Benign landscapes of low-dimensional nonconvex relaxations for orthogonal synchronization on general graphs

(a pre-arXiv preprint)

Andrew D. McRae and Nicolas Boumal

Abstract

Orthogonal synchronization is the problem of estimating n elements $Z_1, \ldots, Z_n$ from the orthogonal group $\mathrm{O}(r)$ given some relative measurements $R_{ij} \approx Z_i Z_j^{-1}$. The least-squares formulation is nonconvex. To avoid its local minima, a Shor-type convex relaxation squares the dimension of the optimization problem from $O(n)$ to $O(n^2)$. Alternatively, Burer–Monteiro-type nonconvex relaxations offer generic guarantees at dimension $O(n^{3/2})$. Below that dimension, the problem instance matters. It has been observed in the robotics literature that nonconvex relaxations of dimension barely higher than the original seem sufficient for practical instances arising in simultaneous localization and mapping (SLAM). We partially explain this. This also has implications for Kuramoto oscillators.

Explicitly, we minimize the least-squares cost function in terms of estimators $Y_1, \ldots, Y_n$. Each Y_i is relaxed to the Stiefel manifold $\mathrm{St}(r, p)$ of $r \times p$ matrices with orthonormal rows. (This reduces to $\mathrm{O}(r)$ with $p = r$.) The available measurements implicitly define a (connected) graph G on n vertices. In the noiseless case, we show that second-order critical points are globally optimal as soon as $p \geq r + 2$ for all connected graphs G. (This implies that Kuramoto oscillators on $\mathrm{St}(r, p)$ synchronize for all $p \geq 2$.) This result is the best possible for general graphs; the previous best known result requires $2p \geq 3(r + 1)$. For $p > r + 2$, our result is robust to modest amounts of noise (depending on p and G). When local minima remain, they still achieve minimax-optimal error rates (within a constant depending on p).

Finally, we partially extend our noiseless landscape results to the complex case (unitary group), showing that there are no spurious local minima when $2p \geq 3r$.

1 Introduction and results

This paper examines the optimization landscape of a class of quadratically constrained quadratic problems (QCQPs) that arise from the *orthogonal group synchronization* problem. This widely-studied problem has many applications, including in simultaneous localization and mapping (SLAM) [1], cryo-electron microscopy (cryo-EM) [2], computer vision [3], and phase retrieval [4].

Mathematically, the orthogonal synchronization problem that we study is the following: Let $G = (V, E)$ be an undirected graph on the vertices $V = \{1, \ldots, n\}$ for some integer $n \geq 1$. Each vertex i is associated with an unknown orthogonal matrix $Z_i \in \mathrm{O}(r)$. We want to estimate $Z_1, \ldots, Z_n$ from (potentially noisy) measurements of the form $R_{ij} = Z_i Z_j^T + \Delta_{ij} \in \mathbf{R}^{r \times r}$ for each edge $(i, j) \in E$, where Δ_{ij} represents measurement error/noise. Since the measurements are relative, estimation can only be done up to a global orthogonal transformation.

1.1 Optimization problem setup and landscape results

A simple least-squares estimate of $Z_1, \ldots, Z_n$ can be obtained from the following optimization problem:

$$\min_{Y \in \mathrm{O}(r)^n} \sum_{(i,j) \in E} \|Y_i - R_{ij} Y_j\|_\mathrm{F}^2. \tag{1}$$

Although the cost function itself is convex in Y, the constraint set $\mathrm{O}(r)^n$ is nonconvex. In general, the problem has spurious local minima in which local search methods (such as gradient descent) can get stuck.

Owing to the orthogonality constraints, the above problem is equivalent to

$$\max_{Y \in \mathrm{O}(r)^n} \sum_{(i,j)\in E} \langle R_{ij}, Y_i Y_j^T \rangle,$$

where $\langle A, B\rangle = \mathrm{tr}(AB^T)$ is the Hilbert–Schmidt (or Frobenius) matrix inner product. To write this more compactly, let $C \in \mathbf{R}^{rn\times rn}$ denote the (incomplete) measurement matrix with blocks[1]

$$C_{ij} = \begin{cases} R_{ij} & (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

With the convention that $R_{ij} = R_{ji}^T$, C is a symmetric matrix. Then, we can rewrite the problem as

$$\max_{Y \in \mathbf{R}^{rn\times r}} \langle C, YY^T \rangle \text{ s.t. } Y_i Y_i^T = I_r, i = 1, \dots, n. \tag{2}$$

The matrix YY^T is positive semidefinite with rank at most r. The classical (Shor) convex relaxation consists in replacing YY^T with a positive semidefinite matrix X, ignoring the rank constraint:

$$\max_{\substack{X \in \mathbf{R}^{rn\times rn} \\ X \succeq 0}} \langle C, X \rangle \text{ s.t. } X_{ii} = I_r, i = 1, \dots, n. \tag{3}$$

Alternatively, we can allow Y to have $p \geq r$ columns. This effectively allows YY^T to have rank up to p, providing for a more gradual relaxation as we increase p. This yields the rank-p Burer–Monteiro relaxation:

$$\max_{Y \in \mathbf{R}^{rn\times p}} \langle C, YY^T \rangle \text{ s.t. } Y_i Y_i^T = I_r, i = 1, \dots, n. \tag{4}$$

We can view the rank $p \geq r$ as a hyperparameter that interpolates the original nonconvex problem (2) (in the case $p = r$) and the full SDP relaxation (3) (when $p \geq rn$).

This paper primarily considers the *optimization landscape* of the rank-relaxed nonconvex problem (4) for various values of p. In particular, we ask: how large does p need to be such that (4) has no spurious local optima? For general cost matrices C, we need $p = O(n)$ to guarantee such a benign landscape (resulting in $O(n^2)$ variables, which is the same order as the full SDP relaxation); with the assumption that C is "generic" (non-pathological), we can reduce this to $p = O(n^{1/2})$ ($O(n^{3/2})$ variables). See Section 1.2 for details. We will show that the particular structure of our choice of C allows us to choose $p = O(1)$, resulting in a similar $O(n)$ number of variables as the original problem (2).

This approach to orthogonal group synchronization has been studied before in, for example, [5]–[7] (see Section 2 for a more thorough literature review). These papers show that by iteratively increasing p, one will eventually arrive at a certifiably globally optimal solution to (3) and (4). However, these works do not address theoretically *how large* of a relaxation dimension p is required. Empirically, though, a very modest increase in dimension ($p = 5$ or 6 when $r = 3$) appears to be sufficient [7].

In this paper, we prove rigorously that this phenomenon occurs in general. In particular, we give conditions under which every *second-order critical point*[2] is a global optimum. We define such a point more precisely in Section 3, but it is, essentially, a point $Y \in \mathbf{R}^{rn\times p}$ where, subject to the constraints, the gradient at Y is zero and all eigenvalues of the Hessian are nonpositive.

We first consider the case with no measurement error (i.e., $R_{ij} = Z_i Z_j^T$). Clearly, a globally optimal solution to the original least-squares problem (1) is

$$Z \coloneqq \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix},$$

and we can trivially recover Z (up to a global orthogonal transformation) by traversing any spanning tree of G and recursively applying the measured relative differences. However, the *landscape* of (1), (2) still in general has spurious local minima, and we want to understand what rank relaxation does. Furthermore, the problem is closely connected to coupled oscillator dynamics (see Section 1.3).

[1]Throughout, indices into a matrix dimension of length rn refer to blocks of r rows or columns. Thus, for $C \in \mathbf{R}^{rn\times rn}$, C_{ij} refers to the (i,j)th $r \times r$ block; for $Y \in \mathbf{R}^{rn\times p}$, Y_i refers to the ith $r \times p$ block of Y, etc.

[2]This is a stronger statement: every local optimum is a second-order critical point.

Theorem 1. *If the measurements are exact, i.e., $R_{ij} = Z_i Z_j^T$ for all $(i, j) \in E$, then, if $p \geq r + 2$, any second-order critical point Y of (4) satisfies $YY^T = ZZ^T$. Equivalently, $Y = ZU$ for some $r \times p$ matrix U satisfying $UU^T = I_r$.*

We next consider the more general noisy case. We cannot expect perfect recovery of Z in this case. Let $\|A\|_{\ell_2}$ and $\|A\|_{\mathrm{F}}$ respectively denote the operator and Frobenius norms of a matrix A. The quality metric we will use for a candidate solution Y of (4) is the correlation $\langle ZZ^T, YY^T\rangle = \|Z^T Y\|_{\mathrm{F}}^2$. The maximum value this can take is $n^2 r$ (because $\|ZZ^T\|_{\ell_2} = n$, and $\operatorname{tr}(YY^T) = nr$), and this value is reached if and only if $YY^T = ZZ^T$ (as in Theorem 1).

Our results depend on how well G is connected. In particular, we use the *algebraic connectivity* of G, which is a spectral property of the graph. Let $L = L(G)$ be the unnormalized graph Laplacian of G, defined as $L = \operatorname{diag}(A\mathbf{1}) - A$, where A is the adjacency matrix of G. L is a positive semidefinite matrix with eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n = \|L\|_{\ell_2}$. Because G is connected, $\lambda_2 > 0$. λ_2 is often called the algebraic connectivity or Fiedler value of G.

Furthermore, we must quantify the measurement error (or noise). We denote by Δ the $rn \times rn$ matrix with the errors Δ_{ij} in the appropriate places for $(i, j) \in E$, setting $\Delta_{ij} = 0$ for $(i, j) \notin E$. Thus Δ is the portion of the cost matrix C that is due to measurement error. Our results are stated in terms of the operator norm $\|\Delta\|_{\ell_2}$.

