Z}\in\mathbb{R}^{n\times n}}\sup_{z\in\{-1,1\}^n}\mathbb{E}_z\frac{1}{n^2}\|\widehat{Z}-zz^{\mathrm{T}}\|_{\mathrm{F}}^2 \\
\ge\ &\inf_{\widehat{Z}\in\mathbb{R}^{n\times n}}\frac{1}{2^n}\sum_{z\in\{-1,1\}^n}\mathbb{E}_z\frac{1}{n^2}\|\widehat{Z}-zz^{\mathrm{T}}\|_{\mathrm{F}}^2 \\
\ge\ &\frac{1}{n^2}\sum_{1\le j\ne k\le n}\frac{1}{2^{n-2}}\sum_{z_{-(j,k)}\in\{-1,1\}^{n-2}} \\
&\inf_{\widehat{T}}\frac{1}{4}\sum_{z_j\in\{-1,1\}}\sum_{z_k\in\{-1,1\}}\mathbb{E}_z|\widehat{T}-z_jz_k|^2,
\end{aligned}$$

where $z_{-(j,k)}$ is a sub-vector of z by excluding the jth and the kth entries. For each $z_{-(j,k)}$, we have

$$\begin{aligned}
&\inf_{\widehat{T}} \sum_{z_j\in\{-1,1\}} \sum_{z_k\in\{-1,1\}} \mathbb{E}_z|\widehat{T}-z_jz_k|^2\\
\geq\ & \inf_{\widehat{T}} \Big(\mathbb{E}_{(z_{-(j,k)},z_j=1,z_k=-1)}|\widehat{T}+1|^2\\
&\quad +\mathbb{E}_{(z_{-(j,k)},z_j=1,z_k=1)}|\widehat{T}-1|^2\Big)\\
\geq\ & 2\int \mathrm{d}\mathbb{P}_{(z_{-(j,k)},z_j=1,z_k=-1)} \wedge \mathrm{d}\mathbb{P}_{(z_{-(j,k)},z_j=1,z_k=1)},
\end{aligned}$$

where the last inequality is due to the classical Le Cam's two-point method. The total variation affinity characterizes the optimal testing error between two simple hypotheses of $z_k=-1$ versus $z_k=1$ with the values of all other parameters are known. By Neyman-Pearson lemma, we have

$$\begin{aligned}
&\int \mathrm{d}\mathbb{P}_{(z_{-(j,k)},z_j=1,z_k=-1)} \wedge \mathrm{d}\mathbb{P}_{(z_{-(j,k)},z_j=1,z_k=1)}\\
\geq\ & \mathbb{P}_{(z_{-(j,k)},z_j=1,z_k=-1)}\left(\frac{\mathrm{d}\mathbb{P}_{(z_{-(j,k)},z_j=1,z_k=1)}}{\mathrm{d}\mathbb{P}_{(z_{-(j,k)},z_j=1,z_k=-1)}}>1\right)\\
=\ & \mathbb{P}_{(z_{-(j,k)},z_j=1,z_k=-1)}\left(\sum_{j\in[n]\setminus\{k\}} z_jA_{jk}Y_{jk}>0\right)\\
=\ & \mathbb{P}\left(\sigma\sum_{j\in[n]\setminus\{k\}} z_jA_{jk}W_{jk}>\sum_{j\in[n]\setminus\{k\}} A_{jk}\right).
\end{aligned}$$

Let $\mathcal{A}$ be the collections of A's that satisfy the conclusions of Lemma 6, and we know that $\mathbb{P}(\mathcal{A})\geq 1-n^{-10}$. Let $\mathbb{P}_A$ be the shorthand of the conditional probability $\mathbb{P}(\cdot|A)$. For each $A\in\mathcal{A}$, a standard Gaussian tail bound implies

$$\begin{aligned}
&\mathbb{P}_A\left(\sigma\sum_{j\in[n]\setminus\{k\}} z_jA_{jk}W_{jk}>\sum_{j\in[n]\setminus\{k\}} A_{jk}\right)\\
&\geq \exp\left(-(1+\delta)\frac{np}{2\sigma^2}\right),
\end{aligned}$$

where $\delta=C\sqrt{\frac{\log n+\sigma^2}{np}}$ for some constant $C>0$. This implies

