

Tightness of SDP and Burer-Monteiro Factorization for Phase Synchronization in High-Noise Regime

Anderson Ye Zhang

University of Pennsylvania

Abstract

We study the difference between the maximum likelihood estimation (MLE) and its semi-definite programming (SDP) relaxation for the phase synchronization problem, where n latent phases are estimated based on pairwise observations corrupted by Gaussian noise at a level σ . While previous studies have established that SDP coincides with the MLE when $\sigma \lesssim \sqrt{n/\log n}$, the behavior in the high-noise regime $\sigma \gtrsim \sqrt{n/\log n}$ remains unclear. We address this gap by quantifying the deviation between the SDP and the MLE in the high-noise regime as $\exp(-c\frac{n}{\sigma^2})$, indicating an exponentially small discrepancy. In fact, we establish more general results for the Burer-Monteiro (BM) factorization that covers the SDP as a special case: it has the exponentially small deviation from the MLE in the high-noise regime and coincides with the MLE when σ is small. To obtain our results, we develop a refined entrywise analysis of the MLE that is beyond the existing ℓ_∞ analysis in literature.

1 Introduction

In this paper, we study the phase synchronization problem [31, 3, 1, 38]. Let $z_1^*, \dots, z_n^* \in \mathbb{C}_1$ be latent parameters where $\mathbb{C}_1 = \{x \in \mathbb{C} : |x| = 1\}$ includes all unit complex numbers. That is, each z_j^* represents an angle in $[0, 2\pi)$ or a phase. The observations are

$$Y_{jk} = z_j^* \overline{z_k^*} + \sigma W_{jk}, \quad 1 \leq j < k \leq n, \quad (1.1)$$

where $\sigma > 0$ is the noise level and $\{W_{jk}\}_{1 \leq j < k \leq n} \in \mathbb{C}$ are the additive noises following the standard complex Gaussian distribution independently. This model can be conveniently expressed in matrix form. By defining $Y_{jj} = 1, W_{jj} = 0$ for all $j \in [n]$ and $Y_{kj} = \overline{Y_{jk}}, W_{kj} = \overline{W_{jk}}$ for all $1 \leq j < k \leq n$, we can rewrite the model as

$$Y = z^* (z^*)^H + \sigma W \in \mathbb{C}^{n \times n}, \quad (1.2)$$

where $z^* \in \mathbb{C}_1^n$ with coordinates z_1^*, \dots, z_n^* . The goal is to estimate the latent vector z^* from the observed matrix Y .

To solve the phase synchronization problem, one natural approach is to use maximum likelihood estimation (MLE) [15, 38]. The MLE can be formulated as the following optimization problem:

$$\hat{z}^{\text{MLE}} = \underset{z \in \mathbb{C}_1^n}{\operatorname{argmax}} \langle Y, zz^H \rangle. \quad (1.3)$$

However, this optimization is over a non-convex set \mathbb{C}_1^n , making it computationally challenging. To overcome this computational difficulty, note that \hat{z}^{MLE} satisfies

$$\hat{z}^{\text{MLE}} (\hat{z}^{\text{MLE}})^H = \underset{Z \in \mathbb{C}^{n \times n} : Z = Z^H, \operatorname{rank}(Z) = 1, Z_{jj} = 1, \forall j \in [n]}{\operatorname{argmax}} \langle Y, Z \rangle, \quad (1.4)$$

which can be relaxed into a convex problem using semi-definite programming (SDP) [2, 32, 20, 12, 35, 16, 19]:

$$\hat{Z}^{\text{SDP}} = \underset{Z \in \mathbb{C}^{n \times n}: Z = Z^H, Z \succeq 0, Z_{jj} = 1, \forall j \in [n]}{\operatorname{argmax}} \langle Y, Z \rangle. \quad (1.5)$$

Here, the optimization is over all n -by- n positive semi-definite complex matrices with diagonal entries equal to 1, which forms a convex set. Compared to (1.4), the formulation in (1.5) relaxes the rank constraint $\operatorname{rank}(Z) = 1$ to $Z \succeq 0$, thus making the feasible set convex.

While the SDP offers computational convenience, its solution is not guaranteed to be $\hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H$. A crucial question is how \hat{Z}^{SDP} differs from $\hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H$. If they coincide, the SDP relaxation is considered tight in the literature. [3] demonstrates that the SDP is tight when $\sigma \lesssim n^{1/4}$. This result is further refined by [38], which shows that the SDP is tight when $\sigma \lesssim \sqrt{n/\log n}$. However, the behavior of the SDP when $\sigma \gtrsim \sqrt{n/\log n}$, referred to as the high-noise regime in this paper, remains unclear. This motivates us to address the following question:

Question 1: How does the SDP differ from the MLE in the high-noise regime where $\sigma \gtrsim \sqrt{n/\log n}$?

In addition to the SDP relaxation, in recent years, the Burer-Monteiro (BM) factorization [8, 9, 7, 4, 28, 23] has drawn increasing attention. For any $m \in \mathbb{N}$, the BM factorization solves the following optimization problem:

$$\hat{Z}^{\text{BM},m} = \underset{Z \in \mathbb{C}^{n \times n}: Z = Z^H, \operatorname{rank}(Z) \leq m, Z \succeq 0, Z_{jj} = 1, \forall j \in [n]}{\operatorname{argmax}} \langle Y, Z \rangle. \quad (1.6)$$

Compared to the SDP, the BM factorization imposes an additional rank constraint. Note that when $m = 1$, the feasible set of the BM factorization is the set of all rank-1 Hermitian matrices with the non-zero eigenvalue being 1. Hence, the BM factorization is equivalent to the MLE in the sense that

$$\hat{Z}^{\text{BM},1} = \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H.$$

When $m \geq n$, the rank constraint is not effective, and the BM factorization is equivalent to the SDP. As a result, the BM factorization can be seen as a more conservative relaxation of the MLE compared to the SDP when $m < n$. In addition, the SDP can be seen as a special case of the BM factorization, such that

$$\hat{Z}^{\text{BM},n} = \hat{Z}^{\text{SDP}}.$$

With the SDP seen as a special case of the BM factorization, the question posed above about the SDP can be further generalized:

Question 2: How does the BM factorization differ from the MLE?

Note that the difference between the BM factorization and the MLE can be quantified by the following normalized squared Frobenius norm: $n^{-2} \|\hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H\|_F^2$, between $\hat{Z}^{\text{BM},m}$ and $\hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H$. Since both are $n \times n$ matrices with entries having modulus at most 1, the quantity is between 0 and 4. Hence, to address Question 2, we aim to establish an upper bound for $n^{-2} \|\hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H\|_F^2$. If it is equal to 0, then $\hat{Z}^{\text{BM},m} = \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H$ in which case we can say the BM factorization is tight.

The main results of this paper are presented below in Theorem 1.1.

Theorem 1.1. Suppose $m \in \mathbb{N} \setminus \{1\}$.

1. There exists some absolute constant $C > 0$ such that the following hold:

$$\mathbb{E} \left(\frac{1}{n^2} \left\| \hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}} (\hat{z}^{\text{MLE}})^{\text{H}} \right\|_{\text{F}}^2 \right) \leq C \exp \left(-\frac{n}{8\sigma^2} \right) + 2n^{-10}. \quad (1.7)$$

2. There exists some absolute constant $C' > 0$ such that if $\sigma \leq \min\{C'\sqrt{n}, \sqrt{n/(9 \log n)}\}$, then the following holds with high probability:

$$\hat{Z}^{\text{BM},m} = \hat{z}^{\text{MLE}} (\hat{z}^{\text{MLE}})^{\text{H}}.$$

Theorem 1.1 first provides an upper bound (1.7) for the expected value of $n^{-2} \|\hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}} (\hat{z}^{\text{MLE}})^{\text{H}}\|_{\text{F}}^2$. Note that $n^{-2} \|\hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}} (\hat{z}^{\text{MLE}})^{\text{H}}\|_{\text{F}}^2$ is random as both $\hat{Z}^{\text{BM},m}$ and \hat{z}^{MLE} depend on the random Gaussian noises $\{W_{jk}\}_{1 \leq j < k \leq n}$. Therefore, we take the expectation to obtain a deterministic upper bound. The upper bound comprises two terms. The first term has an exponential form with $\frac{n}{\sigma^2}$ in the exponent, which can be understood as the signal-to-noise ratio. The second term n^{-10} arises from a high-probability event controlling $\|W\|$ and can be made arbitrarily smaller, thus considered negligible compared to the first term. Ignoring the second term, the bound indicates the difference between $\hat{Z}^{\text{BM},m}$ and $\hat{z}^{\text{MLE}} (\hat{z}^{\text{MLE}})^{\text{H}}$ is exponentially small.

To better understand the exponential error term $\exp(-\frac{n}{8\sigma^2})$ in (1.7), particularly its magnitude, we compare it with the distances to the ground truth $z^*(z^*)^{\text{H}}$ when $\sigma = o(\sqrt{n})$. When $m = n$, the BM factorization becomes the SDP. [16] shows that $n^{-2} \|\hat{Z}^{\text{SDP}} - z^*(z^*)^{\text{H}}\|_{\text{F}}^2$, the difference between the SDP and the ground truth, is of the order $\frac{\sigma^2}{n}$. A similar result is established in [15] for the MLE. In addition, [15] demonstrates that $\frac{\sigma^2}{n}$ is the minimax rate for the estimation of $z^*(z^*)^{\text{H}}$ in the phase synchronization, implying that no estimator can achieve an error much smaller than $\frac{\sigma^2}{n}$. The left panel of Figure 1 visualizes the geometric relationship among these quantities. Note that $\exp(-\frac{n}{8\sigma^2})$ is much smaller than $\frac{\sigma^2}{n}$, especially when $\frac{n}{\sigma^2}$ is large. This reveals that while \hat{Z}^{SDP} and $\hat{z}^{\text{MLE}} (\hat{z}^{\text{MLE}})^{\text{H}}$ are distant from $z^*(z^*)^{\text{H}}$, they are relatively close to each other, indication that relaxing the feasible set in (1.4) to that in (1.6) only slightly alters the solution.

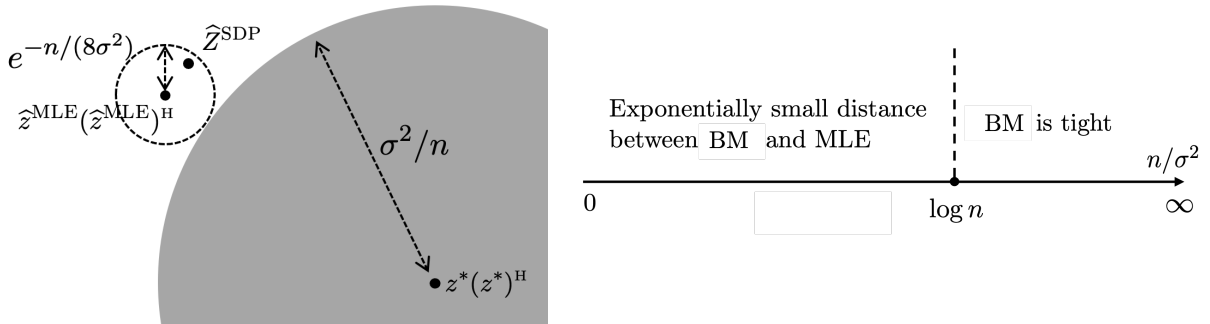


Figure 1: Left: A visualization of the geometric relationship among the SDP, the MLE, and the ground truth when $\sigma = o(\sqrt{n})$. Right: Summary of Theorem 1.1: The distance between the BM factorization and the MLE decays exponentially as $\frac{n}{\sigma^2}$ increases. The distance becomes 0, indicating tightness, when $\frac{n}{\sigma^2} \gtrsim \log n$.

While (1.7) holds for any noise level σ , Theorem 1.1 further provides a high-probability result for the tightness of the BM factorization. When n is large enough, a sufficient condition is $\sigma \leq$

$\sqrt{n/(9 \log n)}$. Compared to the condition $\sigma \lesssim \sqrt{n/\log n}$ for the SDP in the existing literature [38], our result holds for any BM factorization, including the SDP.

In summary, Theorem 1.1 offers comprehensive answers to the two questions posed above. Theorem 1.1 directly addresses Question 2, demonstrating how the BM factorization differs from the MLE as $\frac{n}{\sigma^2}$ increases (see the right panel of Figure 1 for an illustration). For Question 1, since the SDP is a special case of the BM factorization, our results indicate that the SDP differs from the MLE with an exponentially small error in the high-noise regime $\sigma \gtrsim \sqrt{n/\log n}$.

To establish Theorem 1.1, our analysis is beyond that of [38]. [38] connects the SDP and the MLE through a dual certificate. Instead, we leverage a property that both the BM factorization and the MLE are fixed points of certain mappings F_m and F_1 , respectively (see Section 2.2 for their definitions). The fixed-point property applies not just to the SDP but also the BM factorization, allowing us to establish a general framework for the BM factorization that includes the SDP as a special case. More importantly, this enables us to first establish contraction-type results for these mappings, ultimately showing that the difference between the BM factorization and the MLE can be upper bounded by a quantity of the MLE (see Corollary 2.4 for more details). As a result, the remaining proof is about analyzing the MLE. While [38] investigates a similar quantity when $\sigma \lesssim \sqrt{n/\log n}$ by developing an ℓ_∞ norm analysis for the MLE, their approach no longer works in the high-noise regime $\sigma \gtrsim \sqrt{n/\log n}$, due to the fact that the mapping F_1 involves entrywise normalization, which poses analytical challenges. To address this, our strategy is to replace F_1 with a Lipschitz mapping whose fixed points closely approximate \hat{z}^{MLE} and are relatively easier to analyze. This novel strategy, together with the leave-one-out technique developed in [38], allows us to establish the desired exponential bound presented in Theorem 1.1.

Related Literature. Recent years have seen a surge of interest in SDP relaxations for tackling various non-convex optimization problems with underlying low-rank structures. This includes applications in community detection [18, 13], clustering [17, 14], matrix completion [10], and phase retrieval [33]. Among these, phase synchronization is particularly notable for its relatively straightforward structure, making it a prime candidate for in-depth study.

The SDP in phase synchronization can be studied from several perspectives. From a statistical standpoint, as explored in [19, 16], the focus is on using SDP to estimate the true structure $z^*(z^*)^H$ and to evaluate its estimation error and statistical optimality. This involves analyzing the distance between \hat{Z}^{SDP} and the true $z^*(z^*)^H$. Conversely, studies such as [29, 3, 38] concentrate on the tightness of the SDP—specifically, the discrepancy between \hat{Z}^{SDP} and $\hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H$, essentially quantifying the cost of transforming a non-convex optimization problem into a convex one through SDP relaxation. [29] examines the differences between the objective function values, $\langle Y, \hat{Z}^{\text{SDP}} \rangle$ and $\langle Y, \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H \rangle$, while [3, 38] demonstrates that \hat{Z}^{SDP} equals $\hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H$ under low noise conditions. These works have inspired further investigations, such as those by [22, 21], into the tightness of SDP in related problems, like orthogonal group synchronization and the generalized orthogonal Procrustes problem. Our work extends the line of research initiated by [38] by focusing on the tightness of the SDP in the high-noise regime, an area not extensively covered by existing research. By doing so, we contribute to filling a crucial gap in understanding the limits and power of SDP relaxations under more challenging conditions. While our results are limited to the phase synchronization, they could potentially be extended to other synchronization problems such as the orthogonal group synchronization.

While SDP is a convex optimization approach solvable in polynomial time, its scalability issues have prompted the exploration of alternatives such as the BM factorization [8]. Despite its non-convex nature, BM factorization often exhibits surprisingly good performance when applied through local optimization algorithms. This observation has spurred a series of investigations into the con-

ditions under which SDP and BM factorization yield equivalent optima [9, 7, 6]. Furthermore, studies such as [28, 23, 27, 11] examine the landscape and benignness of the BM factorization's optimization process. Unlike much of the existing literature that focuses on comparing BM factorization directly with SDP, our work considers both SDP and BM factorization as relaxations of the MLE. Therefore, we explore the differences between BM factorization and MLE, rather than between BM factorization and SDP.

Regarding the statistical properties of phase synchronization, [15] establishes the minimax rate of $\frac{\sigma^2}{n}$, demonstrating its attainability via the MLE and a generalized power method. Subsequent research by [16, 37] confirms that both the SDP and the eigenvector method [31] are minimax optimal. Phase synchronization serves as a specific instance within the broader framework of group synchronization problems [1], where the elements $\{z_j^*\}_{j \in [n]}$ belong to various groups. The performance of several algorithms, including those mentioned, has also been studied in different synchronization settings such as \mathbb{Z}_2 synchronization [16] and orthogonal group synchronization [22, 37].

Organization. In Section 2, we conduct a deterministic analysis of the difference between the MLE and the BM factorization using their fixed-point properties. Section 3 focuses on analyzing the MLE. We include proofs of main results in Section 5. Due to the page limit, proofs of remaining lemmas are included in the supplementary material.

Notations. Define $\mathbb{N} = \{1, 2, 3, \dots\}$ as the set of natural numbers. For any positive integer n , we write $[n] = \{1, 2, \dots, n\}$ and denote I_n as the $n \times n$ identity matrix. For a complex number $x \in \mathbb{C}$, we use \bar{x} for its complex conjugate, $\text{Re}(x)$ for its real part, and $|x|$ for its modulus. Define $\mathbb{C}_{\leq 1} = \{x \in \mathbb{C} : |x| \leq 1\}$ as the set of complex numbers whose modulus are at most 1. For a complex vector $x = (x_j) \in \mathbb{C}^d$, we denote $\|x\| = (\sum_{j=1}^d |x_j|^2)^{1/2}$ as its Euclidean norm. For a complex matrix $B = (B_{jk}) \in \mathbb{C}^{d_1 \times d_2}$, we use $B^H \in \mathbb{C}^{d_2 \times d_1}$ for its conjugate transpose such that $(B^H)_{jk} = \overline{B_{kj}}$. The Frobenius norm and the operator norm of B are defined by $\|B\|_F := (\sum_{j=1}^{d_1} \sum_{k=1}^{d_2} |B_{jk}|^2)^{1/2}$ and $\|B\| := \sup_{u \in \mathbb{C}^{d_1}, v \in \mathbb{C}^{d_2} : \|u\| = \|v\| = 1} u^H B v$. We use the notation $B \succeq 0$ when B is positive semi-definite. Define B_j as its j th column and $B_{\cdot j}$ as its j th row. For a square matrix B , define $\text{Tr}(B)$ as its trace and $\text{rank}(B)$ as its rank. For two matrices $A = (A_{jk}) \in \mathbb{C}^{d_1 \times d_2}$ and $B = (B_{jk}) \in \mathbb{C}^{d_1 \times d_2}$, define $\langle A, B \rangle = \text{Tr}(A^H B) = \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} \overline{A_{jk}} B_{jk}$ as its Frobenius inner product. For two positive sequences $\{a_n\}$ and $\{b_n\}$, $a_n \lesssim b_n$ and $b_n \gtrsim a_n$ both mean $a_n \leq C b_n$ for some constant $C > 0$ independent of n . We also write $a_n = o(b_n)$ or $\frac{b_n}{a_n} \rightarrow \infty$ when $\limsup_n \frac{a_n}{b_n} = 0$. We use $\mathbb{I}\{\cdot\}$ as the indicator function.

2 A Deterministic Analysis Through Fixed Points

In this section, we show the difference between the MLE and the BM factorization can be upper bounded by a quantity related to the MLE. Our analysis is deterministic, using a fact the estimators are fixed points.

2.1 Introducing an Equivalent Representation of the BM Factorization and a Different Loss Function

We first introduce an equivalent representation of the BM factorization. For any $m \in \mathbb{N} \setminus \{1\}$, define

$$\mathcal{V}_m = \{V = (V_1, \dots, V_n) \in \mathbb{C}^{m \times n} : \|V_j\| = 1, \forall j \in [n]\}$$

as a set containing all $m \times n$ complex matrices with unit norm columns. Note that for any Z that is in the feasible set of (1.6), it can be represented as $Z = V^H V$ for some $V \in \mathcal{V}_m$, and vice versa. Consequently, the BM factorization can be equivalently formulated as

$$\hat{V}^{\text{BM},m} = \operatorname{argmax}_{V \in \mathcal{V}_m} \langle Y, V^H V \rangle, \quad (2.1)$$

with $\hat{Z}^{\text{BM},m} = (\hat{V}^{\text{BM},m})^H \hat{V}^{\text{BM},m}$. Since MLE and SDP are special cases of the BM factorization, we have $\hat{z}^{\text{MLE}} = (\hat{V}^{\text{BM},1})^H$ and $\hat{Z}^{\text{SDP}} = (\hat{V}^{\text{SDP}})^H \hat{V}^{\text{SDP}}$, where we define $\hat{V}^{\text{SDP}} = \hat{V}^{\text{BM},n}$. As we will show in Section 2.2, $\hat{V}^{\text{BM},m}$ is a fixed point of certain mapping, a critical property on which our analysis is built.