We can now state our first noisy landscape result:

Theorem 2. *Suppose $p > r + 2$, and define*

$$C_p := \frac{2p + 3r - 4}{p - r - 2}.$$

Then any second-order critical point Y of (4) satisfies

$$\operatorname{rank}(Y) \leq r + 5C_p^2 \left(\frac{\|\Delta\|_{\ell_2}}{\lambda_2} \right)^2 nr.$$

If, furthermore, $p > r + 5C_p^2 \left(\frac{\|\Delta\|_{\ell_2}}{\lambda_2} \right)^2 nr$, then Y is a globally optimal solution to (4), and YY^T is an optimal solution to the SDP (3).

This result quantitatively bounds how large p needs to be so that (4) has a benign landscape and yields an exact solution to the full SDP relaxation. The bound depends on the effective signal-to-noise ratio $\|\Delta\|_{\ell_2}/\lambda_2$.

When we can ensure that $\operatorname{rank}(Y) = r$ exactly, we obtain a stronger result:

Theorem 3. *Suppose $p > r + 2$, and let C_p be as defined in Theorem 2. If*

$$\|\Delta\|_{\ell_2} < \frac{\lambda_2}{\sqrt{5} C_p \sqrt{nr}}, \tag{5}$$

then any second-order critical point Y of (4) satisfies the following:

- *Y is the unique solution to (4) up to a global orthogonal transformation.*
- *Y has rank r, and, if we write $Y = \widehat{Z} U$ for some $\widehat{Z} \in \mathbf{R}^{nr \times r}$ and $U \in \mathbf{R}^{r \times p}$ such that $UU^T = I_r$, $\widehat{Z}$ is the unique solution of (2) up to a global orthogonal transformation.*
- *YY^T is the unique solution to the SDP (3).*

First, note that if the measurement error Δ is small enough, $p = r + 3$ suffices[3] to obtain a benign landscape. Furthermore, the second-order critical points yield a global solution to the original problem (2), which does not (in general) have a benign landscape.

[3] The noiseless result Theorem 1 can be made robust to noise to obtain a similar landscape result when $p = r + 2$, but the required bound on $\|\Delta\|_{\ell_2}$ would be much worse. See Section 4.1 for details.

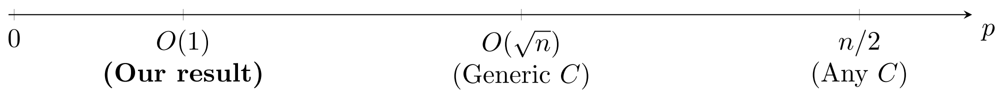


Figure 1: Minimum Burer–Monteiro factorization rank p required to obtain a benign optimization landscape for the relaxed $\mathrm{O}(1) \cong \mathbf{Z}_2$ synchronization problem (7) under various assumptions on the cost matrix C.

1.2 Why is a structured cost matrix necessary?

To obtain our results, it is essential to consider the particular *structure* of the orthogonal group estimation problem: specifically, the graph-based relative rotations $R_{ij} \approx Z_i Z_j^T$ for (i,j) the edges of a (connected) graph G.

To see this, consider what is known about the problem (4) for *general* cost matrices C. We examine the case $r = 1$ because it has been the most carefully studied; however, it is likely that the results cited in this section could easily be extended to $r > 1$. In this case, C is an arbitrary $n \times n$ symmetric matrix, and we consider the SDP

$$\max_{\substack{X \in \mathbf{R}^{n\times n} \\ X \succeq 0}} \ \langle C, X\rangle \text{ s.t. } \operatorname{diag}(X) = I_n \tag{6}$$

and its rank-p factorized version

$$\max_{Y \in \mathbf{R}^{n\times p}} \ \langle C, YY^T\rangle \text{ s.t. } \operatorname{diag}(YY^T) = I_n. \tag{7}$$

The SDP (6) is sometimes called the Max Cut SDP, as it is the standard SDP relaxation of that problem (where C is a graph Laplacian matrix).

The landscape of (7) has been well studied. For general C, it is known that $p > n/2$ guarantees that there are no spurious local minima [8]. For $p \leq n/2$ there exist matrices C such that there are spurious local minima [9]. Thus, in general, we must have p of order n, in which case the number of optimization variables is $O(n^2)$, which is comparable that that in the full SDP relaxation (6). Furthermore, for "generic" C (i.e., all C except in a set of zero Lebesgue measure), it is known that $p \gtrsim \sqrt{n}$ suffices to make the landscape benign [8], but for $p \lesssim \sqrt{n}$, there is a nonzero-measure set of cost matrices for which there are spurious local minima [10]. Thus even when C is generic, we still need $O(n^{3/2})$ optimization variables for a benign landscape.

Our results show that for the cost matrix C arising from the orthogonal group synchronization problem (with sufficiently small noise), $p = O(1)$ suffices to guarantee that (7) has a benign landscape. Therefore, $O(n)$ optimization variables suffice, which is comparable to the number in the original formulation (2).

Figure 1 illustrates the values of p required under various assumptions on C and highlights our results.

1.3 Implications for oscillator network synchronization

Another way to view the problem we have just described is *oscillator synchronization*. We briefly describe this connection here, but see, for example, [11], [12] for more detailed discussion and derivations.

Given a connected graph G defined as before, a simple version of the *Kuramoto* model for an oscillator network on G is the following: we have time-varying angles $\theta_1(t), \dots, \theta_n(t)$ associated with the n vertices, and these angles follow the dynamics model

$$\dot{\theta}_i = \sum_{j \in \mathcal{N}_i} \sin(\theta_j - \theta_i), i = 1, \dots, n, \tag{8}$$

where $\mathcal{N}_i$ is the set of neighbors of vertex i in G. One can easily check that these dynamics are the gradient flow for the following optimization problem:

$$\max_{\theta \in \mathbf{R}^n} \ \frac{1}{2} \sum_{i,j=1}^{n} A_{ij} \cos(\theta_i - \theta_j). \tag{9}$$

Clearly, the optima of (9) are the "synchronized" states $\theta_1 = \cdots = \theta_n \mod 2\pi$. Many papers have studied this model, particularly with the following question in mind: for which graphs G does the dynamical system (8) converge to a synchronized state as $t \to \infty$ for "generic" initial conditions (e.g., with probability 1 if $\theta_1(0), \ldots, \theta_n(0)$ are chosen uniformly at random over $[0, 2\pi)$)?

To connect this problem to our work, note that (9) is a reparametrization of the problem (4) in the case $r = 1, p = 2$ when $Z_i = 1$ for all i and the measurements are exact.[4] There are many examples (see, e.g., [14]) of connected graphs G under which (9) has spurious local optima (and, therefore, the gradient flow (8) can converge to a non-synchronized state even with generic initialization). This implies that (4) *does not* have a benign landscape for $r = 1, p = 2$.

For $r \geq 1$, consider (4) in the case that $Z_1 = \cdots = Z_n = I_r$.[5] Furthermore, assume that there is no measurement error,[6] so $R_{ij} = Z_i Z_j^T = I_r$ for all $(i, j) \in E$. Note, furthermore, that the feasible points $Y = [Y_i]_i$ lie in a product of *Stiefel manifolds* (we explore this connection in much more detail in the proofs of our main results). Let

$$\mathrm{St}(r, p) = \{U \in \mathbf{R}^{r\times p} : UU^T = I_r\}$$

denote the Stiefel manifold in $\mathbf{R}^{r\times p}$.

With these simplifications and notation, (4) becomes (within a factor of 2)

$$\max_{Y \in \mathrm{St}(r,p)^n} \frac{1}{2} \sum_{i,j=1}^{n} A_{ij} \operatorname{tr}(Y_i Y_j^T), \tag{10}$$

where A is the adjacency matrix of G. Again, the global optima are precisely the Y such that $Y_1 = \cdots = Y_n \in \mathrm{St}(r, p)$.

Any trajectory $Y(t)$ of the gradient flow on the constraint manifold satisfies the following system of differential equations:

$$\dot{Y}_i = \mathcal{P}_{T_{Y_i}} \left(\sum_{j \in \mathcal{N}_i} (Y_j - Y_i) \right), i = 1, \ldots, n. \tag{11}$$

T_U denotes the tangent space of $\mathrm{St}(r, p)$ at U (we will characterize this precisely in Section 4), and $\mathcal{P}_{T_U}$ is the (Euclidean) orthogonal projection onto T_U. This dynamics model is (a simple version of) the Kuramoto model for a network of Stiefel-manifold valued oscillators.

Once again, we ask the question: when does the Stiefel manifold valued Kuramoto oscillator network governed by (11) *synchronize* (i.e., converge to a synchronized state $Y_1 = \cdots = Y_n$ as $t \to \infty$ for generic initialization)? The remarkable results of [11], [12], [15] show that for certain manifolds, Kuramoto oscillator networks synchronize for *any* connected graph G. Specifically, the paper [15] shows that for all $p \geq 3$, oscillator networks on $\mathrm{St}(1, p)$ (equivalently on the d-sphere S^d for $d \geq 2$) always synchronize for connected G. This was extended in [11], [12] to show that $\mathrm{St}(r, p)$ oscillator networks synchronize if $2p \geq 3(r + 1)$.

In this paper, we prove that $\mathrm{St}(r, p)$-valued oscillator networks synchronize under the weaker condition $p \geq r + 2$. This is a simple corollary of the noiseless orthogonal group synchronization result Theorem 1.

Corollary 1. *For any connected graph G, the* $\mathrm{St}(r, p)$*-valued Kuramoto oscillator network on G synchronizes if $p \geq r + 2$.*

To prove this from Theorem 1, note that the theorem implies that at every non-optimal critical point of (10), the Hessian has at least one strictly positive[7] eigenvalue. This implies that gradient flow reaches a global optimum for all initializations except possibly on a set of measure zero (see, e.g., [12]).

1.4 The complex case

We can extend the previous (real) orthogonal matrix estimation problem to the complex case. Here, we seek to estimate *unitary* matrices $Z_1, \ldots, Z_n \in \mathrm{U}(r)$ given measurements of the form $R_{ij} = Z_i Z_j^* + \Delta_{ij}$.

