Information and Inference: A Journal of the IMA (2023) **12**, 591–632
https://doi.org/10.1093/imaiai/iaac022
Advance Access publication on 6 August 2022

Optimal orthogonal group synchronization and rotation group synchronization

CHAO GAO†
Department of Statistics, University of Chicago, Chicago, IL 60637, USA
†Corresponding author: Email: chaogao@uchicago.edu

AND

ANDERSON Y. ZHANG
Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104-1686, USA

[Received on 28 September 2021; revised on 25 April 2022; accepted on 27 June 2022]

We study the statistical estimation problem of orthogonal group synchronization and rotation group synchronization. The model is $Y_{ij} = Z_i^* Z_j^{*T} + \sigma W_{ij} \in \mathbb{R}^{d\times d}$ where W_{ij} is a Gaussian random matrix and Z_i^* is either an orthogonal matrix or a rotation matrix, and each Y_{ij} is observed independently with probability p. We analyze an iterative polar decomposition algorithm for the estimation of Z^* and show it has an error of $(1+o(1))\frac{\sigma^2 d(d-1)}{2np}$ when initialized by spectral methods. A matching minimax lower bound is further established that leads to the optimality of the proposed algorithm as it achieves the exact minimax risk.

Keywords: group synchronization; minimax risk; iterative polar decomposition.

1. Introduction

Consider

$$Y_{ij} = Z_i^* Z_j^{*\mathrm{T}} + \sigma W_{ij} \in \mathbb{R}^{d\times d}, \tag{1}$$

for all $1 \le i < j \le n$. We assume

$$Z_i^* \in \mathcal{O}(d) = \{U \in \mathbb{R}^{d\times d} : UU^{\mathrm{T}} = U^{\mathrm{T}}U = I_d\}, \tag{2}$$

for all $i \in [n]$ and $W_{ij} \sim \mathcal{MN}(0, I_d, I_d)$ independently for all $1 \le i < j \le n$.[1] Our goal is to estimate the orthogonal matrices $Z_1^*, \cdots, Z_n^*$. This problem is known as orthogonal group synchronization or $\mathcal{O}(d)$ synchronization. In addition to (2), we also consider a closely related setting that

$$Z_i^* \in \mathcal{SO}(d) = \{U \in \mathcal{O}(d) : \det(U) = 1\}, \tag{3}$$

for all $i \in [n]$. This is known as rotation group synchronization or $\mathcal{SO}(d)$ synchronization. Both $\mathcal{O}(d)$ and $\mathcal{SO}(d)$ synchronizations have found successful applications across a wide range of areas including

[1] A random matrix X follows a matrix Gaussian distribution $\mathcal{MN}(M, \Sigma, \Omega)$, if its density function is proportional to $\exp\left(-\frac{1}{2}\operatorname{Tr}\left(\Omega^{-1}(X-M)^{\mathrm{T}}\Sigma^{-1}(X-M)\right)\right)$.

structural biology, computational genomics, robotics, computer vision and distributed networks. For example, synchronization over $\mathscr{O}(d)$ has been applied to the sensor network localization problem [8]. The problem over $\mathscr{O}(3)$ can be used to solve the graph realization problem [9], and that over $\mathscr{SO}(3)$ plays a central role in cryo-electron microscopy [30, 32] and global motion estimation [1].

Despite a growing literature in application and methodology, theoretical understandings of synchronization over $\mathscr{O}(d)$ or $\mathscr{SO}(d)$ have not been thoroughly explored. In particular, the exact minimax estimation of the $\mathscr{O}(d)$ synchronization under the model (1) still remains an important open problem. In this paper, we carefully characterize the minimax risk with respect to the following loss function,

$$\ell(Z, Z^*) = \min_{B \in \mathscr{O}(d)} \frac{1}{n} \sum_{i=1}^{n} \|Z_i - Z_i^* B\|_{\mathrm{F}}^2, \tag{4}$$

defined for all $Z, Z^* \in \mathscr{O}(d)^n$. Note that the minimization over $B \in \mathscr{O}(d)$ is necessary, since multiplying every Z_i^* by some $B \in \mathscr{O}(d)$ does not change the distribution of the observations. Our result is obtained under a setting that allows the possibility of missing interactions. Instead of observing Y_{ij} for all $1 \leq i < j \leq n$, we assume that each Y_{ij} is observed with probability p. In other words, we only observe (1) on a random graph that $A_{ij} \sim \text{Bernoulli}(p)$ independently for all $1 \leq i < j \leq n$. We summarize the main result of the paper on the $\mathscr{O}(d)$ synchronization as the following theorem.

THEOREM 1.1. For the $\mathscr{O}(d)$ Synchronization (2), assume $\frac{np}{\sigma^2} \to \infty$, $\frac{np}{\log n} \to \infty$ and $2 \leq d = O(1)$. Then, there exists some $\delta = o(1)$ such that

$$\inf_{\widehat{Z} \in \mathscr{O}(d)^n} \sup_{Z \in \mathscr{O}(d)^n} \mathbb{E}_Z \ell(\widehat{Z}, Z) \geq (1 - \delta) \frac{\sigma^2 d(d-1)}{2np}. \tag{5}$$

Moreover, the algorithm $\widehat{Z}$ described in Section 2.4 satisfies

$$\ell(\widehat{Z}, Z^*) \leq (1 + \delta) \frac{\sigma^2 d(d-1)}{2np}, \tag{6}$$

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

Though the result of Theorem 1.1 is stated in asymptotic forms, non-asymptotic versions under the assumptions $\frac{np}{\sigma^2} \geq c_1$ and $\frac{np}{\log n} \geq c_2$ for some sufficiently large constants $c_1, c_2 > 0$ are also presented in the paper, with the exact form of δ explicitly given in Theorem 2.1 and Theorem 3.1. The high-probability upper bound (6) immediately implies an in-expectation upper bound given the boundedness of the loss function that $\ell(Z, Z^*) \leq 4d$ for all $Z, Z^* \in \mathscr{O}(d)^n$. Since $\exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right) = o\left(\frac{\sigma^2 d(d-1)}{2np}\right)$, we have

$$\sup_{Z \in \mathscr{O}(d)^n} \mathbb{E}_Z \ell(\widehat{Z}, Z) \leq (1 + \delta) \frac{\sigma^2 d(d-1)}{2np} + \frac{4d}{n^8}, \tag{7}$$

for some $\delta = o(1)$. According to the proof of Theorem 1.1, the $4dn^{-8}$ in (7) can actually be improved to $4dn^{-C}$ for any constant $C > 0$. Therefore, if we additionally assume that $\sigma^2/p \geq n^{-c}$ for some constant $c > 0$, we will have

$$(1-\delta)\frac{\sigma^2 d(d-1)}{2np} \leq \inf_{\widehat{Z}\in\mathscr{O}(d)^n} \sup_{Z\in\mathscr{O}(d)^n} \mathbb{E}_Z \ell(\widehat{Z}, Z) \leq (1+\delta)\frac{\sigma^2 d(d-1)}{2np},$$

for some $\delta = o(1)$. Hence, $\frac{\sigma^2 d(d-1)}{2np}$ is the exact asymptotic minimax risk for $\mathscr{O}(d)$ synchronization. We remark that $\frac{\sigma^2 d(d-1)}{2np}$ is intuitive to understand, since σ^2 is the noise level, np is the effective sample size, and $\frac{d(d-1)}{2}$ is the degrees of freedom of $\mathscr{O}(d)$. In additional to Theorem 1.1, we also obtain a very similar result for the minimax risk of $\mathscr{SO}(d)$ synchronization. See Theorem 2.2 and Theorem 3.2 for the exact statement.

To achieve the minimax optimality, we consider an iterative polar decomposition procedure that projects the matrix $\sum_{j\in[n]\setminus\{i\}} A_{ij}Y_{ij}Z_j^{(t-1)}$ to the space $\mathscr{O}(d)$ at the tth iteration. This algorithm can be viewed as an approximation to the maximum likelihood estimator, and is known under other names such as generalized power method [4, 14, 23, 27] and projected power method [7] in the literature. We establish a sharp statistical error bound for the evolution of the algorithm and shows that the error decays exponentially to the optimal $\frac{\sigma^2 d(d-1)}{2np}$ as long as the algorithm is initialized by a spectral method [1, 32]. Our lower bound analysis is a careful application of the celebrated van Trees' inequality [17]. It complements the Cramér–Rao lower bound derived by [6] for the set of unbiased estimators.

Let us give some very brief comments on the assumptions of Theorem 1.1. Our paper is focused on the setting where d does not grow with the sample size n. This covers the most interesting applications in the literature for $d = 3$, though an extension of our result to a growing d would also be theoretically interesting. We exclude the case $d = 1$, because $\mathscr{SO}(1)$ is a degenerate set, and the problem over $\mathscr{O}(1) = \{-1, 1\}$ is known as $\mathbb{Z}_2$ synchronization, whose minimax rate has already been derived in the literature [11, 16]. It is interesting to note that the minimax rate of $\mathbb{Z}_2$ synchronization is exponential instead of the polynomial rate of $\mathscr{O}(d)$ synchronization for $d \geq 2$. When $d = O(1)$, the condition $\frac{np}{\sigma^2} \to \infty$ is equivalent to the minimax risk being vanishing. Since the loss function is bounded, the condition $\frac{np}{\sigma^2} \to \infty$ can also be viewed as necessary for the minimax risk to have a nontrivial rate. We remark that a nontrivial estimation when $\frac{np}{\sigma^2} \asymp 1$ is still possible, but that requires a very different technique of approximate message passing (AMP) [28] that does not apply to $\frac{np}{\sigma^2} \to \infty$. Finally, the condition $\frac{np}{\log n} \to \infty$ guarantees that the random graph is connected with high probability so that synchronization up to a global phase ambiguity is possible.

1.1 *Related literature*

One popular method for group synchronization is semi-definite programming (SDP) [1, 32]. The tightness of SDP and other forms of convex relaxation has been studied by [10, 24, 35]. In particular, it is shown by [24] that SDP is tight for $\mathscr{O}(d)$ synchronization when $\sigma^2 \lesssim \sqrt{n}$ in the setting of $p = 1$. The papers [1, 5, 29, 32] have studied spectral methods and its asymptotic error behavior. In terms of statistical estimation error, [23] and [25] have derived error bounds for the generalized power method and the spectral method for ℓ_∞-type loss functions in the setting of $p = 1$ with a general d that can potentially grow. In particular, both rates are $\frac{\sigma^2 \log n}{n}$ when d is bounded by some constant. For partial observations, the analysis of [27] for the generalized power method applied to $p < 1$, but they require $p \gtrsim n^{-1/2}$ for a nontrivial result.

When $d = 2$, $\mathscr{SO}(d)$ synchronization is also known as angular/phase synchronization and has been extensively studied in the literature [31]. The tightness of SDP has been established by [2, 36]. The convergence property and statistical estimation error of the generalized power method are studied by [26, 36]. The work that is mostly related to us is [15] that derives the minimax risk of phase synchronization. The analysis of [15] relies critically on the representation of an $\mathscr{SO}(2)$ element as a unit complex number. However, as soon as $d \geq 3$, elements of the groups $\mathscr{O}(d)$ and $\mathscr{SO}(d)$ are general non-commutative matrices, and the techniques in [15] cannot be applied to derive Theorem 1.1. See Section 2.2 for a detailed discussion.

1.2 *Paper organization*

The rest of the paper is organized as follows. In Section 2, we analyze the error decay of the iterative polar decomposition algorithm and the statistical property of the initialization procedure. This leads to the upper bound results for $\mathscr{O}(d)$ and $\mathscr{SO}(d)$ synchronizations. The lower bounds are derived in Section 3. Finally, Section 4 collects all the technical proofs of the paper.

1.3 *Notation*

For $d \in \mathbb{N}$, we write $[d] = \{1, \dots, d\}$. Given $a, b \in \mathbb{R}$, we write $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For a set S, we use $\mathbb{I}\{S\}$ and $|S|$ to denote its indicator function and cardinality, respectively. The notation $\mathbb{1}_d$ denotes a vector of all ones. For a matrix $B = (B_{ij}) \in \mathbb{R}^{d_1 \times d_2}$, the Frobenius norm, matrix ℓ_∞ norm, and operator norm of B are defined by $\|B\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{d_1} \sum_{j=1}^{d_2} |B_{ij}|^2}$, $\|B\|_{\ell_\infty} = \max_{1 \leq i \leq d_1} \sum_{j=1}^{d_2} |B_{ij}|$ and $\|B\|_{\mathrm{op}} = s_{\max}(B)$, and we use $s_{\min}(B)$ and $s_{\max}(B)$ for the smallest and the largest singular values of B. For $U, V \in \mathbb{R}^{d_1 \times d_2}$, $U \circ V \in \mathbb{R}^{d_1 \times d_2}$ is the Hadamard product $U \circ V = (U_{ij} V_{ij})$, and the trace inner product is $\mathsf{Tr}(UV^{\mathrm{T}}) = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} U_{ij} V_{ij}$. The notation $\det(\cdot)$, $\mathrm{vec}(\cdot)$ and $\otimes$ are used for determinant, vectorization and Kronecker product. For two integers $d_1 \geq d_2$, define $\mathscr{O}(d_1, d_2) = \{U \in \mathbb{R}^{d_1 \times d_2} : U^{\mathrm{T}} U = I_{d_2}\}$ so that $\mathscr{O}(d) = \mathscr{O}(d, d)$. The notation $\mathbb{P}$ and $\mathbb{E}$ are generic probability and expectation operators whose distribution is determined from the context. For two positive sequences $\{a_n\}$ and $\{b_n\}$, $a_n \lesssim b_n$ or $a_n = O(b_n)$ means $a_n \leq C b_n$ for some constant $C > 0$ independent of n. We also write $a_n = o(b_n)$ or $\frac{b_n}{a_n} \to \infty$ when $\limsup_n \frac{a_n}{b_n} = 0$.

2. Optimality of iterative polar decomposition

In this section, we derive the upper bound parts of the main results. We first investigate the $\mathscr{O}(d)$ synchronization in Sections 2.1– 2.4 and then extend the results to the $\mathscr{SO}(d)$ synchronization in Section 2.5.

2.1 *The algorithm*

For a squared matrix $X \in \mathbb{R}^{d \times d}$ that is of full rank, it admits a singular value decomposition (SVD) $X = UDV^{\mathrm{T}}$ with $U, V \in \mathscr{O}(d)$ and D being diagonal. Then, the polar decomposition of X is given by $X = PQ$, where $P = UV^{\mathrm{T}}$ and $Q = VDV^{\mathrm{T}}$. We denote the first factor, which is called generalized phase, by

$$\mathscr{P}(X) = UV^{\mathrm{T}}. \tag{8}$$

It is well known [18] that $\mathscr{P}(X)$ can also be defined by

$$\mathscr{P}(X) = \underset{B\in\mathscr{O}(d)}{\operatorname{argmin}} \|B - X\|_{\mathrm{F}}^2. \tag{9}$$

The operator $\mathscr{P}(\cdot)$ satisfies the following properties:

1. For any $c > 0$, $\mathscr{P}(X) = \mathscr{P}(cX)$.
2. For any $R \in \mathscr{O}(d)$, $\mathscr{P}(RX) = R\mathscr{P}(X)$ and $\mathscr{P}(XR^{\mathrm{T}}) = \mathscr{P}(X)R^{\mathrm{T}}$.
3. Suppose $X = X^{\mathrm{T}}$ and is positive definite, then $\mathscr{P}(X) = I_d$.

The iterative polar decomposition algorithm for the $\mathscr{O}(d)$ synchronization is given by the following iteration,

$$Z_i^{(t)} = \begin{cases} \mathscr{P}\left(\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}Z_j^{(t-1)}\right), & \det\left(\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}Z_j^{(t-1)}\right) \neq 0, \\ Z_i^{(t-1)}, & \det\left(\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}Z_j^{(t-1)}\right) = 0, \end{cases} \tag{10}$$

starting from some initialization $\{Z_i^{(0)}\}_{i\in[n]}$. To understand (10), we can consider the situation where $\{Z_j^*\}_{j\in[n]\backslash\{i\}}$ are known. Then, to estimate Z_i^*, one can apply the maximum likelihood estimator (MLE) that solves the following optimization problem:

$$\min_{Z_i\in\mathscr{O}(d)} \sum_{j\in[n]\backslash\{i\}} A_{ij}\|Y_{ij} - Z_iZ_j^{*T}\|_{\mathrm{F}}^2.$$

With some straightforward arrangement of the objective function and (9), the minimum is achieved by $\mathscr{P}\left(\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}Z_j^*\right)$ as long as $\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}Z_j^*$ has full rank. Thus, the iteration (10) can be thought of as a local MLE step with the unknown $\{Z_j^*\}_{j\in[n]\backslash\{i\}}$ replaced by $\{Z_j^{(t-1)}\}_{j\in[n]\backslash\{i\}}$ from the last step.

2.2 *An oracle perturbation bound*

In this section, we give an oracle analysis for (10) to understand its statistical properties. An iterative algorithm that is similar to (10) has been analyzed by [15] in the context of phase synchronization. However, the argument used in [15] that leads to the correct constant is limited to the phase synchronization and cannot be used for the $\mathscr{O}(d)$ or $\mathscr{SO}(d)$ synchronization. In the following, we first summarize the analysis in [15] for the phase synchronization and then present our new analysis for the $\mathscr{O}(d)$ synchronization to achieve the correct constant.