While the difference between the BM factorization and the MLE can be captured by the normalized squared Frobenius norm $n^{-2} \|\hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H\|_F^2$, we opt to quantify it through the deviation between $\hat{V}^{\text{BM},m}$ and \hat{z}^{MLE} . Consider the following loss function $\ell_m : \mathbb{C}^{m \times n} \times \mathbb{C}^n \rightarrow \mathbb{R}$ defined as

$$\ell_m(V, z) = \min_{a \in \mathbb{C}^m : \|a\|=1} \frac{1}{n} \|V - az^H\|_F^2, \quad (2.2)$$

for any $V \in \mathbb{C}^{m \times n}$ and any $z \in \mathbb{C}^n$. Then the deviation can be measured by $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$. The advantage of studying $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$ instead of $n^{-2} \|\hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H\|_F^2$ is twofold. First, the two quantities are closely related through the following inequality (see Lemma SM5.1 in the supplementary material for its proof):

$$\frac{1}{n^2} \left\| \hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H \right\|_F^2 \leq 2\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}). \quad (2.3)$$

Hence, in order to establish Theorem 1.1, it is sufficient to upper bound $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$. Second and more importantly, in our analysis, we view $\hat{V}^{\text{BM},m}$ and \hat{z}^{MLE} as fixed points of certain mappings, a property that it is more natural to exploit with $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$. As a result, in the remaining part of the paper, we will focus on analyzing and upper bounding $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$.

2.2 Introducing F_1 and F_m : \hat{z}^{MLE} and $\hat{V}^{\text{BM},m}$ Are Their Fixed Points

Our analysis relies on the fact that the MLE and the BM factorization are both fixed points. Define a function $F_1 : \mathbb{C}_1^n \rightarrow \mathbb{C}_1^n$ such that for any $z \in \mathbb{C}_1^n$, the j th coordinate of $F_1(z)$ is

$$[F_1(z)]_j = \begin{cases} \frac{\sum_{k \in [n]} Y_{jk} z_k}{|\sum_{k \in [n]} Y_{jk} z_k|}, & \text{if } \sum_{k \in [n]} Y_{jk} z_k \neq 0, \\ z_j, & \text{o.w.,} \end{cases} \quad \forall j \in [n].$$

It can be written equivalently as

$$[F_1(z)]_j = \begin{cases} \frac{[Yz]_j}{|[Yz]_j|}, & \text{if } [Yz]_j \neq 0, \\ z_j, & \text{o.w.,} \end{cases} \quad \forall j \in [n]. \quad (2.4)$$

Then the MLE is a fixed point of the mapping F_1 , that is,

$$\hat{z}^{\text{MLE}} = F_1(\hat{z}^{\text{MLE}}).$$

To see this, recall from the definition in (1.3) that \hat{z}^{MLE} maximizes the objective $\langle Y, zz^H \rangle$ over all $z \in \mathbb{C}_1^n$. Fix any index $j \in [n]$. We can write the objective at \hat{z}^{MLE} as

$$\langle Y, \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^H \rangle = 2\text{Re} \left(\overline{\hat{z}_j^{\text{MLE}}} \sum_{k \in [n]} Y_{jk} \hat{z}_k^{\text{MLE}} \right) + (\text{terms independent of } \hat{z}_j^{\text{MLE}}).$$

Since \hat{z}^{MLE} is the global maximizer, its j th coordinate must maximize the first term of the above expression over \mathbb{C}_1 . This implies

$$\hat{z}_j^{\text{MLE}} = \underset{z_j \in \mathbb{C}_1}{\text{argmax}} \text{Re} \left(\overline{z_j} \sum_{k \in [n]} Y_{jk} \hat{z}_k^{\text{MLE}} \right) = \frac{\sum_{k \in [n]} Y_{jk} \hat{z}_k^{\text{MLE}}}{\left| \sum_{k \in [n]} Y_{jk} \hat{z}_k^{\text{MLE}} \right|} = [F_1(\hat{z}^{\text{MLE}})]_j,$$

as long as $\sum_{k \in [n]} Y_{jk} \hat{z}_k^{\text{MLE}} \neq 0$. Therefore, every coordinate of \hat{z}^{MLE} satisfies the fixed-point condition, so $\hat{z}^{\text{MLE}} = F_1(\hat{z}^{\text{MLE}})$.

In addition, for any $m \in \mathbb{N} \setminus \{1\}$, define a function $F_m : \mathcal{V}_m \rightarrow \mathcal{V}_m$ such that for any $V \in \mathcal{V}_m$, the j th column of $F_m(V)$ is

$$[F_m(V)]_j = \begin{cases} \frac{\sum_{k \in [n]} \bar{Y}_{jk} V_k}{\left\| \sum_{k \in [n]} \bar{Y}_{jk} V_k \right\|}, & \text{if } \sum_{k \in [n]} \bar{Y}_{jk} V_k \neq 0, \\ V_j, & \text{o.w.,} \end{cases} \quad \forall j \in [n].$$

It can be written equivalently as

$$[F_m(V)]_j = \begin{cases} \frac{[VY^H]_j}{\| [VY^H]_j \|}, & \text{if } [VY^H]_j \neq 0, \\ V_j, & \text{o.w.,} \end{cases} \quad \forall j \in [n].$$

From the definition in (2.1), $\hat{V}^{\text{BM},m}$ maximizes the objective $\langle Y, V^H V \rangle$ over all $V \in \mathcal{V}_m$. Following an argument analogous to the one above for \hat{z}^{MLE} , it can be shown that $\hat{V}^{\text{BM},m}$ is fixed point of the mapping F_m , that is,

$$\hat{V}^{\text{BM},m} = F_m(\hat{V}^{\text{BM},m}).$$

The fact that \hat{z}^{MLE} and $\hat{V}^{\text{BM},m}$ are fixed points of F_1 and F_m opens a door for our analysis. In Section 2.3, we will first establish a contraction-type result for the mappings. Consider any $z \in \mathbb{C}_1^n$ and $V \in \mathcal{V}_m$. The key is to understand how $\ell_m(F_m(V), F_1(z))$ depends on $\ell_m(V, z)$. We aim to establish the following contraction-type result:

$$\ell_m(F_m(V), F_1(z)) \leq (\text{some factor smaller than } 1) \times \ell_m(V, z) + \text{some additive error term.} \quad (2.5)$$

On a high level, it means the mappings jointly have a contraction but with some additive error term. If V, z are fixed points such that $F_m(V) = V$ and $F_1(z) = z$, we have $\ell_m(F_m(V), F_1(z)) = \ell_m(V, z)$, and (2.5) consequently becomes

$$\ell_m(V, z) \leq (\text{some factor smaller than } 1) \times \ell_m(V, z) + \text{some additive error term,} \quad (2.6)$$

an inequality involving $\ell_m(V, z)$ on both sides. Then the term $\ell_m(V, z)$ on the right-hand side can be absorbed into the one on the left-hand side, leading to an upper bound for $\ell_m(V, z)$ with an explicit expression. Since $V^{\text{BM},m}$ and z^{MLE} are fixed points, the derived upper bound holds for $\ell_m(V^{\text{BM},m}, z^{\text{MLE}})$. In Section 2.4, we include the result for $\ell_m(V^{\text{BM},m}, z^{\text{MLE}})$ with discussions.

We would like to clarify that the use of \hat{z}^{MLE} and $\hat{V}^{\text{BM},m}$ as fixed points in this paper is purely analytical: we explicitly exploit their fixed-point properties to quantify the theoretical difference between them. One might be tempted to use the mappings F_1 and F_m for computation—for example, by iteratively applying F_1 starting from an initialization until convergence. This procedure is known as the generalized power method and has been studied in prior work [5, 25, 38]. In particular, convergence of the method is established in the low-noise regime where $\sigma \lesssim \sqrt{n/\log n}$ [38]. However, due to the non-linear nature of the mappings F_1 and F_m , the behavior of these iterative algorithms in the high-noise regime remains less understood. In particular, convergence may not be guaranteed when σ is large, and the conditions required for initialization in such regimes are still unclear. Further discussion on computational aspects is provided in Section 4.1.

2.3 A Contraction-type Result for $\ell_m(F_m(V), F_1(z))$ and $\ell_m(V, z)$

In this section, we aim to establish (2.5). To achieve this, we need to study the two mappings F_1 and F_m . Note that they can both be decomposed into two similar steps, as demonstrated below:

$$\begin{array}{ccccc}
 z & \xrightarrow{\quad\quad\quad} & Yz & \xrightarrow{\quad\quad\quad} & F_1(z) \\
 \uparrow \ell_m(V, z) & \text{matrix} & \uparrow \ell_m(VY^H, Yz) & \text{entrywise} & \uparrow \ell_m(F_m(V), F_1(z)) \\
 & \text{multiplication} & & \text{normalization} & \\
 V & \xrightarrow{\quad\quad\quad} & VY^H & \xrightarrow{\quad\quad\quad} & F_m(V)
 \end{array}$$

In the first step, they involve a matrix multiplication with the data matrix Y such that z becomes Yz and V becomes VY^H . In the second step, they perform an entrywise (or column-wise) normalization such that Yz becomes $F_1(z)$ and VY^H becomes $F_m(V)$. Consequently, our analysis is decomposed into two parts.

For the first part (the matrix multiplication part), Lemma 2.1 shows that $\ell_m(VY^H, Yz)$ can be upper bounded by $\ell_m(V, z)$ up to some factor, provided that V and z are close to the ground truth z^* . The closeness of V to z^* can be measured by $\ell_m(V, z^*)$. Regarding z , we define a loss function in an analogous way. Define a loss function $\ell_1 : \mathbb{C}_1^n \times \mathbb{C}_1^n \rightarrow \mathbb{R}$ such that for any $z, z' \in \mathbb{C}^n$,

$$\ell_1(z', z) = \min_{a \in \mathbb{C}_1} \frac{1}{n} \|z' - az\|^2 = 2 - n^{-1} |(z')^H z|. \quad (2.7)$$

Then, the closeness of z to z^* can be measured by $\ell_1(z, z^*)$. Lemma 2.1 is an extension of Lemma 12 of [38], which proves the vector case that connects $\ell_1(Yz', Yz)$ with $\ell_1(z', z)$ for $z, z' \in \mathbb{C}_1^n$ that are close to z^* . We generalize it to the matrix case.

Lemma 2.1. *Suppose $m \in \mathbb{N} \setminus \{1\}$ and $\epsilon \in (0, 1/2)$. For any $z \in \mathbb{C}_1^n$ such that $\ell_1(z, z^*) \leq \epsilon^2$ and any $V \in \mathcal{V}_m$ such that $\ell_m(V, z^*) \leq \epsilon^2$, we have*

$$\ell_m(VY^H, Yz) \leq n^2 \left(6\epsilon + \frac{\sigma \|W\|}{n} \right)^2 \ell_m(V, z).$$

For the second part (the normalization part), we need to study how the normalization affects the loss function ℓ_m to connect $\ell_m(VY^H, Yz)$ with $\ell_m(F_m(V), F_1(z))$. However, this is not straightforward as the normalization operation is not continuous and, more importantly, does not necessarily have a contraction property. To see this, consider any $t > 0$ and any $x, y \in \mathbb{C}$ such that $|x| = |y| = t$. Then $\left| \frac{x}{|x|} - \frac{y}{|y|} \right| = t^{-1} |x - y|$. If $t > 1$, then x and y get closer after normalization. If $t < 1$ and is close to 0, the distance between $\frac{x}{|x|}$ and $\frac{y}{|y|}$ can be much larger than that between x and y , though

the distance is capped at most 2. The following Lemma 2.2 shows how normalization changes the distance between vectors. In the lemma, (2.8) is a simple case where two vectors x, y are non-zero; (2.9) allows x, y to be zero, which can be used to analyze the normalizations in F_1 and F_m .

Lemma 2.2. *Suppose $m \in \mathbb{N}$. For any vectors $x, y \in \mathbb{C}^m \setminus \{0\}$ and for any $t > 0$, we have*

$$\left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\| \leq \frac{2\|x - y\|}{t} + 2\mathbb{I}\{\|y\| < t\}. \quad (2.8)$$

For any vectors $x, y, u, v \in \mathbb{C}^m$ such that $\|u\|, \|v\| \leq 1$ and for any $t > 0$, we have

$$\left\| \left(\frac{x}{\|x\|} \mathbb{I}\{x \neq 0\} + u \mathbb{I}\{x = 0\} \right) - \left(\frac{y}{\|y\|} \mathbb{I}\{y \neq 0\} + v \mathbb{I}\{y = 0\} \right) \right\| \leq \frac{2\|x - y\|}{t} + 2\mathbb{I}\{\|y\| < t\}. \quad (2.9)$$

Lemma 2.2 (more specifically, (2.9)) can be applied to analyze the difference between each column-wise normalization of VY^H and each coordinate-wise normalization of Yz for any threshold $t > 0$. To be more specific, for the j th normalization of VY^H and Yz , the application of (2.9) leads to two terms, corresponding to the two terms in the upper bound of (2.9): the first term is essentially about the distance between $[VY^H]_j$ and $[Yz]_j$, and the second term is about $\mathbb{I}\{|[Yz]_j| < t\}$. Aggregated over all $j \in [n]$, the first term can be related to $\ell_m(VY^H, Yz)$, which can be further bounded by Lemma 2.1; the second term becomes $\sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < t\}$. This leads to the following theorem, which gives a connection between $\ell_m(F_m(V), F_1(z))$ and $\ell_m(V, z)$.

Theorem 2.3. *Suppose $m \in \mathbb{N} \setminus \{1\}$ and $\epsilon \in (0, 1/2)$. For any $z \in \mathbb{C}_1^n$ such that $\ell_1(z, z^*) \leq \epsilon^2$ and any $V \in \mathcal{V}_m$ such that $\ell_m(V, z^*) \leq \epsilon^2$, we have*

$$\ell_m(F_m(V), F_1(z)) \leq \frac{4n^2}{t^2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right)^2 \ell_m(V, z) + \frac{4}{n} \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < t\}, \quad \forall t > 0. \quad (2.10)$$

If z, V are further assumed to satisfy $z = F_1(z)$ and $V = F_m(V)$, we have

$$\ell_m(V, z) \leq \frac{8}{n} \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < \delta n\}, \quad \forall \delta \geq 2\sqrt{2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right). \quad (2.11)$$

In Theorem 2.3, (2.10) holds for any threshold $t > 0$ and any $z \in \mathbb{C}_1^n$ and $V \in \mathcal{V}_m$ that are close to z^* . With a sufficiently large choice of t , the factor $\frac{4n^2}{t^2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right)^2$ is smaller than 1, leading to the establishment of (2.5). Under this scenario, F_m and F_1 jointly have a contraction-type property: after one iteration, $F_m(V)$ and $F_1(z)$ get closer, up to an additive error $\frac{4}{n} \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < t\}$, compared to V and z , with respect to the loss ℓ_m .

In Theorem 2.3, (2.11) is an immediate consequence of (2.10) if z and V are further assumed to be fixed points, following the argument as in (2.6). (2.11) shows that the distance between V , a fixed point of F_m , and z , a fixed point of F_1 , provided that they are close to z^* , can be upper bounded by $\frac{8}{n} \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < \delta n\}$, a property of z . This property essentially concerns the number of coordinates in Yz whose absolute values are smaller than a certain threshold. If there is no such coordinate, i.e., $\sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < \delta n\} = 0$, then (2.11) leads to $\ell_m(V, z) = 0$.

2.4 Implications on $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$

Since $\hat{V}^{\text{BM},m}$ and \hat{z}^{MLE} are fixed points of F_m and F_1 , respectively, a direct consequence of (2.11) in Theorem 2.3 is the following corollary for $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$.

Corollary 2.4. *Suppose $m \in \mathbb{N} \setminus \{1\}$ and $\epsilon \in (0, 1/2)$. If $\ell_1(\hat{z}^{\text{MLE}}, z^*) \leq \epsilon^2$ and $\ell_m(\hat{V}^{\text{BM},m}, z^*) \leq \epsilon^2$ are satisfied, we have*

$$\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \leq \frac{8}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y \hat{z}^{\text{MLE}}]_j| < \delta n \}, \quad \forall \delta \geq 2\sqrt{2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right).$$

Corollary 2.4 reveals that the distance between $\hat{V}^{\text{BM},m}$ and \hat{z}^{MLE} can be upper bounded by a quantity of the MLE: $\frac{8}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y \hat{z}^{\text{MLE}}]_j| < \delta n \}$, which is essentially the proportion of coordinates in the vector $Y \hat{z}^{\text{MLE}}$ whose absolute value is smaller than δn . Below are some remarks about this quantity and Corollary 2.4.

1) *Connection between the tightness of the BM factorization and the ℓ_∞ norm analysis.* It turns out an ℓ_∞ norm analysis for $W \hat{z}^{\text{MLE}}$ is sufficient to show $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) = 0$, i.e., the tightness of the BM factorization. To see this, recall the decomposition of Y in (1.2). Then for each $j \in [n]$,

$$[Y \hat{z}^{\text{MLE}}]_j = [(z^*(z^*)^H + \sigma W) \hat{z}^{\text{MLE}}]_j = z_j^* ((z^*)^H \hat{z}^{\text{MLE}}) + \sigma [W \hat{z}^{\text{MLE}}]_j.$$

Since $|(z^*)^H \hat{z}^{\text{MLE}}| = n(1 - \ell_1(\hat{z}^{\text{MLE}}, z^*)/2) \geq n(1 - \epsilon^2/2)$ according to (2.7), we have

$$|[Y \hat{z}^{\text{MLE}}]_j| \geq n(1 - \epsilon^2/2) - \sigma |[W \hat{z}^{\text{MLE}}]_j| \geq n(1 - \epsilon^2/2) - \sigma \|W \hat{z}^{\text{MLE}}\|_\infty. \quad (2.12)$$

Hence, if $\sigma \|W \hat{z}^{\text{MLE}}\|_\infty \leq n(1 - \delta - \epsilon^2/2)$ holds, then $\sum_{j \in [n]} \mathbb{I} \{ |[Y \hat{z}^{\text{MLE}}]_j| < \delta n \} = 0$, and consequently $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) = 0$. As a result, it is sufficient to carry out an ℓ_∞ norm analysis for the MLE to establish $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) = 0$, which is achievable when σ is small enough, as shown in [38] for the SDP.

2) *Corollary 2.4 is beyond the tightness of the BM factorization and consequently requires analysis beyond the existing ℓ_∞ norm framework.* Corollary 2.4 is not just about establishing $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) = 0$, the tightness of the BM factorization. In fact, it quantifies the deviation by $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$ and upper bound it by $\frac{8}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y \hat{z}^{\text{MLE}}]_j| < \delta n \}$, the analysis of which is actually beyond the ℓ_∞ norm of $W \hat{z}^{\text{MLE}}$. To see this, from (2.12), we have that for each $j \in [n]$,

$$\begin{aligned} \mathbb{I} \{ |[Y \hat{z}^{\text{MLE}}]_j| < \delta n \} &\leq \mathbb{I} \{ n(1 - \epsilon^2/2) - \sigma |[W \hat{z}^{\text{MLE}}]_j| < \delta n \} \\ &\leq \mathbb{I} \{ \sigma |[W \hat{z}^{\text{MLE}}]_j| > n(1 - \delta - \epsilon^2/2) \}. \end{aligned}$$

Hence, the upper bound in Corollary 2.4 can be further upper bounded by

$$\frac{8}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y \hat{z}^{\text{MLE}}]_j| < \delta n \} \leq \frac{8}{n} \sum_{j \in [n]} \mathbb{I} \{ \sigma |[W \hat{z}^{\text{MLE}}]_j| > n(1 - \delta - \epsilon^2/2) \}, \quad (2.13)$$

which is essentially about the proportion of coordinates in $W \hat{z}^{\text{MLE}}$ that is larger than certain threshold in absolute value. To bound it, we need to study entrywise behavior of $W \hat{z}^{\text{MLE}}$ instead of its ℓ_∞ norm.

3) *Intuition on the exponential bound in Theorem 1.1 for $\frac{8}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y \hat{z}^{\text{MLE}}]_j| < \delta n \}$.* From Corollary 2.4, it is clear that in order to establish our main result Theorem 1.1, it is sufficient to analyze the MLE \hat{z}^{MLE} to provide an upper bound for the quantity $\frac{8}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y \hat{z}^{\text{MLE}}]_j| < \delta n \}$.

However, directly establishing the exponential bound for it is not easy, due to the dependence between \hat{z}^{MLE} and Y (through the noise matrix W). Nevertheless, here we assume they are independent from each other to provide some intuition why it has exponential upper bound. Recall (2.13) holds. Then each coordinate of $W\hat{z}^{\text{MLE}}$ follows a Gaussian distribution as W is a Gaussian matrix. Though they are not completely independent from each other, we can show $\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \{ |\sigma[W\hat{z}^{\text{MLE}}]_j| > n(1 - \delta - \epsilon^2/2) \}$ concentrates on its expected values, $\mathbb{P}(\sigma|[W\hat{z}^{\text{MLE}}]_1| > n(1 - \delta - \epsilon^2/2))$ which is a Gaussian tail probability. With small δ, ϵ , it can be bounded explicitly by $\exp(-\frac{cn}{\sigma^2})$ for some constant $c > 0$. This provides an intuition to explain why the bound in Theorem 1.1 takes this form. On the other hand, this is purely just an intuition as \hat{z}^{MLE} and Y are actually highly dependent on each other.