[4] Explicitly, $\theta_i \mapsto Y_i = [\cos(\theta_i), \sin(\theta_i)]$ so that $\cos(\theta_i - \theta_j) = Y_i Y_j^T$. In this case, the cost matrix C is equal to the adjacency matrix A of G. The differential of the change of variable is surjective, hence it does not introduce spurious critical points [13, Proposition 9.6]; thus the landscapes in θ and in Y are qualitatively the same.

[5] This is without loss of generality as we can always smoothly change variables to bring the ground truth to this position without affecting the landscape: see Section 4 for details.

[6] The simple connection to oscillators outlined in this section is most meaningful in the noiseless case.

[7] The positive eigenvalue is important because we have phrased this as a *maximization* problem. In the more common minimization formulation (which we will use in our analysis), we want a strictly *negative* eigenvalue.

We form our cost matrix $C \in \mathbf{C}^{rn\times rn}$ (C is now *Hermitian*, i.e., $C = C^*$) and consider the relationships between the original unitary group least-squares problem

$$\max_{Y\in\mathbf{C}^{rn\times r}} \langle C, YY^*\rangle \text{ s.t. } Y_iY_i^* = I_r, i = 1,\dots,n, \tag{12}$$

the SDP relaxation

$$\max_{\substack{X\in\mathbf{C}^{rn\times rn}\\ X\succeq 0}} \langle C, X\rangle \text{ s.t. } X_{ii} = I_r, i = 1,\dots,n, \tag{13}$$

and the rank-p relaxation

$$\max_{Y\in\mathbf{C}^{rn\times p}} \langle C, YY^*\rangle \text{ s.t. } Y_iY_i^* = I_r, i = 1,\dots,n. \tag{14}$$

For the oscillator point of view, we denote the complex $r \times p$ Stiefel manifold by

$$\mathrm{St}(r, p, \mathbf{C}) = \{U \in \mathbf{C}^{r\times p} : UU^* = I_r\}.$$

For simplicity, we only consider the noiseless landscape, and we obtain the following result (which, again, has implications for Kuramoto oscillators):

Theorem 4. *Suppose G is connected, and $2p \geq 3r$. Then any second-order critical point Y of (14) satisfies $YY^* = ZZ^*$. Consequently, the $\mathrm{St}(r, p, \mathbf{C})$-valued Kuramoto oscillator network on G synchronizes.*

Due to the 3/2 factor and certain similarities in the proof, this can also be seen as a complex adaptation of the result in [12]. Curiously, the innovations that allow us to improve that previous result in the real case do not easily carry over to the complex case.

As discussed further in Section 2.3, our results for the real case show that we obtain a benign landscape (and all connected oscillator networks synchronize) as soon as the real Stiefel manifold $\mathrm{St}(r, p)$ is *simply connected*. Interestingly, in the complex case, the Stiefel manifold $\mathrm{St}(r, p, \mathbf{C})$ is simply connected as soon as $p \geq r + 1$. We conjecture that this is the correct condition in Theorem 4, but we have been unable to prove it, and it is unclear how to test it empirically.[8]

1.5 Discussion of small-r synchronization conditions

It is interesting to consider the implications of the noiseless landscape results Theorems 1 and 4 and Corollary 1 for small values of the matrix dimension r, as many of these groups are well-known. See Table 1 for a summary.

The group $\mathrm{SO}(2)$, which is one of the two connected components of $\mathrm{O}(2)$, is isomorphic to the circle and thus $\mathrm{U}(1)$, and therefore we can relax estimation over $\mathrm{SO}(2)$ to synchronization of the *complex* Stiefel manifold $\mathrm{St}(2, 1, \mathbf{C}) = S^3$. However, in our framework, to estimate elements of the full group $\mathrm{O}(2)$, we must relax to $\mathrm{St}(4, 2)$, which is a 5-dimensional manifold.

A similar but more complex relationship exists between $\mathrm{U}(2)$ and $\mathrm{SO}(3) \subset \mathrm{O}(3)$. It is well-known that there is a surjective group homomorphism from $\mathrm{SU}(2)$ to $\mathrm{SO}(3)$, where $\mathrm{SU}(2)$ is the subgroup of $\mathrm{U}(2)$ of matrices with determinant 1. Hence, one can synchronize elements of $\mathrm{SO}(3)$ via synchronization on $\mathrm{SU}(2) \subset \mathrm{U}(2)$, which, in turn, can be relaxed to $\mathrm{St}(3, 2, \mathbf{C})$. Again, to synchronize on $\mathrm{O}(3)$, our results require relaxing to $\mathrm{St}(5, 3)$, which has larger dimension than $\mathrm{St}(3, 2, \mathbf{C})$.

1.6 Additional results

Our analysis yields several additional results that, though unrelated to our primary focus on optimization landscapes, may be of independent interest.

[8] The standard counterexamples to benign landscape/synchronization results are cycle graphs (indeed, the paper [16] uses cycle graphs to show that networks of oscillators taking values in a non–simply-connected manifold do not synchronize in general). We do not expect this counterexample to work in the complex case when $p \geq r + 1$, because a cycle graph corresponds geometrically to the unit circle, that is, $\mathrm{U}(1)$, and our result shows that, if $r = 1$, $p \geq 2 = r + 1$ suffices in the complex case. It is not clear how to construct a higher-dimensional equivalent as a candidate counterexample to our conjecture. For example, intuition from homotopy theory would suggest that we try a graph corresponding geometrically to a higher-dimensional sphere.

Group	Field	r	Min. p req.	Orig. dim.	Min. Stiefel dim.
$\mathrm{O}(1) = \mathbf{Z}_2$	$\mathbf{R}$	1	3	0	2
$\mathrm{U}(1) = S^1 = \mathrm{SO}(2)$	$\mathbf{C}$	1	2	1	3
$\mathrm{O}(2)$	$\mathbf{R}$	2	4	1	5
$\mathrm{U}(2)$	$\mathbf{C}$	2	3	4	8
$\mathrm{O}(3)$	$\mathbf{R}$	3	5	3	9
$\mathrm{U}(3)$	$\mathbf{C}$	3	5 (4^*)	9	21 (15^*)

Table 1: Properties of the groups $\mathrm{O}(r)$ and $\mathrm{U}(r)$ for $r \leq 3$ and the relaxations required to guaranteed synchronization. In the last row, * indicates the result of the conjectured $p \geq r + 1$ condition for the complex case.

1.6.1 Error bounds for all second-order critical points

Even if the conditions of Theorems 2 and 3 are not fully satisfied, we can still obtain useful error bounds for all second-order critical points:

Theorem 5. *If*

$$p > r + 2,$$

then any second-order critical point Y of (4) satisfies

$$\langle ZZ^T, YY^T \rangle \geq \left[1 - C_p^2 \frac{\|\Delta\|_{\ell_2}^2}{\lambda_2^2} \right] n^2 r,$$

where C_p is defined in the statement of Theorem 3.

Comparable error bounds have been derived previously for the eigenvector method (see Section 2).

1.6.2 Consequences for the SDP relaxation

Our results have interesting implications for the full SDP relaxation (3). First, we have an exactness result:

Corollary 2. *If*

$$\|\Delta\|_{\ell_2} < \frac{\lambda_2}{2\sqrt{5}\sqrt{nr}},$$

then

- *The SDP relaxation (3) has a unique solution $\widehat{X}$, and $\operatorname{rank}(\widehat{X}) = r$.*
- *The original unrelaxed problem (2) has a unique (up to global orthogonal transformation) solution $\widehat{Z} \in \mathbf{R}^{nr \times r}$.*
- *$\widehat{X} = \widehat{Z}\widehat{Z}^T$.*

Next, we have a general error bound for the SDP relaxation that applies even if the exactness result above does not:

Corollary 3. *Any solution X to (3) satisfies*

$$\langle X, ZZ^T \rangle \geq \left(1 - \frac{4\|\Delta\|_{\ell_2}^2}{\lambda_2^2} \right) n^2 r.$$

These corollaries are proved by applying Theorems 3 and 5 in the limit $p \to \infty$; see Section 4.3 for details.

These corollaries can be combined with Theorems 2 and 3 (which give conditions under which a second-order critical point Y yields an SDP solution) to give some modest improvements, but we do not develop this here.

2 Related work

The literature on orthogonal synchronization is vast, appearing under various names in multiple communities such as robotics, image processing, signal processing, and dynamical systems. We do not attempt to cover this entire literature here, but see [17], [18] for partial surveys.

Many of the tools we use in our analysis have been used in previous papers. We will point this out in our analysis as we introduce these tools.

2.1 Rank relaxation for synchronization

Low-rank factorization of SDPs (which, in our case, correspond to partial rank relaxations of the synchronization problem) have a long history. It is often called Burer–Monteiro factorization after the pioneering work of those authors (e.g., [19], [20]).

The report [5], along with the more general results in [8], provides a theoretical framework for analyzing Burer–Monteiro factorizations (like ours) of SDPs with (block-)diagonal constraints. The papers [6] and [7] propose and develop fast algorithms (generalized to the special Euclidean group case in [6]), showing that the "Riemannian staircase" approach (iteratively increasing the relaxation rank; p in our notation) proposed in [5] will provide an exact solution to the SDP relaxation, but they do nott specify what relaxation rank suffices. Our results provide an upper bound (optimal in the noiseless case) on how much such algorithms must relax the rank constraint.

As far as we are aware, *landscape* results similar to ours have previously only been proved in the complete measurement graph case. For this case, the papers [21], [22] provide error bounds and benign landscape results for rank-relaxed optimization like (4). Specifically, the paper [21] studies rank-2 relaxations for $\mathbf{Z}_2$ ($\mathrm{O}(1)$) synchronization.