As we have mentioned in Section 1, phase synchronization is $\mathscr{SO}(d)$ synchronization with $d = 2$. Since a rotation matrix in $\mathbb{R}^2$ is parametrized by an angle, we can equivalently set up the problem via complex numbers. That is, we have $Y_{ij} = z_i^*\bar{z}_j^* + \sigma W_{ij} \in \mathbb{C}$ for $1 \leq i < j \leq n$. Each z_i^* is a complex number with norm 1 and $\bar{z}_j^*$ stands for the complex conjugate of z_j^*. The noise variable W_{ij} is standard complex Gaussian. Suppose we have partial observations on a random graph $\{A_{ij}\}_{1\leq i<j\leq n}$, the generalized power method [4, 12, 28] is given by the following iteration,

$$z_i^{(t)} = \begin{cases} \frac{\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}z_j^{(t-1)}}{\left|\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}z_j^{(t-1)}\right|}, & \left|\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}z_j^{(t-1)}\right| \neq 0, \\ z_i^{(t-1)}, & \left|\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}z_j^{(t-1)}\right| = 0. \end{cases} \tag{11}$$

It was shown by [15] that (11) achieves the optimal statistical error with a sharp leading constant after sufficient steps of iterations. The key mathematical ingredient in the analysis of [15] is the understanding of a one-step iteration error starting from the truth $z^{(t-1)} = z^*$. That is, we define

$$\check{z}_i = \frac{\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}z_j^*}{\left|\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}z_j^*\right|}, \tag{12}$$

and our goal is to give a sharp bound for $|\check{z}_i - z_i^*|^2$. We can easily rearrange the right hand side of (12) as $\check{z}_i = z_i^* \frac{1+e_i}{|1+e_i|}$, with $e_i = \frac{\sigma \sum_{j\in[n]\backslash\{i\}} A_{ij}W_{ij}\bar{z}_i^* z_j^*}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}$. Then,

$$|\check{z}_i - z_i^*|^2 = \left|\frac{1+e_i}{|1+e_i|} - 1\right|^2 \leq \frac{|\mathrm{Im}(e_i)|^2}{|\mathrm{Re}(1+e_i)|^2}, \tag{13}$$

where the inequality above is by the fact that

$$\left|\frac{x}{|x|} - 1\right| \leq \left|\frac{\mathrm{Im}(x)}{}\mathrm{Re}(x)\right| \quad \text{for any } x \in \mathbb{C} \text{ such that } \mathrm{Re}(x)\ \ 0. \tag{14}$$

For a proof of (14), see Lemma 5.6 in [15]. Since it can be shown that e_i is small, the denominator of (13) is close to 1. The numerical of (13) can be accurately controlled by the Gaussianity of e_i conditioning on the random graph $\{A_{ij}\}_{1\leq i<j\leq n}$. To summarize, the first order behavior of $|\check{z}_i - z_i^*|^2$ is determined by $|\mathrm{Im}(e_i)|^2$, which leads to the optimal error of phase synchronization with a sharp constant in [15].

The above analysis relies on (14) and critically on the representation of an $\mathscr{SO}(2)$ element as a unit complex number. Next, we present our new analysis for the $\mathscr{O}(d)$ synchronization. To understand the statistical property of (10) for the $\mathscr{O}(d)$ synchronization, let us similarly consider an oracle setting with $Z_i^{(t-1)} = Z_i^*$ for all $i \in [n]$. Define

$$\check{Z}_i = \mathscr{P}\left(\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}Z_j^*\right), \tag{15}$$

and our goal is to bound $\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2$. Compared with (12), the formula (15) does not have a closed form anymore, and the inequality (14) that only applies to complex numbers does not have a straightforward extension to general orthogonal matrices.

By the property of $\mathscr{P}(\cdot)$, let us first write write

$$\check{Z}_i = \mathscr{P}\left(\frac{\sum_{j\in[n]\backslash\{i\}} A_{ij} Y_{ij} Z_j^*}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}\right) = \mathscr{P}\left(Z_i^* + E_i\right),$$

where the error matrix is given by

$$E_i = \sigma \frac{\sum_{j\in[n]\backslash\{i\}} A_{ij} W_{ij} Z_j^*}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}.$$

Note that $Z_i^* = \mathscr{P}(Z_i^*)$, and thus, bounding $\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2$ requires a perturbation analysis of the operator $\mathscr{P}(\cdot)$, which is given by the following lemma.

LEMMA 2.1 (Theorem 1 of [22]). Let $X, \widetilde{X} \in \mathbb{R}^{d\times d}$ be two matrices of full rank. Then,

$$\|\mathscr{P}(X) - \mathscr{P}(\widetilde{X})\|_{\mathrm{F}} \leq \frac{2}{s_{\min}(X) + s_{\min}(\widetilde{X})} \|X - \widetilde{X}\|_{\mathrm{F}}.$$

By Lemma 2.1, we have

$$\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2 \leq \left(1 - \|E_i\|_{\mathrm{op}}/2\right)^{-2} \|E_i\|_{\mathrm{F}}^2. \tag{16}$$

Given $\{A_{ij}\}$, the conditional expectation of $\|E_i\|_{\mathrm{F}}^2$ is $\frac{\sigma^2 d^2}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}$. Moreover, $\sum_{j\in[n]\backslash\{i\}} A_{ij}$ concentrates around $(n-1)p$. Therefore, it can be shown that

$$\|E_i\|_{\mathrm{F}}^2 = (1 + o_{\mathbb{P}}(1)) \frac{\sigma^2 d^2}{np}, \tag{17}$$

under appropriate conditions. When $\frac{\sigma^2 d^2}{np} = o(1)$, we have the bound

$$\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2 \leq (1 + o_{\mathbb{P}}(1)) \frac{\sigma^2 d^2}{np}. \tag{18}$$

However, compared with our minimax lower bound in Theorem 3.1, it is clear that the error bound $\frac{\sigma^2 d^2}{np}$ is not optimal. This is due to a naive application of Lemma 2.1. Below, we present an improved analysis of $\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2$. Recall the definition of E_i and the properties of $\mathscr{P}(\cdot)$, and we have

$$\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2 = \|\mathscr{P}(I_d + E_i Z_i^{*\mathrm{T}}) - I_d\|_{\mathrm{F}}^2.$$

The key observation is that we can write I_d as the operator $\mathscr{P}(\cdot)$ applied to any positive definite matrix. With this idea, we have

$$\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2 = \|\mathscr{P}(I_d + E_i Z_i^{*\mathrm{T}}) - \mathscr{P}\,(\text{any positive definite matrix})\,\|_{\mathrm{F}}^2. \tag{19}$$

We shall choose a positive definite matrix whose difference from $I_d + E_i Z_i^*$ is as small as possible and then apply Lemma 2.1. It turns out a correct choice is $I_d + \frac{1}{2}E_i Z_i^* + \frac{1}{2}Z_i^{*\mathrm{T}}E_i^{\mathrm{T}}$. When $\frac{\sigma^2 d^2}{np} = o(1)$, we can view $I_d + \frac{1}{2}E_i Z_i^* + \frac{1}{2}Z_i^{*\mathrm{T}}E_i^{\mathrm{T}}$ as a small perturbation from I_d by (17), and thus, it is positive definite. Apply Lemma 2.1, and we have

$$\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2 = \left\| \mathscr{P}(I_d + E_i Z_i^{*\mathrm{T}}) - \mathscr{P}\left(I_d + \frac{1}{2}E_i Z_i^* + \frac{1}{2}Z_i^{*\mathrm{T}}E_i^{\mathrm{T}}\right)\right\|_{\mathrm{F}}^2$$
$$\leq \left(1 - \|E_i\|_{\mathrm{op}}\right)^{-2} \left\|\frac{1}{2}E_i Z_i^{*\mathrm{T}} - \frac{1}{2}Z_i^* E_i^{\mathrm{T}}\right\|_{\mathrm{F}}^2.$$

Compared with the previous bound (16), the $\|E_i\|_{\mathrm{F}}^2$ in (16) has been improved to $\|\frac{1}{2}E_i Z_i^{*\mathrm{T}} - \frac{1}{2}Z_i^* E_i^{\mathrm{T}}\|_{\mathrm{F}}^2$. To see why this is an improvement, we have two simple observations:

1. The diagonal entries of $\frac{1}{2}E_i Z_i^{*\mathrm{T}} - \frac{1}{2}Z_i^* E_i^{\mathrm{T}}$ are all zero, whereas those of E_i are all nonzero.
2. The off-diagonal entries of $\frac{1}{2}E_i Z_i^{*\mathrm{T}} - \frac{1}{2}Z_i^* E_i^{\mathrm{T}}$ have smaller variance. For any $1 \leq a < b \leq d$, we have $\left(\frac{1}{2}E_i Z_i^{*\mathrm{T}} - \frac{1}{2}Z_i^* E_i^{\mathrm{T}}\right)_{ab} | A \sim \mathscr{N}\left(0, \frac{\sigma^2}{2\sum_{j\in[n]\backslash\{i\}} A_{ij}}\right)$, compared with $(E_i)_{ab}|A \sim \mathscr{N}\left(0, \frac{\sigma^2}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}\right)$.

By direct calculation, the conditional expectation of $\|\frac{1}{2}E_i Z_i^{*\mathrm{T}} - \frac{1}{2}Z_i^* E_i^{\mathrm{T}}\|_{\mathrm{F}}^2$ given $\{A_{ij}\}$ is $\frac{\sigma^2 d(d-1)}{2\sum_{j\in[n]\backslash\{i\}} A_{ij}}$, and it can be shown that

$$\left\|\frac{1}{2}E_i Z_i^{*\mathrm{T}} - \frac{1}{2}Z_i^* E_i^{\mathrm{T}}\right\|_{\mathrm{F}}^2 = (1 + o_{\mathbb{P}}(1))\frac{\sigma^2 d(d-1)}{2np},$$

under appropriate conditions. This leads to the bound

$$\|\check{Z}_i - Z_i^*\|_{\mathrm{F}}^2 \leq (1 + o_{\mathbb{P}}(1))\frac{\sigma^2 d(d-1)}{2np}, \tag{20}$$

which is optimal in view of the minimax lower bound in Theorem 3.1. Compared with d^2 in (18), the factor $\frac{1}{2}d(d-1)$ in (20) is the correct degrees of freedom of the space $\mathscr{O}(d)$.

2.3 *Analysis of the iteration*

Having understood how a one-step iteration (10) would achieve the optimal statistical error bound if it were started from the truth, we are ready to analyze the evolution of (10) starting from an initialization that is close to the truth. Let us first shorthand the formula (10) by

$$Z^{(t)} = f(Z^{(t-1)}).$$

In other words, we have introduced a map $f : \mathscr{O}(d)^n \to \mathscr{O}(d)^n$ such that $f(Z^{(t-1)})_i$ is defined by (10). We characterize the evolution of the loss function (4) through the map f by the following lemma.

LEMMA 2.2. For the $\mathscr{O}(d)$ synchronization (2), assume $\frac{np}{\sigma^2} \geq c_1$ and $\frac{np}{\log n} \geq c_2$ for some sufficiently large constants $c_1, c_2 > 0$ and $2 \leq d \leq C$ for some constant $C > 0$. Then, for any $\gamma \in [0, 1/16)$, we have

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

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

To understand the consequence of Lemma 2.2, we can first do a sanity check by setting $Z = Z^*$. This results in the bound

$$\ell(f(Z^*), Z^*) \leq (1+\delta_2)\frac{\sigma^2 d(d-1)}{2np},$$

which is the one-step iteration error starting from the truth, and thus, the oracle analysis in Section 2.2 is recovered.

More generally, as long as $Z^{(t-1)}$ satisfies $\ell(Z^{(t-1)}, Z^*) \leq \gamma$ for some $\gamma \in (0, 1/16)$, we have

$$\ell(Z^{(t)}, Z^*) \leq \delta_1 \ell(Z^{(t-1)}, Z^*) + (1+\delta_2)\frac{\sigma^2 d(d-1)}{2np}. \tag{21}$$

From (21), we know that $\ell(Z^{(t)}, Z^*) \leq \delta_1\gamma + (1+\delta_2)\frac{\sigma^2 d(d-1)}{2np}$, which can again be bounded by γ under the condition that $\frac{np}{\sigma^2} \geq c_1$ and $\frac{np}{\log n} \geq c_2$ for some sufficiently large constants $c_1, c_2 > 0$. By mathematical induction, we can conclude that (21) holds for all $t \geq 1$ as long as $\ell(Z^{(0)}, Z^*) \leq \gamma$. We can also rearrange the one-step iteration bound (21) into a linear convergence result,

$$\ell(Z^{(t)}, Z^*) \leq \delta_1^t \ell(Z^{(0)}, Z^*) + \frac{1+\delta_2}{1-\delta_1}\frac{\sigma^2 d(d-1)}{2np}, \tag{22}$$

for all $t \geq 1$. The bound (22) implies that $Z^{(t)}$ will eventually be statistically optimal after sufficient steps of iterations. However, it does not imply that $Z^{(t)}$ converges to any fixed object. As is shown by [23, 36] via a leave-one-out argument, the algorithmic convergence of the iterative algorithm requires the condition $\sigma^2 = O\left(\frac{n}{\log n}\right)$ at least when $p = 1$. In comparison, the bound (22) guarantees the statistical optimality without characterizing its convergence property and thus only requires $\frac{np}{\sigma^2}$ to be sufficiently large.

As we have discussed in the introduction section, the two conditions $\frac{np}{\sigma^2} \geq c_1$ and $\frac{np}{\log n} \geq c_2$ are essentially necessary for the result to hold, at least when d is bounded by a constant. If both conditions are slightly strengthened to $\frac{np}{\sigma^2} \to \infty$ and $\frac{np}{\log n} \to \infty$, the same result of Lemma 2.2 will hold with vanishing δ_1 and δ_2.

The proof of Lemma 2.2 also holds more generally for d that can potentially grow. Without assuming $d \leq C$ for some constant $C > 0$, we would obtain the same high probability bound with $\delta_1 = C_1\sqrt{\frac{d\log n + d^2\sigma^2}{np}}$, $\delta_2 = C_2\left(\gamma^2 + \frac{\log n + d\sigma^2}{np}\right)^{1/4}$, and the conditions replaced by $\frac{np}{d^2\sigma^2} \geq c_1$ and $\frac{np}{d\log n} \geq c_2$.

2.4 *Optimal upper bound*

To derive the minimax upper bound for the $\mathscr{O}(d)$ synchronization from Lemma 2.2, we need to construct an initialization $Z^{(0)} \in \mathscr{O}(d)^n$ whose statistical error $\ell(Z^{(0)}, Z^*)$ is sufficiently small. Let us first organize the observations $\{Y_{ij}\}_{1\leq i<j\leq n}$ into a matrix $Y \in \mathbb{R}^{nd\times nd}$. That is,

$$Y = \begin{pmatrix} Y_{11} & \cdots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{n1} & \cdots & Y_{nn} \end{pmatrix}, \tag{23}$$

where $Y_{ji} = Y_{ij}$ for all $1 \leq i < j \leq n$ and $Y_{ii} = I_d$ for all $i \in [n]$. The noise matrices $\{W_{ij}\}_{1\leq i<j\leq n}$ can be organized into $W \in \mathbb{R}^{nd\times nd}$ with the same arrangement as (23), and we set $W_{ji} = W_{ij}$ for all $1 \leq i < j \leq n$ and $W_{ii} = 0$ for all $i \in [n]$. Then, we can write the model (1) as

$$Y = Z^*Z^{*\mathrm{T}} + \sigma W,$$

where $Z^{*\mathrm{T}} = (Z_1^{*\mathrm{T}}, \cdots, Z_n^{*\mathrm{T}}) \in \mathbb{R}^{d\times nd}$. In other words, Y can be viewed as a noisy version of the rank-d matrix $Z^*Z^{*\mathrm{T}}$, and thus, we can use a spectral method to estimate the column space of Z^*. Since we do not observe all Y_{ij}'s, the spectral method can be applied to $(A \otimes \mathbb{1}_d\, \mathbb{1}_d^T) \circ Y$, where $A \in \{0,1\}^{n\times n}$ with $A_{ji} = A_{ij}$ for all $1 \leq i < j \leq n$ and $A_{ii} = 0$ for all $i \in [n]$. Recall that $\otimes$ stands for the matrix Kronecker product and $\circ$ denotes the Hadamard product. To compute $Z^{(0)} \in \mathscr{O}(d)^n$, we first find

$$\widehat{U} = \underset{U\in\mathscr{O}(nd,d)}{\operatorname{argmax}} \operatorname{Tr}\left(U^{\mathrm{T}}\left((A \otimes \mathbb{1}_d\, \mathbb{1}_d^T) \circ Y\right) U\right), \tag{24}$$

and then compute

$$Z_i^{(0)} = \begin{cases} \mathscr{P}(\widehat{U}_i), & \det(\widehat{U}_i) \neq 0, \\ I_d, & \det(\widehat{U}_i) = 0, \end{cases} \tag{25}$$

for all $i \in [n]$. Here, $\widehat{U}_i$ stands for the ith $d \times d$ block of $\widehat{U}$ and thus $\widehat{U}^{\mathrm{T}} = (\widehat{U}_1^{\mathrm{T}}, \cdots, \widehat{U}_n^{\mathrm{T}})$. The error bound of $Z^{(0)}$ is given by the following lemma.

LEMMA 2.3. For the $\mathscr{O}(d)$ synchronization (2), assume $\frac{np}{\log n} \geq c$ for some sufficiently large constants $c > 0$ and $2 \leq d \leq C_1$ for some constant $C_1 > 0$. Then, we have

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

with probability at least $1 - n^{-9}$ for some constant $C > 0$ only depending on C_1.

The error rate $\frac{\sigma^2+1}{np}$ is the sum of two terms. The first term $\frac{\sigma^2}{np}$ is from the additive Gaussian noise in the model (1). The second term $\frac{1}{np}$ is a consequence of the randomness from the graph. It comes from an upper bound $\|A-\mathbb{E}A\|_{\mathrm{op}} \lesssim \sqrt{np}$. In fact, it can be slightly improved by $\|A-\mathbb{E}A\|_{\mathrm{op}} \lesssim \sqrt{n(p \wedge (1-p))}$, which makes a difference when $1-p$ is small. This leads to the second term $\frac{1}{np}$ replaced by $\frac{p\wedge(1-p)}{np^2}$. As a result, when $\sigma^2 = 0$ and $p = 1$, we have $\ell(Z^{(0)}, Z^*) = 0$, i.e. perfect recovery of Z^*.

By Lemma 2.3, we know that $\ell(Z^{(0)}, Z^*)$ is sufficiently small when $\frac{np}{\sigma^2}$ and $\frac{np}{\log n}$ are sufficiently large. Then, we can directly apply Lemma 2.2 and its consequence (22) to obtain the desired upper bound.

Theorem 2.1. For the $\mathscr{O}(d)$ synchronization (2), assume $\frac{np}{\sigma^2} \geq c_1$ and $\frac{np}{\log n} \geq c_2$ for some sufficiently large constants $c_1, c_2 > 0$ and $2 \leq d \leq C_1$ for some constant $C_1 > 0$. Consider the algorithm (10) initialized by (25). We have

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

for all $t \geq \log\left(\frac{1}{\sigma^2}\right)$ with probability at least $1 - 2n^{-9} - \exp\left(-\left(\frac{np}{\sigma^2}\right)^{1/4}\right)$ for some constant $C > 0$ only depending on C_1.