$$\begin{aligned}
&\mathbb{P}\left(\sigma\sum_{j\in[n]\setminus\{k\}} z_jA_{jk}W_{jk}>\sum_{j\in[n]\setminus\{k\}} A_{jk}\right)\\
&\geq \inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sigma\sum_{j\in[n]\setminus\{k\}} z_jA_{jk}W_{jk}>\sum_{j\in[n]\setminus\{k\}} A_{jk}\right)\mathbb{P}(\mathcal{A})\\
&\geq \frac{1}{2}\exp\left(-(1+\delta)\frac{np}{2\sigma^2}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\inf_{\widehat{z}\in\{-1,1\}^n}\sup_{z\in\{-1,1\}^n}\mathbb{E}_z\ell(\widehat{z},z)\\
\geq\ & \frac{1}{2}\inf_{\widehat{Z}\in\mathbb{R}^{n\times n}}\sup_{z\in\{-1,1\}^n}\mathbb{E}_z\frac{1}{n^2}\|\widehat{Z}-zz^{\mathrm{T}}\|_{\mathrm{F}}^2\\
\geq\ & \frac{1}{16}\exp\left(-(1+\delta)\frac{np}{2\sigma^2}\right).
\end{aligned}$$

By absorbing the constant $1/16$ into the exponent, the proof is complete.

F. Proofs of Lemma 2, Lemma 5, and Lemma 4

Proof of Lemma 2: By the definition of $\widehat{Z}=\widehat{V}^{\mathrm{H}}\widehat{V}$, we have $\mathsf{Tr}((A\circ Y)\widehat{Z})\geq\mathsf{Tr}((A\circ Y)z^*z^{*\mathrm{H}})$. Rearranging this inequality, we obtain

$$\begin{aligned}
&\mathsf{Tr}(z^*z^{*\mathrm{H}}(z^*z^{*\mathrm{H}}-\widehat{Z}))\\
&\leq \mathsf{Tr}\left((A\circ Y/p-z^*z^{*\mathrm{H}})(\widehat{Z}-z^*z^{*\mathrm{H}})\right). \qquad (66)
\end{aligned}$$

The right hand side of (66) can be bounded by

$$\begin{aligned}
&\left|\mathsf{Tr}\left((A\circ Y/p-z^*z^{*\mathrm{H}})\widehat{Z}\right)\right|\\
&\quad +\left|\mathsf{Tr}\left((A\circ Y/p-z^*z^{*\mathrm{H}})z^*z^{*\mathrm{H}}\right)\right|\\
\leq\ & \|A\circ Y/p-z^*z^{*\mathrm{H}}\|_{\mathrm{op}}\,\mathsf{Tr}(\widehat{Z})\\
&\quad +\|A\circ Y/p-z^*z^{*\mathrm{H}}\|_{\mathrm{op}}\,\mathsf{Tr}(z^*z^{*\mathrm{H}})\\
=\ & 2n\|A\circ Y/p-z^*z^{*\mathrm{H}}\|_{\mathrm{op}}\\
\leq\ & 2n\left(\frac{1}{p}\|(A-\mathbb{E}A)\circ z^*z^{*\mathrm{H}}\|_{\mathrm{op}}+\frac{\sigma}{p}\|A\circ W\|_{\mathrm{op}}\right).
\end{aligned}$$

By Lemma 6,

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

with probability at least $1-n^{-10}$. By Lemma 7, $\|A\circ W\|_{\mathrm{op}}\leq C_2\sqrt{np}$ with probability at least $1-n^{-10}$. Thus, we have

$$\mathsf{Tr}(z^*z^{*\mathrm{H}}(z^*z^{*\mathrm{H}}-\widehat{Z}))\leq C_3n\sqrt{\frac{(1+\sigma^2)n}{p}}.$$

Define $m=\frac{1}{n}\sum_{j=1}^n\widehat{V}_jz_j^*$. By the inequality $\|x/\|x\|-y/\|y\|\|\leq 2\|x-y\|/\|x\|$, we have

$$\begin{aligned}
\ell(\widehat{V},z^*) &= \min_{a\in\mathbb{C}^n:\|a\|^2=1}\frac{1}{n}\sum_{j=1}^n\|\widehat{V}_jz_j^*-a\|^2\\
&= \min_{a\in\mathbb{C}^n\setminus\{0\}}\frac{1}{n}\sum_{j=1}^n\|\widehat{V}_jz_j^*-a/\|a\|\|^2\\
&\leq \min_{a\in\mathbb{C}^n\setminus\{0\}}\frac{4}{n}\sum_{j=1}^n\|\widehat{V}_jz_j^*-a\|^2\\
&= \frac{4}{n}\sum_{j=1}^n\|\widehat{V}_jz_j^*-m\|^2\\
&= \frac{2}{n^2}\sum_{j=1}^n\sum_{l=1}^n\left(\|\widehat{V}_jz_j^*-m\|^2+\|\widehat{V}_lz_l^*-m\|^2\right)\\
&= \frac{2}{n^2}\sum_{j=1}^n\sum_{l=1}^n\|\widehat{V}_jz_j^*-\widehat{V}_lz_l^*\|^2\\
&= \frac{4}{n^2}\sum_{j=1}^n\sum_{l=1}^n(1-\bar{z}_j^*z_l^*\widehat{V}_j^{\mathrm{H}}\widehat{V}_l)\\
&= \frac{4}{n^2}\mathsf{Tr}(z^*z^{*\mathrm{H}}(z^*z^{*\mathrm{H}}-\widehat{Z})).
\end{aligned}$$