The paper [22] analyzes the landscape of (4) in the same general $\mathrm{O}(r)$ case that we do and is thus the most comparable work to ours. We can directly compare our Theorem 3 with [22, Theorem 3]. In the complete-graph case, $\lambda_2 = n - 1$, so the condition (5) of Theorem 3 becomes $\|\Delta\|_{\ell_2} \lesssim_{p,r} \sqrt{n}$. This prior result [22, Theorem 3], in the adversarial-noise case (i.e., only assuming a bound on $\|\Delta\|_{\ell_2}$), has a comparable requirement. With additional assumptions on Δ, the prior result improves this to $\|\Delta\|_{\ell_2} \lesssim n^{3/4}$, but the techniques used do not easily carry over to our general-graph case. In addition, the paper [22] requires $p > 2r$ for all its results.

To the best of our knowledge, the partial benign landscape result of Theorem 2 (which gives conditions under which (4) has a benign landscape and yields an exact solution to the SDP even when the solution is not necessarily rank-r) is a new concept even in the complete-graph case.

In a different vein, [23] provides a general bound on how well rank-p Burer–Monteiro factorizations for $\mathrm{O}(r)$ optimization can approximate the full SDP relaxation in terms of *objective function value*. Their results bound the approximation error by a term proportional to r/p. Our results show that, in some cases, we obtain perfect approximation for p only slightly larger than r.

An interesting parallel work on a different type of Burer–Monteiro factorization is [24]. This paper shows that if the objective function is *strongly convex* in a semidefinite matrix parameter (and there are no other constraints), the factorized landscape is benign if the factorization rank is larger than that of the unfactorized problem's true solution (within a multiplicative factor depending on the objective function's condition number). The setting is quite different from ours (which has a linear objective and linear constraints), so the results are not directly comparable.

2.2 The spectral approach and previous error bounds

Our work follows a large body of prior results that use spectral properties of the measurement graph. In particular, these results use the eigenvalues and eigenvectors of the graph Laplacian matrix L (or, in some cases, the adjacency matrix A) of the graph G. A particularly important quantity is the *graph connection Laplacian* matrix (defined in [25] for different purposes) which can be formed directly from our observations R_{ij} (this is precisely the matrix $\widehat{L}$ in our notation—see Section 3 for the definition).

The *eigenvector method* was introduced in [26] for the purpose of rotation synchronization; this method directly uses the eigenvectors of the graph connection Laplacian (or, in the original paper, the adjacency matrix equivalent). Closely related to this, the paper [27] studies the relationship between the eigenvalues/vectors of the graph connection Laplacian and the optimal objective function value of (2).

The paper [28], by analyzing eigenvalues of $\widehat{L}$, provides conditions under which the SDP relaxation gives exact recover for $\mathbf{Z}_2$ (i.e., $\mathrm{O}(1)$) synchronization on an Erdős–Rényi graph. The paper [29] considers a robust version of the SDP relaxation that uses the sum of absolute errors rather than least-squares. They provide conditions for exact recovery for an Erdős–Rényi random graph and sparse errors.

The papers [4], [30], [31] and [32, Chapter 6] contain a variety of error bounds for the eigenvector method in terms of the graph Fiedler value $\lambda_2(L)$ and various norms of the measurement error matrix Δ. [31] furthermore extends this to SE(r) synchronization. These results are the most comparable to our error bounds, because they apply to general graphs and measurement errors. The bound in [31], is, within constants, identical to our bounds in Theorem 5 and Corollary 3.

The papers [33]–[35] show that the SDP relaxation, the (rounded) eigenvector method, and generalized power method (eigenvector method followed by iterative refinement) all achieve asymptotically minimax-optimal error with Gaussian noise on Erdős–Rényi random graphs (interestingly, [35], like the older paper [25], analyzes the adjacency matrix leading eigenvector). These results agree with our error bounds within constants, though our results apply to much more general situations.

2.3 Oscillator synchronization literature

Oscillator synchronization is a large field of research. See [36] for a recent broad survey. Our work touches the small subset corresponding to simplified Kuramoto oscillators.

As discussed in Section 1.3, the classical and most well-studied Kuramoto oscillator is that of angular synchronization (i.e., on the unit circle S^1). One line of research studies *which connected graphs* synchronize on S^1. This includes deterministic guarantees based on the density of a graph [14], [37]–[41] and high-probability properties of random graphs [38], [42]–[44]. For example, [41] shows that every graph synchronizes in which every vertex is connected to at least 3/4 of the other vertices. [43] gives more general deterministic conditions based on spectral expander properties and shows that, asymptotically, Erdős–Rényi graphs that are connected also synchronize.

A complementary line of research studies *for which manifolds* do *all* connected oscillator networks synchronize. It was shown in [15], [45] that Kuramoto networks on the sphere S^n synchronize for any $n \geq 2$. This corresponds to the Stiefel manifold for $r = 1$ and $p \geq 3$.

More generally, the papers [16], [46] prove that for any non-simply-connected manifold embedded in Euclidean space, there are connected oscillator networks that do not synchronize (i.e., have spurious local minima). The real Stiefel manifold is simply connected if and only if $p \geq r + 2$ (see, e.g., [47, Example 4.53]). The paper [45] conjectures, based on simulations, that in this case all connected oscillator networks synchronize. Our Corollary 1 proves this for the first time. The previous best result proved this for $2p \geq 3(r + 1)$ [11], [12]. In the case $r = 1$ (where the Stiefel manifold is a sphere), our result and these previous results coincide with each other and with [15], [45]. The complex Stiefel manifold is simply connected if and only if $p \geq r + 1$ (again, see [47, Example 4.53]), so we conjecture that $p \geq r + 1$ is in fact sufficient. Our result Theorem 4 is weaker but agrees in the cases $r = 1, 2$.

3 Key mathematical tools

In this section, we make precise and fill out the mathematical framework for our analysis and results. We only consider the real case in this section; we make the necessary adjustments for the complex case later.

3.1 Graph Laplacian formulation

In each optimization problem we consider, the orthogonality/block diagonal constraint ensures that the $r \times r$ diagonal blocks of the cost matrix C have no effect. We therefore replace C by another matrix that will be more convenient for analysis.

Let $A \in \mathbf{R}^{n \times n}$ be the symmetric matrix with $a_{ij} = \mathbf{1}_{\{(i,j)\in\Omega\}}$. Let L be the graph Laplacian matrix defined by

$$L_{ij} = \begin{cases} \sum_{k \neq i} a_{ik} & i = j \\ -a_{ij} & i \neq j. \end{cases}$$

It is well-known that L is a positive semidefinite (PSD) matrix whose smallest eigenvalue is $\lambda_1(L) = 0$ with corresponding eigenvector $v_1 = \mathbf{1}_n$. The measurement graph is connected if and only if the second-smallest eigenvalue $\lambda_2 = \lambda_2(L) > 0$.

Recall that $Z_1, \dots, Z_n \in \mathrm{O}(r)$ are the ground-truth matrices that we want to estimate. Let

$$D = D(Z) := \begin{bmatrix} Z_1 & & & \\ & Z_2 & & \\ & & \ddots & \\ & & & Z_n \end{bmatrix}.$$

Let $L_Z := D(L \otimes I_r)D^T$, where $\otimes$ is the Kronecker product. Note that the eigenvalues of L_Z are identical to those of L (with multiplicities multiplied by r), and, if G is connected, the r-dimensional subspace corresponding to $\lambda_1(L) = 0$ is precisely the span of the columns of Z.

Let $\widehat{L} = L_Z - \Delta$, where Δ is the symmetric matrix containing the measurement noise blocks Δ_{ij} (with $\Delta_{ij} = 0$ if $(i,j) \notin \Omega$). $\widehat{L}$ was introduced in [25] under the name *graph connection Laplacian.*

Note that $C_{ij} = -\widehat{L}_{ij}$ for $i \neq j$, and the diagonal blocks of the cost matrix have no effect on the optimization landscape of our problems (3) and (4) due to the block diagonal constraint. Therefore, the SDP (3) has the exact same landscape in the variable X as

$$\min_{X \succeq 0} \langle \widehat{L}, X \rangle \text{ s.t. } X_{ii} = I_r, i = 1, \dots, n, \tag{15}$$

and, similarly, the (relaxed) nonconvex problem (4) has the same landscape as

$$\min_{Y \in \mathbf{R}^{rn \times p}} \langle \widehat{L}, YY^T \rangle \text{ s.t. } Y_i Y_i^T = I_r, i = 1, \dots, n. \tag{16}$$

From now on, we will use this formulation.

3.2 Manifold of feasible points

Optimization problems of the form (16) have been well-studied. See, for example, [5], [8] for an overview. We summarize the relevant facts in this section.

First, note that the constraints in (15) and (16) apply to the $r \times r$ diagonal blocks of a matrix. To simplify our notation, we denote the symmetric block-diagonal projection $\mathrm{SBD}\colon \mathbf{R}^{rn \times rn} \to \mathbf{R}^{rn \times rn}$ by

$$\mathrm{SBD}(X)_{ij} = \begin{cases} \frac{X_{ii} + X_{ii}^T}{2} & i = j \\ 0 & i \neq j. \end{cases}$$

We can then write the semidefinite problem (15) more compactly as

$$\min_{X \succeq 0} \langle \widehat{L}, X \rangle \text{ s.t. } \mathrm{SBD}(X) = I_{rn}. \tag{17}$$

Similarly, we can write (16) as

$$\min_{Y \in \mathbf{R}^{rn \times p}} \langle \widehat{L}, YY^T \rangle \text{ s.t. } \mathrm{SBD}(YY^T) = I_{rn}. \tag{18}$$

The *symmetrizing* aspect of the projection operator SBD is, for now, completely redundant but will become useful as we continue.