Remark 2.1. The proof of Lemma 2.3 also holds more generally for d that can potentially grow. When d grows, the initialization has the error bound $\ell(Z^{(0)}, Z^*) \leq C\frac{d^4(d\sigma^2+1)}{np}$ with high probability. As a consequence, without assuming $d \leq C_1$, the result of Theorem 2.1 becomes

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

with high probability under the conditions that $\frac{np}{d\log n}$, $\frac{np}{d^2\sigma^2}$ and $\frac{np}{d^4}$ are sufficiently large. The same iterative algorithm has also been analyzed by [23] for an ℓ_∞ type loss when $p = 1$, and they showed that

$$\min_{B\in\mathscr{O}(d)} \max_{i\in[n]} \|Z_i^{(t)} - Z_i^* B\|_{\mathrm{F}}^2 \leq C\frac{\sigma^2(d^2 + d\log n)}{n},$$

with high probability under the condition that $\frac{n}{\sigma^2(d^2+d\log n)}$ is sufficiently large.

2.5 *Rotation group synchronization*

In this section, we study $\mathscr{SO}(d)$ synchronization, where our goal is to estimate $Z^* \in \mathscr{SO}(d)^n$ from the observations (1) on a random graph. The loss function for this problem is defined by

$$\bar{\ell}(Z, Z^*) = \min_{B\in\mathscr{SO}(d)} \frac{1}{n}\sum_{i=1}^{n} \|Z_i - Z_i^* B\|_{\mathrm{F}}^2,$$

for any $Z, Z^* \in \mathscr{SO}(d)^n$. The $\mathscr{SO}(d)$ synchronization problem requires a slight modification of the iterative algorithm (10). To do this, we first introduce an operator $\bar{\mathscr{P}}(\cdot)$ that maps a $d \times d$ full-rank matrix to $\mathscr{SO}(d)$. For a full-rank squared matrix $X \in \mathbb{R}^{d\times d}$ that admits an SVD $X = UDV^{\mathrm{T}}$, we define

$$\bar{\mathscr{P}}(X) = U \begin{pmatrix} I_{d-1} & 0 \\ 0 & \det(UV^{\mathrm{T}}) \end{pmatrix} V^{\mathrm{T}}.$$

The only difference from $\mathscr{P}(X)$ is the diagonal matrix with the last entry $\det(UV^{\mathrm{T}})$ sandwiched between U and V. It is clear that $\det(UV^{\mathrm{T}}) \in \{-1, 1\}$ and thus $\det(\bar{\mathscr{P}}(X)) = 1$, which implies $\bar{\mathscr{P}}(X) \in \mathscr{SO}(d)$. By [20], $\bar{\mathscr{P}}(X)$ can also be characterized as the solution to an optimization problem. That is,

$$\bar{\mathscr{P}}(X) = \operatorname*{argmin}_{B \in \mathscr{SO}(d)} \|B - X\|_{\mathrm{F}}^2.$$

Then, similar to the motivation behind the iteration (10), we consider an iterative procedure for $\mathscr{SO}(d)$ synchronization,

$$Z_i^{(t)} = \begin{cases} \bar{\mathscr{P}}\left(\sum_{j\in[n]\setminus\{i\}} A_{ij} Y_{ij} Z_j^{(t-1)}\right), & \det\left(\sum_{j\in[n]\setminus\{i\}} A_{ij} Y_{ij} Z_j^{(t-1)}\right) \neq 0, \\ Z_i^{(t-1)}, & \det\left(\sum_{j\in[n]\setminus\{i\}} A_{ij} Y_{ij} Z_j^{(t-1)}\right) = 0. \end{cases} \tag{26}$$

The iteration (26) enjoys a similar convergence property as given by Lemma 2.2 with a good initialization. This result is stated as Lemma 4.7 in Section 4.3. Mathematically, one can show that as long as $\bar{\ell}(Z^{(t-1)}, Z^*) \leq \gamma$ for some sufficiently small γ, the determinant of $\sum_{j\in[n]\setminus\{i\}} A_{ij} Y_{ij} Z_j^{(t-1)}$ is positive for most $i \in [n]$ so that $\bar{\mathscr{P}}(\cdot) = \mathscr{P}(\cdot)$ for those i's. With this argument, the same proof that leads to the conclusion of Lemma 2.2 also characterizes the convergence of (26).

To initialize the iterative procedure (26), we also use a spectral method. Define

$$Z_i^{(0)} = \begin{cases} \bar{\mathscr{P}}(\widehat{U}_i), & \det(\widehat{U}_i) \neq 0, \\ I_d, & \det(\widehat{U}_i) = 0, \end{cases} \tag{27}$$

where $\widehat{U}^{\mathrm{T}} = (\widehat{U}_1^{\mathrm{T}}, \cdots, \widehat{U}_n^{\mathrm{T}})$ is given by (24). It is clear that $Z_i^{(0)} \in \mathscr{SO}(d)$ for all $i \in [n]$. We show in Lemma 4.8 that $\bar{\ell}(Z^{(0)}, Z^*)$ is sufficiently small. Together with the statistical property of the iterative procedure (26), we have the following upper bound result for $\mathscr{SO}(d)$ synchronization.

THEOREM 2.2. For the $\mathscr{SO}(d)$ synchronization (3), assume $\frac{np}{\sigma^2} \geq c_1$ and $\frac{np}{\log n} \geq c_2$ for some sufficiently large constants $c_1, c_2 > 0$ and $2 \leq d \leq C_1$ for some constant $C_1 > 0$. Consider the algorithm (26) initialized by (27). We have

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

for all $t \geq \log\left(\frac{1}{\sigma^2}\right)$ with probability at least $1 - 2n^{-9} - \exp\left(-\left(\left(\frac{np}{\sigma^2}\right)^{1/4}\right)\right)$ for some constant $C > 0$ only depending on C_1.

3. Minimax lower bound

In this section, we derive the lower bound part of the main result. We first consider $\mathscr{O}(d)$ synchronization, and the minimax risk is given by

$$\inf_{\widehat{Z}\in\mathscr{O}(d)^n}\sup_{Z\in\mathscr{O}(d)^n}\mathbb{E}_Z\ell(\widehat{Z},Z)=\inf_{\widehat{Z}\in\mathscr{O}(d)^n}\sup_{Z\in\mathscr{O}(d)^n}\mathbb{E}_Z\left[\min_{B\in\mathscr{O}(d)}\frac{1}{n}\sum_{i=1}^n\|\widehat{Z}_i-Z_iB\|_{\mathrm{F}}^2\right]. \tag{28}$$

Our first step is to apply Lemma 4.6 and lower bound the loss function $\ell(\widehat{Z},Z)$ by

$$\ell(\widehat{Z},Z)\geq\frac{1}{2n^2}\sum_{i=1}^n\sum_{j=1}^n\|\widehat{Z}_i\widehat{Z}_j^{\mathrm{T}}-Z_iZ_j^{\mathrm{T}}\|_{\mathrm{F}}^2. \tag{29}$$

Compared with $\ell(\widehat{Z},Z)$, the right hand side of (29) is a loss that is decomposable across all pairs $(i,j)\in[n]^2$, and therefore, it is sufficient to lower bound the estimation error of each individual $Z_iZ_j^{\mathrm{T}}$ and then to aggregate the results. Following this strategy, we have

$$\begin{aligned}&\inf_{\widehat{Z}\in\mathscr{O}(d)^n}\sup_{Z\in\mathscr{O}(d)^n}\mathbb{E}_Z\ell(\widehat{Z},Z)\\
&\geq\frac{1}{2n^2}\inf_{\widehat{Z}\in\mathscr{O}(d)^n}\sup_{Z\in\mathscr{O}(d)^n}\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}_Z\|\widehat{Z}_i\widehat{Z}_j^{\mathrm{T}}-Z_iZ_j^{\mathrm{T}}\|_{\mathrm{F}}^2\\
&\geq\frac{1}{2n^2}\inf_{\widehat{Z}\in\mathscr{O}(d)^n}\sum_{1\leq i\neq j\leq n}\int\mathbb{E}_Z\|\widehat{Z}_i\widehat{Z}_j^{\mathrm{T}}-Z_iZ_j^{\mathrm{T}}\|_{\mathrm{F}}^2\prod_{k=1}^n\mathrm{d}\Pi(Z_k)\end{aligned} \tag{30}$$

$$\geq\frac{1}{2n^2}\sum_{1\leq i\neq j\leq n}\int\left(\inf_{\widehat{T}}\int\int\mathbb{E}_Z\|\widehat{T}-Z_iZ_j^{\mathrm{T}}\|_{\mathrm{F}}^2\mathrm{d}\Pi(Z_i)\mathrm{d}\Pi(Z_j)\right)\prod_{k\in[n]\backslash\{i,j\}}\mathrm{d}\Pi(Z_k), \tag{31}$$

where Π is some probability distribution supported on $\mathscr{O}(d)$ to be specified later. The inequality (30) lower bounds the minimax risk with a Bayes risk, and (31) is lower bounding infimum of average by average of infimum. Now it suffices to lower bound

$$\inf_{\widehat{T}}\int\int\mathbb{E}_Z\|\widehat{T}-Z_iZ_j^{\mathrm{T}}\|_{\mathrm{F}}^2\mathrm{d}\Pi(Z_i)\mathrm{d}\Pi(Z_j), \tag{32}$$

for each $\{Z_k\}_{k\in[n]\backslash\{i,j\}}$ and each $i\neq j$. The quantity (32) can be understood as the Bayes risk of estimating $Z_iZ_j^{\mathrm{T}}$ given the knowledge of $\{Z_k\}_{k\in[n]\backslash\{i,j\}}$.

To analyze (32), we first need to construct a probability distribution Π on $\mathscr{O}(d)$. Given $\{r_{ab}\}_{1\leq a<b\leq d}$ and $\{s_{ab}\}_{1\leq b\leq a\leq d}$, we can form the following $d\times d$ matrix,

$$Q=\begin{pmatrix} s_{11} & r_{12} & r_{13} & \cdots & r_{1d} \\ s_{21} & s_{22} & r_{23} & \cdots & r_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{d-1,1} & s_{d-1,2} & s_{d-1,3} & \cdots & r_{d-1,d} \\ s_{d1} & s_{d2} & s_{d3} & \cdots & s_{dd} \end{pmatrix}. \tag{33}$$

In other words, $\{r_{ab}\}_{1\leq a<b\leq d}$ are the upper triangular elements of Q, and the lower triangular and the diagonal elements of Q are given by $\{s_{ab}\}_{1\leq b\leq a\leq d}$. We are going to specify the values of $\{s_{ab}\}_{1\leq b\leq a\leq d}$ by $\{r_{ab}\}_{1\leq a<b\leq d}$ so that Q is a matrix of degrees of freedom $\frac{1}{2}d(d-1)$. To do this, let us introduce some new notation. For each integer $a\geq 2$, $Q_{a-1}\in\mathbb{R}^{(a-1)\times(a-1)}$ is the submatrix of Q collecting the first $a-1$ rows and columns. We also define

$$S_{a-1}=\begin{pmatrix} s_{a1} \\ \vdots \\ s_{a,a-1}\end{pmatrix},\quad R_{a-1}=\begin{pmatrix} r_{1a} \\ \vdots \\ r_{a-1,a}\end{pmatrix}\quad\text{and}\quad v_{a-1}=\begin{pmatrix} r_{1,a+1}r_{a,a+1}+\cdots+r_{1d}r_{ad} \\ \vdots \\ r_{a-1,a+1}r_{a,a+1}+\cdots+r_{a-1,d}r_{ad}\end{pmatrix}.$$

All the three vectors above belong to $\mathbb{R}^{a-1}$. The construction of $\{s_{ab}\}_{1\leq b\leq a\leq d}$ is given by the following procedure.

1. We first set $s_{11}=\sqrt{1-\left(r_{12}^2+\cdots+r_{1d}^2\right)}$.
2. Given the values of $\{s_{ab}\}_{1\leq b\leq a\leq k-1}$, we find $s_{k1},s_{k2},\cdots,s_{kk}$ through the equations

$$Q_{k-1}S_{k-1}+s_{kk}R_{k-1}=-v_{k-1} \tag{34}$$

$$\|S_{k-1}\|^2+s_{kk}^2=1-\left(r_{k,k+1}^2+\cdots+r_{kd}^2\right). \tag{35}$$

   Note that the equations above have two sets of real solutions (under the assumption of Lemma 3.1). We take the set of solutions with the larger value of s_{kk}.

After going through the above procedure, the matrix Q in the form of (33) is fully determined by its upper triangular elements $\{r_{ab}\}_{1\leq a<b\leq d}$, and therefore, we can write $Q=Q(r)$. The equation (34) guarantees that the rows of $Q(r)$ are orthogonal to each other and (35) implies that each row of $Q(r)$ is a unit vector. The following lemma characterizes a sufficient condition on $\{r_{ab}\}_{1\leq a<b\leq d}$ that implies the existence of $Q(r)$.

LEMMA 3.1. Assume $\max_{1\leq a<b\leq d}|r_{ab}|\leq\frac{1}{8d^{5/2}}$. Then, $Q(r)$ is well defined and the following properties are satisfied:

1. $Q(r)\in\mathscr{SO}(d)$;
2. $\max_{1\leq b<a\leq d}|s_{ab}|\leq\frac{1}{4d^2}$ and $\min_{a\in[d]}s_{aa}\geq\frac{7}{8}$;
3. $\max_{1\leq a<b\leq d}\max_{u\in[d]}\sqrt{\sum_{v=1}^{u}\left|\frac{\partial s_{uv}}{\partial r_{ab}}\right|^2}\leq 5$.

According to Lemma 3.1, the constructed $Q(r)$ can also be used for deriving the minimax lower bound of $\mathscr{SO}(d)$ synchronization. Moreover, its entries and the derivatives with respect to $\{r_{ab}\}_{1\leq a<b\leq d}$ are well controlled, which means that the matrix $Q(r)$ is smoothly parametrized by r. Let P be a distribution under which $r_{ab} \sim \mu$ independently for all $1 \leq a < b \leq d$ with some μ being a smooth probability density function supported on $\left[-\frac{1}{8d^{5/2}}, \frac{1}{8d^{5/2}}\right]$. To be specific, we can set $\mu(t) \propto \exp\left(-\frac{1}{1-64d^5t^2}\right)\mathbb{I}\{|t| \leq 1/(8d^{5/2})\}$. Then, by letting Π be the induced probability measure of $Q(r)$ with $\{r_{ab}\}_{1\leq a<b\leq d} \sim P$, we can write (32) as

$$\inf_{\widehat{T}} \int\int \mathbb{E}_Z\|\widehat{T} - Z_i(r)Z_j(r')^{\mathrm{T}}\|_{\mathrm{F}}^2 \mathrm{d}P(r)\mathrm{d}P(r'), \tag{36}$$

where we have some slight abuse of notation that $Z_i = Z_i(r) = Q(r)$ and $Z_j = Z_j(r') = Q(r')$. Compared with (32), the distribution P in (36) is a standard probability measure on $\mathbb{R}^{\frac{d(d-1)}{2}}$. The Bayes risk (36) can then be lower bounded via van Trees' inequality [17].

Lemma 3.2. Assume $\frac{np}{\sigma^2} \geq c$ for some sufficiently large constants $c > 0$ and $2 \leq d \leq C_1$ for some constant $C_1 > 0$. Then, we have

$$\inf_{\widehat{T}} \int\int \mathbb{E}_Z\|\widehat{T} - Z_i(r)Z_j(r')^{\mathrm{T}}\|_{\mathrm{F}}^2 \mathrm{d}P(r)\mathrm{d}P(r') \geq \left(1 - C\left(\frac{1}{n} + \frac{\sigma^2}{np}\right)\right)\frac{\sigma^2 d(d-1)}{np},$$

for some constant $C > 0$ only depending on C_1.

The proof of Lemma 3.2, which verifies the technical conditions of [17], is given in Section 4.5. Note that these technical conditions are implied by the conclusion of Lemma 3.1. In view of the inequality (31), we immediately have the following theorem.

Theorem 3.1. Assume $\frac{np}{\sigma^2} \geq c$ for some sufficiently large constants $c > 0$ and $2 \leq d \leq C_1$ for some constant $C_1 > 0$. Then, we have

$$\inf_{\widehat{Z}\in\mathscr{O}(d)^n} \sup_{Z\in\mathscr{O}(d)^n} \mathbb{E}_Z \ell(\widehat{Z}, Z) \geq \left(1 - C\left(\frac{1}{n} + \frac{\sigma^2}{np}\right)\right)\frac{\sigma^2 d(d-1)}{2np},$$

for some constant $C > 0$ only depending on C_1.

The derivation of the minimax lower bound for $\mathscr{SO}(d)$ synchronization follows the same approach. According to Lemma 4.6, the inequality (29) also holds for the loss $\bar{\ell}(\widehat{Z}, Z)$. Moreover, by Lemma 3.1, the construction of $Q(r)$ is already in $\mathscr{SO}(d)$. Then, an analogous result to Lemma 3.2 also holds with some straightforward modification. This leads to the following theorem.

Theorem 3.2. Assume $\frac{np}{\sigma^2} \geq c$ for some sufficiently large constants $c > 0$ and $2 \leq d \leq C_1$ for some constant $C_1 > 0$. Then, we have

$$\inf_{\widehat{Z}\in\mathscr{SO}(d)^n} \sup_{Z\in\mathscr{SO}(d)^n} \mathbb{E}_Z \bar{\ell}(\widehat{Z}, Z) \geq \left(1 - C\left(\frac{1}{n} + \frac{\sigma^2}{np}\right)\right)\frac{\sigma^2 d(d-1)}{2np},$$

for some constant $C > 0$ only depending on C_1.

4. Proofs

4.1 *Some auxiliary lemmas*

LEMMA 4.1. Assume $\frac{np}{\log n} > c$ for some sufficiently large constant $c > 0$. Then, we have

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

and

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

with probability at least $1 - n^{-10}$ for some constant $C > 0$.

Proof. The first result is a direct application of union bound and Bernstein's inequality. The second result is Theorem 5.2 of [21]. □

LEMMA 4.2. Assume $\frac{np}{\log n} > c$ for some sufficiently large constant $c > 0$. Then, we have

$$\|(A \otimes \mathbb{1}_d\, \mathbb{1}_d^T) \circ W\|_{\mathrm{op}} \le C\sqrt{dnp},$$

with probability at least $1 - n^{-10}$ for some constant $C > 0$.