Therefore, we have $\ell(\widehat{V},z^*)\leq 4C_3\sqrt{\frac{(1+\sigma^2)}{np}}$, and the proof is complete. □

Proof of Lemma 5: Following the same argument in the proof of Lemma 2, we have

$$\mathsf{Tr}(z^*z^{*\mathrm{T}}(z^*z^{*\mathrm{T}} - \widehat{Z})) \le Cn\sqrt{\frac{(1+\sigma^2)n}{p}},$$

with probability at least $1 - n^{-9}$ and $\ell(\widehat{V}, z^*) \le \frac{4}{n^2}\mathsf{Tr}(z^*z^{*\mathrm{T}}(z^*z^{*\mathrm{T}} - \widehat{Z}))$. Then, we obtain the bound $\ell(\widehat{V}, z^*) \le 4C\sqrt{\frac{(1+\sigma^2)}{np}}$, and the proof is complete. □

Proof of Lemma 4: Let $\mathcal{A}$ be the collections of A's that satisfy the conclusions of Lemma 6, and we know that $\mathbb{P}(\mathcal{A}) \ge 1 - n^{-10}$. Let $\mathbb{P}_A$ be the shorthand of the conditional probability $\mathbb{P}(\cdot|A)$. For each $A \in \mathcal{A}$, a standard Gaussian tail bound implies

$$\begin{aligned}
&\frac{8}{n}\sum_{j=1}^{n}\mathbb{P}_A\left(|U_j| > 1-\delta\right)\\
\le\ & \frac{16}{n}\sum_{j=1}^{n}\exp\left(-\frac{(1-\delta)^2(n-1)^2p^2}{2\sigma^2\sum_{k\in[n]\setminus\{j\}}A_{jk}}\right)\\
\le\ & \exp\left(-(1-\bar{\delta})\frac{np}{2\sigma^2}\right),
\end{aligned}$$

where $\bar{\delta} = C\left(\delta + \sqrt{\frac{\log n}{np}}\right)$ for some constant $C > 0$. Therefore,

$$\begin{aligned}
&\mathbb{P}\left(\frac{8}{n}\sum_{j=1}^{n}\mathbb{I}\{|U_j| > 1-\delta\}\right.\\
&\qquad\left. > \exp\left(-\left(1-\bar{\delta}-\sqrt{\frac{2\sigma^2}{np}}\right)\frac{np}{2\sigma^2}\right)\right)\\
\le\ & \sup_{A\in\mathcal{A}}\mathbb{P}_A\left(\frac{8}{n}\sum_{j=1}^{n}\mathbb{I}\{|U_j| > 1-\delta\}\right.\\
&\qquad\left. > \exp\left(-\left(1-\bar{\delta}-\sqrt{\frac{2\sigma^2}{np}}\right)\frac{np}{2\sigma^2}\right)\right) + \mathbb{P}(\mathcal{A}^c)\\
\le\ & \exp\left(-\sqrt{\frac{np}{\sigma^2}}\right) + n^{-10},
\end{aligned}$$

by Markov's inequality. This immediately implies the first conclusion. For the second conclusion, it is easy to see that when $\left(1-\bar{\delta}-\sqrt{\frac{2\sigma^2}{np}}\right)\frac{np}{2\sigma^2} > \log n$, we have $\frac{8}{n}\sum_{j=1}^{n}\mathbb{I}\{|U_j| > 1-\delta\} \le \frac{1}{n}$, and thus the value of $\frac{8}{n}\sum_{j=1}^{n}\mathbb{I}\{|U_j| > 1-\delta\}$ has to be 0. □

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*.

[3] 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.

[4] T. Gao and Z. Zhao, "Multi-frequency phase synchronization," in *Proc. Int. Conf. Mach. Learn.*, vol. 97, 2020, pp. 2132–2141.