The feasible points of (18) form a smooth submanifold $\mathcal{M}$ of $\mathbf{R}^{nr \times p}$ (more precisely, $\mathcal{M}$ is a product of Stiefel manifolds). The most important object for us to understand is the *tangent space* T_Y at a point $Y \in \mathcal{M}$. The tangent space consists of the directions in which the constraint operator $Y \mapsto \mathrm{SBD}(YY^T)$ has zero derivative:

$$T_Y = \{\dot{Y} \in \mathbf{R}^{nr \times p} : \mathrm{SBD}(Y\dot{Y}^T + \dot{Y}Y^T) = 0\}.$$

If we write

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \text{ and } \dot{Y} = \begin{bmatrix} \dot{Y}_1 \\ \dot{Y}_2 \\ \vdots \\ \dot{Y}_n \end{bmatrix},$$

then $\dot{Y} \in T_Y(\mathcal{M})$ if and only if $\dot{Y}_i Y_i^T + Y_i \dot{Y}_i^T = 0 \in \mathbf{R}^{r \times r}$ for all $i = 1, \dots, n$.

The orthogonal projection of an arbitrary $W \in \mathbf{R}^{nr\times p}$ onto T_Y is given by

$$\mathcal{P}_{T_Y}(W) = W - \mathrm{SBD}(WY^T)Y.$$

For a matrix Y to be a local minimum of (18), it must satisfy two necessary first- and second-order criticality conditions (see [8, Definition 2.3]). Let $S(Y) := \widehat{L} - \mathrm{SBD}(\widehat{L}YY^T)$.

- First order condition: $S(Y)Y = 0$. This is proportional to the T_Y-projected gradient of the objective function.
- Second order condition: for all $\dot{Y} \in T_Y$, $\langle S(Y)\dot{Y}, \dot{Y}\rangle \geq 0$. This says that the Riemannian Hessian of the objective function is positive semidefinite.

A point Y that satisfies the first condition is a *first-order critical point*. A point Y that satisfies both conditions is a *second-order critical point*. Note that the second-order criticality of a feasible point Y is independent of whether we write the problem as (4) or (18) (in fact, $S(Y)$ is unchanged if we replace $\widehat{L}$ by $-C$).

3.3 Dual certificates and global optimality

A natural question that remains is the following: how do we know if a candidate solution Y to (18) is, in fact, globally optimal? This is a well-studied problem (again, see [8] for further discussion and references), and here we list some facts that will be useful throughout our analysis.

To study optimality conditions for (18), it is useful to consider the full SDP relaxation (17); optimality conditions of SDPs are very well-understood. To show that a feasible point X of (17) is optimal, it suffices to provide a *dual certificate*. For (17), a dual certificate is a matrix S of the form

$$S = \widehat{L} - \mathrm{SBD}(\Lambda)$$

for some $rn \times rn$ matrix Λ with the conditions that (a) $S \succeq 0$ and (b) $\langle S, X\rangle = 0$ (or $SX = 0$, which is equivalent because $S, X \succeq 0$). If such a matrix S exists, we have, for any feasible point $X' \succeq 0$,

$$\begin{aligned}
\langle \widehat{L}, X'\rangle - \langle \widehat{L}, X\rangle &= \langle \widehat{L}, X' - X\rangle \\
&= \langle S, X' - X\rangle + \langle \mathrm{SBD}(\Lambda), X' - X\rangle \\
&= \langle S, X'\rangle - \underbrace{\langle S, X\rangle}_{=0} + \langle \Lambda, \underbrace{\mathrm{SBD}(X' - X)}_{=0}\rangle \\
&= \langle S, X'\rangle \\
&\geq 0.
\end{aligned}$$

Thus X is optimal. Furthermore, if X' is also optimal (i.e., equality holds above), we must have $SX' = 0$; we will use this fact later.

To show that Y is a globally optimal solution to (18), it clearly suffices to show that YY^T is an optimal solution to (17). Thus we want to find a dual certificate proving optimality of YY^T. A natural candidate is the matrix $S(Y) = \widehat{L} - \mathrm{SBD}(\widehat{L}YY^T)$ presented in Section 3.2. First-order criticality implies that $S(Y)YY^T = 0$. Thus it remains to show that $S(Y) \succeq 0$. We will do this with the following well-known but remarkable result (see, e.g., [8, Proposition 3.1]):

Lemma 1. *If Y is a second-order critical point of (18) that is rank-deficient (i.e., $\mathrm{rank}(Y) < p$), then $S(Y) \succeq 0$, and consequently YY^T is an optimal solution to the SDP (17), and Y is a globally optimal solution to (18).*

This can be proved by noting that if u is any unit-norm vector in the null space of Y, then, for any $z \in \mathbf{R}^{rn}$, $\dot{Y}(z) := zu^T \in T_Y$, and second-order criticality implies

$$\langle S(Y)z, z\rangle = \langle S(Y)\dot{Y}(z), \dot{Y}(z)\rangle \geq 0.$$

Thus the key step in showing that Y is globally optimal is to show that it is *rank-deficient*.

4 Proofs (real case)

We (again) denote by $\|A\|_{\ell_2}$, $\|A\|_{\mathrm{F}}$, and $\|A\|_*$ the operator, Frobenius, and nuclear norms of a matrix A.

Our proof strategy is as follows: first, we show that all second-order critical points satisfy certain error bounds (Theorems 1 and 5). In particular, we combine the PSDness of the Riemannian Hessian (i.e., $\langle S(Y)\dot{Y}, \dot{Y}\rangle \geq 0$ for all $\dot{Y} \in T_Y$) with a *random* choice of tangent vector $\dot{Y}$. Choosing $\dot{Y}$ randomly and taking an expectation simplifies many computations and circumvents some of the difficulty of choosing a good descent direction $\dot{Y}$.

Second, to prove our landscape results Theorems 2 and 3, the critical step is to show that second-order critical points have low rank (after which we apply Lemma 1). To do this, we show (using the error bounds previously derived) that the matrix $S(Y)$ has *high* rank and then apply the first-order criticality condition $S(Y)Y = 0$.

To simplify the notation in our proofs, we will, without loss of generality, assume that $Z_i = I_r$ for all $i = 1, \ldots, n$. Therefore, $L_Z = L \otimes I_r$. To see why we can do this, recall that $D = D(Z)$ is the orthogonal matrix whose $r \times r$ diagonal blocks are the Z_i's. Let $F(Y) = \langle \widehat{L}, YY^T\rangle$ be the objective function of (18). Let $\overline{F}(\overline{Y}) = F(D\overline{Y}) = \langle D^T\widehat{L}D, \overline{Y}\overline{Y}^T\rangle$ be the composition of F with the map $\pi \colon \mathcal{M} \to \mathcal{M}$ given by $\overline{Y} \mapsto D\overline{Y}$.

Note that $D^T\widehat{L}D = L \otimes I_r - \overline{\Delta}$, where $\overline{\Delta} \coloneqq D^T\Delta D$. The map π is clearly a Riemannian isometry on $\mathcal{M}$, so [13, Proposition 9.6] implies that Y is, respectively, a first-order critical point, second-order critical point, or global optimum of (18) if and only if $\overline{Y} = \pi^{-1}(Y) = D^TY$ is, respectively, a first-order critical point, second-order critical point, or global optimum of

$$\min_{\overline{Y} \in \mathbf{R}^{rn \times p}} \langle L \otimes I_r - \overline{\Delta}, \overline{Y}\overline{Y}^T\rangle \text{ s.t. } \mathrm{SBD}(\overline{Y}\overline{Y}^T) = I_{nr}.$$

Note furthermore that $\|\overline{\Delta}\|_{\ell_2} = \|\Delta\|_{\ell_2}$.

Now, assuming every $Z_i = I_r$, we analyze critical points of (18). Recall from Section 3 that, setting $S(Y) = \widehat{L} - \mathrm{SBD}(\widehat{L}YY^T)$, a feasible point Y is first- and second-order critical if $S(Y)Y = 0$ and, for every $\dot{Y} \in T_Y$, $\langle S(Y), \dot{Y}\dot{Y}^T\rangle \geq 0$. From here on, assume Y is such a point.

Recall that $\dot{Y} \in T_Y$ if and only if, for each i, $\dot{Y}_iY_i^T + Y_i\dot{Y}_i^T = 0$ for $i = 1, \ldots, n$. In other words, each $Y_i\dot{Y}_i^T$ must be skew-symmetric. This is true if and only we can write

$$\dot{Y}_i = \Gamma_i(I - Y_i^TY_i) + S_iY_i$$

for some $r \times p$ matrix Γ_i and some *skew-symmetric* $r \times r$ matrix S_i. The first term is the row-wise orthogonal projection of Γ_i onto $\mathrm{null}(Y_i)$. If we choose $S_i = \Gamma_iY_i^T - Y_i\Gamma_i^T$, we obtain

$$\dot{Y}_i = \Gamma_i - Y_i\Gamma_i^TY_i = Y_i(Y_i^T\Gamma_i - \Gamma_i^TY_i).$$

In fact, this last formulation covers all possible $\dot{Y}_i$, because any $r \times r$ skew-symmetric matrix S_i can be written in the given form, and the components of Γ_i in $\mathrm{range}(Y_i^T)$ and $\mathrm{null}(Y_i)$ can be chosen independently.

We choose a common $\Gamma_i = \Gamma$, where Γ is a random $r \times p$ matrix whose entries are i.i.d. standard normal random variables. This results in a random $\dot{Y} \in T_Y$. Because the second-order criticality inequality holds for *any* $\dot{Y} \in T_Y$, we can take an expectation to obtain

$$\langle S(Y), \mathbf{E}\,\dot{Y}\dot{Y}^T\rangle = \mathbf{E}\langle S(Y), \dot{Y}\dot{Y}^T\rangle \geq 0.$$

We can calculate

$$\begin{aligned}
\mathbf{E}\,\dot{Y}_i\dot{Y}_j^T &= \mathbf{E}(\Gamma - Y_i\Gamma^TY_i)(\Gamma - Y_j\Gamma^TY_j)^T \\
&= \mathbf{E}\big(\Gamma\Gamma^T - Y_i\Gamma^TY_i\Gamma^T - \Gamma Y_j^T\Gamma Y_j^T + Y_i\Gamma^TY_iY_j^T\Gamma Y_j^T\big) \\
&= (p-2)I_r + \mathrm{tr}(Y_iY_j^T)Y_iY_j^T.
\end{aligned} \tag{19}$$

We have used the facts that (1) for any $r \times p$ matrix U, $\mathbf{E}\,\Gamma^TU\Gamma^T = U^T$, and (2) for any $r \times r$ matrix B, $\mathbf{E}\,\Gamma^TB\Gamma = \mathrm{tr}(B)I_p$. When $i = j$, we have $\mathbf{E}\,\dot{Y}_i\dot{Y}_i^T = (p + r - 2)I_r$. This implies $\mathrm{SBD}(\mathbf{E}\,\dot{Y}\dot{Y}^T) = (p + r - 2)I_{nr}$.