To conclude this section, note that Corollary 2.4 requires \hat{z}^{MLE} and $\hat{V}^{\text{BM},m}$ to be close to the ground truth z^* . The following lemma shows that $\ell_m(\hat{V}^{\text{BM},m}, z^*)$ and $\ell_1(\hat{z}^{\text{MLE}}, z^*)$ are upper bounded by $\frac{8\sigma\|W\|}{n}$. Hence, when σ is small such that $\frac{8\sigma\|W\|}{n} \leq \epsilon^2$, the assumptions needed in Corollary 2.4 are satisfied, and then the conclusion established therein holds.

Lemma 2.5. *For any $m \in \mathbb{N} \setminus \{1\}$, we have $\ell_m(\hat{V}^{\text{BM},m}, z^*) \leq \frac{8\sigma\|W\|}{n}$. In addition, the same upper bound holds for $\ell_1(\hat{z}^{\text{MLE}}, z^*)$.*

3 Analysis on MLE

From Corollary 2.4, it is evident that to establish our main result, Theorem 1.1, it suffices to analyze the MLE \hat{z}^{MLE} and provide an upper bound for the quantity $\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y\hat{z}^{\text{MLE}}]_j| < \delta n \}$. As demonstrated in Section 2.4, if \hat{z}^{MLE} is independent of Y , an exponential upper bound can be established immediately. However, the dependence between them complicates the problem, requiring a delicate analysis. In Section 3.1, we provide an overview of our analysis, which can be decomposed into two steps detailed in Sections 3.2 and 3.3. The main results are given in Section 3.4.

3.1 High-level Idea of Our Analysis

In this section, we present the high-level idea of our analysis, which is quite technical and involved. It consists of the following two steps.

Step 1: Approximate \hat{z}^{MLE} by fixed points of Lipschitz mappings. Recall that \hat{z}^{MLE} is a fixed point of F_1 . As discussed in Section 2, the normalization in F_1 complicates the analysis. To address this, note that for any $x \in \mathbb{C}$, the normalization operation $x \rightarrow x/|x|$ can be approximated by a function $x \rightarrow x/\max\{|x|, t\}$ for some tuning parameter $t > 0$ (denoted as $g_t(\cdot)$ in (3.8)).

The advantage of using this approximated mapping is two-fold. First, the approximation error can be controlled, as $|x/|x| - x/\max\{|x|, t\}| \leq \mathbb{I}\{|x| < t\}$. Second and more importantly, the mapping is Lipschitz (see Lemma 3.1). With the help of this mapping, we define another Lipschitz mapping $G(\cdot, \cdot, \cdot)$ (see (3.9) for its definition) whose fixed points are used to approximate \hat{z}^{MLE} . Specifically, in Lemma 3.6, we show that for a suitable δ , we have

$$\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y\hat{z}^{\text{MLE}}]_j| < \delta n \} \leq \frac{9}{n} \sum_{j \in [n]} \mathbb{I} \{ |z_j^*|(z^*)^H \hat{z}^{\text{MLE}}| + \sigma[Wz]_j| < 2\delta n \},$$

where z is any fixed point of $G(\cdot, |(z^*)^H \hat{z}^{\text{MLE}}|, 2\delta n)$. In this way, we reduce the problem to a fixed point analysis for $G(\cdot, (z^*)^H \hat{z}^{\text{MLE}}, 2\delta n)$.

However, note that in $G(\cdot, |(z^*)^H \hat{z}^{\text{MLE}}|, 2\delta n)$, the quantity $|(z^*)^H \hat{z}^{\text{MLE}}|$ is still random as it involves \hat{z}^{MLE} . To completely remove \hat{z}^{MLE} from the above expression, we approximate $|(z^*)^H \hat{z}^{\text{MLE}}|$ by a grid of scalars. In this way, the problem becomes: given some s, t , how to upper bound $\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \left| z_j^* s + \sigma [Wz]_j \right| < \text{some threshold} \right\}$, or more conveniently

$$\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \sigma \left| [Wz]_j \right| > \text{some threshold} \right\}, \quad (3.1)$$

where z is a fixed point of $G(\cdot, s, t)$.

Step 2: Leave-one-out analysis for fixed points of $G(\cdot, s, t)$. Upper bounding (3.1) is still challenging because the mapping $G(\cdot, s, t)$ involves the noise matrix W for any given s, t . Consequently, z , a fixed point of $G(\cdot, s, t)$, also depends on W . Note that for each $j \in [n]$, $[Wz]_j = W_j \cdot z$. The key to decoupling this dependence is to approximate z by some quantity $z^{(-j)}$ that is close but independent of W_j , such that $W_j \cdot z \approx W_j \cdot z^{(-j)}$, which follows a Gaussian distribution. As a result, (3.1) becomes

$$\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \sigma \left| W_j \cdot z^{(-j)} \right| \geq \text{some threshold} \right\}, \quad (3.2)$$

which can be analyzed and leads to the desired exponential bound.

Now, the problem becomes finding the desired $z^{(-j)}$ for z . To achieve this, we use the idea of leave-one-out. Let $W^{(-j)}$ be a matrix equal to W but with its j th row and column zeroed out. Let $G^{(-j)}$ be a function equal to G but using $W^{(-j)}$ instead of W , and let $z^{(-j)}$ be its fixed point. By definition, $z^{(-j)}$ is independent of W_j . On the other hand, since W and $W^{(-j)}$ only differ by one column and one row, intuitively, the two functions G and $G^{(-j)}$ do not differ much, and their fixed points are consequently close.

To establish the closeness of z and $z^{(-j)}$ rigorously, we use the following method. Let $z^{(0)} = z^*$. Starting from it, we apply $G(\cdot, s, t)$ iteratively to obtain a sequence of vectors $z^{(0)}, z^{(1)}, z^{(2)}, \dots$. We can show that this sequence converges to a fixed point. So we let $z = z^{(\infty)}$. Instead of $G(\cdot, s, t)$, we can apply $G^{(-j)}(\cdot, s, t)$ iteratively, which leads to another sequence $z^{(0, -j)} = z^*, z^{(1, -j)}, z^{(2, -j)}, \dots, z^{(-j)} = z^{(\infty, -j)}$. It is evident that $\|z^{(0, -j)} - z^{(0)}\| = 0$. Using mathematical induction, we can show (see Lemma 3.10) that

$$\left\| z^{(T)} - z^{(T, -j)} \right\| \leq 3, \forall T \in \mathbb{N}. \quad (3.3)$$

Hence, the same result holds for the limit, which is $\|z - z^{(-j)}\|$.

The details of these two steps are included in Sections 3.2 and 3.3. Together, they lead to the establishment of exponential bounds in Section 3.4, which is the main result of this paper.

We conclude this section by discussing techniques used in [38] for the small σ regime, explaining why they fail in the high σ regime, and how our approach connects with and diverges from theirs. In [38], the main technical difficulty lies in controlling $\|W \hat{z}^{\text{MLE}}\|_\infty$, which is challenging due to the dependence between W and \hat{z}^{MLE} . To decouple this dependence, they construct a sequence of vectors $x^{(0)}, x^{(1)}, x^{(2)}, \dots$, where $x^{(0)}$ is the leading eigenvector of the data matrix and the subsequent vectors are obtained by iteratively applying F_1 . Additionally, they construct a leave-one-out counterpart that is independent of W_j : $x^{(0, -j)}, x^{(1, -j)}, x^{(2, -j)}, \dots$. When σ is small, they show the

sequences satisfy:

$$\ell_1(x^{(T)}, x^{(T,-j)}) \lesssim \frac{1}{n}, \quad \|Wx^{(T)}\|_\infty \lesssim \sqrt{n \log n}, \quad \ell_1(x^{(T)}, z^*) \lesssim 1, \quad (3.4)$$

for all $T \geq 0$, and the limit of $x^{(T)}$ is the MLE \hat{z}^{MLE} . As a result, $\|W\hat{z}^{\text{MLE}}\|_\infty \lesssim \sqrt{n \log n}$ holds, which leads to the tightness of the SDP in the small σ regime.

The key in [38]'s analysis is the ℓ_∞ norm result (3.4). Note that $x^{(T+1)} = F_1(x^{(T)})$ involves a column-wise normalization of $Yx^{(T)}$. As discussed in Section 2, this normalization is difficult to analyze. However, once (3.4) holds, under the assumption that σ is small, one can show the norm of each column $[Yx^{(T)}]_j$ is of the order n . Hence, the normalization $[Yx^{(T)}]_j / |[Yx^{(T)}]_j|$ is $[Yx^{(T)}]_j$ multiplied by a factor close to 1. This makes it easy to connect the error of $x^{(T+1)}$ with $x^{(T)}$. However, this argument no longer works when σ is large, as the norms of columns $Yx^{(T)}$ are no longer guaranteed to be of the order n . For a column $[Yx^{(T)}]_j$ with a norm close to 0, its normalization $[Yx^{(T)}]_j / |[Yx^{(T)}]_j|$ differs dramatically from itself. This makes connecting the error of $x^{(T+1)}$ with that of $x^{(T)}$ difficult, leading to the failure of their analysis.

Compared to [38]'s analysis, our key novelty is in Step 1, where we bypass the normalization step of F_1 by replacing F_1 with a smoother mapping G . In Step 2, we follow the same idea of constructing leave-one-out sequences as in [38]. However, due to the use of G instead of F_1 , our analysis is much simpler, as we only need to establish (3.3) instead of the ℓ_∞ bound (3.4). Additionally, our sequences start from the ground truth z^* , instead of the eigenvector of the data matrix as in (3.4). This allows us to avoid the analysis of the eigenvector needed in [38]. It is also worth mentioning that [38] relates the SDP with the MLE through a dual certificate. We avoid this by simply using the fact that they are fixed points, which enables us to study not only the SDP but also the BM factorization more generally.

3.2 Step 1: Approximate \hat{z}^{MLE} by Fixed Points of Lipschitz Mappings

As outlined in Section 3.1, in this step, we are going to approximate \hat{z}^{MLE} by fixed points of a Lipschitz mapping G in order to upper bound $\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j| < \delta n\}$ by (3.1). Before introducing g_t and G , we first introduce an auxiliary mapping F'_1 that is closely related to F_1 such that \hat{z}^{MLE} is its fixed point.

Recall the definition of the function F_1 in (2.4). Note that for any $z \in \mathbb{C}_1^n$, $Yz = (z^*(z^*)^H + \sigma W)z$ and consequently $[Yz]_j = z_j^*(z^*)^H z + \sigma [Wz]_j$ for any $j \in [n]$. Then (2.4) can be written equivalently as

$$[F_1(z)]_j = \begin{cases} \frac{z_j^*(z^*)^H z + \sigma [Wz]_j}{|z_j^*(z^*)^H z + \sigma [Wz]_j|}, & \text{if } z_j^*(z^*)^H z + \sigma [Wz]_j \neq 0, \\ z_j, & \text{o.w.,} \end{cases} \quad \forall j \in [n].$$

Define a function $F'_1 : \mathbb{C}_1^n \times \mathbb{C} \rightarrow \mathbb{C}_1^n$ such that for any $z \in \mathbb{C}_1^n$ and $s \in \mathbb{C}$, the j th coordinate of $F'_1(z)$ is

$$[F'_1(z, s)]_j = \begin{cases} \frac{z_j^* s + \sigma [Wz]_j}{|z_j^* s + \sigma [Wz]_j|}, & \text{if } z_j^* s + \sigma [Wz]_j \neq 0, \\ z_j, & \text{o.w.,} \end{cases} \quad \forall j \in [n]. \quad (3.5)$$

Since $\hat{z}^{\text{MLE}} = F_1(\hat{z}^{\text{MLE}})$, it is easy to verify that \hat{z}^{MLE} is a fixed point of $F'(\cdot, (z^*)^H \hat{z}^{\text{MLE}})$, i.e., $\hat{z}^{\text{MLE}} = F'_1(\hat{z}^{\text{MLE}}, (z^*)^H \hat{z}^{\text{MLE}})$. For simplicity, denote

$$\hat{s} = (z^*)^H \hat{z}^{\text{MLE}}. \quad (3.6)$$

Then z^{MLE} satisfies

$$\hat{z}^{\text{MLE}} = F'_1(\hat{z}^{\text{MLE}}, \hat{s}). \quad (3.7)$$

The difference between these two mappings $F_1(\cdot)$ and $F_1(\cdot, \hat{s})$ is that: in $F_1(z)$, z appears twice in its numerator $z_j^*(z^*)^H z + \sigma[Wz]_j$; on the contrary, in $F'_1(z, \hat{s})$, z only appears once in its numerator $z_j^* \hat{s} + \sigma[Wz]_j$, despite that \hat{s} depends on z . Hence, the mapping $F'_1(\cdot, \hat{s})$ is less complicated than $F_1(\cdot)$ and is relatively easier to analyze.

Now we are ready to introduce g_t , an approximation of the normalization mapping $x \rightarrow x/|x|$, as we outline in Section 3.1. For any $t \in \mathbb{R}$ such that $t > 0$, define a function $g_t : \mathbb{C} \rightarrow \mathbb{C}$ as $g_t(x) = \frac{x}{\max\{|x|, t\}}$ for any $x \in \mathbb{C}$. That is, for any $x \in \mathbb{C}$,

$$g_t(x) = \begin{cases} \frac{x}{|x|}, & \text{if } |x| > t, \\ \frac{x}{t}, & \text{o.w..} \end{cases} \quad (3.8)$$

The following lemma shows g_t is Lipschitz.

Lemma 3.1. *For any $x, y \in \mathbb{C}$, we have $|g_t(x) - g_t(y)| \leq \frac{|x-y|}{t}, \forall t > 0$.*

With g_t , define $G : \mathbb{C}_{\leq 1}^n \times \mathbb{C} \times \mathbb{R} \rightarrow \mathbb{C}_{\leq 1}^n$ such that for any $z \in \mathbb{C}_{\leq 1}^n, s \in \mathbb{C}, t > 0$, the j th coordinate of $G(z, s, t)$ is

$$[G(z, s, t)]_j = g_t(z_j^* s + \sigma[Wz]_j) = g_t([z^* s + \sigma Wz]_j), \forall j \in [n]. \quad (3.9)$$

The following Lemma 3.2 gives a list of properties G has. First, since g_t is Lipschitz, $G(\cdot, s, t)$ is also Lipschitz. With a suitable choice of t , it is a contraction mapping, and consequently has a unique fixed point which can be achieved by iteratively applying the function starting from z^* . The last property shows that the sensitivity of the fixed point with respect to s is well-controlled.

Lemma 3.2. *The function $G(\cdot, \cdot, \cdot)$ has the following properties:*

1. *For any $x, y \in \mathbb{C}^n$ and for any $s \in \mathbb{C}, t > 0$, we have*

$$\|G(x, s, t) - G(y, s, t)\| \leq t^{-1} \sigma \|W\| \|x - y\|.$$

2. *For any $s \in \mathbb{C}, t \geq 2\sigma \|W\|$, and for any $z^{(0)} \in \mathbb{C}_{\leq 1}^n$, define $z^{(T)} = G(z^{(T-1)}, s, t)$ for all $T \in \mathbb{N}$. Then*

$$\|z^{(T+1)} - z^{(T)}\| \leq \frac{1}{2} \|z^{(T)} - z^{(T-1)}\|, \forall T \in \mathbb{N}.$$

3. *For any $s \in \mathbb{C}, t \geq 2\sigma \|W\|$, $G(\cdot, s, t)$ has exactly one fixed point. That is, there exists one and only one $z \in \mathbb{C}_{\leq 1}^n$ such that $z = G(z, s, t)$. In addition, z can be achieved by iteratively applying $G(\cdot, s, t)$ starting from z^* . That is, let $z^{(0)} = z^*$ and define $z^{(T)} = G(z^{(T-1)}, s, t)$ for all $T \in \mathbb{N}$. We have $z = \lim_{T \rightarrow \infty} G(z^{(T)}, s, t)$.*

4. *For any $s \in \mathbb{C}, s' \in \mathbb{C}, t \geq 2\sigma \|W\|$, let z be the fixed point of $G(\cdot, s, t)$ and let z' be the fixed point of $G(\cdot, s', t)$. We have $\|z - z'\|^2 \leq 4nt^{-2} |s - s'|^2$ and*

$$\|(z^* s + \sigma Wz) - (z^* s' + \sigma Wz')\|^2 \leq 4n |s - s'|^2.$$

In addition, since g_t approximates the normalization mapping $x \rightarrow x/|x|$, $G(\cdot, s, t)$ can be seen as a Lipschitz function that approximates $F'_1(\cdot, s)$. In fact, the approximation error can be bounded by the following lemma.

Lemma 3.3. *For any $z \in \mathbb{C}_{\leq 1}^n$, $s \in \mathbb{C}$, and $t > 0$, we have*

$$\|F'_1(z, s) - G(z, s, t)\|^2 \leq 4 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* s + \sigma[Wz]_j| < t \}^2.$$

With Lemmas 3.2 and 3.3, the following lemma shows that \hat{z}^{MLE} can be approximated by the fixed point of $G(\cdot, \hat{s}, t)$.

Lemma 3.4. *For any $t \geq 4\sigma \|W\|$, with $z \in \mathbb{C}_{\leq 1}^n$ being the fixed point of $G(\cdot, \hat{s}, t)$, we have*

$$\|\hat{z}^{\text{MLE}} - z\|^2 \leq 32 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \}.$$

With Lemma 3.4, the quantity $\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y\hat{z}^{\text{MLE}}]_j| < \delta n \}$ can be upper bounded by a similar quantity associated with the fixed point of G .

Lemma 3.5. *For any $\delta \geq \frac{2\sigma\|W\|}{n}$, with $z \in \mathbb{C}_{\leq 1}^n$ being the fixed point of $G(\cdot, \hat{s}, 2\delta n)$, we have*

$$\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y\hat{z}^{\text{MLE}}]_j| < \delta n \} \leq \frac{9}{n} \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < 2\delta n \}.$$

Note that in Lemma 3.5, the function $G(\cdot, \hat{s}, 2\delta n)$ depends on \hat{s} and consequently depend on \hat{z}^{MLE} . We want to further decouple the dependence so that we only need to study $G(\cdot, s, t)$ for some fixed s . To achieve this, in Lemma 3.6, we first show that $G(\cdot, \hat{s}, 2\delta n)$ can be replaced by $G(\cdot, |\hat{s}|, 2\delta n)$. That is, the phase information in $G(\cdot, \hat{s}, 2\delta n)$ is not important. What matters is its magnitude.

Lemma 3.6. *For any $\delta \geq \frac{2\sigma\|W\|}{n}$, with $z \in \mathbb{C}_{\leq 1}^n$ being the fixed point of $G(\cdot, |\hat{s}|, 2\delta n)$, we have*

$$\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y\hat{z}^{\text{MLE}}]_j| < \delta n \} \leq \frac{9}{n} \sum_{j \in [n]} \mathbb{I} \{ |z_j^* |\hat{s}| + \sigma[Wz]_j| < 2\delta n \}.$$

Note that $|\hat{s}|$ is a real number. Once we have Lemma 3.6, we then approximate $|\hat{s}|$ by points on a grid $\{s_0, s_1, s_2, \dots\} \subset \mathbb{R}$. Consequently, the fixed point of $G(\cdot, |\hat{s}|, 2\delta n)$ can be approximated by those of $G(\cdot, s_k, 2\delta n)$ where $k = 0, 1, 2, \dots$, leading to the following Proposition 3.7.

Proposition 3.7. *Suppose $\epsilon \in (0, 1/2)$, $h > 0$, and $\delta \geq \frac{2\sigma\|W\|}{n}$. Assume \hat{z}^{MLE} satisfies $\ell_1(\hat{z}^{\text{MLE}}, z^*) \leq \epsilon^2$. For each $k = 0, 1, 2, \dots, \lceil n\epsilon/h \rceil$, define $s_k = n - kh \in \mathbb{R}$ and let $z_{s_k} \in \mathbb{C}_{\leq 1}^n$ be the fixed point of $G(\cdot, s_k, 2\delta n)$. Then*

$$\begin{aligned} & \frac{1}{n} \sum_{j \in [n]} \mathbb{I} \{ |[Y\hat{z}^{\text{MLE}}]_j| < \delta n \} \\ & \leq 9 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \{ |z_j^* s_k + \sigma[Wz_{s_k}]_j| < 4\delta n \} \right) + \frac{9h^2}{\delta^2 n^2} \mathbb{I} \{ h > \delta\sqrt{n} \}. \end{aligned} \quad (3.10)$$

Compared Lemma 3.6, Proposition 3.7 avoids the appearance of \hat{s} by using a grid $\{s_0, s_1, \dots, s_{\lceil n\epsilon/h \rceil}\}$. Note that we can show $\hat{s} \in [(1-\epsilon)n, n]$ under the assumption $\ell_1(\hat{z}^{\text{MLE}}, z^*) \leq \epsilon^2$. Hence, we only need to discretize the interval $[(1-\epsilon)n, n]$. In Proposition 3.7, h is the distance among the points in the grid, a parameter can be optimized later. The use of the grid instead of \hat{s} comes with costs, reflected in the two term in (3.10). Let $\hat{k} \in \{0, 1, \dots, \lceil n\epsilon/h \rceil\}$ be the index such that $s_{\hat{k}}$ is the one closest to \hat{s} in the grid. First, note that $s_{\hat{k}}$ is still random as it depends on \hat{s} . To deal with, we upper bound the error associated with $s_{\hat{k}}$, $\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \left| z_j^* s_{\hat{k}} + \sigma [W z_{s_{\hat{k}}}]_j \right| < 4\delta n \right\}$, by a summation of errors $\sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \left| z_j^* s_k + \sigma [W z_{s_k}]_j \right| < 4\delta n \right\} \right)$, as each one is non-negative. Second, the approximation error $|\hat{s} - s_{\hat{k}}| \leq h$ results in the second term in (3.10). Nevertheless, the cost of having the summation of all indexes and the approximation error turns out to be negligible with a suitable choice of h .