[5] Y. Zhong and N. Boumal, "Near-optimal bounds for phase synchronization," *SIAM J. Optim.*, vol. 28, no. 2, pp. 989–1016, Apr. 2018.

[6] C. Gao and A. Y. Zhang, "Exact minimax estimation for phase synchronization," *IEEE Trans. Inf. Theory*, vol. 67, no. 12, pp. 8236–8247, Dec. 2021.

[7] A. S. Bandeira, Y. Khoo, and A. Singer, "Open problem: Tightness of maximum likelihood semidefinite relaxations," 2014, *arXiv:1404.2655*.

[8] S. Ling, "Solving orthogonal group synchronization via convex and low-rank optimization: Tightness and landscape analysis," 2020, *arXiv:2006.00902*.

[9] N. Boumal, "Nonconvex phase synchronization," *SIAM J. Optim.*, vol. 26, no. 4, pp. 2355–2377, 2016.

[10] Y. Lu and H. H. Zhou, "Statistical and computational guarantees of Lloyd's algorithm and its variants," 2016, *arXiv:1612.02099*.

[11] C. Gao and A. Y. Zhang, "Iterative algorithm for discrete structure recovery," 2019, *arXiv:1911.01018*.

[12] Y. Fei and Y. Chen, "Achieving the Bayes error rate in synchronization and block models by SDP, robustly," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3929–3953, Jun. 2020.

[13] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Math. Program.*, vol. 95, no. 2, pp. 329–357, Feb. 2003.

[14] 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.

[15] P.-W. Wang, W.-C. Chang, and J. Z. Kolter, "The mixing method: Low-rank coordinate descent for semidefinite programming with diagonal constraints," 2017, *arXiv:1706.00476*.

[16] M. A. Erdogdu, A. Ozdaglar, P. A. Parrilo, and N. D. Vanli, "Convergence rate of block-coordinate maximization Burer–Monteiro method for solving large SDPs," 2018, *arXiv:1807.04428*.

[17] A. Y. Zhang and H. H. Zhou, "Minimax rates of community detection in stochastic block models," *Ann. Statist.*, vol. 44, no. 5, pp. 2252–2280, 2016.

[18] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation. III," *SIAM J. Numer. Anal.*, vol. 7, no. 1, pp. 1–46, Mar. 1970.

[19] A. S. Bandeira, "Random Laplacian matrices and convex relaxations," *Found. Comput. Math.*, vol. 18, no. 2, pp. 345–379, Apr. 2018.

[20] E. Abbe, J. Fan, K. Wang, and Y. Zhong, "Entrywise eigenvector analysis of random matrices with low expected rank," *Ann. Statist.*, vol. 48, no. 3, pp. 1452–1474, Jun. 2020.

[21] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2788–2797, May 2016.

[22] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming: Extensions," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5918–5937, Oct. 2016.

[23] A. Perry and A. S. Wein, "A semidefinite program for unbalanced multisection in the stochastic block model," in *Proc. Int. Conf. Sampling Theory Appl. (SampTA)*, Jul. 2017, pp. 64–67.

[24] A. A. Amini and E. Levina, "On semidefinite relaxations for the block model," *Ann. Statist.*, vol. 46, no. 1, pp. 149–179, Feb. 2018.

[25] Y. Chen, X. Li, and J. Xu, "Convexified modularity maximization for degree-corrected stochastic block models," *Ann. Statist.*, vol. 46, no. 4, pp. 1573–1602, Aug. 2018.

[26] X. Li, Y. Chen, and J. Xu, "Convex relaxation methods for community detection," 2018, *arXiv:1810.00315*.

[27] J. Lei *et al.*, "Consistency of spectral clustering in stochastic block models," *Ann. Statist.*, vol. 43, no. 1, pp. 215–237, 2015.

[28] 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.

[29] 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.

[30] P.-Å. Wedin, "Perturbation bounds in connection with singular value decomposition," *BIT Numer. Math.*, vol. 12, no. 1, pp. 99–111, 1972.

Chao Gao received the B.S. degree in statistics from Zhejiang University, Hangzhou, China, in 2010, and the M.A. and Ph.D. degrees in statistics from Yale University, New Haven, CT, USA, in 2016. He has been an Assistant Professor of statistics at The University of Chicago since 2016. He was a recipient of the NSF-CAREER Award in 2019, the Tweedie New Researcher Award from the Institute of Mathematical Statistics (IMS) in 2021, and the Sloan Research Fellowship in 2022.

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 at the Department of Statistics, The University of Chicago, in 2018. He has been an Assistant Professor of statistics and data science at the University of Pennsylvania since 2019.