Using a random $\dot{Y} \in T_Y$ has appeared before in [22], [23]. Those papers chose (in our notation) $\dot{Y}_i = \Gamma(I - Y_i^TY_i)$, which is only the portion of Γ in $\mathrm{null}(Y_i)$. Another similar approach to ours appears in [11], [12], where, rather than considering a random choice of $\dot{Y}$, the authors analyze the quadratic form $\Gamma \mapsto \langle S(Y), \dot{Y}(\Gamma)\dot{Y}(\Gamma)^T \rangle$ with $\dot{Y}_i(\Gamma) = \mathcal{P}_{T_{Y_i}}(\Gamma)$. Their results do not use randomness and scale differently the portions of each $\dot{Y}_i$ that are in $\operatorname{null}(Y_i)$ and $\operatorname{range}(Y_i^T)$ (note that this is related to the choice of metric on the Stiefel manifold).

4.1 Noiseless case

In Theorem 1, we have $\Delta = 0$, so $\widehat{L} = L_Z = L \otimes I_r$, and $S(Y) = L_Z - \operatorname{SBD}(L_Z YY^T)$.

We then calculate

$$
\begin{aligned}
\langle \operatorname{SBD}(\widehat{L}YY^T), \mathbf{E}\,\dot{Y}\dot{Y}^T \rangle &= \langle \operatorname{SBD}(L_Z YY^T), \mathbf{E}\,\dot{Y}\dot{Y}^T \rangle \\
&= \langle L_Z YY^T, \operatorname{SBD}(\mathbf{E}\,\dot{Y}\dot{Y}^T) \rangle \\
&= (p + r - 2) \operatorname{tr}(L_Z YY^T) \\
&= (p + r - 2) \sum_{i,j=1}^{n} L_{ij} \operatorname{tr}(Y_i Y_j^T).
\end{aligned}
$$

Next, by (19),

$$
\begin{aligned}
\langle L_Z, \mathbf{E}\,\dot{Y}\dot{Y}^T \rangle &= \sum_{i,j=1}^{n} L_{ij} \langle I_r, \mathbf{E}\,\dot{Y}_i\dot{Y}_j^T \rangle \\
&= \sum_{i,j=1}^{n} L_{ij}((p-2)r + \operatorname{tr}^2(Y_i Y_j^T)) \\
&= \sum_{i,j=1}^{n} L_{ij} \operatorname{tr}^2(Y_i Y_j^T),
\end{aligned}
$$

where the last equality follows from the fact that $\sum_i L_{ij} = 0$ for all j.

To proceed, note that

$$
\begin{aligned}
Y_i Y_j^T + Y_j Y_i^T &= Y_i Y_i^T + Y_i (Y_j - Y_i)^T + Y_j Y_j^T + Y_j (Y_i - Y_j)^T \\
&= 2 I_r - (Y_i - Y_j)(Y_i - Y_j)^T,
\end{aligned}
$$

so

$$
\operatorname{tr}(Y_i Y_j^T) = \frac{1}{2} \operatorname{tr}(Y_i Y_j^T + Y_j Y_i^T) = r - \frac{1}{2} \|Y_i - Y_j\|_{\mathrm{F}}^2. \tag{20}
$$

We then have (again using the fact that all rows and columns of L sum to zero)

$$
\begin{aligned}
\langle L_Z, \mathbf{E}\,\dot{Y}\dot{Y}^T \rangle &= \sum_{i,j=1}^{n} L_{ij} \left(r - \frac{1}{2} \|Y_i - Y_j\|_{\mathrm{F}}^2 \right)^2 \\
&= -r \sum_{i,j=1}^{n} L_{ij} \|Y_i - Y_j\|_{\mathrm{F}}^2 + \frac{1}{4} \sum_{i,j=1}^{n} L_{ij} \|Y_i - Y_j\|_{\mathrm{F}}^4 \\
&= 2r \sum_{i,j=1}^{n} L_{ij} \operatorname{tr}(Y_i Y_j^T) - \frac{1}{4} \sum_{i,j=1}^{n} A_{ij} \|Y_i - Y_j\|_{\mathrm{F}}^4.
\end{aligned}
$$

The condition $\langle S(Y), \mathbf{E}\,\dot{Y}\dot{Y}^T \rangle \geq 0$ then implies

$$
(p - r - 2) \sum_{i,j=1}^{n} L_{ij} \operatorname{tr}(Y_i Y_j^T) + \frac{1}{4} \sum_{i,j=1}^{n} A_{ij} \|Y_i - Y_j\|_{\mathrm{F}}^4 \leq 0. \tag{21}
$$

The second term on the left-hand side is nonnegative. If $p \geq r + 2$, the first term is also; this can be seen by noting either that it is the inner product of PSD matrices or that (from, e.g., (20))

$$
\sum_{i,j=1}^{n} L_{ij} \operatorname{tr}(Y_i Y_j^T) = \frac{1}{2} \sum_{i,j=1}^{n} A_{ij} \|Y_i - Y_j\|_{\mathrm{F}}^2 \geq 0.
$$

Therefore, if $p \geq r+2$, both terms on the left-hand side of (21) must be 0, so $Y_i = Y_j$ for all $(i,j) \in E$. If G is connected, this implies that the Y_i's are identical, and therefore

$$\langle ZZ^T, YY^T \rangle = \|Z^T Y\|_{\mathrm{F}}^2 = \|nY_1\|_{\mathrm{F}}^2 = n^2 r.$$

This finishes the proof of Theorem 1.

The robustness to (small) perturbations comes from the quartic terms in (21) (compare this to the proof of Theorem 3 in Section 4.2; in that proof, the quartic terms are dominated by the quadratic terms arising when $p > r+2$ and are ignored).

4.2 Noisy case

We now consider the case where the noise matrix $\Delta \neq 0$. We will, in fact, prove Theorem 5 first and then use it in the proof of Theorem 3.

To aid our analysis, we write $Y = ZR + W$, where $R = \frac{1}{n} Z^T Y \in \mathbf{R}^{r \times p}$ and W is orthogonal to Z (i.e., $Z^T W = \sum_i W_i = 0$). This gives $Y_i = R + W_i$. Note that

$$\langle ZZ^T, YY^T \rangle = \|Z^T Y\|_{\mathrm{F}}^2 = \|nR\|_{\mathrm{F}}^2 = n\|ZR\|_{\mathrm{F}}^2 = n(\|Y\|_{\mathrm{F}}^2 - \|W\|_{\mathrm{F}}^2) = n^2 r - n\|W\|_{\mathrm{F}}^2. \tag{22}$$

4.2.1 Error bound for all second-order critical points

The second-order criticality condition is now

$$\langle L_Z - \Delta, \dot{Y}\dot{Y}^T \rangle - \langle \mathrm{SBD}((L_Z - \Delta)YY^T), \dot{Y}\dot{Y}^T \rangle \geq 0.$$

We use the same $\dot{Y}$ as in the previous section. We have already analyzed the terms involving L_Z in Section 4.1, so it remains to analyze the terms involving Δ.

Note that

$$Y_i Y_j^T = I_r - Y_i(Y_i - Y_j)^T = I_r - Y_i(W_i - W_j)^T. \tag{23}$$

Furthermore, because we have assumed that every $Z_i = I_r$, we have

$$\sum_{i,j=1}^n \langle \Delta_{ij}, I_r \rangle = \sum_{i,j=1}^n \langle \Delta_{ij}, Z_i Z_j^T \rangle = \langle \Delta, ZZ^T \rangle. \tag{24}$$

From (19), we calculate

$$\begin{aligned}
\langle \Delta, \mathbf{E}\,\dot{Y}\dot{Y}^T \rangle &= \sum_{i,j=1}^n \langle \Delta_{ij}, (p-2)I_r + \mathrm{tr}(Y_i Y_j^T) Y_i Y_j^T \rangle \\
&\overset{\text{(a)}}{=} (p-2)\langle \Delta, ZZ^T \rangle + \sum_{i,j=1}^n \left(r - \frac{1}{2}\|Y_i - Y_j\|_{\mathrm{F}}^2 \right) \langle \Delta_{ij}, Y_i Y_j^T \rangle \\
&\overset{\text{(b)}}{=} (p-2)\langle \Delta, ZZ^T \rangle + r \sum_{i,j=1}^n \langle \Delta_{ij}, I_r - Y_i(W_i - W_j)^T \rangle - \frac{1}{2} \sum_{i,j=1}^n \|W_i - W_j\|_{\mathrm{F}}^2 \langle \Delta_{ij}, Y_i Y_j^T \rangle \\
&\overset{\text{(c)}}{=} (p+r-2)\langle \Delta, ZZ^T \rangle - \sum_{i,j=1}^n \left\langle \Delta_{ij}, \frac{1}{2}\|W_i - W_j\|_{\mathrm{F}}^2 Y_i Y_j^T + r Y_i (W_i - W_j)^T \right\rangle.
\end{aligned}$$

Equality (a) uses (20) and (24). Equality (b) uses the decomposition (23). Equality (c) again uses (24).