The following Corollary 3.8 simplifies the first term in (3.10) in order to make the analysis in Section 3.3 easier. For each $j \in [n]$, note that $\left| z_j^* s_k + \sigma [W z_{s_k}]_j \right| \geq \left| z_j^* s_k \right| - \left| \sigma [W z_{s_k}]_j \right| = s_k - \sigma \left| [W z_{s_k}]_j \right|$. Consequently, we have $\mathbb{I} \left\{ \left| z_j^* s_k + \sigma [W z_{s_k}]_j \right| < 4\delta n \right\} \leq \mathbb{I} \left\{ s_k - \sigma \left| [W z_{s_k}]_j \right| < 4\delta n \right\} = \mathbb{I} \left\{ \sigma \left| [W z_{s_k}]_j \right| > s_k - 4\delta n \right\}$, leading to the corollary.

Corollary 3.8. *Under the same conditions as in Proposition 3.7, we have*

$$\begin{aligned} & \frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \left| [Y \hat{z}^{\text{MLE}}]_j \right| < \delta n \right\} \\ & \leq 9 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \sigma \left| [W z_{s_k}]_j \right| > s_k - 4\delta n \right\} \right) + \frac{9h^2}{\delta^2 n^2} \mathbb{I} \{ h > \delta \sqrt{n} \}. \end{aligned}$$

With Corollary 3.8, we boil down the problem of upper bounding $\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \left| [Y \hat{z}^{\text{MLE}}]_j \right| < \delta n \right\}$ into a problem of analyzing the fixed point of $G(\cdot, s, t)$ for given s, t . Specifically, let z be the fixed point, we want to analyze (3.1), which is the focus of the next section.

3.3 Step 2: Leave-One-Out Analysis for Fixed Points of $G(\cdot, s, t)$

In the step, we are going to study the fixed point of $G(\cdot, s, t)$ to provide an upper bound for (3.1). As we outlined in Section 3.1, the key is to decouple the dependence between W and z in the quantity Wz using the leave-one-out technique.

For any $s \in \mathbb{C}, t > 0$, define $z^{(0)} = z^*$ and

$$z^{(T)} = G(z^{(T-1)}, s, t), \forall T \in \mathbb{N}. \quad (3.11)$$

For any $j \in [n]$, define $W^{(-j)} \in \mathbb{C}^{n \times n}$ such that

$$W_{k,l}^{(-j)} = \begin{cases} W_{k,l}, & \forall k \neq j \text{ and } l \neq j, \\ 0, & \text{o.w..} \end{cases}$$

We refer $W^{(-j)}$ as a leave-one-out counterpart of W , as compared to W , it zeros out its j th column and row. As a result, it and its functions are independent of W_j . Define $G^{(-j)} : \mathbb{C}_{\leq 1}^n \times \mathbb{C} \times \mathbb{R} \rightarrow \mathbb{C}_{\leq 1}^n$ such that for any $z \in \mathbb{C}_{\leq 1}^n$, the j th coordinate of $G^{(-j)}(z, s, t)$ is

$$[G^{(-j)}(z, s, t)]_j = g_t([z^* s + \sigma W^{(-j)} z]_j), \forall j \in [n].$$

Define $z^{(0,-j)} = z^*$ and

$$z^{(T,-j)} = G^{(-j)}(z^{(T-1,-j)}, s, t), \forall T \in \mathbb{N}. \quad (3.12)$$

That is, $G^{(-j)}(z, s, t)$ is a counterpart of $G(z, s, t)$ that uses $W^{(-j)}$ instead of W . Consequently, the sequence $\{z^{(T,-j)}\}_{T \geq 0}$ is independent of W_j .

Note that the existence and uniqueness of the limit $z^{(\infty)}$ for the sequence $\{z^{(T)}\}_{T \geq 0}$ is guaranteed as long as $t \geq 2\sigma \|W\|$, according to the properties of G in Lemma 3.2. Similar properties hold for $G^{(-j)}$ (see Lemma SM3.1 in the supplementary material), with which we can also show the existence and uniqueness of the limit for the sequence $\{z^{(T,-j)}\}_{T \geq 0}$ when $t \geq 2\sigma \|W\|$. These lead to the following lemma about the limits and fixed points.

Lemma 3.9. *For any $s \in \mathbb{C}$, $t \geq 2\sigma \|W\|$, and $j \in [n]$, let $z \in \mathbb{C}_{\leq 1}^n$ be the fixed point of $G(\cdot, s, t)$ and let $z^{(-j)} \in \mathbb{C}_{\leq 1}^n$ be the fixed point of $G^{(-j)}(\cdot, s, t)$. Then $z = \lim_{T \rightarrow \infty} z^{(T)}$, and $z^{(-j)} = \lim_{T \rightarrow \infty} z^{(T,-j)}$.*

Note that $\|z^{(0)} - z^{(0,-j)}\| = 0$. With mathematical induction, the following lemma shows $z^{(T)}$ and $z^{(T,-j)}$ are uniformly close for all $T \in \mathbb{N}$ and so are the limits.

Lemma 3.10. *Under the same conditions as in Lemma 3.9, we have $\|z^{(T)} - z^{(T,-j)}\| \leq 3, \forall T \in \mathbb{N}$. As a consequence, $\|z - z^{(-j)}\| \leq 3, \forall j \in [n]$.*

Note that both z and $z^{(-j)}$ are length- n vectors in $\mathbb{C}_{\leq 1}^n$. Lemma 3.10 means that they are pretty close to each other and consequently one can be approximated by the other one, leading to the following proposition.

Proposition 3.11. *Under the same conditions as in Lemma 3.9, we have*

$$\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\{\sigma |Wz|_j \geq |s| - r\} \leq \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\sigma \left|W_j \cdot z^{(-j)}\right| \geq |s| - r - 3\sigma \|W\|\right\}, \quad \forall r \in \mathbb{R}.$$

In Proposition 3.11, the left-hand and right-hand sides of the display correspond to (3.1) and (3.2), respectively. Note that for each $j \in [n]$, $z^{(-j)}$ is independent of W_j . In this way, we manage to decouple the dependence in Wz . The cost of replacing z by $z^{(-j)}$ is $3\sigma \|W\|$, which means the threshold in (3.2) is slightly smaller than that in (3.1).

3.4 Exponential Bounds

In Sections 3.2 and 3.3, we carry out detailed analysis for the MLE to upper bound the quantity $\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\{|[Y \hat{z}^{\text{MLE}}]_j| < \delta n\}$, which is the main term appearing in $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$, the distance between the MLE and the BM factorization in Corollary 2.4. With Corollary 3.8 and Proposition 3.11, Corollary 2.4 leads to the following lemma regarding $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$.

Lemma 3.12. *Suppose $m \in \mathbb{N} \setminus \{1\}$, $\epsilon \in (0, 1/2)$, $h > 0$, and $\delta \geq 2\sqrt{2} \left(6\epsilon + \frac{\sigma \|W\|}{n}\right)$. Assume $\ell_1(\hat{z}^{\text{MLE}}, z^*) \leq \epsilon^2$ and $\ell_m(\hat{V}^{\text{BM},m}, z^*) \leq \epsilon^2$ are satisfied. For each $k = 0, 1, 2, \dots, \lceil n\epsilon/h \rceil$, define $s_k = n - kh \in \mathbb{R}$ and let $z_{s_k}^{(-j)} \in \mathbb{C}_{\leq 1}^n$ be the fixed point of $G^{(-j)}(\cdot, s_k, 2\delta n)$ for each $j \in [n]$. Then we have*

$$\begin{aligned} & \ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \\ & \leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\sigma \left|W_j \cdot z_{s_k}^{(-j)}\right| > \left(1 - \epsilon - 4\delta - \frac{3\sigma \|W\|}{n}\right) n - h\right\} \right) + \frac{72h^2}{\delta^2 n^2} \mathbb{I}\{h > \delta\sqrt{n}\}. \end{aligned}$$

Lemma 3.12 provides a deterministic upper bound for the difference between the MLE and the BM factorization, $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$, as all the analysis carried out so far is completely deterministic. Despite being complicated, it makes it ready for us to obtain explicit expression for using the fact that each entry of W is Gaussian. To achieve this, first note that Lemma 3.12 involves $\|W\|$ which requires an upper bound. It is known in the literature that there exists some absolute constant $C_0 > 0$ such that

$$\mathbb{P}(\|W\| \leq C_0\sqrt{n}) \geq 1 - n^{-10}. \quad (3.13)$$

Such concentration result is standard and is a direct consequence of Proposition 2.4 of [30]. Regarding the display in Lemma 3.12, if we take expectations on both sides, then the indicator functions on its right-hand side become tail probabilities of Gaussian distributions, which are exponentially small. If we do not take expectations, then when σ is small enough, all the indicator functions are equal to 0 with high probability, leading to the tightness $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) = 0$. In both cases, we need to pick a suitable h so that the second term in the display of the lemma is negligible. In this way, we obtain the following exponential bound in Theorem 3.13. Theorem 1.1 is its immediate consequence.

Theorem 3.13. *Suppose $m \in \mathbb{N} \setminus \{1\}$. There exist constants $C_1, C_2 > 0$ that only depend on C_0 such that:*

1. *When $\frac{n}{\sigma^2} \geq C_1$, we have*

$$\mathbb{E}\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \leq \exp\left(-\frac{n}{8\sigma^2}\right) + n^{-10}. \quad (3.14)$$

2. *When $\frac{n}{\sigma^2} \geq \max\{C_2, 9 \log n\}$, we have $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) = 0$ with probability at least $1 - n^{-1}$.*

In (3.14), the term n^{-10} comes from (3.13). Since in (3.13), by increasing C_0 , we can replace n^{-10} by n^{-c} where $c > 0$ can be sufficiently large, the n^{-10} term in (3.14) can be consequently replaced by the much smaller n^{-c} . Consequently, we view n^{-10} as a negligible term in (3.14) compared to its first term. Despite the upper bound in (3.14) only holds for σ such that $\frac{n}{\sigma^2} \geq C_1$, it can be restated so that it holds for all $\sigma > 0$. This is because if $\frac{n}{\sigma^2} < C_1$, due to the fact that $\mathbb{E}\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$ is at most 1, it can be upper bounded by $\exp\left(\frac{C_1}{8}\right) \exp\left(-\frac{n}{8\sigma^2}\right)$. Hence, there exists some constant $C_3 > 0$ such that $\mathbb{E}\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \leq C_3 \exp\left(-\frac{n}{8\sigma^2}\right) + n^{-10}$ for all $\sigma > 0$. Once we establish it, by the connection between $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$ and $n^{-2}\|\hat{Z}^{\text{BM},m} - \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^{\text{H}}\|_{\text{F}}^2$ in (2.3), we immediately establish (1.7), the first part of Theorem 1.1. With Markov inequality, the in-expectation upper bound (3.14) can be converted into a in-probability upper bounded, resulting in the tightness result, the second part of Theorem 3.13, which is also the second part of Theorem 1.1.

To conclude this section, we reflect on the connections and distinctions between the analysis presented in this paper and our prior work [15, 16]. These earlier studies focus on quantifying the deviations of the MLE and the SDP solutions from the ground truth z^* , leveraging the properties of fixed points. For instance, [16] measures the distance between the SDP and z^* through $\ell_n(\hat{V}^{\text{SDP}}, z^*)$, and establishes that this distance is bounded by a certain quantity of z^* , similar to the results of Corollary 2.4. In contrast, the current work expands these frameworks to analyze $\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}})$. First, our analysis leverages fixed points of both mappings F_m and F_1 , instead of using a single mapping as in previous studies, to derive Corollary 2.4. Second and more importantly, the current analysis confronts the challenge of handling $\frac{8}{n} \sum_{j \in [n]} \mathbb{I}\{|[Y \hat{z}^{\text{MLE}}]_j| < \delta n\}$ within the upper bound in the high-noise regime where $\sigma \gtrsim \sqrt{n/\log n}$, unlike in our earlier work where we only need to

control a less complicated quantity of z^* . To tackle this, we have dedicated an entire Section 3 to this issue, introducing a novel approach that involves replacing F_1 with a Lipschitz mapping. This marks a significant methodological advancement over the more straightforward analysis conducted in our previous work.

The techniques developed in this paper may potentially extend to other low-rank optimization problems. One natural direction is orthogonal group synchronization, which generalizes phase synchronization by replacing phases with d -dimensional orthogonal matrices as latent variables. In this setting, analogues of the MLE, SDP, and BM factorization can be defined, along with corresponding fixed-point mappings F_1 and F_m . This structural similarity suggests that our analysis could, in principle, be adapted to this matrix-valued setting. Beyond synchronization problems, the core ideas might be also applicable to problems such as community detection in stochastic block models or k -means clustering. Exploring these directions remains an interesting topic for future research.

4 Discussions

This paper investigates the distance between the MLE and the BM factorization in the phase synchronization, specifically focusing on the high-noise regime where $\sigma \gtrsim \sqrt{n/\log n}$. While this study contributes to our understanding of this specific problem, the BM factorization presents a wide array of interesting and important open questions, many of which are beyond the scope of our current analysis. Others present challenges that our current analytical framework is not equipped to solve. In the following discussions, we explore some of these unresolved issues, highlighting the challenges they present and suggesting potential avenues for future research.

4.1 Computation and Optimization Landscape

This paper focuses on analyzing $\hat{Z}^{\text{BM},m}$, the global solution to the BM factorization (1.6). A natural question is whether this solution can be computed efficiently in practice, especially given that the problem is non-convex. Interestingly, despite this non-convexity, the optimization landscape can sometimes be well-behaved. For example, it is known [9, 7] that when m is large enough ($m > \sqrt{2n}$), all second-order critical points coincide with global ones. More recent works [11, 23, 34, 28, 27] have shown that similar guarantees can still hold for smaller values of m when the noise level σ is not too large ($\sigma \lesssim \sqrt{n/\log n}$). In such cases, simple local algorithms like starting from the leading eigenvectors and applying the mapping F_m iteratively can reliably find the global solution $\hat{Z}^{\text{BM},m}$.

However, when the problem becomes noisier ($\sigma \gtrsim \sqrt{n/(\log n)}$) and m is small, the landscape may no longer be so favorable. In such scenarios, [29] establishes upper bounds on the discrepancies between the global and local maximum. While our study is centered on $\hat{Z}^{\text{BM},m}$, the theoretical framework can be potentially extended to any fixed points of F_m under certain conditions, as evidenced by the bounds developed in Theorem 2.3. Nevertheless, the framework’s applicability is limited and may not be further extended to second-order critical points due to its reliance on the specific properties of fixed points associated with F_m .

4.2 Role of m

Our main results hold for any $m \geq 2$, and interestingly, the bound we establish does not depend on m . At first glance, this might seem surprising. In the BM factorization formulation (1.6), the size of the feasible set increases with m . In fact, when $m = 1$, this set matches exactly with that of the MLE in (1.4). Because the feasible set gets larger as m increases, one might intuitively expect

the BM factorization solution $\hat{Z}^{\text{BM},m}$ to deviate further from the MLE solution $\hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^{\text{H}}$. However, our theoretical bound in (1.7) does not capture this potential dependence on m . This lack of dependence is not due to a fundamental property of the problem, but rather an artifact of our analysis. Our approach rewrites $\hat{Z}^{\text{BM},m}$ as $(\hat{V}^{\text{BM},m})^{\text{H}}\hat{V}^{\text{BM},m}$, and relies on the fact that both $\hat{V}^{\text{BM},m}$ and the MLE are fixed points of the mappings, F_m and F_1 . We then show that these mappings behave like contractions, which allows us to bound the distance between their fixed points. However, in this analysis, the parameter m does not directly influence the contraction-type properties (see Lemmas 2.1 and 2.2), and therefore does not appear in Theorem 2.3 and subsequent analysis.

Incorporating m in the upper bound would require a more refined analysis that is beyond our current analysis framework. A possible direction for understanding the role of m is to build on recent advances in the analysis of optimization landscapes for synchronization problems [23, 27, 11, 26]. These works examine associated Laplacian matrices to derive conditions on m under which the landscape is benign. Borrowing ideas from this line of work could provide valuable insights into how the choice of m influences the behavior of the BM factorization.

4.3 Tightness of the SDP

This paper does not introduce new theoretical bounds for the tightness of the SDP. Instead, we confirm that the condition $\sigma \lesssim \sqrt{n/\log n}$ remains sufficient for ensuring tightness, consistent with earlier results by [38]. The broader question—under what conditions a non-convex problem can be exactly solved via a convex relaxation—is an important topic in optimization and continues to motivate a great deal of research. In addition to phase synchronization, similar sufficient conditions for SDP tightness have been established in related problems such as orthogonal group synchronization [36, 20] and the generalized orthogonal Procrustes problem [24]. However, what remains less understood is whether such conditions are also necessary. Addressing this question would be of substantial interest.

In the context of phase synchronization, [3] provides some empirical insights. Specifically, its Figure 2 suggests that the SDP remains tight when $\sigma \leq \sqrt{n}/3$. However, the accompanying text in [3] interprets these results to suggest that σ might be allowed to grow at a rate of $\sqrt{n}/\text{polylog}(n)$ to maintain the tightness, where $\text{polylog}(n)$ means some polynomial in $\log n$. This interpretation is echoed in the follow-up work [38]. There is a subtle discrepancy between the visual and textual implications drawn from these studies.

While our result in (1.7) holds for the high-noise regime $\sigma \gtrsim \sqrt{n/\log n}$, it provides only an upper bound that may not be sharp. If it turns out to be sharp, it would imply that $\sigma \lesssim \sqrt{n/\log n}$ is both a necessary and sufficient condition for tightness. To rigorously address this question, it would be important to prove that the SDP fails to be tight when σ exceeds a certain threshold, or to establish a lower bound on the discrepancy $n^{-2}\|\hat{Z}^{\text{SDP}} - \hat{z}^{\text{MLE}}(\hat{z}^{\text{MLE}})^{\text{H}}\|_{\text{F}}^2$. These remain open problems and are worthy of further investigation.

4.4 \sqrt{n} Regime

Our results require that $\frac{n}{\sigma^2}$ must exceed a certain threshold, as reflected in Theorem 3.13. Nonetheless, exploring the regime where σ of the order \sqrt{n} is also interesting and important. In this regime, [19] studies the asymptotic performance of the SDP in synchronization problems. It reveals that the SDP is able to achieve a near-optimal performance for Bayesian estimation of z^* , using the cavity method from spin-glass theory. Our current analytical framework does not extend to this regime because we require $\frac{n}{\sigma^2}$ to be sufficiently large to establish the necessary contraction-type

results in Section 2. To thoroughly understand the performance of the BM factorization in the \sqrt{n} regime, further development of the methodologies presented in [19] may be required.

4.5 Eigenvector Method for Phase Synchronization

While our study focuses on the SDP and the BM factorization for the phase synchronization, alternative methods such as the eigenvector approach [31] also merit consideration. The eigenvector method, which involves computing the leading eigenvector of the data matrix Y followed by entry-wise normalization to ensure unit coordinates, is notable for its computational simplicity. In terms of performance, [37] demonstrates that the eigenvector method achieves the estimation rate of $\frac{\sigma^2}{n}$ for z^* , which is comparable to that of the MLE and the SDP. However, evaluating how closely the eigenvector method approximates the MLE within our current analytical framework presents challenges. Unlike the BM factorization, the eigenvector method is not naturally compatible with our fixed-point analysis, which is central to our theory. As a result, it would require distinct analytical techniques that are outside the scope of this paper.

5 Proofs

In this section, we give proofs of main results: Theorem 2.3, Lemma 3.5, Proposition 3.7, Proposition 3.11, and Theorem 3.13. Due to the page limit, we include proofs of the remaining lemmas in the supplementary material.

Proof of Theorem 2.3. By the definition of ℓ_m in (2.2), there exists $a \in \mathbb{C}^m$ such that $\|a\| = 1$ and $\ell_m(VY^H, Yz) = n^{-1} \|VY^H - a(Yz)^H\|_F^2$. From Lemma 2.1, we have

$$\|VY^H - a(Yz)^H\|_F^2 = n\ell_m(VY^H, Yz) \leq n^3 \left(6\epsilon + \frac{\sigma \|W\|}{n}\right)^2 \ell_m(V, z). \quad (5.1)$$

On the other hand, we have

$$\ell_m(F_m(V), F_1(z)) \leq n^{-1} \|F_m(V) - a(F_1(z))^H\|_F^2 = n^{-1} \|F_m(V) - F_m(az^H)\|_F^2, \quad (5.2)$$

where the last equality is due to the fact that $F_m(az^H) = a(F_1(z))^H$.