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

$$\mathscr{A} = \left\{ \max_{i\in[n]} \sum_{j\in[n]\setminus\{i\}} A_{ij} \le 2np \right\}.$$

Under the assumption $\frac{np}{\log n} > c$, we have $\mathbb{P}(\mathscr{A}^c) \le n^{-11}$ by Bernstein's inequality and a union bound argument. It is clear that the largest row ℓ_2 norm of $A \otimes \mathbb{1}_d\, \mathbb{1}_d^T$ is bounded by $\sqrt{2dnp}$ under the event $\mathscr{A}$. By Corollary 3.9 of [3], we have

$$\sup_{A\in\mathscr{A}} \mathbb{P}_A \left(\|(A \otimes \mathbb{1}_d\, \mathbb{1}_d^T) \circ W\|_{\mathrm{op}} > C_1\sqrt{dnp} + t \right) \le e^{-t^2/4},$$

for some constant $C_1 > 0$. This implies that $\sup_{A\in\mathscr{A}} \mathbb{P}_A \left(\|(A \otimes \mathbb{1}_d\, \mathbb{1}_d^T) \circ W\|_{\mathrm{op}} > C_2\sqrt{dnp} \right) \le n^{-11}$ for some constant $C_2 > 0$. Thus, we have

$$\begin{aligned}
&\mathbb{P} \left(\|(A \otimes \mathbb{1}_d\, \mathbb{1}_d^T) \circ W\|_{\mathrm{op}} > C_2\sqrt{dnp} \right) \\
\le\ & \mathbb{P}(\mathscr{A}^c) + \sup_{A\in\mathscr{A}} \mathbb{P}_A \left(\|(A \otimes \mathbb{1}_d\, \mathbb{1}_d^T) \circ W\|_{\mathrm{op}} > C_2\sqrt{dnp} \right) \\
\le\ & 2n^{-11},
\end{aligned}$$

which implies the desired result. □

LEMMA 4.3. Assume $\frac{np}{\log n} > c$ for some sufficiently large constant $c > 0$. Consider independent random matrices $X_{ij} \sim \mathscr{MN}(0, I_d, I_d)$ for $1 \le i < j \le n$. Write $X_{ji} = X_{ij}$ for $1 \le i < j \le n$. Then, we have

$$\sum_{i=1}^{n} \left\| \sum_{j\in[n]\backslash\{i\}} A_{ij} \left(X_{ij} - X_{ji} \right) \right\|_{\mathrm{F}}^{2} \le 2d(d-1)n(n-1)p + C\left(d^2\sqrt{n^2 p \log n} + d\sqrt{n^3 p^2 \log n} \right),$$

and

$$\sum_{i=1}^{n} \left\| \sum_{j\in[n]\backslash\{i\}} A_{ij} X_{ij} \right\|_{\mathrm{F}}^{2} \le d^2 n(n-1)p + C\left(d^2\sqrt{n^2 p \log n} + d\sqrt{n^3 p^2 \log n} \right),$$

with probability at least $1 - n^{-10}$ for some constant $C > 0$.

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

$$\mathscr{A} = \left\{ \max_{i\in[n]} \sum_{j\in[n]\backslash\{i\}} A_{ij} \le 2np, \sum_{i=1}^{n} \sum_{j\in[n]\backslash\{i\}} A_{ij} \le n(n-1)p + 5\sqrt{n^2 p \log n} \right\}.$$

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

$$g(X) = \sqrt{ \sum_{i=1}^{n} \left\| \sum_{j\in[n]\backslash\{i\}} A_{ij} \left(X_{ij} - X_{ji} \right) \right\|_{\mathrm{F}}^{2} }.$$

Then, for any $A \in \mathscr{A}$ and any X' and X'', we have

$$|g(X') - g(X'')| \le \sqrt{ \sum_{i=1}^{n} \left\| \sum_{j\in[n]\backslash\{i\}} A_{ij} \left(X'_{ij} - X''_{ij} - X'_{ji} + X''_{ji} \right) \right\|_{\mathrm{F}}^{2} } \tag{37}$$

$$\le \sqrt{ \sum_{i=1}^{n} \left(\sum_{j\in[n]\backslash\{i\}} A_{ij} \right) \left(\sum_{j\in[n]\backslash\{i\}} \left\| X'_{ij} - X''_{ij} - X'_{ji} + X''_{ji} \right\|_{\mathrm{F}}^{2} \right) } \tag{38}$$

$$\leq \sqrt{2np}\sqrt{\sum_{i=1}^{n}\sum_{j\in[n]\backslash\{i\}}\|X'_{ij}-X''_{ij}-X'_{ji}+X''_{ji}\|_{\mathrm{F}}^2}$$

$$\leq 2\sqrt{2np}\sqrt{\sum_{i=1}^{n}\sum_{j\in[n]\backslash\{i\}}\|X'_{ij}-X''_{ij}\|_{\mathrm{F}}^2} \tag{39}$$

$$= 4\sqrt{np}\sqrt{\sum_{i=1}^{n}\sum_{j\in[n]:j>i}\|X'_{ij}-X''_{ij}\|_{\mathrm{F}}^2}.$$

The bounds (37) and (39) are due to triangle inequality. The inequality (38) can be viewed as a generalization of Cauchy–Schwarz, since

$$\begin{aligned}
&\left\|\sum_{j\in[n]\backslash\{i\}} A_{ij}\left(X'_{ij}-X''_{ij}-X'_{ji}+X''_{ji}\right)\right\|_{\mathrm{F}}^2 \\
= &\sup_{K\in\mathbb{R}^{d\times d}:\|K\|_{\mathrm{F}}=1}\left|\left\langle K, \sum_{j\in[n]\backslash\{i\}} A_{ij}\left(X'_{ij}-X''_{ij}-X'_{ji}+X''_{ji}\right)\right\rangle\right|^2 \\
= &\sup_{K\in\mathbb{R}^{d\times d}:\|K\|_{\mathrm{F}}=1}\left|\sum_{j\in[n]\backslash\{i\}} A_{ij}\left\langle K, X'_{ij}-X''_{ij}-X'_{ji}+X''_{ji}\right\rangle\right|^2 \\
\leq &\sup_{K\in\mathbb{R}^{d\times d}:\|K\|_{\mathrm{F}}=1}\left(\sum_{j\in[n]\backslash\{i\}} A_{ij}\right)\left(\sum_{j\in[n]\backslash\{i\}}\left\langle K, X'_{ij}-X''_{ij}-X'_{ji}+X''_{ji}\right\rangle^2\right) \\
\leq &\left(\sum_{j\in[n]\backslash\{i\}} A_{ij}\right)\left(\sum_{j\in[n]\backslash\{i\}}\|X'_{ij}-X''_{ij}-X'_{ji}+X''_{ji}\|_{\mathrm{F}}^2\right).
\end{aligned}$$

To summarize, we have shown that $g(X)$ is Lipschitz with respect to $\{X_{ij}\}_{1\leq i<j\leq n}$, and the Lipschitz constant is bounded by $4\sqrt{np}$. By a standard Gaussian concentration inequality for Lipschitz functions [33], we have

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

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

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

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

$$\begin{aligned}\mathbb{E}(g(X)|A) &\leq \sqrt{\sum_{i=1}^{n}\mathbb{E}\left(\left\|\sum_{j\in[n]\backslash\{i\}}A_{ij}\left(X_{ij}-X_{ji}\right)\right\|_{\mathrm{F}}^{2}\middle|A\right)}\\ &= \sqrt{2d(d-1)\sum_{i=1}^{n}\sum_{j\in[n]\backslash\{i\}}A_{ij}}\\ &\leq \sqrt{2d(d-1)\left(n(n-1)p+5\sqrt{n^{2}p\log n}\right)},\end{aligned}$$

for any $A \in \mathscr{A}$. Thus,

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

with probability at least $1-3n^{-11}$, and the first desired bound is implied by squaring both sides of the above inequality. The second bound can be proved by a similar argument, and we omit the details. □

LEMMA 4.4. Consider independent $X_j \sim \mathscr{MN}(0, I_d, I_d)$ and $E_j \sim \text{Bernoulli}(p)$. Then,

$$\mathbb{P}\left(\left\|\sum_{j=1}^{n}E_{j}X_{j}\right\|_{\mathrm{op}}/p>t\right)\leq 81^{d}\exp\left(-\min\left(\frac{pt^{2}}{144n},\frac{pt}{6}\right)\right),$$

for any $t > 0$.

Proof. Apply a standard discretization trick [34], and there exists a subset $\mathscr{U} \subset \{u \in \mathbb{R}^d : \|u\| = 1\}$ with cardinality bound $|\mathscr{U}| \leq 9^d$ that satisfies

$$\left\|\sum_{j=1}^{n}E_{j}X_{j}\right\|_{\mathrm{op}}\leq 3\max_{u,v\in\mathscr{U}}\sum_{j=1}^{n}E_{j}u^{T}X_{j}v.$$

Note that for any $u, v \in \mathscr{U}$, we have $u^T X_j v \sim \mathscr{N}(0,1)$. By Lemma 13 of [13], we have

$$\mathbb{P}\left(3\sum_{j=1}^{n}E_{j}u^{T}X_{j}v/p>t\right)\leq\exp\left(-\min\left(\frac{pt^{2}}{144n},\frac{pt}{6}\right)\right).$$

Hence,

$$\begin{aligned}\mathbb{P}\left(\left\|\sum_{j=1}^n E_jX_j\right\|_{\mathrm{op}}/p > t\right) &\leq \mathbb{P}\left(3\max_{u,v\in\mathscr{U}}\sum_{j=1}^n E_ju^TX_jv/p > t\right)\\ &\leq \sum_{u,v\in\mathscr{U}}\mathbb{P}\left(3\sum_{j=1}^n E_ju^TX_jv/p > t\right)\\ &\leq 81^d\exp\left(-\min\left(\frac{pt^2}{144n},\frac{pt}{6}\right)\right),\end{aligned}$$

which is the desired bound. □

LEMMA 4.5 (Corollary 2.14 of [19]). Consider $X, \widetilde{X} \in \mathbb{R}^{d\times d}$ with X being full rank. Then,

$$\frac{|\det(X)-\det(\widetilde{X})|}{|\det(X)|} \leq \left(\|X^{-1}\|_{\mathrm{op}}\|X-\widetilde{X}\|_{\mathrm{op}}+1\right)^d - 1.$$

LEMMA 4.6. For any $Z, Z^* \in \mathscr{O}(d)^n$, we have

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|Z_iZ_j^{\mathrm{T}} - Z_i^*Z_j^{*\mathrm{T}}\|_{\mathrm{F}}^2 \leq 2\ell(Z,Z^*). \tag{40}$$

For any $Z, Z^* \in \mathscr{SO}(d)^n$, we have

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|Z_iZ_j^{\mathrm{T}} - Z_i^*Z_j^{*\mathrm{T}}\|_{\mathrm{F}}^2 \leq 2\bar{\ell}(Z,Z^*). \tag{41}$$

For any $Z, Z^* \in \mathbb{R}^{nd\times d}$ such that $Z/\sqrt{n}, Z^*/\sqrt{n} \in \mathscr{O}(nd, d)$, we have

$$\min_{B\in\mathscr{O}(d)}\frac{1}{n}\sum_{i=1}^n\|Z_i - Z_i^*B\|_{\mathrm{F}}^2 \leq \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|Z_iZ_j^{\mathrm{T}} - Z_i^*Z_j^{*\mathrm{T}}\|_{\mathrm{F}}^2. \tag{42}$$

In all inequalities above, we have $Z^{\mathrm{T}} = (Z_1^{\mathrm{T}}, \cdots, Z_n^{\mathrm{T}})$ and $Z^{*\mathrm{T}} = (Z_1^{*\mathrm{T}}, \cdots, Z_n^{*\mathrm{T}})$, where Z_i and Z_i^* are the ith block sub-matrices of size $d\times d$ of Z and Z^*, respectively.

Proof. Consider any $Z, Z^* \in \mathscr{O}(d)^n$. By direct expansion, we can write

$$\ell(Z,Z^*) = 2\left(d - \max_{B\in\mathscr{O}(d)}\mathsf{Tr}\left(\frac{1}{n}\sum_{i=1}^n Z_i^{\mathrm{T}}Z_i^*B\right)\right),$$

and

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|Z_iZ_j^{\mathrm{T}}-Z_i^*Z_j^{*\mathrm{T}}\|_{\mathrm{F}}^2=2\left(d-\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{F}}^2\right). \tag{43}$$

Since

$$\max_{B\in\mathscr{O}(d)}\mathsf{Tr}\left(\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*B\right)\le\max_{B\in\mathbb{R}^{d\times d}:\|B\|_{\mathrm{F}}^2=d}\mathsf{Tr}\left(\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*B\right)\le\sqrt{d}\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{F}},$$

we have

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|Z_iZ_j^{\mathrm{T}}-Z_i^*Z_j^{*\mathrm{T}}\|_{\mathrm{F}}^2 &= 2\left(\sqrt{d}+\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{F}}\right)\left(\sqrt{d}-\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{F}}\right)\\ &\le 4\sqrt{d}\left(\sqrt{d}-\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{F}}\right)\\ &\le 2\ell(Z,Z^*).\end{aligned}$$

The proof of (40) is complete. The inequality (41) can be proved with the same argument.

Finally, we prove (42). For any $Z,Z^*\in\mathbb{R}^{nd\times d}$ such that $Z/\sqrt{n},Z^*/\sqrt{n}\in\mathscr{O}(nd,d)$, we have $\min_{B\in\mathscr{O}(d)}\frac{1}{n}\sum_{i=1}^{n}\|Z_i-Z_i^*B\|_{\mathrm{F}}^2 = 2\left(d-\max_{B\in\mathscr{O}(d)}\mathsf{Tr}\left(\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*B\right)\right)$ and the identity (43) continues to hold. Suppose the matrix $\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*$ admits an SVD $\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*=UDV^{\mathrm{T}}$. Then,

$$\max_{B\in\mathscr{O}(d)}\mathsf{Tr}\left(\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*B\right)\ge\mathsf{Tr}(UDV^{\mathrm{T}}VU^{\mathrm{T}})=\mathsf{Tr}(D)\ge\frac{\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\|_{\mathrm{F}}^2}{\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\|_{\mathrm{op}}}.$$

This implies

$$\begin{aligned}\min_{B\in\mathscr{O}(d)}\frac{1}{n}\sum_{i=1}^{n}\|Z_i-Z_i^*B\|_{\mathrm{F}}^2 &\le 2\left(d-\frac{\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{F}}^2}{\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{op}}}\right)\\ &\le 2\left(d-\frac{\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{F}}^2}{\left\|Z/\sqrt{n}\right\|_{\mathrm{op}}\left\|Z^*/\sqrt{n}\right\|_{\mathrm{op}}}\right)\\ &\le 2\left(d-\left\|\frac{1}{n}\sum_{i=1}^{n}Z_i^{\mathrm{T}}Z_i^*\right\|_{\mathrm{F}}^2\right)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|Z_iZ_j^{\mathrm{T}}-Z_i^*Z_j^{*\mathrm{T}}\right\|_{\mathrm{F}}^2.\end{aligned}$$

The proof is complete. □

4.2 *Proof of Lemma 2.2*

For any $Z \in \mathscr{O}(d)^n$ such that $\ell(Z, Z^*) \leq \gamma$, we define

$$\widetilde{Z}_i = \frac{\sum_{j\in[n]\backslash\{i\}} A_{ij}Y_{ij}Z_j}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}, \tag{44}$$

for each $i \in [n]$. Then, write $\widehat{Z} = f(Z)$. It is clear that $\widehat{Z}_i = \mathscr{P}(\widetilde{Z}_i)$ for i such that $\widetilde{Z}_i$ is full rank. The condition $\ell(Z, Z^*) \leq \gamma$ implies that there exists some $B \in \mathscr{O}(d)$ such that $\ell(Z, Z^*) = \frac{1}{n}\sum_{i=1}^n \|Z_i - Z_i^*B\|_{\mathrm{F}}^2 \leq \gamma$. With $Y_{ij} = Z_i^*Z_j^{*\mathrm{T}} + \sigma W_{ij}$, we obtain the following expansion:

$$\begin{aligned} Z_i^{*\mathrm{T}}\widetilde{Z}_iB^{\mathrm{T}} &= I_d + \frac{\sum_{j\in[n]\backslash\{i\}} A_{ij}Z_j^{*\mathrm{T}}(Z_j - Z_j^*B)B^{\mathrm{T}}}{\sum_{j\in[n]\backslash\{i\}} A_{ij}} \\ &\quad + \frac{\sigma\sum_{j\in[n]\backslash\{i\}} A_{ij}Z_i^{*\mathrm{T}}W_{ij}(Z_j - Z_j^*B)B^{\mathrm{T}}}{\sum_{j\in[n]\backslash\{i\}} A_{ij}} + \frac{\sigma\sum_{j\in[n]\backslash\{i\}} A_{ij}Z_i^{*\mathrm{T}}W_{ij}Z_j^*}{\sum_{j\in[n]\backslash\{i\}} A_{ij}} \\ &= I_d + \frac{1}{n-1}\sum_{j=1}^n Z_j^{*\mathrm{T}}(Z_j - Z_j^*B)B^{\mathrm{T}} - \frac{1}{n-1}Z_i^{*\mathrm{T}}(Z_i - Z_i^*B)B^{\mathrm{T}} \\ &\quad + \frac{\sum_{j\in[n]\backslash\{i\}} A_{ij}Z_j^{*\mathrm{T}}(Z_j - Z_j^*B)B^{\mathrm{T}}}{\sum_{j\in[n]\backslash\{i\}} A_{ij}} - \frac{1}{n-1}\sum_{j\in[n]\backslash\{i\}} Z_j^{*\mathrm{T}}(Z_j - Z_j^*B)B^{\mathrm{T}} \\ &\quad + \frac{\sigma\sum_{j\in[n]\backslash\{i\}} A_{ij}Z_i^{*\mathrm{T}}W_{ij}(Z_j - Z_j^*B)B^{\mathrm{T}}}{\sum_{j\in[n]\backslash\{i\}} A_{ij}} + \frac{\sigma\sum_{j\in[n]\backslash\{i\}} A_{ij}Z_i^{*\mathrm{T}}W_{ij}Z_j^*}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}. \end{aligned}$$

Next, we define

$$\widetilde{Q} = I_d + \frac{1}{n-1}\sum_{j=1}^n Z_j^{*\mathrm{T}}(Z_j - Z_j^*B)B^{\mathrm{T}}. \tag{45}$$