Next, again using (23) and (24),

$$\begin{aligned}
\langle \mathrm{SBD}(\Delta YY^T), \mathbf{E}\,\dot{Y}\dot{Y}^T \rangle &= (p+r-2)\,\mathrm{tr}(\Delta YY^T) \\
&= (p+r-2) \sum_{i,j=1}^n \langle \Delta_{ij}, Y_i Y_j^T \rangle \\
&= (p+r-2) \sum_{i,j=1}^n \langle \Delta_{ij}, I_r - Y_i(W_i - W_j)^T \rangle \\
&= (p+r-2)\langle \Delta, ZZ^T \rangle - (p+r-2) \sum_{i,j=1}^n \langle \Delta_{ij}, Y_i(W_i - W_j)^T \rangle.
\end{aligned}$$

The difference is

$$\begin{aligned}
&\langle \mathrm{SBD}(\Delta Y Y^T), \mathbf{E} \dot{Y}\dot{Y}^T\rangle - \langle \Delta, \mathbf{E} \dot{Y}\dot{Y}^T\rangle \\
&= \sum_{i,j=1}^n \left\langle \Delta_{ij}, \frac{1}{2}\|W_i - W_j\|_{\mathrm{F}}^2 Y_i Y_j^T - (p-2) Y_i (W_i - W_j)^T \right\rangle \\
&= \langle \Delta, (Q \otimes \mathbf{1}_r \mathbf{1}_r^T) \circ Y Y^T \rangle - (p-2)\langle \Delta, H^{(1)}\rangle + (p-2)\langle \Delta, H^{(2)}\rangle,
\end{aligned}$$

where $Q_{ij} = \frac{1}{2}\|W_i - W_j\|_{\mathrm{F}}^2$, and, with block indices, $H^{(1)}_{ij} = Y_i W_i^T$, and $H^{(2)}_{ij} = Y_i W_j^T$.

To bound these nuclear norms, note first that $H^{(2)} = Y W^T$, so $\|H^{(2)}\|_* \le \|Y\|_{\mathrm{F}} \|W\|_{\mathrm{F}} = \sqrt{nr}\|W\|_{\mathrm{F}}$. Furthermore,

$$H^{(1)} = \begin{bmatrix} Y_1 W_1^T \\ \vdots \\ Y_n W_n^T \end{bmatrix} Z^T.$$

Because each Y_i has operator norm 1, the left factor in the above expression has Frobenius norm $\le \|W\|_{\mathrm{F}}$. $\|Z\|_{\mathrm{F}} = \sqrt{nr}$, so $\|H^{(1)}\|_* \le \sqrt{nr}\|W\|_{\mathrm{F}}$ also.

The final nuclear norm bound is

$$\begin{aligned}
\|(Q \otimes \mathbf{1}_r \mathbf{1}_r^T) \circ Y Y^T\|_* &\overset{\text{(a)}}{\le} \|(Q \otimes \mathbf{1}_r \mathbf{1}_r^T)\|_* \\
&\overset{\text{(b)}}{=} r\|Q\|_* \\
&= r\left\| \left[\frac{1}{2}\|W_i - W_j\|_{\mathrm{F}}^2\right]_{ij} \right\|_* \\
&\le r\left(\frac{1}{2}\left\|[\|W_i\|_{\mathrm{F}}^2]_{ij}\right\|_* + \frac{1}{2}\left\|[\|W_j\|_{\mathrm{F}}^2]_{ij}\right\|_* + \|[\langle W_i, W_j\rangle]_{ij}\|_* \right) \\
&\overset{\text{(c)}}{=} r\left(\sqrt{n\sum_{i=1}^n \|W_i\|_{\mathrm{F}}^4} + \sum_{i=1}^n \langle W_i, W_i\rangle \right) \\
&\overset{\text{(d)}}{\le} 2r\sqrt{nr\sum_{i=1}^n \|W_i\|_{\mathrm{F}}^2} + r\|W\|_{\mathrm{F}}^2 \\
&= 2r\sqrt{nr}\|W\|_{\mathrm{F}} + r\|W\|_{\mathrm{F}}^2 \\
&\le 3r\sqrt{nr}\|W\|_{\mathrm{F}}.
\end{aligned}$$

Inequality (a) follows from the basic inequality of singular values of Hadamard products [48, Theorem 5.6.2], using the fact that each row of Y has norm 1. Equality (b) follows from the eigenvalue characterization of Kronecker products. Equality (c) uses the facts that the first two matrices in the previous line are rank-1 and that the third is PSD. Equality (d) follows from the fact that every $\|W_i\|_{\mathrm{F}}^2 \le 4r$.

Putting all the nuclear norm bounds together, we obtain, by Hölder's inequality for matrix inner products,

$$\langle \mathrm{SBD}(\Delta Y Y^T), \mathbf{E} \dot{Y}\dot{Y}^T\rangle - \langle \Delta, \mathbf{E} \dot{Y}\dot{Y}^T\rangle \le (2p + 3r - 4)\|\Delta\|_{\ell_2}\sqrt{nr}\|W\|_{\mathrm{F}}.$$

Combining this with the previous section's calculations, we obtain

$$(p - r - 2)\sum_{i,j} L_{ij}\langle Y_i, Y_j\rangle + \sum_{i,j=1}^n A_{ij}\|W_i - W_j\|_{\mathrm{F}}^4 \le (2p + 3r - 4)\|\Delta\|_{\ell_2}\sqrt{nr}\|W\|_{\mathrm{F}}. \tag{25}$$

Ignoring the nonnegative quartic term and using the fact that

$$\sum_{i,j} L_{ij}\langle Y_i, Y_j\rangle = \sum_{i,j} L_{ij}\langle W_i, W_j\rangle \ge \lambda_2\|W\|_{\mathrm{F}}^2,$$

we get

$$(p - r - 2)\lambda_2\|W\|_{\mathrm{F}}^2 \le (2p + 3r - 4)\|\Delta\|_{\ell_2}\sqrt{nr}\|W\|_{\mathrm{F}}.$$

Therefore,

$$\|W\|_{\mathrm{F}}^2 \le \left(\frac{(2p + 3r - 4)\|\Delta\|_{\ell_2}}{(p - r - 2)\lambda_2}\right)^2 nr = C_p^2 nr\frac{\|\Delta\|_{\ell_2}^2}{\lambda_2^2}.$$

Combining this with the identities (22) finishes the proof of Theorem 5.

4.2.2 Bound on solution rank

To show that a second-order critical point Y has low rank, recall that the first-order criticality condition $S(Y)Y = 0$, and, therefore, it suffices to show that $S(Y)$ has high rank. Recall

$$S(Y) = \widehat{L} - \mathrm{SBD}(\widehat{L}YY^T) = L_Z - \Delta - \mathrm{SBD}(L_Z YY^T) + \mathrm{SBD}(\Delta YY^T).$$

If G is connected (i.e., $\lambda_2 > 0$), then L_Z has an r-dimensional null space, and its remaining $(n-1)r$ eigenvalues are at least λ_2.

If $\|\Delta\|_{\ell_2} < \lambda_2$, simple singular value counting (same as eigenvalue counting, since the matrices here are symmetric) gives, for any $c \in (0,1)$,

$$\begin{aligned}
\mathrm{rank}(Y) &\le \dim \mathrm{null}(S(Y)) \\
&\le r + |\{\ell : |\lambda_\ell(\mathrm{SBD}(L_Z YY^T) + \mathrm{SBD}(\Delta YY^T))| \ge \lambda_2 - \|\Delta\|_{\ell_2}\}| \\
&\le r + |\{\ell : |\lambda_\ell(\mathrm{SBD}(L_Z YY^T))| \ge c(\lambda_2 - \|\Delta\|_{\ell_2})\}| \\
&\quad + |\{\ell : |\lambda_\ell(\mathrm{SBD}(\Delta YY^T))| \ge (1-c)(\lambda_2 - \|\Delta\|_{\ell_2})\}| \\
&\le r + \frac{\|\mathrm{SBD}(L_Z YY^T)\|_*}{c(\lambda_2 - \|\Delta\|_{\ell_2})} + \frac{\|\mathrm{SBD}(\Delta YY^T)\|_{\mathrm{F}}^2}{(1-c)^2(\lambda_2 - \|\Delta\|_{\ell_2})^2}.
\end{aligned} \tag{26}$$

Again, $\lambda_\ell(A)$ denotes the ℓth eigenvalue of a symmetric matrix A (arranged in increasing order, though order is irrelevant here).

To bound this last quantity, first note that, because $\|Y_i\|_{\ell_2} = 1$ for each i,

$$\|\mathrm{SBD}(\Delta YY^T)\|_{\mathrm{F}}^2 \le \|\Delta Y\|_{\mathrm{F}}^2 \le \|\Delta\|_{\ell_2}^2 \|Y\|_{\mathrm{F}}^2 = nr\|\Delta\|_{\ell_2}^2. \tag{27}$$

Letting $\mathrm{sym}(A) = A + A^T$ denote the symmetric part of a matrix, We can (crudely) bound $\|\Delta_i Y\|_{\ell_2} \le \|\Delta\|_{\ell_2}\|Y\|_{\ell_2} = \sqrt{n}\|\Delta\|_{\ell_2}$. Next, note that

$$\begin{aligned}
(\mathrm{SBD}(L_Z YY^T))_{ii} &= \sum_{j=1}^n L_{ij}(Y_i Y_j^T + Y_j Y_i^T) \\
&= \sum_{j=1}^n L_{ij}\left(I_r - \frac{1}{2}(Y_i - Y_j)(Y_i - Y_j)^T\right) \\
&= \frac{1}{2}\sum_{j=1}^n A_{ij}(Y_i - Y_j)(Y_i - Y_j)^T \\
&\succeq 0.
\end{aligned}$$

Therefore, $\mathrm{SBD}(L_Z YY^T) \succeq 0$, so

$$\begin{aligned}
\|\mathrm{SBD}(L_Z YY^T)\|_* &= \mathrm{tr}(\mathrm{SBD}(L_Z YY^T)) \\
&= \sum_{i,j=1}^n L_{ij}\,\mathrm{tr}(Y_i Y_j^T) \\
&\overset{\text{(i)}}{\le} C_p\|\Delta\|_{\ell_2}\sqrt{nr}\|W\|_{\mathrm{F}} \\
&\overset{\text{(ii)}}{\le} C_p^2 nr \frac{\|\Delta\|_{\ell_2}^2}{\lambda_2}.
\end{aligned} \tag{28}$$

Inequality (i) comes from (25) in the proof of Theorem 5. Inequality (ii) follows from the final result of Theorem 5.