Consider any $t > 0$. For any $j \in [n]$, recall V_j , $[VY^H]_j$, and $[F_m(V)]_j$ are the j th columns of V , VY^H , and $F_m(V)$, respectively. Note that $[F_m(V)]_j$ and $[F_m(az^H)]_j$ can be written as

$$\begin{aligned} [F_m(V)]_j &= \frac{[VY^H]_j}{\|[VY^H]_j\|} \mathbb{I}\{[VY^H]_j \neq 0\} + V_j \mathbb{I}\{[VY^H]_j = 0\} \\ \text{and } [F_m(az^H)]_j &= \frac{[(az^H)Y^H]_j}{\|[(az^H)Y^H]_j\|} \mathbb{I}\{[(az^H)Y^H]_j \neq 0\} + a\bar{z}_j \mathbb{I}\{[(az^H)Y^H]_j = 0\}. \end{aligned}$$

By applying (2.9) of Lemma 2.2, we have

$$\begin{aligned} \|[F_m(V)]_j - [F_m(az^H)]_j\| &\leq \frac{2\|[VY^H]_j - [a(Yz)^H]_j\|}{t} + 2\mathbb{I}\{\|[(az^H)Y^H]_j\| < t\} \\ &= \frac{2\|[VY^H]_j - [a(Yz)^H]_j\|}{t} + 2\mathbb{I}\{|[Yz]_j| < t\}, \end{aligned}$$

where the last equality is due to $\|[(az^H)Y^H]_j\| = |[(Yz)^H]_j| = |[Yz]_j|$ as $\|a\| = 1$.

Summing over all $j \in [n]$, we have

$$\begin{aligned}
\|F_m(V) - F_m(az^H)\|_F^2 &= \sum_{j \in [n]} \|[F_m(V)]_j - [F_m(az^H)]_j\|^2 \\
&\leq \sum_{j \in [n]} \left(\frac{2\|[VY^H]_j - [a(Yz)^H]_j\|}{t} + 2\mathbb{I}\{|[Yz]_j| < t\} \right)^2 \\
&\leq \sum_{j \in [n]} \left(\frac{4\|[VY^H]_j - [a(Yz)^H]_j\|^2}{t^2} + 4\mathbb{I}\{|[Yz]_j| < t\} \right) \\
&= 4t^{-2} \|VY^H - a(Yz)^H\|_F^2 + 4 \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < t\}.
\end{aligned}$$

Together with (5.1) and (5.2), we have

$$\begin{aligned}
\ell_m(F_m(V), F_1(z)) &\leq n^{-1} \left(4t^{-2} \|VY^H - a(Yz)^H\|_F^2 + 4 \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < t\} \right) \\
&\leq n^{-1} \left(4t^{-2} n^3 \left(6\epsilon + \frac{\sigma \|W\|}{n} \right)^2 \ell_m(V, z) + 4 \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < t\} \right) \\
&\leq \frac{4n^2}{t^2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right)^2 \ell_m(V, z) + \frac{4}{n} \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < t\},
\end{aligned}$$

which proves (2.10).

To prove (2.11), set $t = \delta n$. Since $z = F_1(z)$ and $V = F_m(V)$, we have $\ell_m(V, z) = \ell_m(F_m(V), F_1(z))$, and the above display can be written as

$$\left(1 - \frac{4}{\delta^2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right)^2 \right) \ell_m(V, z) \leq \frac{4}{n} \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < \delta n\}.$$

Since $\frac{4}{\delta^2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right)^2 \leq 1/2$, we have $\ell_m(V, z) \leq \frac{8}{n} \sum_{j \in [n]} \mathbb{I}\{|[Yz]_j| < \delta n\}$. \square

Proof of Lemma 3.5. For any $t \geq 4\sigma \|W\|$, let $z \in \mathbb{C}_{\leq 1}^n$ be the fixed point of $G(\cdot, \hat{s}, t)$. By Lemma 3.4, we have $\|\hat{z}^{\text{MLE}} - z\|^2 \leq 32 \sum_{j \in [n]} \mathbb{I}\left\{ \left| z_j^* \hat{s} + \sigma[Wz]_j \right| < t \right\}$. For any $j \in [n]$, note that

$$[Y\hat{z}^{\text{MLE}}]_j = [(z^*(z^*)^H + \sigma W)\hat{z}^{\text{MLE}}]_j = z_j^* \hat{s} + \sigma[W\hat{z}^{\text{MLE}}]_j.$$

and $[z^* \hat{s} + \sigma Wz]_j = z_j^* \hat{s} + \sigma[Wz]_j$. Then $[Y\hat{z}^{\text{MLE}}]_j - [z^* \hat{s} + \sigma Wz]_j = \sigma[W(\hat{z}^{\text{MLE}} - z)]_j$. Hence,

$$\begin{aligned}
\sum_{j \in [n]} |[Y\hat{z}^{\text{MLE}}]_j - [z^* \hat{s} + \sigma Wz]_j|^2 &= \sum_{j \in [n]} |\sigma[W(\hat{z}^{\text{MLE}} - z)]_j|^2 = \sigma^2 \|W(\hat{z}^{\text{MLE}} - z)\|^2 \\
&\leq \sigma^2 \|W\|^2 \|\hat{z}^{\text{MLE}} - z\|^2.
\end{aligned}$$

Then, for any $r > 0$, we have

$$\begin{aligned}
\sum_{j \in [n]} \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j - [z^* \hat{s} + \sigma Wz]_j| > r\} &\leq r^{-2} \sum_{j \in [n]} |[Y\hat{z}^{\text{MLE}}]_j - [z^* \hat{s} + \sigma Wz]_j|^2 \\
&\leq r^{-2} \sigma^2 \|W\|^2 \|\hat{z}^{\text{MLE}} - z\|^2.
\end{aligned}$$

Note that for any $a, b \in \mathbb{R}$, we have

$$\begin{aligned} \mathbb{I}\{|a| < r\} &= \mathbb{I}\{|a - b + b| < r\} \leq \mathbb{I}\{|b| - |a - b| < r\} \\ &= \mathbb{I}\{|b| - |a - b| < r, |a - b| > r\} + \mathbb{I}\{|b| - |a - b| < r, |a - b| \leq r\} \\ &\leq \mathbb{I}\{|a - b| > r\} + \mathbb{I}\{|b| < 2r\}. \end{aligned} \quad (5.3)$$

Hence, for each $j \in [n]$,

$$\mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j| < r\} \leq \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j - [z^*\hat{s} + \sigma Wz]_j| > r\} + \mathbb{I}\{|[z^*\hat{s} + \sigma Wz]_j| < 2r\}.$$

Summing over all $j \in [n]$, we have

$$\begin{aligned} \sum_{j \in [n]} \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j| < r\} &\leq \sum_{j \in [n]} \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j - [z^*\hat{s} + \sigma Wz]_j| > r\} + \sum_{j \in [n]} \mathbb{I}\{|[z^*\hat{s} + \sigma Wz]_j| < 2r\} \\ &\leq r^{-2} \sigma^2 \|W\|^2 \|\hat{z}^{\text{MLE}} - z\|^2 + \sum_{j \in [n]} \mathbb{I}\{|[z^*\hat{s} + \sigma Wz]_j| < 2r\} \\ &= r^{-2} \sigma^2 \|W\|^2 \|\hat{z}^{\text{MLE}} - z\|^2 + \sum_{j \in [n]} \mathbb{I}\{|z_j^* \hat{s} + \sigma [Wz]_j| < 2r\}. \end{aligned}$$

As a consequence,

$$\sum_{j \in [n]} \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j| < r\} \leq 32r^{-2} \sigma^2 \|W\|^2 \sum_{j \in [n]} \mathbb{I}\{|z_j^* \hat{s} + \sigma [Wz]_j| < t\} + \sum_{j \in [n]} \mathbb{I}\{|z_j^* \hat{s} + \sigma [Wz]_j| < 2r\}.$$

Consider any $\delta \geq \frac{2\sigma\|W\|}{n}$. Set $t = 2\delta n$ and $r = \delta n$, we have

$$\begin{aligned} \sum_{j \in [n]} \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j| < \delta n\} &\leq \frac{32\sigma^2 \|W\|^2}{\delta^2 n^2} \sum_{j \in [n]} \mathbb{I}\{|z_j^* \hat{s} + \sigma [Wz]_j| < 2\delta n\} + \sum_{j \in [n]} \mathbb{I}\{|z_j^* \hat{s} + \sigma [Wz]_j| < 2\delta n\} \\ &\leq 9 \sum_{j \in [n]} \mathbb{I}\{|z_j^* \hat{s} + \sigma [Wz]_j| < 2\delta n\}. \end{aligned}$$

Multiplying n^{-1} on both sides, we complete the proof. \square

Proof of Proposition 3.7. Let $a \in \mathbb{C}_1$ such that $\|\hat{z}^{\text{MLE}} - az^*\|^2 = n\ell_1(\hat{z}^{\text{MLE}}, z^*) \leq n\epsilon^2$. Then

$$\begin{aligned} |\hat{s}| &= |(z^*)^H \hat{z}^{\text{MLE}}| = |(z^*)^H (\hat{z}^{\text{MLE}} - az^*) + (z^*)^H (az^*)| \geq |(z^*)^H (az^*)| - |(z^*)^H (\hat{z}^{\text{MLE}} - az^*)| \\ &\geq n - \sqrt{n} \|\hat{z}^{\text{MLE}} - az^*\| \geq (1 - \epsilon)n. \end{aligned}$$

Then $|\hat{s}| \in [(1 - \epsilon)n, n] \subset [s_{\lceil n\epsilon/h \rceil}, s_0]$. Define $\hat{k} = \operatorname{argmin}_{0 \leq k \leq \lceil n\epsilon/h \rceil} |\hat{s}| - s_k$. Then $|\hat{s}| - s_{\hat{k}} \leq h$.

Consider any $\delta \geq \frac{2\sigma\|W\|}{n}$ and let $z \in \mathbb{C}_{\leq 1}^n$ be the fixed point of $G(z, |\hat{s}|, 2\delta n)$. By the 4th property of Lemma 3.2, we have

$$\left\| (z^* |\hat{s}| + \sigma Wz) - (z^* s_{\hat{k}} + \sigma Wz_{s_{\hat{k}}}) \right\|^2 \leq 4n |\hat{s}| - s_{\hat{k}}|^2 \leq 4nh^2.$$

Note that we have the following fact: $\sum_{j \in [n]} \mathbb{I}\{|x_j| > t\} \leq t^{-2} \|x\|^2 \mathbb{I}\{\|x\| > t\}$ for any $x \in \mathbb{R}^n$ and any $t > 0$. This is because if $\|x\| \leq t$, then $\sum_{j \in [n]} \mathbb{I}\{|x_j| > t\} = 0$; if $\|x\| > t$, then

$\sum_{j \in [n]} \mathbb{I}\{|x_j| > t\} \leq \sum_{j \in [n]} t^{-2} |x_j|^2 \mathbb{I}\{|x_j| > t\} \leq \sum_{j \in [n]} t^{-2} |x_j|^2 = t^{-2} \|x\|^2$. Hence,

$$\begin{aligned} & \sum_{j \in [n]} \mathbb{I}\left\{\left|[z^*|\hat{s}| + \sigma Wz]_j - [z^*s_{\hat{k}} + \sigma Wz_{s_{\hat{k}}}]_j\right| > 2\delta n\right\} \\ & \leq (2\delta n)^{-2} \left\| (z^*|\hat{s}| + \sigma Wz) - (z^*s_{\hat{k}} + \sigma Wz_{s_{\hat{k}}}) \right\|^2 \mathbb{I}\left\{\left\| (z^*|\hat{s}| + \sigma Wz) - (z^*s_{\hat{k}} + \sigma Wz_{s_{\hat{k}}}) \right\| > 2\delta n\right\} \\ & \leq (2\delta n)^{-2} (4nh^2) \mathbb{I}\left\{\sqrt{4nh^2} > 2\delta n\right\} = \frac{h^2}{\delta^2 n} \mathbb{I}\{h > \delta\sqrt{n}\}. \end{aligned}$$

Using (5.3) and the above display, we have

$$\begin{aligned} & \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|z_j^*|\hat{s}| + \sigma [Wz]_j\right| < 2\delta n\right\} \\ & = \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|z_j^*s_{\hat{k}} + \sigma [Wz_{s_{\hat{k}}}]_j\right| < 4\delta n\right\} + \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|[z^*|\hat{s}| + \sigma Wz]_j - [z^*s_{\hat{k}} + \sigma Wz_{s_{\hat{k}}}]_j\right| > 2\delta n\right\} \\ & \leq \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|z_j^*s_{\hat{k}} + \sigma [Wz_{s_{\hat{k}}}]_j\right| < 4\delta n\right\} + \frac{h^2}{\delta^2 n^2} \mathbb{I}\{h > \delta\sqrt{n}\} \\ & \leq \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|z_j^*s_k + \sigma [Wz_{s_k}]_j\right| < 4\delta n\right\} \right) + \frac{h^2}{\delta^2 n^2} \mathbb{I}\{h > \delta\sqrt{n}\}, \end{aligned}$$

where the last inequality is due to $\hat{k} \in \{0, 1, 2, \dots, \lceil n\epsilon/h \rceil\}$. By applying Lemma 3.6, we have

$$\begin{aligned} & \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j| < \delta n\} \\ & \leq \frac{9}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|z_j^*|\hat{s}| + \sigma [Wz]_j\right| < 4\delta n\right\} \\ & \leq 9 \left(\sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|z_j^*s_k + \sigma [Wz_{s_k}]_j\right| < 4\delta n\right\} \right) + \frac{h^2}{\delta^2 n^2} \mathbb{I}\{h > \delta\sqrt{n}\} \right) \\ & \leq 9 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|z_j^*s_k + \sigma [Wz_{s_k}]_j\right| < 4\delta n\right\} \right) + \frac{9h^2}{\delta^2 n^2} \mathbb{I}\{h > \delta\sqrt{n}\}. \end{aligned}$$

□

Proof of Proposition 3.11. For any $j \in [n]$, we have

$$\begin{aligned} |[Wz]_j| &= |[Wz^{(-j)}]_j + [W(z - z^{(-j)})]_j| = |W_{j \cdot} z^{(-j)} + [W(z - z^{(-j)})]_j| \\ &\leq |W_{j \cdot} z^{(-j)}| + |[W(z - z^{(-j)})]_j| \leq |W_{j \cdot} z^{(-j)}| + \|W(z - z^{(-j)})\| \\ &\leq |W_{j \cdot} z^{(-j)}| + \|W\| \|z - z^{(-j)}\| \leq |W_{j \cdot} z^{(-j)}| + 3\|W\|, \end{aligned}$$

where the last inequality is due to Lemma 3.10. Hence, for any $r \in \mathbb{R}$, $\mathbb{I}\{\sigma|[Wz]_j| \geq |s| - r\} \leq \mathbb{I}\{\sigma|W_{j \cdot} z^{(-j)}| \geq |s| - r - 3\sigma\|W\|\}$. Summing over all $j \in [n]$, we have

$$\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\{\sigma|[Wz]_j| \geq |s| - r\} \leq \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\{\sigma|W_{j \cdot} z^{(-j)}| \geq |s| - r - 3\sigma\|W\|\}.$$

□

Proof of Theorem 3.13. Consider any $m \in \mathbb{N} \setminus \{1\}$. Recall the upper bound for $\|W\|$ in (3.13). We first prove (3.14). Since

$$\ell_m(V, \hat{z}^{\text{MLE}}) = \ell_m(V, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0 \sqrt{n} \} + \ell_m(V, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| > C_0 \sqrt{n} \},$$

we have

$$\begin{aligned} \mathbb{E} \ell_m(V, \hat{z}^{\text{MLE}}) &= \mathbb{E} (\ell_m(V, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0 \sqrt{n} \}) + \mathbb{E} (\ell_m(V, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| > C_0 \sqrt{n} \}) \\ &\leq \mathbb{E} (\ell_m(V, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0 \sqrt{n} \}) + \mathbb{E} \mathbb{I} \{ \|W\| > C_0 \sqrt{n} \} \\ &\leq \mathbb{E} (\ell_m(V, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0 \sqrt{n} \}) + n^{-10}, \end{aligned} \quad (5.4)$$

where the last inequality is due to (3.13), we focus on analyzing $\mathbb{E} (\ell_m(V, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0 \sqrt{n} \})$.

Assume $\|W\| \leq C_0 \sqrt{n}$. We first perform some deterministic analysis on $\ell_m(V, \hat{z}^{\text{MLE}})$. From Lemma 2.5, we have $\ell_1(\hat{z}^{\text{MLE}}, z^*), \ell_m(\hat{V}^{\text{BM},m}, z^*) \leq 8\sigma \|W\|/n \leq 8C_0\sigma/\sqrt{n}$. Set $\epsilon = \left(\frac{8C_0\sigma}{\sqrt{n}}\right)^{\frac{1}{2}}$ so that $\ell_1(\hat{z}^{\text{MLE}}, z^*), \ell_m(\hat{V}^{\text{BM},m}, z^*) \leq \epsilon^2$. Assume $\frac{C_0\sigma}{\sqrt{n}} < \frac{1}{16}$. Then $\epsilon < 1/2$ and

$$2\sqrt{2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right) = 2\sqrt{2} \left(6 \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}} + \frac{\sigma \|W\|}{n} \right) = 48 \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}} + 2\sqrt{2} \frac{C_0\sigma}{\sqrt{n}} \leq 49 \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}}.$$

Hence, by setting $\delta = 49 \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}}$, we have $\delta \geq 2\sqrt{2} \left(6\epsilon + \frac{\sigma \|W\|}{n} \right)$. In this way, the conditions required in Lemma 3.12 are satisfied, which leads to

$$\begin{aligned} \ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) &\leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \left| \sigma W_{j, z_{s_k}^{(-j)}} \right| > \left(1 - \epsilon - 4\delta - \frac{3\sigma \|W\|}{n} \right) n - h \right\} \right) + \frac{72h^2}{\delta^2 n^2} \mathbb{I} \{ h > \delta \sqrt{n} \} \\ &\leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \left| \sigma W_{j, z_{s_k}^{(-j)}} \right| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right\} \right) + \frac{72h^2}{\delta^2 n^2} \mathbb{I} \{ h > \delta \sqrt{n} \}. \end{aligned}$$

Since the above inequality holds under the assumption $\|W\| \leq C_0 \sqrt{n}$, we can write it as

$$\begin{aligned} \ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0 \sqrt{n} \} &\leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \left| \sigma W_{j, z_{s_k}^{(-j)}} \right| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right\} \right) + \frac{72h^2}{\delta^2 n^2} \mathbb{I} \{ h > \delta \sqrt{n} \}. \end{aligned} \quad (5.5)$$

Take expectation on both sides of (5.5). Then, we have

$$\begin{aligned} \mathbb{E} \ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0 \sqrt{n} \} &\leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{P} \left(\left| \sigma W_{j, z_{s_k}^{(-j)}} \right| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right) \right) + \frac{72h^2}{\delta^2 n^2} \mathbb{I} \{ h > \delta \sqrt{n} \} \\ &\leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{P} \left(\left| \sigma W_{j, z_{s_k}^{(-j)}} \right| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right) \right) + \frac{72h^2}{\delta^2 n^2}. \end{aligned}$$

For each j and k , due to the independence between W_j and $z_{s_k}^{(-j)}$, $W_j \cdot z_{s_k}^{(-j)}$ is complex Gaussian with zero mean and variance $\|z_{s_k}^{(-j)}\|^2 \leq n$. Then $|W_j \cdot z_{s_k}^{(-j)}|$ is Gaussian with zero mean and variance $\|z_{s_k}^{(-j)}\|^2$. Let $1 - \Phi(x)$ be cumulative distribution function of the standard normal. That is, $\Phi(x) = \int_{-\infty}^x 1/\sqrt{2\pi} \exp(-u^2/2) du$. Then

$$\begin{aligned} \mathbb{P}\left(\sigma \left|W_j \cdot z_{s_k}^{(-j)}\right| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}}\right)n - h\right) &= \Phi\left(\frac{1}{\sigma \|z_{s_k}^{(-j)}\|} \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} - \frac{h}{n}\right)n\right) \\ &\leq \Phi\left(\left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} - \frac{h}{n}\right) \frac{\sqrt{n}}{\sigma}\right), \end{aligned}$$

which is invariant of j or k . As a result,

$$\begin{aligned} &\mathbb{E}\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I}\{\|W\| \leq C_0\sqrt{n}\} \\ &\leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \Phi\left(\left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} - \frac{h}{n}\right) \frac{\sqrt{n}}{\sigma}\right) + \frac{72h^2}{\delta^2 n^2} \\ &\leq 72 \left\lceil \frac{n\epsilon}{h} \right\rceil \Phi\left(\left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} - \frac{h}{n}\right) \frac{\sqrt{n}}{\sigma}\right) + \frac{72h^2}{\delta^2 n^2}. \end{aligned} \quad (5.6)$$

Recall that $\epsilon = \left(\frac{8C_0\sigma}{\sqrt{n}}\right)^{\frac{1}{2}}$ and $\delta = 49 \left(\frac{C_0\sigma}{\sqrt{n}}\right)^{\frac{1}{2}}$. Set $h = n \exp\left(-\frac{n}{8\sigma^2}\right)$. Then (5.6) becomes

$$\begin{aligned} &\mathbb{E}\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I}\{\|W\| \leq C_0\sqrt{n}\} \\ &\leq 72 \exp\left(\frac{n}{8\sigma^2}\right) \Phi\left(\left(1 - \left((196 + 2\sqrt{2}) \left(\frac{C_0\sigma}{\sqrt{n}}\right)^{\frac{1}{2}} + \frac{3C_0\sigma}{\sqrt{n}} + \exp\left(-\frac{n}{8\sigma^2}\right)\right)\right) \frac{\sqrt{n}}{\sigma}\right) \\ &\quad + \frac{1}{100C_0} \left(\frac{n}{\sigma^2}\right)^{\frac{1}{2}} \exp\left(-\frac{n}{4\sigma^2}\right). \end{aligned}$$

Note that $\Phi(x) \leq \frac{2}{\sqrt{\pi}} \exp\left(-\frac{x^2}{2}\right)$ for any $x > 0$. Then there exists some constant $C_1 > 0$ that only depends on C_0 , such that if $\frac{n}{\sigma^2} \geq C_1$, we have

$$(196 + 2\sqrt{2}) \left(\frac{C_0\sigma}{\sqrt{n}}\right)^{\frac{1}{2}} + \frac{3C_0\sigma}{\sqrt{n}} + \exp\left(-\frac{n}{8\sigma^2}\right) \leq \frac{1}{4}, \quad (5.7)$$

and consequently,

$$\begin{aligned} &\mathbb{E}\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I}\{\|W\| \leq C_0\sqrt{n}\} \\ &\leq 72 \exp\left(\frac{n}{8\sigma^2}\right) \Phi\left(\frac{3\sqrt{n}}{4\sigma}\right) + \frac{1}{100C_0} \left(\frac{n}{\sigma^2}\right)^{\frac{1}{2}} \exp\left(-\frac{n}{4\sigma^2}\right) \\ &\leq 72 \exp\left(\frac{n}{8\sigma^2}\right) \frac{2}{\sqrt{\pi}} \exp\left(-\frac{9n}{32\sigma^2}\right) + \frac{1}{100C_0} \left(\frac{n}{\sigma^2}\right)^{\frac{1}{2}} \exp\left(-\frac{n}{4\sigma^2}\right) \\ &= \frac{144}{\sqrt{\pi}} \exp\left(-\frac{5n}{32\sigma^2}\right) + 2 \exp\left(-\frac{n}{6\sigma^2}\right) \leq \exp\left(-\frac{n}{8\sigma^2}\right). \end{aligned}$$

The proof for (3.14) is complete with (5.4).