It is clear that

$$\|\widetilde{Q} - I_d\|_{\mathrm{F}} \leq \frac{1}{n-1}\sum_{j=1}^n \|Z_j - Z_j^*B\|_{\mathrm{F}} \leq \frac{n}{n-1}\sqrt{\frac{1}{n}\sum_{j=1}^n \|Z_j - Z_j^*B\|_{\mathrm{F}}^2} \leq 2\sqrt{\gamma}, \tag{46}$$

and therefore, $\widetilde{Q}$ is full rank and we can define $Q = \mathscr{P}(\widetilde{Q})$. With these definitions and the above expansion of $Z_i^{*\mathrm{T}}\widetilde{Z}_iB^{\mathrm{T}}$, we can write

$$Z_i^{*\mathrm{T}}\widetilde{Z}_iB^{\mathrm{T}}Q^{\mathrm{T}} = \widetilde{Q}Q^{\mathrm{T}} + S_i, \tag{47}$$

where $S_i = -\frac{1}{n-1} Z_i^{*\mathrm{T}}(Z_i - Z_i^* B)B^{\mathrm{T}}Q^{\mathrm{T}} + F_i + G_i + H_i$, with

$$\begin{aligned} F_i &= \frac{\sum_{j\in[n]\backslash\{i\}} A_{ij} Z_j^{*\mathrm{T}}(Z_j - Z_j^* B)B^{\mathrm{T}}Q^{\mathrm{T}}}{\sum_{j\in[n]\backslash\{i\}} A_{ij}} - \frac{1}{n-1}\sum_{j\in[n]\backslash\{i\}} Z_j^{*\mathrm{T}}(Z_j - Z_j^* B)B^{\mathrm{T}}Q^{\mathrm{T}}, \\ G_i &= \frac{\sigma \sum_{j\in[n]\backslash\{i\}} A_{ij} Z_i^{*\mathrm{T}} W_{ij}(Z_j - Z_j^* B)B^{\mathrm{T}}Q^{\mathrm{T}}}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}, \\ H_i &= \frac{\sigma \sum_{j\in[n]\backslash\{i\}} A_{ij} Z_i^{*\mathrm{T}} W_{ij} Z_j^* Q^{\mathrm{T}}}{\sum_{j\in[n]\backslash\{i\}} A_{ij}}. \end{aligned}$$

For the first term of (47), it is clear that $\widetilde{Q}Q^{\mathrm{T}}$ is a symmetric matrix by the definition of $\mathscr{P}(\cdot)$ given in (8). We also have

$$\left\| \frac{1}{n-1} Z_i^{*\mathrm{T}}(Z_i - Z_i^* B)B^{\mathrm{T}}Q^{\mathrm{T}} \right\|_{\mathrm{F}} = \frac{1}{n-1}\left\| Z_i - Z_i^* B \right\|_{\mathrm{F}} \leq \frac{\sqrt{n}}{n-1}\sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\| Z_i - Z_i^* B\right\|_{\mathrm{F}}^2} \leq \sqrt{\gamma}. \tag{48}$$

As long as $\|F_i\|_{\mathrm{op}} \vee \|G_i\|_{\mathrm{op}} \vee \|H_i\|_{\mathrm{op}} \leq \rho$, by (46), (47) and (48), we have

$$s_{\min}(Z_i^{*\mathrm{T}}\widetilde{Z}_i B^{\mathrm{T}}Q^{\mathrm{T}}) \geq s_{\min}(\widetilde{Q}) - \|S_i\|_{\mathrm{op}} \geq 1 - 3(\rho + \sqrt{\gamma}). \tag{49}$$

When $3(\rho + \sqrt{\gamma}) < 1$, $Z_i^{*\mathrm{T}}\widetilde{Z}_i B^{\mathrm{T}}Q^{\mathrm{T}}$ has full rank, and so does $\widetilde{Z}_i$, and thus, we have $\widehat{Z}_i = \mathscr{P}(\widetilde{Z}_i)$. Now we can apply a perturbation analysis to (47),

$$\begin{aligned} \|\widehat{Z}_i - Z_i^* QB\|_{\mathrm{F}}^2 &= \|\mathscr{P}(\widetilde{Z}_i) - Z_i^* QB\|_{\mathrm{F}}^2 \\ &= \|\mathscr{P}(Z_i^{*\mathrm{T}}\widetilde{Z}_i B^{\mathrm{T}}Q^{\mathrm{T}}) - I_d\|_{\mathrm{F}}^2 \end{aligned} \tag{50}$$

$$= \left\| \mathscr{P}(\widetilde{Q}Q^{\mathrm{T}} + S_i) - \mathscr{P}\left(\widetilde{Q}Q^{\mathrm{T}} + \frac{1}{2}S_i + \frac{1}{2}S_i^{\mathrm{T}}\right) \right\|_{\mathrm{F}}^2 \tag{51}$$

$$\leq \frac{1}{[1 - 3(\rho + \sqrt{\gamma})]^2} \left\| \frac{1}{2}S_i - \frac{1}{2}S_i^{\mathrm{T}} \right\|_{\mathrm{F}}^2. \tag{52}$$

The equality (50) is by the properties of $\mathscr{P}(\cdot)$ listed in Section 2.1. We then used the fact that $\widetilde{Q}Q^{\mathrm{T}} + \frac{1}{2}S_i + \frac{1}{2}S_i^{\mathrm{T}}$ is symmetric and positive definite so that $\mathscr{P}\left(\widetilde{Q}Q^{\mathrm{T}} + \frac{1}{2}S_i + \frac{1}{2}S_i^{\mathrm{T}}\right) = I_d$, which leads to (51). The inequality (52) is by Lemma 2.1, with $s_{\min}(\widetilde{Q}Q^{\mathrm{T}} + S_i)$ lower bounded by (49) and $s_{\min}\left(\widetilde{Q}Q^{\mathrm{T}} + \frac{1}{2}S_i + \frac{1}{2}S_i^{\mathrm{T}}\right)$ is lower bounded similarly (which also implies that $\widetilde{Q}Q^{\mathrm{T}} + \frac{1}{2}S_i + \frac{1}{2}S_i^{\mathrm{T}}$ is positive definite required by (51)). The perturbation analysis has been done under the condition

$\|F_i\|_{\mathrm{op}} \vee \|G_i\|_{\mathrm{op}} \vee \|H_i\|_{\mathrm{op}} \leq \rho$. Without this condition, we have

$$\begin{aligned}
\|\widehat{Z}_i - Z_i^* QB\|_{\mathrm{F}}^2 \;\leq\; & \frac{1}{[1-3(\rho+\sqrt{\gamma})]^2} \left\| \frac{1}{2}S_i - \frac{1}{2}S_i^{\mathrm{T}} \right\|_{\mathrm{F}}^2 \mathbb{I}\{\|F_i\|_{\mathrm{op}} \vee \|G_i\|_{\mathrm{op}} \vee \|H_i\|_{\mathrm{op}} \leq \rho\} \\
& + 4d\mathbb{I}\{\|F_i\|_{\mathrm{op}} \vee \|G_i\|_{\mathrm{op}} \vee \|H_i\|_{\mathrm{op}} > \rho\} \\
\leq\; & \frac{1+\eta}{[1-3(\rho+\sqrt{\gamma})]^2} \left\| \frac{1}{2}H_i - \frac{1}{2}H_i^{\mathrm{T}} \right\|_{\mathrm{F}}^2 + \frac{3(1+\eta^{-1})}{[1-3(\rho+\sqrt{\gamma})]^2} \left(\|F_i\|_{\mathrm{F}}^2 + \|G_i\|_{\mathrm{F}}^2 \right) \\
& + \frac{3(1+\eta^{-1})}{[1-3(\rho+\sqrt{\gamma})]^2} \left\| \frac{1}{n-1} Z_i^{*\mathrm{T}}(Z_i - Z_i^* B)B^{\mathrm{T}}Q^{\mathrm{T}} \right\|_{\mathrm{F}}^2 \\
& + 4d\mathbb{I}\{\|F_i\|_{\mathrm{op}} > \rho\} + 4d\mathbb{I}\{\|G_i\|_{\mathrm{op}} > \rho\} + 4d\mathbb{I}\{\|H_i\|_{\mathrm{op}} > \rho\},
\end{aligned} \tag{53}$$

where we have used the inequality $\|X + X'\|_{\mathrm{F}}^2 \leq (1+\eta)\|X\|_{\mathrm{F}}^2 + (1+\eta^{-1})\|X'\|_{\mathrm{F}}^2$ for any matrices $X, X' \in \mathbb{R}^{d\times d}$. The above bound holds deterministically for any $\eta \in (0,1)$ and any $\rho > 0$ that satisfies $3(\rho+\sqrt{\gamma}) < 1$. The specific values of η and ρ will be determined later.

In the next step of the proof, we will need to analyze F_i, G_i and H_i. This requires a few high probability events. First, by Lemma 4.4, we have

$$\sum_{i=1}^n \mathbb{P}\left(\frac{2\sigma}{np} \left\| \sum_{j\in[n]\setminus\{i\}} A_{ij}W_{ij}Z_j^* \right\|_{\mathrm{op}} > \rho \right) \leq 81^d n \exp\left(-\min\left(\frac{\rho^2 np}{576\sigma^2}, \frac{\rho np}{12\sigma} \right) \right).$$

By Markov inequality, we have

$$\sum_{i=1}^n \mathbb{I}\left\{ \frac{2\sigma}{np} \left\| \sum_{j\in[n]\setminus\{i\}} A_{ij}W_{ij}Z_j^* \right\|_{\mathrm{op}} > \rho \right\} \leq \frac{\sigma^2}{\rho^2 p} \exp\left(-\sqrt{\frac{\rho^2 np}{\sigma^2}} \right), \tag{54}$$

with probability at least

$$\begin{aligned}
& 1 - 81^d \frac{\rho^2 np}{\sigma^2} \left[\exp\left(-\frac{\rho^2 np}{576\sigma^2} + \sqrt{\frac{\rho^2 np}{\sigma^2}} \right) + \exp\left(-\frac{\rho np}{\sigma^2} + \sqrt{\frac{\rho^2 np}{\sigma^2}} \right) \right] \\
\geq\; & 1 - 81^d \frac{2\rho^2 np}{\sigma^2} \exp\left(-3\sqrt{\frac{\rho^2 np}{\sigma^2}} \right) \\
\geq\; & 1 - 81^d \exp\left(-2\sqrt{\frac{\rho^2 np}{\sigma^2}} \right) \\
\geq\; & 1 - \exp\left(-\sqrt{\frac{\rho^2 np}{\sigma^2}} \right).
\end{aligned}$$

The above set of inequalities requires that ρ satisfies $\sqrt{\frac{\rho^2 np}{\sigma^2}} > 2304 + 5d$. Then, by Lemma 4.1, Lemma 4.2 and Lemma 4.3, we know that

$$\min_{i\in[n]} \sum_{j\in[n]\backslash\{i\}} A_{ij} \geq (n-1)p - C\sqrt{np\log n}, \tag{55}$$

$$\max_{i\in[n]} \sum_{j\in[n]\backslash\{i\}} A_{ij} \leq (n-1)p + C\sqrt{np\log n}, \tag{56}$$

$$\|A - \mathbb{E}A\|_{\mathrm{op}} \leq C\sqrt{np}, \tag{57}$$

$$\|(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ W\|_{\mathrm{op}} \leq C\sqrt{dnp}, \tag{58}$$

$$\sum_{i=1}^n \left\| \sum_{j\in[n]\backslash\{i\}} A_{ij}\left(Z_i^{*\mathrm{T}} W_{ij} Z_j^* - Z_j^{*\mathrm{T}} W_{ji} Z_i^*\right) \right\|_{\mathrm{F}}^2 \leq 2d(d-1)n^2 p\left(1 + C\sqrt{\frac{\log n}{n}}\right), \tag{59}$$

$$\sum_{i=1}^n \left\| \sum_{j\in[n]\backslash\{i\}} A_{ij} W_{ij} Z_j^* \right\|_{\mathrm{F}}^2 \leq d^2 n^2 p\left(1 + C\sqrt{\frac{\log n}{n}}\right), \tag{60}$$

all hold with probability at least $1-n^{-9}$ for some constant $C > 0$. We conclude that the events (54), (55), (56), (57), (58), (59) and (60) hold simultaneously with probability at least $1 - n^{-9} - \exp\left(-\sqrt{\frac{\rho^2 np}{\sigma^2}}\right)$. These events will be assumed from now on.

Analysis of F_i. By triangle inequality, (55) and (56), we have

$$\begin{aligned}
\|F_i\|_{\mathrm{F}} &\leq \frac{\left\|\sum_{j\in[n]\backslash\{i\}}(A_{ij}-p)Z_j^{*\mathrm{T}}(Z_j - Z_j^* B)\right\|_{\mathrm{F}}}{\sum_{j\in[n]\backslash\{i\}} A_{ij}} \\
&\quad + \left|\frac{p}{\sum_{j\in[n]\backslash\{i\}} A_{ij}} - \frac{1}{n-1}\right| \left\| \sum_{j\in[n]\backslash\{i\}} Z_j^{*\mathrm{T}}(Z_j - Z_j^* B)\right\|_{\mathrm{F}} \\
&\leq \frac{2}{np}\left\| \sum_{j\in[n]\backslash\{i\}} (A_{ij}-p)Z_j^{*\mathrm{T}}(Z_j - Z_j^* B)\right\|_{\mathrm{F}} \\
&\quad + \frac{2\left|\sum_{j\in[n]\backslash\{i\}}(A_{ij}-p)\right|}{n^2 p} \sum_{j\in[n]\backslash\{i\}} \|Z_j - Z_j^* B\|_{\mathrm{F}} \\
&\leq \frac{2}{np}\left\| \sum_{j\in[n]\backslash\{i\}} (A_{ij}-p)Z_j^{*\mathrm{T}}(Z_j - Z_j^* B)\right\|_{\mathrm{F}} + \frac{C_1\sqrt{p\log n}}{np}\sqrt{\sum_{i=1}^n \|Z_i - Z_i^* B\|_{\mathrm{F}}^2}.
\end{aligned}$$

Using (57), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\|F_i\|_{\mathrm{F}}^2 &\leq \frac{8}{n^3p^2}\sum_{i=1}^{n}\left\|\sum_{j\in[n]\backslash\{i\}}(A_{ij}-p)Z_j^{*\mathrm{T}}(Z_j-Z_j^*B)\right\|_{\mathrm{F}}^2+2C_1^2\frac{\log n}{np}\ell(Z,Z^*)\\ &\leq \frac{8}{n^3p^2}\|A-\mathbb{E}A\|_{\mathrm{op}}^2\sum_{i=1}^{n}\|Z_i-Z_i^*B\|_{\mathrm{F}}^2+2C_1^2\frac{\log n}{np}\ell(Z,Z^*)\\ &\leq C_2\frac{\log n}{np}\ell(Z,Z^*).\end{aligned} \tag{61}$$

This bound also implies

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{\|F_i\|_{\mathrm{op}}>\rho\}\leq\frac{1}{n\rho^2}\sum_{i=1}^{n}\|F_i\|_{\mathrm{F}}^2\leq C_2\frac{\log n}{\rho^2 np}\ell(Z,Z^*). \tag{62}$$

Analysis of G_i. By (55) and (58), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\|G_i\|_{\mathrm{F}}^2 &\leq \frac{2\sigma^2}{n^3p^2}\sum_{i=1}^{n}\left\|\sum_{j\in[n]\backslash\{i\}}A_{ij}W_{ij}(Z_j-Z_j^*B)\right\|_{\mathrm{F}}^2\\ &\leq \frac{2\sigma^2}{n^3p^2}\|(A\otimes \mathbb{1}_d\,\mathbb{1}_d^T)\circ W\|_{\mathrm{op}}^2\sum_{i=1}^{n}\|Z_i-Z_i^*B\|_{\mathrm{F}}^2\\ &\leq C_3\frac{\sigma^2 d}{np}\ell(Z,Z^*),\end{aligned} \tag{63}$$

and thus,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{\|G_i\|_{\mathrm{op}}>\rho\}\leq\frac{1}{n\rho^2}\sum_{i=1}^{n}\|G_i\|_{\mathrm{F}}^2\leq C_3\frac{\sigma^2 d}{\rho^2 np}\ell(Z,Z^*). \tag{64}$$

Analysis of H_i. First, by (54) and (55), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{\|H_i\|_{\mathrm{op}}>\rho\} &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\left\{\frac{2\sigma}{np}\left\|\sum_{j\in[n]\backslash\{i\}}A_{ij}W_{ij}Z_j^*\right\|_{\mathrm{op}}>\rho\right\}\\ &\leq \frac{\sigma^2}{\rho^2 np}\exp\left(-\sqrt{\frac{\rho^2 np}{\sigma^2}}\right)\\ &\leq \exp\left(-\frac{1}{2}\sqrt{\frac{\rho^2 np}{\sigma^2}}\right).\end{aligned} \tag{65}$$

Next, to bound $\|\frac{1}{2}H_i - \frac{1}{2}H_i^{\mathrm{T}}\|_{\mathrm{F}}^2$, we introduce the notation $E_i = \frac{\sigma \sum_{j\in[n]\setminus\{i\}} A_{ij} Z_i^{*\mathrm{T}} W_{ij} Z_j^*}{\sum_{j\in[n]\setminus\{i\}} A_{ij}}$ and thus, we can write $H_i = E_i Q^{\mathrm{T}}$. We then have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \left\|\frac{1}{2}H_i - \frac{1}{2}H_i^{\mathrm{T}}\right\|_{\mathrm{F}}^2 &= \frac{1}{n}\sum_{i=1}^n \left\|\frac{1}{2}E_i - \frac{1}{2}E_i^{\mathrm{T}} + \frac{1}{2}E_i(Q^{\mathrm{T}} - I_d) - \frac{1}{2}(Q - I_d)E_i^{\mathrm{T}}\right\|_{\mathrm{F}}^2 \\
&\leq (1+\eta)\frac{1}{n}\sum_{i=1}^n \left\|\frac{1}{2}E_i - \frac{1}{2}E_i^{\mathrm{T}}\right\|_{\mathrm{F}}^2 + (1+\eta^{-1})\frac{1}{n}\sum_{i=1}^n \|E_i(Q^{\mathrm{T}} - I_d)\|_{\mathrm{F}}^2 \\
&\leq (1+\eta)\frac{1}{n}\sum_{i=1}^n \left\|\frac{1}{2}E_i - \frac{1}{2}E_i^{\mathrm{T}}\right\|_{\mathrm{F}}^2 + (1+\eta^{-1})\|Q - I_d\|_{\mathrm{F}}^2 \frac{1}{n}\sum_{i=1}^n \|E_i\|_{\mathrm{F}}^2.
\end{aligned}$$

By (55) and (59), we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \left\|\frac{1}{2}E_i - \frac{1}{2}E_i^{\mathrm{T}}\right\|_{\mathrm{F}}^2 \\
&\leq \frac{\sigma^2}{4\left(np - 2C\sqrt{np\log n}\right)^2}\frac{1}{n}\sum_{i=1}^n \left\|\sum_{j\in[n]\setminus\{i\}} A_{ij}\left(Z_i^{*\mathrm{T}} W_{ij} Z_j^* - Z_j^{*\mathrm{T}} W_{ji} Z_i^*\right)\right\|_{\mathrm{F}}^2 \\
&\leq \left(1 + C_5\sqrt{\frac{\log n}{np}}\right)\frac{\sigma^2 d(d-1)}{2np}.
\end{aligned}$$