To prove the second part of the theorem, note that from (5.5), we can also get a in-probability bound. Set $h = \delta\sqrt{n}$. Then (5.5) becomes

$$\begin{aligned} & \ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0\sqrt{n} \} \\ & \leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \sigma |W_{j, z_{s_k}^{(-j)}}| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right\} \right). \end{aligned}$$

Using Markov inequality, we have

$$\begin{aligned} & \mathbb{P} \left(\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0\sqrt{n} \} > 0 \right) \\ & \leq \mathbb{P} \left(72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \sigma |W_{j, z_{s_k}^{(-j)}}| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right\} \right) > 0 \right) \\ & = \mathbb{P} \left(\sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \sigma |W_{j, z_{s_k}^{(-j)}}| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right\} \right) > 0 \right) \\ & = \mathbb{P} \left(\sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \sigma |W_{j, z_{s_k}^{(-j)}}| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right\} \right) \geq \frac{1}{n} \right) \\ & \leq n \mathbb{E} \left(\sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I} \left\{ \sigma |W_{j, z_{s_k}^{(-j)}}| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right\} \right) \right) \\ & = \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\sum_{j \in [n]} \mathbb{P} \left(\sigma |W_{j, z_{s_k}^{(-j)}}| > \left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} \right) n - h \right) \right), \end{aligned}$$

Then by the same simplification as used in the derivation of (5.6), we have

$$\mathbb{P} \left(\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0\sqrt{n} \} > 0 \right) \leq n \left\lceil \frac{n\epsilon}{h} \right\rceil \Phi \left(\left(1 - \epsilon - 4\delta - \frac{3C_0\sigma}{\sqrt{n}} - \frac{h}{n} \right) \frac{\sqrt{n}}{\sigma} \right).$$

Recall that $\epsilon = \left(\frac{8C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}}$, $\delta = 49 \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}}$, and $h = \delta\sqrt{n}$, we have

$$\begin{aligned} & \mathbb{P} \left(\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I} \{ \|W\| \leq C_0\sqrt{n} \} > 0 \right) \\ & \leq n \left\lceil \frac{2\sqrt{2n}}{49} \right\rceil \Phi \left(\left(1 - (196 + 2\sqrt{2}) \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}} - \frac{3C_0\sigma}{\sqrt{n}} - \frac{49}{\sqrt{n}} \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}} \right) \frac{\sqrt{n}}{\sigma} \right) \\ & \leq n \left\lceil \frac{2\sqrt{2n}}{49} \right\rceil \Phi \left(\left(1 - \left(196 + 2\sqrt{2} + \frac{49}{\sqrt{n}} \right) \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}} - \frac{3C_0\sigma}{\sqrt{n}} \right) \frac{\sqrt{n}}{\sigma} \right). \end{aligned}$$

Similar to (5.7), there exists some constant $C_2 > 0$ that only depends on C_0 such that if $\frac{n}{\sigma^2} \geq C_2$, we have

$$\left(196 + 2\sqrt{2} + \frac{49}{\sqrt{n}} \right) \left(\frac{C_0\sigma}{\sqrt{n}} \right)^{\frac{1}{2}} + \frac{3C_0\sigma}{\sqrt{n}} \leq \frac{1}{4}.$$

Then we have

$$\begin{aligned} \mathbb{P}\left(\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I}\{\|W\| \leq C_0\sqrt{n}\} > 0\right) &\leq n \left\lceil \frac{2\sqrt{2n}}{49} \right\rceil \Phi\left(\frac{3\sqrt{n}}{4\sigma}\right) \\ &\leq n \left\lceil \frac{2\sqrt{2n}}{49} \right\rceil \frac{2}{\sqrt{\pi}} \exp\left(-\frac{9n}{32\sigma^2}\right) \leq \frac{1}{2} n^{\frac{3}{2}} \exp\left(-\frac{9n}{32\sigma^2}\right) \leq \frac{1}{2} n^{-1}, \end{aligned}$$

where the last inequality holds under the assumption $\frac{n}{\sigma^2} \geq 9 \log n$. Hence,

$$\begin{aligned} &\mathbb{P}\left(\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) > 0\right) \\ &\leq \mathbb{P}\left(\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I}\{\|W\| \leq C_0\sqrt{n}\} > 0\right) + \mathbb{P}\left(\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I}\{\|W\| > C_0\sqrt{n}\} > 0\right) \\ &\leq \mathbb{P}\left(\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \mathbb{I}\{\|W\| \leq C_0\sqrt{n}\} > 0\right) + \mathbb{P}(\|W\| > C_0\sqrt{n}) \\ &\leq \frac{1}{2} n^{-1} + n^{-10} \leq n^{-1}. \end{aligned}$$

□

References

- [1] Emmanuel Abbe, Laurent Massoulié, Andrea Montanari, Allan Sly, and Nikhil Srivastava. Group synchronization on grids. *arXiv preprint arXiv:1706.08561*, 2017.
- [2] Mica Arie-Nachimson, Shahar Z Kovalsky, Ira Kemelmacher-Shlizerman, Amit Singer, and Ronen Basri. Global motion estimation from point matches. In *2012 Second international conference on 3D imaging, modeling, processing, visualization & transmission*, pages 81–88. IEEE, 2012.
- [3] Afonso S Bandeira, Nicolas Boumal, and Amit Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163(1-2):145–167, 2017.
- [4] Afonso S Bandeira, Nicolas Boumal, and Vladislav Voroninski. On the low-rank approach for semidefinite programs arising in synchronization and community detection. In *Conference on learning theory*, pages 361–382. PMLR, 2016.
- [5] Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.
- [6] Nicolas Boumal, Pierre-Antoine Absil, and Coralia Cartis. Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33, 2019.
- [7] Nicolas Boumal, Vlad Voroninski, and Afonso Bandeira. The non-convex burer-monteiro approach works on smooth semidefinite programs. *Advances in Neural Information Processing Systems*, 29, 2016.
- [8] Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical programming*, 95(2):329–357, 2003.

- [9] Samuel Burer and Renato DC Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical programming*, 103(3):427–444, 2005.
- [10] Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.
- [11] Faniriana Rakoto Endor and Irène Waldspurger. Benign landscape for burer-monteiro factorizations of maxcut-type semidefinite programs. *arXiv preprint arXiv:2411.03103*, 2024.
- [12] Yifeng Fan, Yuehaw Khoo, and Zhizhen Zhao. Joint community detection and rotational synchronization via semidefinite programming. *arXiv preprint arXiv:2105.06031*, 2021.
- [13] Yingjie Fei and Yudong Chen. Achieving the bayes error rate in synchronization and block models by sdp, robustly. *IEEE Transactions on Information Theory*, 66(6):3929–3953, 2020.
- [14] Yingjie Fei and Yudong Chen. Hidden integrality and semirandom robustness of sdp relaxation for sub-gaussian mixture model. *Mathematics of Operations Research*, 47(3):2464–2493, 2022.
- [15] Chao Gao and Anderson Y Zhang. Exact minimax estimation for phase synchronization. *IEEE Transactions on Information Theory*, 67(12):8236–8247, 2021.
- [16] Chao Gao and Anderson Y Zhang. Sdp achieves exact minimax optimality in phase synchronization. *IEEE Transactions on Information Theory*, 68(8):5374–5390, 2022.
- [17] Christophe Giraud and Nicolas Verzelen. Partial recovery bounds for clustering with the relaxed k -means. *Mathematical Statistics and Learning*, 1(3):317–374, 2019.
- [18] Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Transactions on Information Theory*, 62(10):5918–5937, 2016.
- [19] Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi. Phase transitions in semidefinite relaxations. *Proceedings of the National Academy of Sciences*, 113(16):E2218–E2223, 2016.
- [20] Shuyang Ling. Solving orthogonal group synchronization via convex and low-rank optimization: Tightness and landscape analysis. *arXiv preprint arXiv:2006.00902*, 2020.
- [21] Shuyang Ling. Near-optimal bounds for generalized orthogonal procrustes problem via generalized power method. *arXiv preprint arXiv:2112.13725*, 2021.
- [22] Shuyang Ling. Improved performance guarantees for orthogonal group synchronization via generalized power method. *SIAM Journal on Optimization*, 32(2):1018–1048, 2022.
- [23] Shuyang Ling. Local geometry determines global landscape in low-rank factorization for synchronization. *arXiv preprint arXiv:2311.18670*, 2023.
- [24] Shuyang Ling. Near-optimal bounds for generalized orthogonal procrustes problem via generalized power method. *Applied and Computational Harmonic Analysis*, 66:62–100, 2023.
- [25] Huikang Liu, Man-Chung Yue, and Anthony Man-Cho So. On the estimation performance and convergence rate of the generalized power method for phase synchronization. *SIAM Journal on Optimization*, 27(4):2426–2446, 2017.

- [26] Andrew D McRae. Benign landscapes for synchronization on spheres via normalized laplacian matrices. *arXiv preprint arXiv:2503.18801*, 2025.
- [27] Andrew D McRae, Pedro Abdalla, Afonso S Bandeira, and Nicolas Boumal. Nonconvex landscapes for \mathbb{Z}_2 synchronization and graph clustering are benign near exact recovery thresholds. *arXiv preprint arXiv:2407.13407*, 2024.
- [28] Andrew D McRae and Nicolas Boumal. Benign landscapes of low-dimensional relaxations for orthogonal synchronization on general graphs. *SIAM Journal on Optimization*, 34(2):1427–1454, 2024.
- [29] Song Mei, Theodor Misiakiewicz, Andrea Montanari, and Roberto Imbuzeiro Oliveira. Solving sdps for synchronization and maxcut problems via the grothendieck inequality. In *Conference on learning theory*, pages 1476–1515. PMLR, 2017.
- [30] Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pages 1576–1602. World Scientific, 2010.
- [31] Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011.
- [32] Amit Singer and Yoel Shkolnisky. Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM journal on imaging sciences*, 4(2):543–572, 2011.
- [33] Irene Waldspurger, Alexandre d’Aspremont, and Stéphane Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149:47–81, 2015.
- [34] Irene Waldspurger and Alden Waters. Rank optimality for the burer–monteiro factorization. *SIAM journal on Optimization*, 30(3):2577–2602, 2020.
- [35] Lanhui Wang and Amit Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference: A Journal of the IMA*, 2(2):145–193, 2013.
- [36] Joong-Ho Won, Teng Zhang, and Hua Zhou. Orthogonal trace-sum maximization: Tightness of the semidefinite relaxation and guarantee of locally optimal solutions. *SIAM Journal on Optimization*, 32(3):2180–2207, 2022.
- [37] Anderson Ye Zhang. Exact minimax optimality of spectral methods in phase synchronization and orthogonal group synchronization. *The Annals of Statistics*, 52(5):2112–2138, 2024.
- [38] Yiqiao Zhong and Nicolas Boumal. Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016, 2018.

SUPPLEMENTARY MATERIAL: TIGHTNESS OF SDP AND BURER-MONTEIRO FACTORIZATION FOR PHASE SYNCHRONIZATION IN HIGH-NOISE REGIME

BY Anderson Ye Zhang

University of Pennsylvania

SM1 Proofs of Lemmas in Section 2

We defer the proof of Lemma 2.1 to Section SM5 as the lemma is a direct generalization of Lemma 12 of [38] and our proof follows theirs.

Proof of Lemma 2.2. To prove (2.8), let $\theta \in [0, \pi]$ be the angle between x and y . By the cosine formula of triangles, we have $\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\cos(\theta)$ and $\|x/\|x\| - y/\|y\|\|^2 = 2 - 2\cos(\theta)$. Consider the following scenarios.

- If $\|x\|, \|y\| \geq t$, since $\|x\|^2 + \|y\|^2 \geq 2\|x\|\|y\|$, we have

$$\|x - y\|^2 \geq 2\|x\|\|y\|(1 - \cos(\theta)) \geq 2t^2(1 - \cos(\theta)) = t^2\|x/\|x\| - y/\|y\|\|^2.$$

Hence, $\|x/\|x\| - y/\|y\|\| \leq \|x - y\|/t$.

- If $\|y\| \geq t > \|x\|$ and $\cos(\theta) \geq 0$, define a function $f(a, b) = a^2 + b^2 - 2ab\cos(\theta)$ for $a, b \in \mathbb{R}$. Note that for any $1 \geq a > 0, b \geq 1$, we have $f(a, b) \geq 1 - \cos^2(\theta)$. This is because $f(a, b) \geq \min_{b' \geq 1} f(a, b') = f(a, 1) = a^2 + 1 - 2a\cos(\theta) \geq \min_{1 \geq a' > 0} f(a', 1) = f(\cos(\theta), 1) = 1 - \cos^2(\theta)$. Hence,

$$\begin{aligned} \frac{2\|x - y\|^2}{t^2} &= 2 \left(\left(\frac{\|x\|}{t} \right)^2 + \left(\frac{\|y\|}{t} \right)^2 - \frac{\|x\|}{t} \frac{\|y\|}{t} \cos(\theta) \right) \\ &\geq 2(1 - \cos^2(\theta)) \\ &\geq 2(1 - \cos(\theta)) \\ &= \left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\|^2. \end{aligned}$$

Hence, $\|x/\|x\| - y/\|y\|\| \leq \sqrt{2}\|x - y\|/t$.

- If $\|y\| \geq t > \|x\|$ and $\cos(\theta) < 0$, we have $\|x - y\|^2 \geq \|y\|^2 \geq t^2$ and $\|x/\|x\| - y/\|y\|\| \leq 2$. Hence, $\|x/\|x\| - y/\|y\|\| \leq 2\|x - y\|/t$.
- If $\|y\| < t$, we have $\|x/\|x\| - y/\|y\|\| \leq 2 = 2\mathbb{I}\{\|y\| < t\}$.

The proof of (2.8) is complete.

To prove (2.9), we only need to consider scenarios $x = 0$ or $y = 0$, as otherwise (2.9) is reduced to (2.8). If $y = 0$, we have

$$\begin{aligned} &\left\| \left(\frac{x}{\|x\|} \mathbb{I}\{x \neq 0\} + u \mathbb{I}\{x = 0\} \right) - \left(\frac{y}{\|y\|} \mathbb{I}\{y \neq 0\} + v \mathbb{I}\{y = 0\} \right) \right\| \\ &= \left\| \left(\frac{x}{\|x\|} \mathbb{I}\{x \neq 0\} + u \mathbb{I}\{x = 0\} \right) - v \right\| \leq 2 = 2\mathbb{I}\{\|y\| < t\}. \end{aligned}$$

If $x = 0$ and $y \neq 0$, we have

$$\begin{aligned} & \left\| \left(\frac{x}{\|x\|} \mathbb{I}\{x \neq 0\} + u \mathbb{I}\{x = 0\} \right) - \left(\frac{y}{\|y\|} \mathbb{I}\{y \neq 0\} + v \mathbb{I}\{y = 0\} \right) \right\| \\ &= \left\| u - \frac{y}{\|y\|} \right\| \leq 2 = 2\mathbb{I}\{\|y\| \geq t\} + 2\mathbb{I}\{\|y\| < t\} = 2\mathbb{I}\{\|x - y\| \geq t\} + 2\mathbb{I}\{\|y\| < t\} \\ &\leq \frac{2\|x - y\|}{t} + 2\mathbb{I}\{\|y\| < t\}. \end{aligned}$$

The proof of (2.9) is complete. \square

Proof of Lemma 2.5. Consider any $m \in \mathbb{N} \setminus \{1\}$. For simplicity, we write $\hat{Z}^{\text{BM},m}$ as \hat{Z} so that $\hat{Z} = (\hat{V}^{\text{BM},m})^{\text{H}} \hat{V}^{\text{BM},m}$.

First, we are going to show

$$\ell(\hat{V}^{\text{BM},m}, z^*) \leq \frac{4}{n^2} \text{Tr}(z^* z^{*\text{H}} (z^* z^{*\text{H}} - \hat{Z})). \quad (\text{SM1.1})$$

Define $b = n^{-1} \sum_{j=1}^n \hat{V}_j^{\text{BM},m} z_j^* = n^{-1} \hat{V}^{\text{BM},m} z^* \in \mathbb{C}^m$. If $b = 0$, we have

$$\begin{aligned} \text{Tr}(z^* z^{*\text{H}} (z^* z^{*\text{H}} - \hat{Z})) &= \text{Tr}(z^* (z^*)^{\text{H}} z^* (z^*)^{\text{H}}) - \text{Tr}(z^* z^{*\text{H}} (\hat{V}^{\text{BM},m})^{\text{H}} \hat{V}^{\text{BM},m}) \\ &= n \text{Tr}(z^* z^{*\text{H}}) - \text{Tr}(z^* (nb)^{\text{H}} \hat{V}^{\text{BM},m}) \\ &= n^2. \end{aligned}$$

Note that $\ell(\hat{V}^{\text{BM},m}, z^*) \leq n^{-1} \sum_{j \in [n]} 4 = 4$. Then (SM1.1) holds. In the following, we assume $b \neq 0$. From Lemma 2.2, we have for any $x, y \in \mathbb{C}^m$ such that $x \neq 0$ and $\|y\| = 1$, $\|x/\|x\| - y\| \leq 2\|x - y\|$. Hence, we have

$$\begin{aligned} \ell(\hat{V}^{\text{BM},m}, z^*) &= \min_{a \in \mathbb{C}^n: \|a\|^2=1} \frac{1}{n} \sum_{j=1}^n \|\hat{V}_j^{\text{BM},m} z_j^* - a\|^2 \\ &= \min_{a \in \mathbb{C}^n \setminus \{0\}} \frac{1}{n} \sum_{j=1}^n \|\hat{V}_j^{\text{BM},m} z_j^* - a/\|a\|\|^2 \\ &\leq \min_{a \in \mathbb{C}^n \setminus \{0\}} \frac{4}{n} \sum_{j=1}^n \|\hat{V}_j^{\text{BM},m} z_j^* - a\|^2. \end{aligned}$$

Since the minimum of the above display is achieved when a is the arithmetic mean of $\{\hat{V}_j^{\text{BM},m} z_j^*\}_{j \in [n]}$, i.e., b , we have

$$\begin{aligned} \ell(\hat{V}^{\text{BM},m}, z^*) &\leq \frac{4}{n} \sum_{j=1}^n \|\hat{V}_j^{\text{BM},m} z_j^* - b\|^2 \\ &= \frac{2}{n^2} \sum_{j=1}^n \sum_{l=1}^n \left(\|\hat{V}_j^{\text{BM},m} z_j^* - b\|^2 + \|\hat{V}_l^{\text{BM},m} z_l^* - b\|^2 \right) \\ &= \frac{2}{n^2} \sum_{j=1}^n \sum_{l=1}^n \|\hat{V}_j^{\text{BM},m} z_j^* - \hat{V}_l^{\text{BM},m} z_l^*\|^2 \\ &= \frac{4}{n^2} \sum_{j=1}^n \sum_{l=1}^n (1 - \bar{z}_j^* z_l^* (\hat{V}_j^{\text{BM},m})^{\text{H}} \hat{V}_l^{\text{BM},m}) \\ &= \frac{4}{n^2} \text{Tr}(z^* z^{*\text{H}} (z^* z^{*\text{H}} - \hat{Z})). \end{aligned}$$

Therefore, (SM1.1) holds.