By Lemma 2.1 and (46), we have

$$\|Q - I_d\|_{\mathrm{F}} \leq 2\|\widetilde{Q} - I_d\|_{\mathrm{F}} \leq 4\sqrt{\gamma}.$$

By (55) and (60), we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \|E_i\|_{\mathrm{F}}^2 &\leq \frac{\sigma^2}{\left(np - 2C\sqrt{np\log n}\right)^2}\frac{1}{n}\sum_{i=1}^n \left\|\sum_{j\in[n]\setminus\{i\}} A_{ij} Z_i^{*\mathrm{T}} W_{ij} Z_j^*\right\|_{\mathrm{F}}^2 \\
&\leq \frac{2\sigma^2 d^2}{np}.
\end{aligned}$$

We combine the three bounds above and obtain

$$\frac{1}{n}\sum_{i=1}^n \|\frac{1}{2}H_i - \frac{1}{2}H_i^{\mathrm{T}}\|_{\mathrm{F}}^2 \leq (1+\eta)\left(1 + C_5\sqrt{\frac{\log n}{np}}\right)\frac{\sigma^2 d(d-1)}{2np} + (1+\eta^{-1})\frac{16\gamma\sigma^2 d^2}{np}. \tag{66}$$

Finally, we also have the bound

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{1}{n-1}Z_i^{*\mathrm{T}}(Z_i - Z_i^*B)B^{\mathrm{T}}Q^{\mathrm{T}}\right\|_{\mathrm{F}}^2 \leq \frac{1}{(n-1)^2}\ell(Z,Z^*). \tag{67}$$

Now we can plug the bounds (61), (62), (63), (64), (65), (66) and (67) into (53), and we have

$$\begin{aligned}\ell(\widehat{Z},Z^*) &\leq \frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i - Z_i^*QB\|_{\mathrm{F}}^2 \\ &\leq \left(1+C_6\left(\rho+\sqrt{\gamma}+\eta+\eta^{-1}\gamma+\sqrt{\frac{\log n}{np}}\right)\right)\frac{\sigma^2 d(d-1)}{2np} \\ &\quad +4d\exp\left(-\frac{1}{2}\sqrt{\frac{\rho^2 np}{\sigma^2}}\right)+C_6\left(\eta^{-1}+d\rho^{-2}\right)\frac{\log n+\sigma^2 d}{np}\ell(Z,Z^*).\end{aligned}$$

So far we have required $\eta \in (0,1)$, $\rho > 0$, $3(\rho+\sqrt{\gamma}) < 1$ and $\sqrt{\frac{\rho^2 np}{\sigma^2}} > 2304 + 5d$. Set

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

Under the conditions that $\gamma < 16^{-1}$, $\frac{d\log n+\sigma^2 d^2}{np}$ upper bounded by a sufficiently small constant, all the requirements are satisfied. With this choice, we have

$$\ell(Z,Z^*) \leq \left(1+C_7\left(\gamma^2+\frac{\log n+\sigma^2 d}{np}\right)^{1/4}\right)\frac{\sigma^2 d(d-1)}{2np} + C_7\sqrt{\frac{d\log n+\sigma^2 d^2}{np}}\ell(Z,Z^*).$$

Note that this inequality has been derived from the events (54), (55), (56), (57), (58), (59) and (60) and $\ell(Z,Z^*) \leq \gamma$. Thus, it holds uniformly over all $Z \in \mathscr{O}(d)^n$ such that $\ell(Z,Z^*) \leq \gamma$ with probability at least $1 - n^{-9} - \exp\left(-\left((\frac{np}{\sigma^2}\right)^{1/4}\right)$. The proof is complete.

4.3 *Proof of Theorem 2.2*

We first characterize the evolution of $\bar{\ell}(Z^{(t)}, Z^*)$ through the map

$$Z^{(t)} = \bar{f}(Z^{(t-1)}),$$

where $\bar{f} : \mathscr{SO}(d)^n \to \mathscr{SO}(d)^n$ is defined by (26).

LEMMA 4.7. For the $\mathscr{SO}(d)$ synchronization (3), assume $\frac{np}{\sigma^2} \geq c_1$ and $\frac{np}{\log n} \geq c_2$ for some sufficiently large constants $c_1, c_2 > 0$ and $d \leq C$ for some constant $C > 0$. Then, for any $\gamma \in [0, 1/(32d^2))$, we have

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

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

Proof. The proof follows the same argument as that of Lemma 2.2, and we only point out the difference. For any $Z \in \mathscr{SO}(d)^n$ such that $\bar{\ell}(Z,Z^*) \leq \gamma$, define $\widetilde{Z}$ according to (44). The condition $\bar{\ell}(Z,Z^*) \leq \gamma$ implies that there exists some $B \in \mathscr{SO}(d)$ such that $\bar{\ell}(Z,Z^*) = \frac{1}{n}\sum_{i=1}^n \|Z_i - Z_i^* B\|_{\mathrm{F}}^2 \leq \gamma$. Recall the definition of $\widetilde{Q}$ in (45) and the bound $\|\widetilde{Q} - I_d\|_{\mathrm{F}} \leq 2\sqrt{\gamma}$ that it satisfies according to (46). By Lemma 4.5, we have

$$|\det(\widetilde{Q}) - 1| \leq \left(\|\widetilde{Q} - I_d\|_{\mathrm{F}} + 1\right)^d - 1 \leq (2\sqrt{\gamma}+1)^d - 1 < 1.$$

The last inequality is by the condition $2\sqrt{\gamma} \leq e^{d^{-1}\log 2} - 1$. Then, we can conclude that $\det(\widetilde{Q}) > 0$. By the definition of $\bar{\mathscr{P}}(\cdot)$, we then have $Q = \mathscr{P}(\widetilde{Q}) = \bar{\mathscr{P}}(\widetilde{Q}) \in \mathscr{SO}(d)$. Recall the definitions of S_1, F_i, G_i and H_i in the proof of Lemma 2.2. By (47), we have

$$\|Z_i^{*\mathrm{T}}\widetilde{Z}_i B^{\mathrm{T}} - I_d\|_{\mathrm{op}} \leq \|Z_i^{*\mathrm{T}}\widetilde{Z}_i B^{\mathrm{T}} Q^{\mathrm{T}} - \widetilde{Q}Q^{\mathrm{T}}\|_{\mathrm{op}} + \|\widetilde{Q} - I_d\|_{\mathrm{op}} \leq 3(\rho + \sqrt{\gamma}),$$

as long as $\|F_i\|_{\mathrm{op}} \vee \|G_i\|_{\mathrm{op}} \vee \|H_i\|_{\mathrm{op}} \leq \rho$. By Lemma 4.5, we have

$$|\det(Z_i^{*\mathrm{T}}\widetilde{Z}_i B^{\mathrm{T}}) - 1| \leq \left(\|Z_i^{*\mathrm{T}}\widetilde{Z}_i B^{\mathrm{T}} - I_d\|_{\mathrm{op}} + 1\right)^d - 1 \leq (3(\rho+\sqrt{\gamma})+1)^d - 1 < 1.$$

The last inequality requires $3(\rho + \sqrt{\gamma}) < e^{d^{-1}\log 2} - 1$. We then have $\det(Z_i^{*\mathrm{T}}\widetilde{Z}_i B^{\mathrm{T}}) > 0$, which implies $\det(\widetilde{Z}_i) > 0$. Thus, $\bar{\mathscr{P}}(\widetilde{Z}_i) = \mathscr{P}(\widetilde{Z}_i)$. Together with the fact $Q \in \mathscr{SO}(d)$ that we have just established, the remaining arguments in the proof of Lemma 2.2 can be directly applied to obtain the desired conclusion. The only subtle difference is that here we require $3(\rho + \sqrt{\gamma}) < e^{d^{-1}\log 2} - 1$ compared with $3(\rho + \sqrt{\gamma}) < 1$ in the proof of Lemma 2.2. As a result, a sufficient condition for γ is $\gamma < 1/(32d^2)$ compared with $\gamma < 1/16$ in Lemma 2.2. This detail does not affect the result given the assumption that $d \leq C$. □

Then, we establish the error bound for the initialization procedure.

LEMMA 4.8. For the $\mathscr{SO}(d)$ synchronization (3), assume $\frac{np}{\log n} \geq c$ for some sufficiently large constants $c > 0$ and $d \leq C_1$ for some constant $C_1 > 0$. Then, for $Z^{(0)}$ defined in (27), we have

$$\bar{\ell}(Z^{(0)}, Z^*) \leq C\frac{\sigma^2+1}{np},$$

with probability at least $1 - n^{-9}$ for some constant $C > 0$ only depending on C_1.

The proof of Lemma 4.8 will be given in Section 4.4.

Proof of Theorem 2.2. By Lemma 4.7 and the same argument that leads to (22), we have

$$\bar{\ell}(Z^{(t)}, Z^*) \leq \delta_1^t \bar{\ell}(Z^{(0)}, Z^*) + \frac{1+\delta_2}{1-\delta_1}\frac{\sigma^2 d(d-1)}{2np},$$

for all $t \geq 1$, as long as $\bar{\ell}(Z^{(0)}, Z^*)$ is sufficiently small. The initial error $\bar{\ell}(Z^{(0)}, Z^*)$ is controlled by Lemma 4.8, and thus the desired result follows. □

4.4 *Proofs of Lemma 2.3 and Lemma 4.8*

We first state a property of the operator $\bar{\mathscr{P}}(\cdot)$.

Lemma 4.9. Consider a full-rank $X = (X_1, \cdots, X_d) \in \mathbb{R}^{d\times d}$, where $X_a \in \mathbb{R}^d$ is the ath column of X. Define $\widetilde{X} = (X_1, \cdots, X_{d-1}, -X_d)$ by changing the sign of the last column of X. Then, we have $\bar{\mathscr{P}}(X) = \bar{\mathscr{P}}(\widetilde{X})$.

Proof. Suppose X admits an SVD $X = UDV^{\mathrm{T}}$. Define $\widetilde{V}^{\mathrm{T}}$ by changing the sign of the last column of V^{T}. Then, the SVD of $\widetilde{X}$ is $\widetilde{X} = UD\widetilde{V}^{\mathrm{T}}$, and we have

$$\bar{\mathscr{P}}(X) = U\begin{pmatrix}1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \det(UV^{\mathrm{T}})\end{pmatrix}V^{\mathrm{T}} = U\begin{pmatrix}1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & -\det(UV^{\mathrm{T}})\end{pmatrix}\widetilde{V}^{\mathrm{T}} = \bar{\mathscr{P}}(\widetilde{X}),$$

where the last equality is by $-\det(UV^{\mathrm{T}}) = \det(U\widetilde{V}^{\mathrm{T}})$. □

Proofs of Lemma 2.3 and Lemma 4.8. For any $Z \in \mathbb{R}^{nd\times d}$ such that $Z/\sqrt{n} \in \mathscr{O}(nd, d)$, we have

$$\|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T)\circ Y - ZZ^{\mathrm{T}}\|_{\mathrm{F}}^2 = \|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T)\circ Y\|_{\mathrm{F}}^2 + n^2 d - 2\,\mathsf{Tr}\left((p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T)\circ Y)ZZ^{\mathrm{T}}\right).$$

Therefore, minimizing $\|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T)\circ Y - ZZ^{\mathrm{T}}\|_{\mathrm{F}}^2$ is equivalent to maximizing $\mathsf{Tr}\left((p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T)\circ Y)ZZ^{\mathrm{T}}\right)$. For $\mathscr{O}(d)$ synchronization, we can thus write $Z^{(0)}$ as

$$Z_i^{(0)} = \begin{cases}\mathscr{P}(\widehat{Z}_i), & \det(\widehat{Z}_i) \neq 0,\\ I_d, & \det(Z_i) = 0,\end{cases}$$

where

$$\widehat{Z} = \mathop{\mathrm{argmin}}_{Z\in\mathbb{R}^{nd\times d}: Z/\sqrt{n}\in\mathscr{O}(nd,d)} \|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T)\circ Y - ZZ^{\mathrm{T}}\|_{\mathrm{F}}^2.$$

For $\mathscr{SO}(d)$ synchronization, $Z^{(0)}$ is

$$Z_i^{(0)} = \begin{cases} \bar{\mathscr{P}}(\widehat{Z}_i), & \det(\widehat{Z}_i) \neq 0, \\ I_d, & \det(Z_i) = 0, \end{cases}$$

where $\widehat{Z}$ has the same definition.

Our first step is to derive a bound for $\widehat{Z}$. For both $\mathscr{O}(d)$ synchronization and $\mathscr{SO}(d)$ synchronization, it is clear that $Z^*/\sqrt{n} \in \mathscr{O}(nd, d)$. By the definition of $\widehat{Z}$, we have

$$\|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - \widehat{Z}\widehat{Z}^{\mathrm{T}}\|_{\mathrm{F}}^2 \leq \|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - Z^* Z^{*\mathrm{T}}\|_{\mathrm{F}}^2.$$

After rearrangement, we obtain

$$\|\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}\|_{\mathrm{F}}^2 \leq 2 \left| \mathsf{Tr}\left((\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}})(p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - Z^* Z^{*\mathrm{T}}) \right) \right|.$$

This implies

$$\|\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}\|_{\mathrm{F}} \leq 2 \left| \mathsf{Tr}\left(\frac{\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}}{\|\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}\|_{\mathrm{F}}} (p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - Z^* Z^{*\mathrm{T}}) \right) \right|.$$

Note that $\frac{\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}}{\|\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}\|_{\mathrm{F}}}$ is a matrix of rank at most $2d$, and thus it admits an eigendecomposition $\frac{\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}}{\|\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}\|_{\mathrm{F}}} = \sum_{j=1}^{2d} \lambda_j u u_j^{\mathrm{T}}$ with the eigenvalues satisfying $\sum_{j=1}^{2d} \lambda_j^2 = 1$. Then, we have

$$\begin{aligned} & \left| \mathsf{Tr}\left(\frac{\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}}{\|\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}\|_{\mathrm{F}}} (p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - Z^* Z^{*\mathrm{T}}) \right) \right| \\ \leq\ & \sum_{j=1}^{2d} |\lambda_j| \left| u_j^{\mathrm{T}} (p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - Z^* Z^{*\mathrm{T}}) u_j \right| \\ \leq\ & \|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - Z^* Z^{*\mathrm{T}}\|_{\mathrm{op}} \sum_{j=1}^{2d} |\lambda_j| \\ \leq\ & \sqrt{2d} \|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - Z^* Z^{*\mathrm{T}}\|_{\mathrm{op}}. \end{aligned}$$

Hence,

$$\begin{aligned} & \|\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^* Z^{*\mathrm{T}}\|_{\mathrm{F}} \\ \leq\ & 2\sqrt{2d} \|p^{-1}(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y - Z^* Z^{*\mathrm{T}}\|_{\mathrm{op}} \\ \leq\ & 2\sqrt{2d} \frac{1}{p} \|((A - \mathbb{E}A) \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Z^* Z^{*\mathrm{T}}\|_{\mathrm{op}} + 2\sqrt{2d} \frac{\sigma}{p} \|(A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ W\|_{\mathrm{op}}. \end{aligned}$$

For the first term, we have

$$
\begin{aligned}
&\|((A-\mathbb{E}A)\otimes \mathbb{1}_d\, \mathbb{1}_d^T)\circ Z^*Z^{*\mathrm{T}}\|_{\mathrm{op}} \\
=\;& \max_{u^{\mathrm{T}}=(u_1^{\mathrm{T}},\cdots,u_n^{\mathrm{T}}):\sum_{i=1}^n \|u_i\|^2=1} \left|\sum_{i=1}^n\sum_{j=1}^n (A_{ij}-\mathbb{E}A_{ij})u_i^{\mathrm{T}}Z_iZ_j^{\mathrm{T}}u_j\right| \\
\leq\;& \max_{u^{\mathrm{T}}=(u_1^{\mathrm{T}},\cdots,u_n^{\mathrm{T}}):\sum_{i=1}^n \|u_i\|^2=1} \left|\sum_{i=1}^n\sum_{j=1}^n (A_{ij}-\mathbb{E}A_{ij})u_i^{\mathrm{T}}u_j\right| \\
\leq\;& \|(A-\mathbb{E}A)\otimes I_d\|_{\mathrm{op}} \\
\leq\;& \|A-\mathbb{E}A\|_{\mathrm{op}} \\
\leq\;& C_1\sqrt{np},
\end{aligned}
$$

with probability at least $1-n^{-10}$ by Lemma 4.1. The second term can be bounded by Lemma 4.2. That is,

$$
\|(A\otimes \mathbb{1}_d\, \mathbb{1}_d^T)\circ W\|_{\mathrm{op}} \leq C_2\sqrt{npd},
$$

with probability at least $1-n^{-10}$. Combining the above bounds, we have

$$
\frac{1}{n^2}\|\widehat{Z}\widehat{Z}^{\mathrm{T}} - Z^*Z^{*\mathrm{T}}\|_{\mathrm{F}}^2 \leq C_3\frac{d\left(1+\sigma^2 d\right)}{np},
$$

with probability at least $1-2n^{-10}$. Apply Lemma 4.6, and we have

$$
\min_{B\in\mathscr{O}(d)}\frac{1}{n}\sum_{i=1}^n\|\widehat{Z}_i - Z_i^*B\|_{\mathrm{F}}^2 \leq C_3\frac{d\left(1+\sigma^2 d\right)}{np}. \tag{68}
$$

Next, let us consider the setting of $\mathscr{O}(d)$ synchronization. For any $B\in\mathscr{O}(d)$, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\|Z_i^{(0)}-Z_i^*B\|_{\mathrm{F}}^2 \\
=\;& \frac{1}{n}\sum_{i=1}^n\|Z_i^{(0)}-Z_i^*B\|_{\mathrm{F}}^2\mathbb{I}\{\det(\widehat{Z}_i)\neq 0\} + \frac{1}{n}\sum_{i=1}^n\|Z_i^{(0)}-Z_i^*B\|_{\mathrm{F}}^2\mathbb{I}\{\det(\widehat{Z}_i)=0\} \\
\leq\;& \frac{1}{n}\sum_{i=1}^n\|\mathscr{P}(\widehat{Z}_i)-Z_i^*B\|_{\mathrm{F}}^2\mathbb{I}\{\det(\widehat{Z}_i)\neq 0\} \\
&+\frac{1}{n}\sum_{i=1}^n\|Z_i^{(0)}-Z_i^*B\|_{\mathrm{F}}^2\mathbb{I}\left\{\left\|\widehat{Z}_i - Z_i^*B\right\|_{\mathrm{F}} \geq e^{d^{-1}\log 2}-1\right\}
\end{aligned} \tag{69}
$$

$$\leq \frac{4}{n}\sum_{i=1}^{n}\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}}^2 + \frac{4d}{n}\sum_{i=1}^{n}\frac{\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}}^2}{\left(e^{d^{-1}\log 2}-1\right)^2} \tag{70}$$

$$\leq C_4 d^3 \frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}}^2,$$

where (69) is because $\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}} < e^{d^{-1}\log 2} - 1$ implies $\det(\widehat{Z}_i) \neq 0$ by Lemma 4.5, and (70) is by Lemma 2.1 and Markov inequality. Taking minimum on both sides and applying (68), we have

$$\min_{B\in\mathcal{O}(d)} \frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)} - Z_i^* B\|_{\mathrm{F}}^2 \leq C_5 \frac{d^4\left(1+\sigma^2 d\right)}{np},$$

with probability at least $1 - 2n^{-10}$.

Finally, we consider the setting of $\mathcal{SO}(d)$ synchronization. We know that $Z_i^* \in \mathcal{SO}(d)$ for all $i \in [n]$. For any $B \in \mathcal{SO}(d)$, it is clear that we also have $Z_i^* B \in \mathcal{SO}(d)$, which implies $\det(Z_i^* B) > 0$. Then,

$$\frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)} - Z_i^* B\|_{\mathrm{F}}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)} - Z_i^* B\|_{\mathrm{F}}^2 \mathbb{I}\{\det(\widehat{Z}_i) > 0\} + \frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)} - Z_i^* B\|_{\mathrm{F}}^2 \mathbb{I}\{\det(\widehat{Z}_i) \leq 0\}$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}\|\mathcal{P}(\widehat{Z}_i) - Z_i^* B\|_{\mathrm{F}}^2 \mathbb{I}\{\det(\widehat{Z}_i) > 0\} \tag{71}$$

$$+\frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)} - Z_i^* B\|_{\mathrm{F}}^2 \mathbb{I}\{\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}} \geq e^{d^{-1}\log 2} - 1\} \tag{72}$$

$$\leq \frac{4}{n}\sum_{i=1}^{n}\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}}^2 + \frac{4d}{n}\sum_{i=1}^{n}\frac{\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}}^2}{\left(e^{d^{-1}\log 2}-1\right)^2}$$

$$\leq C_4 d^3 \frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}}^2.$$

In (71) we have used the fact that $Z_i^{(0)} = \bar{\mathcal{P}}(\widehat{Z}_i) = \mathcal{P}(\widehat{Z}_i)$ when $\det(\widehat{Z}_i) > 0$, and (72) is because $\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}} < e^{d^{-1}\log 2} - 1$ implies $\det(\widehat{Z}_i) > 0$ by Lemma 4.5 and we know the fact that $\det(Z_i^* B) > 0$. Taking minimum on both sides of the inequality, we have

$$\min_{B\in\mathcal{SO}(d)} \frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)} - Z_i^* B\|_{\mathrm{F}}^2 \leq C_4 d^3 \min_{B\in\mathcal{SO}(d)} \frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i - Z_i^* B\|_{\mathrm{F}}^2. \tag{73}$$

For any $B \in \mathcal{O}(d)\backslash\mathcal{SO}(d)$, we can define $\widetilde{B}$ by changing the sign of the last column of B, and then we have $\widetilde{B} \in \mathcal{SO}(d)$. We also define $\widetilde{Z}_i$ by changing the sign of the last column of $\widehat{Z}_i$. Note that $\det(Z_i^* \widetilde{B}) > 0$, and thus,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)}-Z_i^*\widetilde{B}\|_{\mathrm{F}}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)}-Z_i^*\widetilde{B}\|_{\mathrm{F}}^2\mathbb{I}\{\det(\widetilde{Z}_i)>0\}+\frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)}-Z_i^*\widetilde{B}\|_{\mathrm{F}}^2\mathbb{I}\{\det(\widetilde{Z}_i)\le 0\} \\
&\le \frac{1}{n}\sum_{i=1}^{n}\|\mathscr{P}(\widetilde{Z}_i)-Z_i^*\widetilde{B}\|_{\mathrm{F}}^2\mathbb{I}\{\det(\widetilde{Z}_i)>0\} \qquad (74)\\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)}-Z_i^*\widetilde{B}\|_{\mathrm{F}}^2\mathbb{I}\{\|\widetilde{Z}_i-Z_i^*\widetilde{B}\|_{\mathrm{F}}\ge e^{d^{-1}\log 2}-1\} \qquad (75)\\
&\le \frac{4}{n}\sum_{i=1}^{n}\|\widetilde{Z}_i-Z_i^*\widetilde{B}\|_{\mathrm{F}}^2+\frac{4d}{n}\sum_{i=1}^{n}\frac{\|\widetilde{Z}_i-Z_i^*\widetilde{B}\|_{\mathrm{F}}^2}{\left(e^{d^{-1}\log 2}-1\right)^2} \\
&\le C_4 d^3\frac{1}{n}\sum_{i=1}^{n}\|\widetilde{Z}_i-Z_i^*\widetilde{B}\|_{\mathrm{F}}^2 \\
&= C_4 d^3\frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i-Z_i^*B\|_{\mathrm{F}}^2. \qquad (76)
\end{aligned}
$$

To see (74), we first apply Lemma 4.9 to obtain $Z_i^{(0)}=\bar{\mathscr{P}}(\widehat{Z}_i)=\bar{\mathscr{P}}(\widetilde{Z}_i)$, and then we have $\bar{\mathscr{P}}(\widetilde{Z}_i)=\mathscr{P}(\widetilde{Z}_i)$ when $\det(\widetilde{Z}_i)>0$. The bound (75) is obtained because $\|\widetilde{Z}_i-Z_i^*\widetilde{B}\|_{\mathrm{F}}<e^{d^{-1}\log 2}-1$ implies $\det(\widetilde{Z}_i)>0$ by Lemma 4.5 and we also know the fact that $\det(Z_i^*\widetilde{B})>0$. The last equality (76) is a direct consequence of the definitions of $\widetilde{Z}_i$ and $\widetilde{B}$. Taking minimum on both sides of the inequality, we have

$$
\min_{B\in\mathscr{SO}(d)}\frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)}-Z_i^*B\|_{\mathrm{F}}^2\le C_4d^3\min_{B\in\mathscr{O}(d)\setminus\mathscr{SO}(d)}\frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i-Z_i^*B\|_{\mathrm{F}}^2. \qquad (77)
$$

Combining (73), (77) and (68), we have

$$
\begin{aligned}
&\min_{B\in\mathscr{SO}(d)}\frac{1}{n}\sum_{i=1}^{n}\|Z_i^{(0)}-Z_i^*B\|_{\mathrm{F}}^2 \\
&\le \min\left(C_4d^3\min_{B\in\mathscr{SO}(d)}\frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i-Z_i^*B\|_{\mathrm{F}}^2, C_4d^3\min_{B\in\mathscr{O}(d)\setminus\mathscr{SO}(d)}\frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i-Z_i^*B\|_{\mathrm{F}}^2\right) \\
&= C_4d^3\min_{B\in\mathscr{O}(d)}\frac{1}{n}\sum_{i=1}^{n}\|\widehat{Z}_i-Z_i^*B\|_{\mathrm{F}}^2 \\
&\le C_4\frac{d^4\left(1+\sigma^2 d\right)}{np},
\end{aligned}
$$

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

4.5 *Proofs of Lemma 3.1, Lemma 3.2 and Theorem 3.2*

Proof of Lemma 3.1. We use a mathematical induction argument. First, since

$$1-\left(r_{12}^2+\cdots+r_{1d}^2\right)\geq 1-\frac{1}{64d^4},$$

it is clear that s_{11} is well defined and satisfies $s_{11}\geq\frac{7}{8}$. Next, we study s_{21} and s_{22}. The equation (34) can be written as

$$s_{21}=-\frac{r_{13}r_{23}+\cdots+r_{1d}r_{2d}}{s_{11}}-\frac{r_{12}}{s_{11}}s_{22}. \tag{78}$$

We plug (78) into (35) and obtain a quadratic equation of s_{22}. A sufficient condition for a quadratic equation $ax^2+bx+c=0$ to have two real solutions is $ac<0$. For the quadratic equation of s_{22}, this condition is

$$1-\left(r_{23}^2+\cdots+r_{2d}^2\right)-\frac{(r_{13}r_{23}+\cdots+r_{1d}r_{2d})^2}{s_{11}^2}>0. \tag{79}$$

Since $s_{11}\geq\frac{7}{8}$ and $\max_{1\leq a<b\leq d}|r_{ab}|\leq\frac{1}{8d^{5/2}}$, (79) clearly holds, and thus, s_{22} is well defined. By (78), s_{21} is also well defined. We know from (35) that $|s_{22}|\leq 1$, and therefore, by (78), we have the bound

$$|s_{21}|\leq\frac{|r_{13}r_{23}+\cdots+r_{1d}r_{2d}|+|r_{12}|}{7/8}\leq\frac{1}{56d^4}+\frac{1}{7d^{5/2}}\leq\frac{1}{4d^2},$$

and thus,

$$s_{22}^2=1-s_{21}^2-(r_{23}^2+\cdots+r_{2d}^2)\geq 1-\frac{1}{16d^3},$$

which implies $s_{22}\geq\frac{7}{8}$.

Suppose $\max_{1\leq b<a\leq k-1}|s_{ab}|\leq\frac{1}{4d^2}$ and $\min_{a\in[k-1]}s_{aa}\geq\frac{7}{8}$, now we study $s_{k1},s_{k2},\cdots,s_{kk}$. Define $\widetilde{Q}_{k-1}\in\mathbb{R}^{(k-1)\times(k-1)}$ to be a diagonal matrix with the same diagonal elements as Q_{k-1}. Then, we have

$$\|Q_{k-1}-\widetilde{Q}_{k-1}\|_{\mathrm{op}}\leq\|Q_{k-1}-\widetilde{Q}_{k-1}\|_{\ell_\infty}\leq\frac{1}{4d},$$

which implies

$$s_{\min}(Q_{k-1})\geq s_{\min}(\widetilde{Q}_{k-1})-\|Q_{k-1}-\widetilde{Q}_{k-1}\|_{\mathrm{op}}\geq\frac{7}{8}-\frac{1}{4d}\geq\frac{3}{4}. \tag{80}$$

Thus, Q_{k-1} is invertible and we can write (34) as

$$S_{k-1}=-s_{kk}Q_{k-1}^{-1}R_{k-1}-Q_{k-1}^{-1}v_{k-1}. \tag{81}$$

We plug (81) into (35) and obtain a quadratic equation of s_{kk}. Similar to (79), a sufficient condition for the quadratic equation to have two real solutions is

$$1 - \|Q_{k-1}^{-1} v_{k-1}\|^2 - \left(r_{k,k+1}^2 + \cdots + r_{kd}^2\right) > 0. \tag{82}$$

By (80) and $\max_{1 \le a < b \le d} |r_{ab}| \le \frac{1}{8d^{5/2}}$, we have $\|Q_{k-1}^{-1} v_{k-1}\|^2 \le (16/9)\|v_{k-1}\|^2 \le \frac{16/9}{64^2 d^7}$ and $r_{k,k+1}^2 + \cdots + r_{kd}^2 \le \frac{1}{64d^4}$, and therefore (82) holds. Thus, s_{kk} is well defined. By (81), $s_{k1}, \cdots, s_{k,k-1}$ are also well defined. We know from (35) that $|s_{kk}| \le 1$, and therefore, by (81), we have the bound

$$\|S_{k-1}\| \le \|Q_{k-1}^{-1} R_{k-1}\| + \|Q_{k-1}^{-1} v_{k-1}\| \le \frac{4}{3}\left(\|R_{k-1}\| + \|v_{k-1}\|\right) \le \frac{1}{6d^2} + \frac{1}{48d^{7/2}} \le \frac{1}{4d^2}.$$

We also have

$$s_{kk}^2 = 1 - \|S_{k-1}\|^2 - \left(r_{k,k+1}^2 + \cdots + r_{kd}^2\right) \ge 1 - \frac{1}{16d^4} - \frac{1}{64d^4},$$

which implies $s_{kk} \ge \frac{7}{8}$. To summarize, we have shown that $s_{k1}, s_{k2}, \cdots, s_{kk}$ are well defined. Moreover, we have $\max_{1 \le b < a \le k} |s_{ab}| \le \frac{1}{4d^2}$ and $\min_{a \in [k]} s_{aa} \ge \frac{7}{8}$. Hence, we conclude that $Q(r)$ is well defined, $Q(r) \in \mathscr{O}(d)$ and $\max_{1 \le b < a \le d} |s_{ab}| \le \frac{1}{4d^2}$ and $\min_{a \in [d]} s_{aa} \ge \frac{7}{8}$.

To prove $Q(r) \in \mathscr{SO}(d)$, it suffices to show $\det(Q(r)) > 0$. We define a diagonal matrix $\widetilde{Q}(r) \in \mathbb{R}^{d \times d}$ that has the same diagonal elements as $Q(r)$. We know that $\det(\widetilde{Q}(r)) \ge \left(\frac{7}{8}\right)^d > 0$. Since $\max_{1 \le b < a \le d} |s_{ab}| \le \frac{1}{4d^2}$ and $\max_{1 \le a < b \le d} |r_{ab}| \le \frac{1}{8d^{5/2}}$, we have $\|Q(r) - \widetilde{Q}(r)\|_{\rm op} \le (4d)^{-1}$. By Lemma 4.5 and, we have

$$\frac{|\det(Q(r)) - \det(\widetilde{Q}(r))|}{|\det(\widetilde{Q}(r))|} \le \left(\frac{8}{7}\|Q(r) - \widetilde{Q}(r)\|_{\rm op} + 1\right)^d - 1 \le \left(\frac{8}{7}\frac{1}{4d} + 1\right)^d - 1 < 1,$$

and therefore, we have $\det(Q(r)) > 0$, which implies $Q(r) \in \mathscr{SO}(d)$.

Finally, we analyze the derivative of $Q(r)$ with respect to each r_{ab}. This is also done via a mathematical induction argument. First, by the formula of s_{11}, we have $\left|\frac{\partial s_{11}}{\partial r_{ab}}\right| \le \frac{1}{7d^{5/2}}$. Suppose $\sqrt{\left\|\frac{\partial S_{l-1}}{\partial r_{ab}}\right\|^2 + \left|\frac{\partial s_{ll}}{\partial r_{ab}}\right|^2} \le 5$ for all $l \in [k-1]$, now we study $\frac{\partial S_{k-1}}{\partial r_{ab}}$ and $\frac{\partial s_{kk}}{\partial r_{ab}}$. We take derivatives of both sides of (34) and (35) with respect to r_{ab} and obtain

$$Q_k \begin{pmatrix} \frac{\partial S_{k-1}}{\partial r_{ab}} \\ \frac{\partial s_{kk}}{\partial r_{ab}} \end{pmatrix} = \begin{pmatrix} -\frac{\partial v_{k-1}}{\partial r_{ab}} - \frac{\partial Q_{k-1}}{\partial r_{ab}} S_{k-1} - s_{kk}\frac{\partial R_{k-1}}{\partial r_{ab}} \\ -\frac{1}{2}\frac{\partial (r_{k,k+1}^2 + \cdots + r_{kd}^2)}{\partial r_{ab}} \end{pmatrix}. \tag{83}$$

By the definitions of v_{k-1} and R_{k-1}, we have $\left\|\frac{\partial v_{k-1}}{\partial r_{ab}}\right\| \le \frac{1}{8d^{5/2}}$ and $\left\|s_{kk}\frac{\partial R_{k-1}}{\partial r_{ab}}\right\| \le 1$. By the condition $\sqrt{\left\|\frac{\partial S_{l-1}}{\partial r_{ab}}\right\|^2 + \left|\frac{\partial s_{ll}}{\partial r_{ab}}\right|^2} \le 5$ for all $l \in [k-1]$, we have

$$\left\|\frac{\partial Q_{k-1}}{\partial r_{ab}}\right\|_{\mathrm{op}} \le \left\|\frac{\partial Q_{k-1}}{\partial r_{ab}}\right\|_{\ell_\infty} \le \sqrt{d} \max_{1\le l\le k-1} \sqrt{\left\|\frac{\partial S_{l-1}}{\partial r_{ab}}\right\|^2 + \left|\frac{\partial s_{ll}}{\partial r_{ab}}\right|^2} + 1 \le 5\sqrt{d}+1,$$

which implies

$$\left\|\frac{\partial Q_{k-1}}{\partial r_{ab}} S_{k-1}\right\| \le \left\|\frac{\partial Q_{k-1}}{\partial r_{ab}}\right\|_{\mathrm{op}} \|S_{k-1}\| \le (5\sqrt{d}+1)\sqrt{\frac{1}{4d}}.$$

We also have $\left|\frac{1}{2}\frac{\partial(r_{k,k+1}^2+\cdots+r_{kd}^2)}{\partial r_{ab}}\right| \le \frac{1}{8d^{5/2}}$. By (80) and (83), we have

$$\left\|\begin{pmatrix} \frac{\partial S_{k-1}}{\partial r_{ab}} \\ \frac{\partial s_{kk}}{\partial r_{ab}} \end{pmatrix}\right\|^2 \le \frac{16}{9}\left\|\begin{pmatrix} -\frac{\partial v_{k-1}}{\partial r_{ab}} - \frac{\partial Q_{k-1}}{\partial r_{ab}} S_{k-1} - s_{kk}\frac{\partial R_{k-1}}{\partial r_{ab}} \\ -\frac{1}{2}\frac{\partial(r_{k,k+1}^2+\cdots+r_{kd}^2)}{\partial r_{ab}} \end{pmatrix}\right\|^2 \le 25,$$

and therefore, $\sqrt{\left\|\frac{\partial S_{l-1}}{\partial r_{ab}}\right\|^2 + \left|\frac{\partial s_{ll}}{\partial r_{ab}}\right|^2} \le 5$ also holds for $l = k$. The proof is complete. □

Proof of Lemma 3.2. Without loss of generality, we consider the problem with $i = 1$ and $j = 2$. Let us first understand the likelihood function for the problem. Given the knowledge of $Z_3, \cdots, Z_n$, we can decompose the likelihood function as

$$p((A \otimes \mathbb{1}_d \mathbb{1}_d^T) \circ Y, A) = p(A)p(A_{12}Y_{12}|A)\left(\prod_{i=3}^n p(A_{1i}Y_{1i}|A)p(A_{2i}Y_{2i}|A)\right) \times \prod_{3\le i<j\le n} p(A_{ij}Y_{ij}|A).$$

Note that the part of the above decomposition that depends on Z_1 or Z_2 is

$$p(A_{12}Y_{12}|A)\left(\prod_{i=3}^n p(A_{1i}Y_{1i}|A)p(A_{2i}Y_{2i}|A)\right), \tag{84}$$

which is proportional to the product of the density functions of $\mathcal{MN}(Z_1Z_2^{\mathrm{T}}, \sigma^2 A_{12}I_d, I_d)$, $\mathcal{MN}\left(\left(\sum_{i=3}^n A_{1i}\right)Z_1, \sigma^2\left(\sum_{i=3}^n A_{1i}\right)I_d, I_d\right)$ and $\mathcal{MN}\left(\left(\sum_{i=3}^n A_{2i}\right)Z_2, \sigma^2\left(\sum_{i=3}^n A_{2i}\right)I_d, I_d\right)$.