Now it remains to upper bound $\text{Tr}(z^* z^{*H} (z^* z^{*H} - \hat{Z}))$. By the definition (1.6), we have $\text{Tr}(Y \hat{Z}) \geq \text{Tr}(Y z^* z^{*H})$. Rearranging this inequality, we obtain $\text{Tr}(Y(\hat{Z} - z^* z^{*H})) \geq 0$. With (1.2), we have

$$\begin{aligned} \text{Tr}(z^* z^{*H} (z^* z^{*H} - \hat{Z})) &\leq \text{Tr}\left((Y - z^* z^{*H})(\hat{Z} - z^* z^{*H})\right) \\ &= \sigma \text{Tr}\left(W(\hat{Z} - z^* z^{*H})\right) \\ &\leq \sigma \left| \text{Tr}\left(W \hat{Z}\right) \right| + \sigma |\text{Tr}(W z^* z^{*H})| \\ &\leq \sigma \|W\| \text{Tr}(\hat{Z}) + \sigma \|W\| \text{Tr}(z^* z^{*H}) \\ &= 2n\sigma \|W\|. \end{aligned}$$

Here, the last inequality is due to the following facts. For any two matrices $A, B \in \mathbb{C}^{n \times n}$, $\text{Tr}(AB) \leq \|A\| \|B\|_*$, where $\|B\|_*$ is the nuclear norm of B that is equal to the summation of all its singular values. If B is further assumed to be positive semi-definite, we have $\|B\|_* = \text{Tr}(B)$. In our setting, \hat{Z} is positive semi-definite as $\min_{u \in \mathbb{C}^n} u^H \hat{Z} u = \min_{u \in \mathbb{C}^n} u^H (\hat{V}^{\text{BM},m})^H \hat{V}^{\text{BM},m} u \geq 0$, and so is $z^* (z^*)^H$.

Consequently, we have $\ell(\hat{V}^{\text{BM},m}, z^*) \leq \frac{8\sigma\|W\|}{n}$. The upper bound for $\ell_1(\hat{z}^{\text{MLE}}, z^*)$ can be established following the same steps as above and hence its proof is omitted. \square

SM2 Proofs of Lemmas in Section 3.2

Proof of Lemma 3.1. Consider the following scenarios. If $|x|, |y| \leq t$, we have $|g_t(x) - g_t(y)| = \frac{|x-y|}{t}$ by definition. If $|x|, |y| \geq t$, then

$$|g_t(x) - g_t(y)| = \left| \frac{x}{|x|} - \frac{y}{|y|} \right|,$$

Let $\theta \in [0, \pi]$ be the angle between x and y on the complex plane. By the cosine formula of triangles, we have $|x - y|^2 = |x|^2 + |y|^2 - 2|x||y|\cos(\theta)$ and $|g_t(x) - g_t(y)|^2 = 2 - 2\cos(\theta)$. Since $|x|^2 + |y|^2 \geq 2|x||y|$, we have

$$|x - y|^2 \geq 2|x||y|(1 - \cos(\theta)) \geq 2t^2(1 - \cos(\theta)) = t^2 |g_t(x) - g_t(y)|^2,$$

which yields the desired result. If $|x| \geq t > |y|$, then

$$|g_t(x) - g_t(y)| = \left| \frac{x}{|x|} - \frac{y}{t} \right|.$$

By using the cosine formula again, we have $|g_t(x) - g_t(y)|^2 = 1 + \frac{|y|^2}{t^2} - 2\frac{|y|}{t}\cos(\theta)$ and $\left|\frac{x}{t} - \frac{y}{t}\right|^2 = \frac{|x|^2}{t^2} + \frac{|y|^2}{t^2} - 2\frac{|x||y|}{t^2}\cos(\theta)$. Then,

$$\begin{aligned} \frac{|x - y|^2}{t^2} - |g_t(x) - g_t(y)|^2 &= \left| \frac{x}{t} - \frac{y}{t} \right|^2 - |g_t(x) - g_t(y)|^2 \\ &= \frac{|x|^2}{t^2} - 1 - 2\frac{|x||y|}{t^2}\cos(\theta) + 2\frac{|y|}{t}\cos(\theta) \\ &= \left(\frac{|x|}{t} - 1 \right) \left(\frac{|x|}{t} + 1 \right) - 2\left(\frac{|x|}{t} - 1 \right) \frac{|y|}{t}\cos(\theta) \\ &= \left(\frac{|x|}{t} - 1 \right) \left(\frac{|x|}{t} + 1 - 2\frac{|y|}{t}\cos(\theta) \right) \\ &\geq 0, \end{aligned}$$

where the last inequality is due to that $\frac{|x|}{t} \geq 1 > \frac{|y|}{t} \geq 0$ and $\cos(\theta) \leq 1$. The scenario $|y| \geq t > |x|$ can be proved similarly. \square

Proof of Lemma 3.2. We prove the properties sequentially.

1. Recall the definition of G in (3.9). For any $j \in [n]$, by Lemma 3.1, we have

$$\begin{aligned} |[G(x, s, t)]_j - [G(y, s, t)]_j| &= |g_t(z_j^* s + \sigma[Wx]_j) - g_t(z_j^* s + \sigma[Wy]_j)| \\ &\leq t^{-1} |(z_j^* s + \sigma[Wx]_j) - (z_j^* s + \sigma[Wy]_j)| \\ &= t^{-1} \sigma |[W(x - y)]_j|. \end{aligned}$$

Summing over all $j \in [n]$, we have

$$\begin{aligned} \|G(x, s, t) - G(y, s, t)\|^2 &\leq \sum_{j \in [n]} |[G(x, s, t)]_j - [G(y, s, t)]_j|^2 \\ &\leq t^{-2} \sigma^2 \sum_{j \in [n]} |[W(x - y)]_j|^2 \\ &= t^{-2} \sigma^2 \|W(x - y)\|^2 \\ &\leq t^{-2} \sigma^2 \|W\|^2 \|x - y\|^2. \end{aligned}$$

2. Using the first property, for any $T \in \mathbb{N}$, we have

$$\begin{aligned} \|z^{(T+1)} - z^{(T)}\| &= \|G(z^{(T)}, s, t) - G(z^{(T-1)}, s, t)\| \\ &\leq t^{-1} \sigma \|W\| \|z^{(T)} - z^{(T-1)}\| \\ &\leq \frac{1}{2} \|z^{(T)} - z^{(T-1)}\|, \end{aligned}$$

where the last inequality is due to the assumption $t \geq 2\sigma \|W\|$.

3. Consider the sequence $z^{(0)} = z^*$ and $z^{(T)} = G(z^{(T-1)}, s, t)$ for all $T \in \mathbb{N}$. By the second property, we have $\|z^{(T+1)} - z^{(T)}\| \leq \frac{1}{2} \|z^{(T)} - z^{(T-1)}\|$ for all $T \in \mathbb{N}$. Note that $\{z^{(T)}\}$ is a sequence in $\mathbb{C}_{\leq 1}^n$, a complete metric space under $\|\cdot\|$. Hence, the sequence converges to a limit $z^{(\infty)} \in \mathbb{C}_{\leq 1}^n$ which satisfies $z^{(\infty)} = G(z^{(\infty)}, s, t)$. Hence, $z^{(\infty)}$ is a fixed point of $G(\cdot, s, t)$. Now we have proved the existence of the fixed point. To prove the uniqueness, note that if there exists another $z' \in \mathbb{C}_{\leq 1}^n$ such that $z' = G(z', s, t)$, we have

$$\|z^{(\infty)} - z'\| = \|G(z^{(\infty)}, s, t) - G(z', s, t)\| \leq t^{-1} \sigma \|z^{(\infty)} - z'\| \leq \|z^{(\infty)} - z'\| / 2,$$

by the first property. Hence, $\|z^{(\infty)} - z'\| = 0$ which means $z^{(\infty)} = z'$.

4. For any $j \in [n]$, we have

$$\begin{aligned} |[z^* s + \sigma Wz]_j - [z^* s' + \sigma Wz']_j| &\leq |z_j^* s - z_j^* s'| + \sigma |[W(z - z')]_j| \\ &\leq |s - s'| + \sigma |[W(z - z')]_j|. \end{aligned}$$

Summing over all $j \in [n]$, we have

$$\begin{aligned} \|(z^*s + \sigma Wz) - (z^*s' + \sigma Wz')\|^2 &\leq \sum_{j \in [n]} (|s - s'| + \sigma |W(z - z')_j|)^2 \\ &\leq \sum_{j \in [n]} (2|s - s'|^2 + 2\sigma^2 |W(z - z')_j|^2) \\ &\leq 2n |s - s'|^2 + 2\sigma^2 \|W\|^2 \|z - z'\|^2. \end{aligned} \quad (\text{SM2.1})$$

Note that for any $j \in [n]$, we have $z_j = [G(z, s, t)]_j = g_t([z^*s + \sigma Wz]_j)$ and similarly $z'_j = g_t([z^*s' + \sigma Wz']_j)$. Hence, by Lemma 3.1, we have

$$|z_j - z'_j| \leq t^{-1} |[z^*s + \sigma Wz]_j - [z^*s' + \sigma Wz']_j|.$$

Summing over all $j \in [n]$, by (SM2.1), we have

$$\begin{aligned} \|z - z'\|^2 &\leq t^{-2} \|(z^*s + \sigma Wz) - (z^*s' + \sigma Wz')\|^2 \\ &\leq 2nt^{-2} |s - s'|^2 + 2\sigma^2 t^{-2} \|W\|^2 \|z - z'\|^2 \\ &\leq 2nt^{-2} |s - s'|^2 + \frac{1}{2} \|z - z'\|^2, \end{aligned}$$

where the last inequality is due to the assumption $t \geq 2\sigma \|W\|$. After rearrangement, we have $\|z - z'\|^2 \leq 4nt^{-2} |s - s'|^2$. From (SM2.1), we have

$$\begin{aligned} \|(z^*s + \sigma Wz) - (z^*s' + \sigma Wz')\|^2 &\leq 2n |s - s'|^2 + 2\sigma^2 \|W\|^2 (4nt^{-2} |s - s'|^2) \\ &\leq 4n |s - s'|^2, \end{aligned}$$

where the last inequality is by $t \geq 2\sigma \|W\|$. □

Proof of Lemma 3.3. Consider any $j \in [n]$. If $|z_j^*s + \sigma[Wz]_j| \geq t$, we have $[G(z, s, t)]_j = g_t(z_j^*s + \sigma[Wz]_j) = (z_j^*s + \sigma[Wz]_j)/|z_j^*s + \sigma[Wz]_j| = [F'_1(z, s)]_j$. If $|z_j^*s + \sigma[Wz]_j| \geq t$ is not satisfied, we have $|[F'_1(z, s)]_j| = 1$ and $|[G(z, s, t)]_j| \leq 1$. Hence,

$$\begin{aligned} |[F'_1(z, s)]_j - [G(z, s, t)]_j| &= |[F'_1(z, s)]_j - [G(z, s, t)]_j| \mathbb{I}\{|z_j^*s + \sigma[Wz]_j| < t\} \\ &\leq 2\mathbb{I}\{|z_j^*s + \sigma[Wz]_j| < t\}. \end{aligned}$$

Summing over all $j \in [n]$, we have

$$\|F'_1(z, s) - G(z, s, t)\|^2 = \sum_{j \in [n]} |[F'_1(z, s)]_j - [G(z, s, t)]_j|^2 \leq 4 \sum_{j \in [n]} \mathbb{I}\{|z_j^*s + \sigma[Wz]_j| < t\}.$$
□

Proof of Lemma 3.4. Consider any $t > 0$. From (3.7), we have $\hat{z}^{\text{MLE}} = F'_1(\hat{z}^{\text{MLE}}, \hat{s})$. Then

$$\begin{aligned} \|\hat{z}^{\text{MLE}} - z\| &= \|F'_1(\hat{z}^{\text{MLE}}, \hat{s}) - G(z, \hat{s}, t)\| \\ &\leq \|F'_1(\hat{z}^{\text{MLE}}, \hat{s}) - F'_1(z, \hat{s})\| + \|F'_1(z, \hat{s}) - G(z, \hat{s}, t)\| \\ &\leq \|F'_1(\hat{z}^{\text{MLE}}, \hat{s}) - F'_1(z, \hat{s})\| + \sqrt{4 \sum_{j \in [n]} \mathbb{I}\{|z_j^*\hat{s} + \sigma[Wz]_j| < t\}}, \end{aligned}$$

where the last inequality is due to Lemma 3.3. Hence,

$$\|\hat{z}^{\text{MLE}} - z\|^2 \leq 2 \|F'_1(\hat{z}^{\text{MLE}}, \hat{s}) - F'_1(z, \hat{s})\|^2 + 8 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \}.$$

Recall the definition of F'_1 in (3.5). For any $j \in [n]$, note that $[F'_1(\hat{z}^{\text{MLE}}, \hat{s})]_j$ and $[F'_1(z, \hat{s})]_j$ can be written as

$$\begin{aligned} [F'_1(\hat{z}^{\text{MLE}}, \hat{s})]_j &= \frac{z_j^* \hat{s} + \sigma[W\hat{z}^{\text{MLE}}]_j}{|z_j^* \hat{s} + \sigma[W\hat{z}^{\text{MLE}}]_j|} \mathbb{I} \{ z_j^* \hat{s} + \sigma[W\hat{z}^{\text{MLE}}]_j \neq 0 \} + \hat{z}_j^{\text{MLE}} \mathbb{I} \{ z_j^* \hat{s} + \sigma[W\hat{z}^{\text{MLE}}]_j = 0 \}, \\ [F'_1(z, \hat{s})]_j &= \frac{z_j^* \hat{s} + \sigma[Wz]_j}{|z_j^* \hat{s} + \sigma[Wz]_j|} \mathbb{I} \{ z_j^* \hat{s} + \sigma[Wz]_j \neq 0 \} + z_j \mathbb{I} \{ z_j^* \hat{s} + \sigma[Wz]_j = 0 \}. \end{aligned}$$

By applying (2.9) of Lemma 2.2,

$$\begin{aligned} |[F'_1(\hat{z}^{\text{MLE}}, \hat{s})]_j - [F'_1(z, \hat{s})]_j| &\leq \frac{2 \left| (z_j^* \hat{s} + \sigma[W\hat{z}^{\text{MLE}}]_j) - (z_j^* \hat{s} + \sigma[Wz]_j) \right|}{t} + 2 \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \} \\ &\leq \frac{2\sigma | [W(\hat{z}^{\text{MLE}} - z)]_j |}{t} + 2 \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \}. \end{aligned}$$

Summing over all $j \in [n]$, we have

$$\begin{aligned} \|F'_1(\hat{z}^{\text{MLE}}, \hat{s}) - F'_1(z, \hat{s})\|^2 &= \sum_{j \in [n]} |[F'_1(\hat{z}^{\text{MLE}}, \hat{s})]_j - [F'_1(z, \hat{s})]_j|^2 \\ &= \sum_{j \in [n]} \left(\frac{4\sigma^2 |[W(\hat{z}^{\text{MLE}} - z)]_j|^2}{t^2} + 4 \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \} \right) \\ &= \frac{4\sigma^2}{t^2} \|W(\hat{z}^{\text{MLE}} - z)\|^2 + 4 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \} \\ &\leq \frac{4\sigma^2}{t^2} \|W\|^2 \|\hat{z}^{\text{MLE}} - z\|^2 + 4 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \}. \end{aligned}$$

Hence,

$$\begin{aligned} \|\hat{z}^{\text{MLE}} - z\|^2 &\leq 2 \left(\frac{4\sigma^2}{t^2} \|W\|^2 \|\hat{z}^{\text{MLE}} - z\|^2 + 4 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \} \right) \\ &\quad + 8 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \} \\ &= \frac{8\sigma^2}{t^2} \|W\|^2 \|\hat{z}^{\text{MLE}} - z\|^2 + 16 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \}. \end{aligned}$$

When $t \geq 4\sigma \|W\|$, we have $\frac{8\sigma^2}{t^2} \|W\|^2 \leq 1/2$ and the above display leads to $\|\hat{z}^{\text{MLE}} - z\|^2 \leq 32 \sum_{j \in [n]} \mathbb{I} \{ |z_j^* \hat{s} + \sigma[Wz]_j| < t \}$. \square

Proof of Lemma 3.6. Recall the definitions of G in (3.9) and g_t in (3.8). Note that for any $t > 0, a \in \mathbb{C}_1, x \in \mathbb{C}$, we have $ag_t(x) = g_t(ax)$. Hence, for any $z \in \mathbb{C}_{\leq 1}^n, s \in \mathbb{C}, t > 0, a \in \mathbb{C}_1$, and $j \in [n]$, we have $a[G(z, s, t)]_j = ag_t([z^*s + \sigma Wz]_j) = g_t(a[z^*s + \sigma Wz]_j) = g_t([z^*(as) + \sigma W(az)]_j)$. As a result,

$$\text{if } z = G(z, s, t), \text{ then } az = G(az, as, t).$$

This means that a fixed point of $G(\cdot, s, t)$ is also a fixed point of $G(\cdot, as, t)$.

Recall the definition of \hat{s} in (3.6). We only need to study the case that $\hat{s} \neq 0$ as otherwise $G(\cdot, |\hat{s}|, \cdot) = G(\cdot, \hat{s}, \cdot)$ and Lemma 3.6 is identical to Lemma 3.5. Since $\hat{s} \neq 0, \hat{s}/|\hat{s}| \in \mathbb{C}_1$ is well-defined. For any $\delta \geq \frac{2\sigma\|W\|}{n}$, let $z \in \mathbb{C}_{\leq 1}^n$ be the fixed point of $G(\cdot, |\hat{s}|, 2\delta n)$. Then we have $\frac{\hat{s}}{|\hat{s}|}z \in \mathbb{C}_{\leq 1}^n$ and

$$\frac{\hat{s}}{|\hat{s}|}z = G\left(\frac{\hat{s}}{|\hat{s}|}z, \frac{\hat{s}}{|\hat{s}|}|\hat{s}|, 2\delta n\right) = G\left(\frac{\hat{s}}{|\hat{s}|}z, \hat{s}, 2\delta n\right).$$

That is, $\frac{\hat{s}}{|\hat{s}|}z$ is the fixed point of $G(\cdot, \hat{s}, 2\delta n)$. By Lemma 3.5, we have

$$\begin{aligned} \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\{|[Y\hat{z}^{\text{MLE}}]_j| < \delta n\} &\leq \frac{9}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|z_j^*\hat{s} + \sigma \left[W \frac{\hat{s}}{|\hat{s}|}z\right]_j\right| < 2\delta n\right\} \\ &= \frac{9}{n} \sum_{j \in [n]} \mathbb{I}\left\{\left|\frac{\hat{s}}{|\hat{s}|} \left(z_j^*|\hat{s}| + \sigma [Wz]_j\right)\right| < 2\delta n\right\} \\ &= \frac{9}{n} \sum_{j \in [n]} \mathbb{I}\left\{|z_j^*|\hat{s}| + \sigma [Wz]_j| < 2\delta n\right\}. \end{aligned}$$

□

SM3 Proofs of Lemmas in Section 3.3

The following lemma is a counterpart of Lemma 3.2 but for $G^{(-j)}$ instead of G . Then Lemma 3.9 is the direct consequence of the third properties of Lemmas 3.2 and SM3.1.