With the parametrization $Z_1 = Z_1(r) = Q(r)$ and $Z_2 = Z_2(r) = Q(r')$, we can write $T = T(r, r') = Z_1(r)Z(r')^{\mathrm{T}}$, and the logarithm of (84) as $\ell(r, r')$. For the simplicity of notation, we re-index $\{r_{ab}\}_{1\le a<b\le d}$ and $\{r'_{ab}\}_{1\le a<b\le d}$ by $\{r_l\}_{1\le l\le d(d-1)/2}$ and $\{r'_l\}_{1\le l\le d(d-1)/2}$, respectively. The order of the re-indexing

process is not important. Then, the information matrix is given by

$$B = \mathbb{E}\begin{pmatrix} \nabla_r \ell(r,r')\nabla_r \ell(r,r')^{\mathrm{T}} & \nabla_r \ell(r,r')\nabla_{r'} \ell(r,r')^{\mathrm{T}} \\ \nabla_{r'} \ell(r,r')\nabla_r \ell(r,r')^{\mathrm{T}} & \nabla_{r'} \ell(r,r')\nabla_{r'} \ell(r,r')^{\mathrm{T}} \end{pmatrix},$$

where the expectation is induced by the distribution (84). Define the Jacobians $G_1 = \frac{\partial \mathrm{vec}(Z_1(r))}{\partial r} \in \mathbb{R}^{\frac{d(d-1)}{2}\times d^2}$ and $G_2 = \frac{\partial \mathrm{vec}(Z_2(r'))}{\partial r'} \in \mathbb{R}^{\frac{d(d-1)}{2}\times d^2}$. By direct calculation, we can write $B = B_1 + B_2 \in \mathbb{R}^{d(d-1)\times d(d-1)}$, where

$$B_1 = \frac{p}{\sigma^2}\begin{pmatrix} G_1G_1^{\mathrm{T}} & G_1G_2^{\mathrm{T}} \\ G_2G_1^{\mathrm{T}} & G_2G_2^{\mathrm{T}} \end{pmatrix},$$

and

$$B_2 = \frac{(n-2)p}{\sigma^2}\begin{pmatrix} G_1G_1^{\mathrm{T}} & 0 \\ 0 & G_2G_2^{\mathrm{T}} \end{pmatrix}.$$

Define

$$F = \left((Z_2 \otimes I_d)G_1^{\mathrm{T}} \ \ (I_d \otimes Z_1)G_2^{\mathrm{T}}\right) \in \mathbb{R}^{d^2\times d(d-1)}.$$

By Equation (11) of the paper [17], we have

$$\begin{aligned} &\inf_{\widehat{T}} \int\int \mathbb{E}_Z \|\widehat{T} - Z_1(r)Z_2(r')^{\mathrm{T}}\|_{\mathrm{F}}^2 \mathrm{d}P(r)\mathrm{d}P(r') \\ &\geq \int\int \mathsf{Tr}(J(r,r'))\mathrm{d}P(r)\mathrm{d}P(r') - I(P), \end{aligned} \tag{85}$$

where $J(r,r') = FB^{-1}F^{\mathrm{T}}$ and $I(P)$ is the information of the distribution P that will be elaborated later.

To analyze (85), we first need to show B is invertible so that $J(r,r')$ is well defined. For any unit vector $v \in \mathbb{R}^{\frac{d(d-1)}{2}}$, we have

$$v^{\mathrm{T}}G_1G_1^{\mathrm{T}}v = \left\|\sum_{l=1}^{\frac{d(d-1)}{2}} v_l \frac{\partial \mathrm{vec}(Z_1(r))}{\partial r_l}\right\|^2 \geq \left\|\sum_{l=1}^{\frac{d(d-1)}{2}} v_l \frac{\partial r}{\partial r_l}\right\|^2 = \sum_{l=1}^{\frac{d(d-1)}{2}} v_l^2 = 1.$$

The inequality above is because r can be viewed as a sub-vector of $\mathrm{vec}(Z_1(r))$. Recall the definition of $\{s_{ab}\}_{1\leq b\leq a\leq d}$ in the parametrization of $Q(r)$, and we also have

$$
\begin{aligned}
v^{\mathrm{T}} G_1 G_1^{\mathrm{T}} v &= \left\| \sum_{l=1}^{\frac{d(d-1)}{2}} v_l \frac{\partial \mathrm{vec}(Z_1(r))}{\partial r_l} \right\|^2 \\
&= \left\| \sum_{l=1}^{\frac{d(d-1)}{2}} v_l \frac{\partial r}{\partial r_l} \right\|^2 + \sum_{1 \le b \le a \le d} \left(\sum_{l=1}^{\frac{d(d-1)}{2}} v_l \frac{\partial s_{ab}}{\partial r_l} \right)^2 \\
&\le 1 + \sum_{1 \le b \le a \le d} \sum_{l=1}^{\frac{d(d-1)}{2}} \left| \frac{\partial s_{ab}}{\partial r_l} \right|^2 \\
&\le 1 + \frac{25d^3}{2},
\end{aligned}
$$

where the last inequality is by Lemma 3.1. Therefore, we have

$$
1 \le s_{\min}(G_1 G_1^{\mathrm{T}}) \le s_{\max}(G_1 G_1^{\mathrm{T}}) \le 1 + \frac{25d^3}{2},
$$

and the same bounds also apply to $G_2 G_2^{\mathrm{T}}$. We also have

$$
\|G_1 G_2^{\mathrm{T}}\|_{\mathrm{op}} \le \left(\max_{\|v\|=1} \|G_1 v\| \right) \left(\max_{\|v\|=1} \|G_2 v\| \right) \le 1 + \frac{25d^3}{2}.
$$

With the above bounds, we immediately have

$$
\|B_2\|_{\mathrm{op}} \le \frac{(n-2)p}{\sigma^2} \left(1 + \frac{25d^3}{2} \right), \tag{86}
$$

$$
s_{\min}(B_2) \ge \frac{(n-2)p}{\sigma^2}, \tag{87}
$$

$$
\|B_1\|_{\mathrm{op}} \le \frac{2p}{\sigma^2} \left(1 + \frac{25d^3}{2} \right). \tag{88}
$$

Therefore, under the condition that d is bounded by a constant, we know that both $B_1 + B_2$ and B_2 are invertible when n is sufficiently large.

Now we study the first term of (85). We can lower bound $\mathsf{Tr}(J(r, r'))$ by

$$
\begin{aligned}
\mathsf{Tr}(J(r, r')) &\ge \mathsf{Tr}(F B_2^{-1} F^{\mathrm{T}}) - |\mathsf{Tr}(F((B_1 + B_2)^{-1} - B_2^{-1}) F^{\mathrm{T}})| \\
&\ge \mathsf{Tr}(F B_2^{-1} F^{\mathrm{T}}) \left(1 - \|B_2^{1/2} (B_1 + B_2)^{-1} B_2^{1/2} - I_2\|_{\mathrm{F}} \right).
\end{aligned}
$$

By the definitions of B_2 and F, we have

$$\begin{aligned}\mathsf{Tr}(FB_2^{-1}F^{\mathrm{T}}) &= \frac{\sigma^2}{(n-2)p}\mathsf{Tr}\left((Z_2\otimes I_d)G_1^{\mathrm{T}}(G_1G_1^{\mathrm{T}})^{-1}G_1(Z_2^{\mathrm{T}}\otimes I_d)\right)\\ &\quad+\frac{\sigma^2}{(n-2)p}\mathsf{Tr}\left((Z_1\otimes I_d)G_2^{\mathrm{T}}(G_2G_2^{\mathrm{T}})^{-1}G_2(Z_1^{\mathrm{T}}\otimes I_d)\right)\\ &= \frac{\sigma^2 d(d-1)}{(n-2)p}.\end{aligned}$$

By (86), (87) and (88), we have

$$\|B_2^{1/2}(B_1+B_2)^{-1}B_2^{1/2}-I_2\|_{\mathrm{F}} \le \|B_2\|_{\mathrm{op}}\|B_2^{-1}\|_{\mathrm{op}}\|(B_1+B_2)^{-1}\|_{\mathrm{op}}\|B_1\|_{\mathrm{F}} \le C_1\frac{d^7}{n}.$$

Hence, the first term of (85) has the following lower bound,

$$\int\int \mathsf{Tr}(J(r,r'))\mathrm{d}P(r)\mathrm{d}P(r') \ge \left(1-\frac{C_2}{n}\right)\frac{\sigma^2 d(d-1)}{(n-2)p}, \tag{89}$$

for some constant C_2 only depending on the bound of d.

Finally, we need to give an upper bound for $I(P)$. Let $\lambda(\cdot)$ be the density function of P, and then $I(P)$ is defined by

$$I(P) = \int\sum_{ikl}\frac{1}{\lambda(r)\lambda(r')}\left(\frac{\partial}{\partial r_k}K_{ik}(r,r')\lambda(r)\lambda(r')\right)\left(\frac{\partial}{\partial r_l}K_{il}(r,r')\lambda(r)\lambda(r')\right)\mathrm{d}r\mathrm{d}r',$$

where $K(r,r') = FB^{-1}$. Given the definition of $\lambda(\cdot)$, we have the bound

$$I(P) \le C_3\left(\max_{r,r'}\max_{i,k}\left|\frac{\partial}{\partial r_k}K_{ik}(r,r')\right| + \max_{r,r'}\max_{i,k}|K_{ik}(r,r')|\right)^2,$$

where C_3 is some constant only depending on the bound of d and the maximum is taken over all r and r' that satisfy $\max_{1\le a<b\le d}|r_{ab}| \le \frac{1}{8d^{5/2}}$ and $\max_{1\le a<b\le d}|r'_{ab}| \le \frac{1}{8d^{5/2}}$. Though this bound can be computed explicitly using formulas of matrix derivatives, we omit the long and tedious details here. Intuitively, each entry of $K(r,r')$ is a smooth function of r and r', and the orders of $\frac{\partial}{\partial r_k}K_{ik}(r,r')$ and $K_{ik}(r,r')$ only depend on that of B_2, since the contribution of B_1 is negligible. We thus have $I(P) \le C_4\left(\frac{\sigma^2}{np}\right)^2$ for some constant C_4 only depending on the bound of d. By (85) and (89), we have

$$\inf_{\widehat{T}}\int\int \mathbb{E}_Z\|\widehat{T}-Z_1(r)Z_2(r')^{\mathrm{T}}\|_{\mathrm{F}}^2\mathrm{d}P(r)\mathrm{d}P(r') \ge \left(1-C_5\left(\frac{1}{n}+\frac{\sigma^2}{np}\right)\right)\frac{\sigma^2 d(d-1)}{np},$$

and the proof is complete. □

Proof of Theorem 3.2. By Lemma 4.6 and the same argument that leads to (31), we have

$$\inf_{\widehat{Z}\in\mathscr{SO}(d)^n}\sup_{Z\in\mathscr{SO}(d)^n}\mathbb{E}_Z\bar{\ell}(\widehat{Z},Z)$$
$$\geq \frac{1}{2n^2}\sum_{1\leq i\neq j\leq n}\int\left(\inf_{\widehat{T}}\int\int\mathbb{E}_Z\|\widehat{T}-Z_iZ_j^{\mathrm{T}}\|_{\mathrm{F}}^2\mathrm{d}\Pi(Z_i)\mathrm{d}\Pi(Z_j)\right)\prod_{k\in[n]\backslash\{i,j\}}\mathrm{d}\Pi(Z_k).$$

Since $\mathrm{supp}(\Pi)\subset\mathscr{SO}(d)$ by Lemma 3.1, the conclusion of Lemma 3.2 also applies here, and thus, we have

$$\inf_{\widehat{T}}\int\int\mathbb{E}_Z\|\widehat{T}-Z_i(r)Z_j(r')^{\mathrm{T}}\|_{\mathrm{F}}^2\mathrm{d}P(r)\mathrm{d}P(r')\geq\left(1-C\left(\frac{1}{n}+\frac{\sigma^2}{np}\right)\right)\frac{\sigma^2d(d-1)}{np}.$$

This leads to the desired result. □

Data Availability Statement

No new data were generated or analyzed in support of this work.

Funding

NSF CAREER Award [DMS-1847590 to CG]; NSF Grant [CCF-1934931 to CG, DMS-2112988 to AZ].

References

1. Arie-Nachimson, M., Kovalsky, S. Z., Kemelmacher-Shlizerman, I., Singer, A. & Basri, R. (2012) Global motion estimation from point matches. *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*.IEEE, pp. 81–88.
2. Bandeira, A. S., Boumal, N. & Singer, A. (2017) Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Math. Programming*, **163**, 145–167.
3. Bandeira, A. S. & Van Handel, R. (2016) Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, **44**, 2479–2506.
4. Boumal, N. (2016) Nonconvex phase synchronization. *SIAM J. Optim.*, **26**, 2355–2377.
5. Boumal, N., Singer, A. & Absil, P.-A. (2013) Robust estimation of rotations from relative measurements by maximum likelihood. *52nd IEEE Conference on Decision and Control*.IEEE, pp. 1156–1161.
6. Boumal, N., Singer, A., Absil, P.-A. & Blondel, V. D. (2014) Cramér–Rao bounds for synchronization of rotations. *Inform. Inference*, **3**, 1–39.
7. Chen, Y. & Candès, E. J. (2018) The projected power method: an efficient algorithm for joint alignment from pairwise differences. *Comm. Pure Appl. Math.*, **71**, 1648–1714.
8. Cucuringu, M., Lipman, Y. & Singer, A. (2012) Sensor network localization by eigenvector synchronization over the Euclidean group. *ACM Trans. Sensor Netw.*, **8**, 1–42.
9. Cucuringu, M., Singer, A. & Cowburn, D. (2012) Eigenvector synchronization, graph rigidity and the molecule problem. *Inform. Inference*, **1**, 21–67.
10. Fan, Y., Khoo, Y. & Zhao, Z. (2021) Joint community detection and rotational synchronization via semidefinite programming. arXiv preprint arXiv:2105.06031.
11. Fei, Y. & Chen, Y. (2020) Achieving the Bayes error rate in synchronization and block models by SDP, robustly. *IEEE Trans. Inform. Theory*, **66**, 3929–3953.

12. Filbir, F., Krahmer, F. & Melnyk, O. (2020) On recovery guarantees for angular synchronization. arXiv preprint arXiv:2005.02032.
13. Gao, C., Lu, Y., Ma, Z. & Zhou, H. H. (2016) Optimal estimation and completion of matrices with biclustering structures. *J. Mach. Learn. Res.*, **17**, 5602–5630.
14. Gao, C. & Zhang, A. Y. (2019) *Iterative algorithm for discrete structure recovery*. arXiv preprint arXiv:1911.01018.
15. Gao, C. & Zhang, A. Y. (2020) Exact minimax estimation for phase synchronization. arXiv preprint arXiv:2010.04345.
16. Gao, C. & Zhang, A. Y. (2021) SDP achieves exact minimax optimality in phase synchronization. arXiv preprint arXiv:2101.02347.
17. Gill, R. D., Levit, B. Y., et al. (1995) Applications of the van trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, **1**, 59–79.
18. Gower, J. C., Dijksterhuis, G. B., et al. (2004) *Procrustes Problems*, vol. **30**.Oxford University Press on Demand.
19. Ipsen, I. C. & Rehman, R. (2008) Perturbation bounds for determinants and characteristic polynomials. *SIAM J. Matrix Anal. Appl.*, **30**, 762–776.
20. Kabsch, W. (1978) A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A*, **34**, 827–828.
21. Lei, J. & Rinaldo, A. (2015) Consistency of spectral clustering in stochastic block models. *Ann. Statist.*, **43**, 215–237.
22. Li, R.-C. (1995) New perturbation bounds for the unitary polar factor. *SIAM J. Matrix Anal. Appl.*, **16**, 327–332.
23. Ling, S. (2020a) Improved performance guarantees for orthogonal group synchronization via generalized power method. arXiv preprint arXiv:2012.00470.
24. Ling, S. (2020b) Solving orthogonal group synchronization via convex and low-rank optimization: tightness and landscape analysis. arXiv preprint arXiv:2006.00902.
25. Ling, S. (2022) Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Appl. Comput. Harmon. Anal.*.
26. Liu, H., Yue, M.-C. & Man-Cho So, A. (2017) On the estimation performance and convergence rate of the generalized power method for phase synchronization. *SIAM J. Optim.*, **27**, 2426–2446.
27. Liu, H., Yue, M.-C. & So, A. M.-C. (2020) A unified approach to synchronization problems over subgroups of the orthogonal group. arXiv preprint arXiv:2009.07514.
28. Perry, A., Wein, A. S., Bandeira, A. S. & Moitra, A. (2018) Message-passing algorithms for synchronization problems over compact groups. *Comm. Pure Appl. Math.*, **71**, 2275–2322.
29. Romanov, E. & Gavish, M. (2020) The noise-sensitivity phase transition in spectral group synchronization over compact groups. *Appl. Comput. Harmon. Anal.*, **49**, 935–970.
30. Shkolnisky, Y. & Singer, A. (2012) Viewing direction estimation in cryo-em using synchronization. *SIAM J. Imaging Sci.*, **5**, 1088–1110.
31. Singer, A. (2011) Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.*, **30**, 20–36.
32. Singer, A. & Shkolnisky, Y. (2011) Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM J. Imaging Sci.*, **4**, 543–572.
33. Talagrand, M. (1995) Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, **81**, 73–205.
34. Vershynin, R. (2010) Introduction to the non-asymptotic analysis of random matrices. arXiv preprint arXiv:1011.3027.
35. Wang, L. & Singer, A. (2013) Exact and stable recovery of rotations for robust synchronization. *Inform. Inference*, **2**, 145–193.
36. Zhong, Y. & Boumal, N. (2018) Near-optimal bounds for phase synchronization. *SIAM J. Optim.*, **28**, 989–1016.