Lemma SM3.1. *Consider any $j \in [n]$. The function $G^{(-j)}(\cdot, \cdot, \cdot)$ has the following properties:*

1. *For any $x, y \in \mathbb{C}^n$ and for any $s \in \mathbb{C}, t > 0$, we have*

$$\left\|G^{(-j)}(x, s, t) - G^{(-j)}(y, s, t)\right\| \leq t^{-1}\sigma\|W\|\|x - y\|^2.$$

2. *For any $s \in \mathbb{C}, t \geq 2\sigma\|W\|$, and for any $z^{(0, -j)} \in \mathbb{C}_{\leq 1}^n$, define $z^{(T, -j)} = G^{(-j)}(z^{(T-1, -j)}, s, t)$ for all $T \in \mathbb{N}$. Then*

$$\left\|z^{(T+1, -j)} - z^{(T, -j)}\right\| \leq \frac{1}{2} \left\|z^{(T, -j)} - z^{(T-1, -j)}\right\|, \forall T \in \mathbb{N}.$$

3. *For any $s \in \mathbb{C}, t \geq 2\sigma\|W\|$, $G(\cdot, s, t)$ has exactly one fixed point. That is, there exists one and only one $z \in \mathbb{C}_{\leq 1}^n$ such that $z = G^{(-j)}(z, s, t)$. In addition, z can be achieved by iteratively applying $G^{(-j)}(\cdot, s, t)$ starting from z^* . That is, let $z^{(0, -j)} = z^*$ and define $z^{(T, -j)} = G^{(-j)}(z^{(T-1, -j)}, s, t)$ for all $T \in \mathbb{N}$. We have $z = \lim_{T \rightarrow \infty} G^{(-j)}(z^{(T, -j)}, s, t)$.*

Proof. Note that $\|W^{(-j)}\| \leq \|W\|$ since $W^{(-j)}$ is obtained from W by zeroing out the j th row and column. With this, the lemma can be proved following the exact same argument as in the proof of Lemma 3.2, and hence is omitted here. \square

Proof of Lemma 3.10. Consider any $T \in \mathbb{N}$. For any $k \in [n]$, by Lemma 3.1, we have

$$\begin{aligned}
& \left| z_k^{(T)} - z_k^{(T,-j)} \right| \\
&= \left| [G(z^{(T-1)}, s, t)]_k - [G^{(-j)}(z^{(T-1,-j)}, s, t)]_k \right| \\
&= \left| g_t([z^*s + \sigma W z^{(T-1)}]_k) - g_t([z^*s + \sigma W^{(-j)} z^{(T-1,-j)}]_k) \right| \\
&\leq t^{-1} \left| [z^*s + \sigma W z^{(T-1)}]_k - [z^*s + \sigma W^{(-j)} z^{(T-1,-j)}]_k \right| \\
&= t^{-1} \sigma \left| [W z^{(T-1)}]_k - [W^{(-j)} z^{(T-1,-j)}]_k \right| \\
&= t^{-1} \sigma \left| [W z^{(T-1)}]_k - [W^{(-j)} z^{(T-1)}]_k + [W^{(-j)} z^{(T-1)}]_k - [W^{(-j)} z^{(T-1,-j)}]_k \right| \\
&= t^{-1} \sigma \left| [(W - W^{(-j)}) z^{(T-1)}]_k + [W^{(-j)}(z^{(T-1)} - z^{(T-1,-j)})]_k \right| \\
&\leq t^{-1} \sigma \left| [(W - W^{(-j)}) z^{(T-1)}]_k \right| + t^{-1} \sigma \left| [W^{(-j)}(z^{(T-1)} - z^{(T-1,-j)})]_k \right|.
\end{aligned}$$

If $k \neq j$, we have $[(W - W^{(-j)}) z^{(T-1)}]_k = W_{kj} z_j^{(j)}$. Then the above display becomes

$$\begin{aligned}
\left| z_k^{(T)} - z_k^{(T,-j)} \right| &\leq t^{-1} \sigma \left| W_{kj} z_j^{(j)} \right| + t^{-1} \sigma \left| [W^{(-j)}(z^{(T-1)} - z^{(T-1,-j)})]_k \right| \\
&\leq t^{-1} \sigma |W_{kj}| + t^{-1} \sigma \left| [W^{(-j)}(z^{(T-1)} - z^{(T-1,-j)})]_k \right|,
\end{aligned}$$

where in the last inequality we use $|z_j^{(j)}| \leq 1$ as $z^{(j)} \in \mathbb{C}_{\leq 1}^n$. Summing over all $k \in [n]$ such that $k \neq j$, we have

$$\begin{aligned}
& \sum_{k \in [n]: k \neq j} \left| z_k^{(T)} - z_k^{(T,-j)} \right|^2 \\
&\leq \sum_{k \in [n]: k \neq j} \left(2t^{-2} \sigma^2 |W_{kj}|^2 + 2t^{-2} \sigma^2 \left| [W^{(-j)}(z^{(T-1)} - z^{(T-1,-j)})]_k \right|^2 \right) \\
&\leq \sum_{k \in [n]} \left(2t^{-2} \sigma^2 |W_{kj}|^2 + 2t^{-2} \sigma^2 \left| [W^{(-j)}(z^{(T-1)} - z^{(T-1,-j)})]_k \right|^2 \right) \\
&= 2t^{-2} \sigma^2 \|W_j\|^2 + 2t^{-2} \sigma^2 \left\| W^{(-j)}(z^{(T-1)} - z^{(T-1,-j)}) \right\|^2 \\
&\leq 2t^{-2} \sigma^2 \|W_j\|^2 + 2t^{-2} \sigma^2 \left\| W^{(-j)} \right\|^2 \left\| z^{(T-1)} - z^{(T-1,-j)} \right\|^2 \\
&\leq 2t^{-2} \sigma^2 \|W_j\|^2 + 2t^{-2} \sigma^2 \|W\|^2 \left\| z^{(T-1)} - z^{(T-1,-j)} \right\|^2,
\end{aligned}$$

where in the last inequality, $\|W^{(-j)}\| \leq \|W\|$ due to that $W^{(-j)}$ is obtained from W by zeroing out

its j th row and column. On the other hand, $|z_j^{(T)} - z_j^{(T,-j)}| \leq 2$. Hence,

$$\begin{aligned} \|z^{(T)} - z^{(T,-j)}\|^2 &\leq 4 + \sum_{k \in [n]: k \neq j} |z_k^{(T)} - z_k^{(T,-j)}|^2 \\ &\leq 4 + 2t^{-2}\sigma^2 \|W_j\|^2 + 2t^{-2}\sigma^2 \|W\|^2 \|z^{(T-1)} - z^{(T-1,-j)}\|^2 \\ &\leq 4 + 2t^{-2}\sigma^2 \|W\|^2 + 2t^{-2}\sigma^2 \|W\|^2 \|z^{(T-1)} - z^{(T-1,-j)}\|^2, \end{aligned}$$

where in the last inequality we use a fact that the operator norm of matrix is greater or equal to the norm of each column. When $t \geq 2\sigma \|W\|$, we have $2t^{-2}\sigma^2 \|W\|^2 \leq 1/2$ and

$$\|z^{(T)} - z^{(T,-j)}\|^2 \leq \frac{9}{2} + \frac{1}{2} \|z^{(T-1)} - z^{(T-1,-j)}\|^2.$$

Note that $\|z^{(0)} - z^{(0,-j)}\|^2 = 0$, by mathematical induction, it is easy to verify $\|z^{(T)} - z^{(T,-j)}\|^2 \leq 9, \forall T \in \mathbb{N}$. Let $T \rightarrow \infty$, we have $\|z - z^{(-j)}\|^2 \leq 9$. \square

SM4 Proofs of Lemmas in Section 3.4

Proof of Lemma 3.12. From Corollary 2.4, we have

$$\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \leq \frac{8}{n} \sum_{j \in [n]} \mathbb{I}\{|[Y \hat{z}^{\text{MLE}}]_j| < \delta n\}.$$

For each $k = 0, 1, 2, \dots, \lceil n\epsilon/h \rceil$, let $z_{s_k} \in \mathbb{C}_{\leq 1}^n$ be the fixed point of $G(\cdot, s_k, 2\delta n)$. Then by Corollary 3.8, we have

$$\begin{aligned} &\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \\ &\leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{ \sigma |[W z_{s_k}]_j| > s_k - 4\delta n \right\} \right) + \frac{72h^2}{\delta^2 n^2} \mathbb{I}\{h > \delta\sqrt{n}\}. \end{aligned}$$

Since $2\delta n > 2\sigma \|W\|$, for each $k = 0, 1, 2, \dots, \lceil n\epsilon/h \rceil$, Proposition 3.11 can be applied, leading to

$$\begin{aligned} \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{ \sigma |[W z_{s_k}]_j| > s_k - 4\delta n \right\} &\leq \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{ \sigma |W_{j \cdot} z_{s_k}^{(-j)}| > s_k - 4\delta n - 3\sigma \|W\| \right\} \\ &\leq \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{ \sigma |W_{j \cdot} z_{s_k}^{(-j)}| > (1 - \epsilon)n - h - 4\delta n - 3\sigma \|W\| \right\} \\ &= \frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{ \sigma |W_{j \cdot} z_{s_k}^{(-j)}| > \left(1 - \epsilon - 4\delta - \frac{3\sigma \|W\|}{n}\right) n - h \right\}, \end{aligned}$$

where in the last inequality, we use $\min_{0 \leq k \leq \lceil n\epsilon/h \rceil} s_k \geq n - (n\epsilon/h + 1)h = (1 - \epsilon)n - h$. Hence, we have

$$\begin{aligned} &\ell_m(\hat{V}^{\text{BM},m}, \hat{z}^{\text{MLE}}) \\ &\leq 72 \sum_{0 \leq k \leq \lceil n\epsilon/h \rceil} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\left\{ \sigma |W_{j \cdot} z_{s_k}^{(-j)}| > \left(1 - \epsilon - 4\delta - \frac{3\sigma \|W\|}{n}\right) n - h \right\} \right) + \frac{72h^2}{\delta^2 n^2} \mathbb{I}\{h > \delta\sqrt{n}\}. \end{aligned}$$

\square

SM5 Auxiliary Lemmas and Proofs

The following lemma is a generalization of Lemma 11 of [16].

Lemma SM5.1. *Consider any $m \in \mathbb{N} \setminus \{1\}$. For any $V \in \mathcal{V}_m$ and any $z \in \mathbb{C}_1^n$, we have*

$$\frac{1}{n^2} \|V^H V - z z^H\|_F^2 \leq 2\ell_m(V, z).$$

Proof. Lemma 11 of [16] only considers the case where $m = n$. However, its proof holds for any $m \geq 2$, which we include here for completeness. By definition, we have

$$\ell_m(V, z) = 2 - \max_{a \in \mathbb{C}^n: \|a\|^2=1} \left(a^H \left(\frac{1}{n} \sum_{j=1}^n z_j V_j \right) + \left(\frac{1}{n} \sum_{j=1}^n z_j V_j \right)^H a \right) = 2 \left(1 - \left\| \frac{1}{n} \sum_{j=1}^n z_j V_j \right\| \right).$$

In addition, we have

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

Therefore, $n^{-2} \|V^H V - z z^H\|_F^2 \leq \ell_m(V, z) (2 - \frac{1}{2}\ell_m(V, z)) \leq 2\ell_m(V, z)$, and the proof is complete. \square

Proof of Lemma 2.1. We follow the proof of Lemma 12 of [38]. We first decompose V and z into orthogonal components:

$$V = a(z^*)^H + \sqrt{n}A \text{ and } z = bz^* + \sqrt{n}\beta, \quad (\text{SM5.1})$$

where $a \in \mathbb{C}^m$, $A \in \mathbb{C}^{m \times n}$, $b \in \mathbb{C}$, $\beta \in \mathbb{C}^n$ and $Az^* = 0$, $\beta^H z^* = 0$. Note the decomposition on V is always possible as $V = Vz^*(z^*)^H + V(I_n - z^*(z^*)^H)$ and $a = Vz^*$, $\sqrt{n}A = V(I_n - z^*(z^*)^H)$. By the definition of the loss ℓ_m in (2.2), there exists some $d \in \mathbb{C}^m$ such $\|d\| = 1$ and $\ell_m(V, z) = n^{-1} \|V - dz^H\|_F^2$. With the decomposition (SM5.1), it means

$$\begin{aligned} n\ell_m(V, z) &= \|V - dz^H\|_F^2 \\ &= \left\| (a(z^*)^H + \sqrt{n}A) - d(bz^* + \sqrt{n}\beta)^H \right\|_F^2 \\ &= \left\| (a - d\bar{b})(z^*)^H + \sqrt{n}(A - d\beta^H) \right\|_F^2 \\ &= \left\| (a - d\bar{b})(z^*)^H \right\|_F^2 + \left\| \sqrt{n}(A - d\beta^H) \right\|_F^2 \\ &= n \|a - d\bar{b}\|^2 + n \|A - d\beta^H\|_F^2. \end{aligned} \quad (\text{SM5.2})$$

where the third equation is due to the orthogonality $(A - d\beta^H)z^* = 0$. Then

$$\|A - d\beta^H\|_F \leq \sqrt{\ell_m(V, z)}. \quad (\text{SM5.3})$$

We also have

$$\begin{aligned}
\|VY^H - d(Yz)^H\|_F &= \|V(z^*(z^*)^H + \sigma W)^H - dz^H(z^*(z^*)^H + \sigma W)^H\|_F \\
&\leq \|(V - dz^H)z^*(z^*)^H\|_F + \|\sigma(V - dz^H)W\|_F \\
&\leq \|(a(z^*)^H - d\bar{b}(z^*)^H)z^*(z^*)^H\|_F + \sigma\|W\|\|V - dz^H\|_F \\
&\leq n\sqrt{n}\|a - d\bar{b}\| + \sigma\|W\|\sqrt{n}\sqrt{\ell_m(V, z)},
\end{aligned} \tag{SM5.4}$$

where the second inequality is due to the fact that $\|B_1 B_2\|_F \leq \|B_1\|_F \|B_2\|_{\text{op}}$ for any two matrices B_1, B_2 . If

$$\|a - d\bar{b}\| \leq 6\epsilon\|A - d\beta^H\|_F \tag{SM5.5}$$

holds, (SM5.4) and (SM5.3) leads to

$$\begin{aligned}
\ell_m(VY^H, Yz) &\leq \frac{1}{n}\|VY^H - d(Yz)^H\|_F^2 \\
&\leq \frac{1}{n}\left(6\epsilon n\sqrt{n}\|A - d\beta^H\|_F + \sigma\|W\|\sqrt{n}\sqrt{\ell_m(V, z)}\right)^2 \\
&\leq \frac{1}{n}\left(6\epsilon n\sqrt{n}\sqrt{\ell_m(V, z)} + \sigma\|W\|\sqrt{n}\sqrt{\ell_m(V, z)}\right)^2 \\
&= n^2\left(6\epsilon + \frac{\sigma\|W\|}{n}\right)^2 \ell_m(V, z),
\end{aligned}$$

which yields the desired result. The remaining proof is devoted to establishing (SM5.5).

Note that

$$\begin{aligned}
\ell_m(V, z^*) &= \min_{u \in \mathbb{C}^m: \|u\|=1} n^{-1} \|a(z^*)^H + \sqrt{n}A - u(z^*)^H\|_F^2 \\
&= \min_{u \in \mathbb{C}^m: \|u\|=1} n^{-1} \left(\|(a - u)(z^*)^H\|_F^2 + \|\sqrt{n}A\|_F^2 \right) \\
&= \min_{u \in \mathbb{C}^m: \|u\|=1} \|a - u\|^2 + \|A\|_F^2.
\end{aligned}$$

Since $\ell_m(V, z^*) \leq \epsilon^2 < 1/4$, we have $\|A\|_F^2 \leq \epsilon^2$, $\|a\| \neq 0$ and $\min_{u \in \mathbb{C}^m: \|u\|=1} \|a - u\|^2 = \|a - a/\|a\|\|^2 = (1 - \|a\|)^2$. Together with $1 = n^{-1}\|V\|_F^2 = n^{-1}\|a(z^*)^H\|_F^2 + n^{-1}\|\sqrt{n}A\|_F^2 = \|a\|^2 + \|A\|_F^2$, we have

$$\ell_m(V, z^*) = (1 - \|a\|)^2 + 1 - \|a\|^2 = 2 - 2\|a\|.$$

Then $\ell_m(V, z^*) \leq \epsilon^2$ leads to $1 \geq \|a\| \geq 1 - \epsilon^2/2$. Similarly for z , we have $\|\beta\|^2 \leq \epsilon^2$, $1 \geq |b| \geq 1 - \epsilon^2/2$ and $1 = |b|^2 + \|\beta\|^2$. Since $\epsilon < 1/2$, we have $\|a\| + |b| > 1$, and consequently $|\|a\| - |b|| \leq |\|a\| - |b||(\|a\| + |b|) = |\|a\|^2 - |b|^2|$. Since $\|a\|^2 + \|A\|_F^2 = |b|^2 + \|\beta\|^2$, we have $|\|a\|^2 - |b|^2| = |\|\beta\|^2 - \|A\|_F^2|$. Together with $\|A\|_F^2, \|\beta\|^2 \leq \epsilon^2$, we have

$$\begin{aligned}
|\|a\| - |b|| &\leq |\|\beta\|^2 - \|A\|_F^2| = |\|\beta\| - \|A\|_F|(\|\beta\| + \|A\|_F) \\
&\leq 2\epsilon|\|\beta\| - \|A\|_F| \leq 2\epsilon\|A - d\beta^H\|_F.
\end{aligned} \tag{SM5.6}$$

Note that

$$\begin{aligned}
\|a - d\bar{b}\| &= \left\| a - \frac{a}{\|a\|} |b| + \frac{a}{\|a\|} \frac{b}{|b|} \bar{b} - d\bar{b} \right\| \\
&\leq \left\| a - \frac{a}{\|a\|} |b| \right\| + \left\| \left(\frac{a}{\|a\|} \frac{b}{|b|} - d \right) \bar{b} \right\| \\
&= \| |a| - |b| \| + \left\| \frac{a}{\|a\|} \frac{b}{|b|} - d \right\| |b| \\
&\leq 2\epsilon \|A - d\beta^H\|_F + \left\| \frac{a}{\|a\|} \frac{b}{|b|} - d \right\|,
\end{aligned}$$

where in the last inequality we use $|b| \leq 1$. Hence, to establish (SM5.5), we only need to show

$$\left\| \frac{a}{\|a\|} \frac{b}{|b|} - d \right\| \leq 4\epsilon \|A - d\beta^H\|_F. \quad (\text{SM5.7})$$

To prove (SM5.7), define $d_0 = \frac{a}{\|a\|} \frac{b}{|b|} \in \mathbb{C}^m$. Then $\|d_0\| = 1$. Similar to (SM5.2), we have $\|V - d_0 z^H\|_F^2 = n \|a - d_0 \bar{b}\|^2 + n \|A - d_0 \beta^H\|_F^2$. By the definition of d , $\|V - d z^H\|_F^2 \leq \|V - d_0 z^H\|_F^2$, which leads to

$$\|a - d\bar{b}\|^2 + \|A - d\beta^H\|_F^2 \leq \|a - d_0 \bar{b}\|^2 + \|A - d_0 \beta^H\|_F^2.$$

Note that $d_0 \bar{b} = a \frac{|b|}{\|a\|}$ is proportional to a and $\|d_0 \bar{b}\| = \|d\bar{b}\| = |b|$. Let $\theta \in [0, \pi]$ be the angle between a and $d\bar{b}$ in \mathbb{C}^m . By the cosine formula of triangles, we have

$$\begin{aligned}
\|a - d\bar{b}\|^2 &= \|a\|^2 + \|d\bar{b}\|^2 - 2\|a\| \|d\bar{b}\| \cos(\theta) = \|a\|^2 + |b|^2 - 2\|a\| |b| \cos(\theta) \\
\|a - d_0 \bar{b}\|^2 &= \left\| a - a \frac{|b|}{\|a\|} \right\|^2 = \|a\|^2 + |b|^2 - 2\|a\| |b| \\
\text{and } \|d - d_0\|^2 &= \|d\|^2 + \|d_0\|^2 - 2\|d\| \|d_0\| \cos(\theta) = 2(1 - \cos(\theta)).
\end{aligned} \quad (\text{SM5.8})$$

Hence, $\|a - d\bar{b}\|^2 - \|a - d_0 \bar{b}\|^2 = 2\|a\| |b|(1 - \cos(\theta))$. By the triangle inequality, $\|A - d_0 \beta^H\|_F - \|A - d\beta^H\|_F \leq \|(d_0 - d)\beta^H\|_F = \|d_0 - d\| \|\beta\| \leq \epsilon \|d_0 - d\|$ where in the last inequality we use $\|\beta\| \leq \epsilon$. Then,

$$\begin{aligned}
2\|a\| |b|(1 - \cos(\theta)) &\leq \|A - d_0 \beta^H\|_F^2 - \|A - d\beta^H\|_F^2 \\
&= (\|A - d_0 \beta^H\|_F - \|A - d\beta^H\|_F) (\|A - d_0 \beta^H\|_F + \|A - d\beta^H\|_F) \\
&\leq \epsilon \|d_0 - d\| (\epsilon \|d_0 - d\| + 2\|A - d\beta^H\|_F).
\end{aligned}$$

By (SM5.8), it becomes $\|a\| |b| \|d_0 - d\|^2 \leq \epsilon \|d_0 - d\| (\epsilon \|d_0 - d\| + 2\|A - d\beta^H\|_F)$, which further leads to

$$(\epsilon^{-1} \|a\| |b| - \epsilon) \|d_0 - d\| \leq 2\|A - d\beta^H\|_F.$$

Since $\|a\|, |b| \geq 1 - \epsilon^2/2$, we have $\epsilon^{-1} \|a\| |b| - \epsilon \geq \epsilon^{-1}(1 - \epsilon^2/2)^2 - \epsilon \geq \epsilon^{-1}(1 - \epsilon^2) - \epsilon = \epsilon^{-1}(1 - 2\epsilon^2) > (2\epsilon)^{-1}$ where the last inequality is due to $\epsilon < 1/2$. Hence, $(2\epsilon)^{-1} \|d_0 - d\| \leq 2\|A - d\beta^H\|_F$, which establishes (SM5.7). The proof of the lemma is complete. \square