Plugging (27) and (28) into (26), we obtain

$$\mathrm{rank}(Y) \le r + \left(\frac{C_p^2}{c\lambda_2(\lambda_2 - \|\Delta\|_{\ell_2})} + \frac{1}{(1-c)^2(\lambda_2 - \|\Delta\|_{\ell_2})^2}\right)\|\Delta\|_{\ell_2}^2 nr.$$

Choosing $c = 1/2$, we obtain, if $\|\Delta\|_{\ell_2} \le \lambda_2/4$,

$$\begin{aligned}
\mathrm{rank}(Y) &\le r + \left(\frac{8}{3}\frac{C_p^2}{\lambda_2^2} + \frac{64}{9}\frac{1}{\lambda_2^2}\right)\|\Delta\|_{\ell_2}^2 nr \\
&\le r + 5\left(\frac{C_p\|\Delta\|_{\ell_2}}{\lambda_2}\right)^2 nr.
\end{aligned}$$

In the last inequality, we use the fact that $C_p \geq 2$. Combined with the fact that $\operatorname{rank}(Y) \leq rn$, the above bound holds regardless of whether $\|\Delta\|_{\ell_2} \leq \lambda_2/4$. This completes the rank-bound portion of Theorem 2. If p is larger than this rank bound, then Y is rank-deficient, and we apply Lemma 1 to obtain the rest of the result.

4.2.3 Solution uniqueness

To prove Theorem 3, note that, under the assumptions, Theorem 2 implies that $\operatorname{rank}(Y) = r$. Furthermore, the proof of Theorem 2 give us that $\operatorname{rank}(S(Y)) = nr - r$.

Lemma 1 again implies that YY^T solves the SDP (17), and $S(Y) \succeq 0$ is its dual certificate. To prove that YY^T is the *unique* solution, note that because Y has rank r, we can write $Y = \widetilde{Z}U^T$, where $\widetilde{Z} \in \mathrm{O}(r)^n$, and $U \in \mathbf{R}^{r\times p}$ satisfies $UU^T = I_r$. Then $YY^T = \widetilde{Z}\widetilde{Z}^T$.

Furthermore, *any* optimal solution X of the SDP has rank at most r (because, as discussed in Section 3.3, X must satisfy $S(Y)X = 0$). Therefore, $YY^T = \widetilde{Z}\widetilde{Z}^T$ is a maximal-rank solution.

To show that this solution is unique, we use the result [49, Theorem 2.4] (a result that first appeared in [50]). Because each $\widetilde{Z}_i \in \mathrm{O}(r)$, it is clear that the linear map $\mathcal{A}_{\widetilde{Z}} \colon S^{r\times r} \to S^{nr\times nr}$ defined by

$$\mathcal{A}_{\widetilde{Z}}(H) = \operatorname{SBD}(\widetilde{Z}H\widetilde{Z}^T)$$

has trivial null space. Therefore, by [49, Theorem 2.4], $YY^T = \widetilde{Z}\widetilde{Z}^T$ is the unique solution to the SDP.

The fact that $YY^T = \widetilde{Z}\widetilde{Z}^T$ is the optimal solution to (17) (and therefore of (3)) implies that $\widetilde{Z}$ and Y are, respectively, global optima of (2) and (4). For any other global optima Z' of (2) and Y' of (4), $Z'Z'^T$ and $Y'Y'^T$ are feasible points of (3) with the same objective function value as $\widetilde{Z}\widetilde{Z}^T = YY^T$. Therefore, by the uniqueness of the SDP solution, we have $Z'Z'^T = Y'Y'^T = YY^T = \widetilde{Z}\widetilde{Z}^T$, implying that $Z' = \widetilde{Z}$ and $Y' = Y$ up to global orthogonal transformations. This completes the proof of Theorem 3.

4.3 Proof of SDP results

Corollaries 2 and 3 can be proved by taking the results of Theorems 3 and 5, respectively, and taking $p \to \infty$.

To make this precise, let X be an optimal solution to (3) and (17). For any $p \geq nr$, there exists $Y \in \mathbf{R}^{nr\times p}$ such that $X = YY^T$. The fact that X is feasible implies that Y is feasible for (4) and (18). Furthermore, the optimality of X implies that Y is a global optimum and therefore is a second-order critical point [8, Proposition 2.4]. We then apply Theorems 3 and 5 and take $p \to \infty$, noting that the quantity $C_p \to 2$.

5 Extension of proofs to the complex case

In the complex case, we can adapt the previous section's argument. Strictly speaking, to use standard differential geometry machinery such as that in [8], we should consider the complex Stiefel manifold $\mathrm{St}(p, r, \mathbf{C})$ by separating its elements into their real and imaginary parts and working with the resulting *real* manifold. This adds considerable complexity to the calculations. However, everything works out the same if we simply substitute complex quantities for real in the previous arguments, so for simplicity of presentation, we proceed in this naïve manner. We replace transposes by their Hermitian counterparts and use the convention that $\langle A, B\rangle = \operatorname{tr}(AB^*)$.

We only consider the noiseless case. We once again assume, without loss of generality, that $Z_1 = \cdots = Z_n = I_r$. We once again choose

$$\dot{Y}_i = \Gamma - Y_i\Gamma^* Y_i,$$

where now Γ is an $r \times p$ matrix of i.i.d. *complex* standard normal random variables. Now, by a similar calculation to before,

$$\begin{aligned}\mathbf{E}\,\dot{Y}_i\dot{Y}_j^* &= \mathbf{E}(\Gamma\Gamma^* - Y_i\Gamma^*Y_i\Gamma^* - \Gamma Y_j^*\Gamma Y_j^* + Y_i\Gamma^*Y_iY_j^*\Gamma Y_j^*)\\ &= pI_r + \operatorname{tr}(Y_iY_j^*)Y_iY_j^*.\end{aligned}$$

The first difference from the real case is that the terms with two factors of Γ or Γ^* have zero expectation, because each entry in these terms is a polynomial in standard complex normal random variables and thus is radially symmetric.

Then, we have

$$\begin{aligned}\langle \mathrm{SBD}(L_Z YY^*), \mathbf{E}\dot{Y}\dot{Y}^*\rangle &= \sum_{j=1}^n \Re\langle (L_Z YY^*)_{jj}, \mathbf{E}\dot{Y}_j\dot{Y}_j^*\rangle \\ &= \sum_{i,j=1}^n L_{ij}\Re\langle Y_iY_j^*, (p+r)I_r\rangle \\ &= (p+r)\sum_{i,j=1}^n L_{ij}\langle Y_i, Y_j\rangle\end{aligned}$$

Next,

$$\begin{aligned}\langle L_Z, \mathbf{E}\dot{Y}\dot{Y}^*\rangle &= \sum_{i,j=1}^n L_{ij}\Re\langle I_r, \mathbf{E}\dot{Y}_i\dot{Y}_j^*\rangle \\ &= \sum_{i,j=1}^n L_{ij}(pr + (\mathrm{tr}(Y_iY_j^*)^*)^2) \\ &= \sum_{i,j=1}^n L_{ij}\left[\Re^2\langle Y_i, Y_j\rangle - \Im^2\langle Y_i, Y_j\rangle\right] \\ &= \frac{1}{4}\sum_{i,j=1}^n L_{ij}\left[\mathrm{tr}^2(Y_iY_j^* + Y_jY_i^*) - |\mathrm{tr}(Y_iY_j^* - Y_jY_i^*)|^2\right].\end{aligned}$$

Here, we see a significant difference from the real case: we must consider the imaginary part of $\mathrm{tr}(Y_iY_j^*)$ when this quantity is squared. The optimal way to handle this is unclear, but one way is the bound

$$\begin{aligned}|\mathrm{tr}(Y_iY_j^* - Y_jY_i^*)|^2 &\le r\|Y_iY_j^* - Y_jY_i^*\|_{\mathrm{F}}^2 \\ &= 4r\|Y_iY_j^*\|_{\mathrm{F}}^2 - r\|Y_iY_j^* + Y_jY_i^*\|_{\mathrm{F}}^2 \\ &\le 4r\|Y_iY_j^*\|_{\mathrm{F}}^2 - \mathrm{tr}^2(Y_iY_j^* + Y_jY_i^*).\end{aligned}$$

Because $L_{ij} \le 0$ for $i \ne j$ (and both sides of the above inequality are zero when $i = j$), we then obtain

$$\begin{aligned}(p+r)\sum_{i,j=1}^n L_{ij}\langle Y_i, Y_j\rangle &\le \langle L_Z, \mathbf{E}\dot{Y}\dot{Y}^*\rangle \\ &\le \sum_{i,j=1}^n L_{ij}\left(\frac{1}{2}\mathrm{tr}^2(Y_iY_j^* + Y_jY_i^*) - r\|Y_iY_j^*\|_{\mathrm{F}}^2\right) \\ &\le 4r\sum_{i,j=1}^n L_{ij}\langle Y_i, Y_j\rangle - \frac{1}{2}\sum_{i,j=1}^n A_{ij}\|Y_i - Y_j\|_{\mathrm{F}}^4,\end{aligned}$$

where we have used the fact that $\sum L_{ij}\|Y_iY_j\|_{\mathrm{F}}^2 = \sum L_{ij}\langle Y_i^*Y_i, Y_j^*Y_j\rangle \ge 0$. This inequality requires $p \ge 3r$ to obtain perfect recovery.

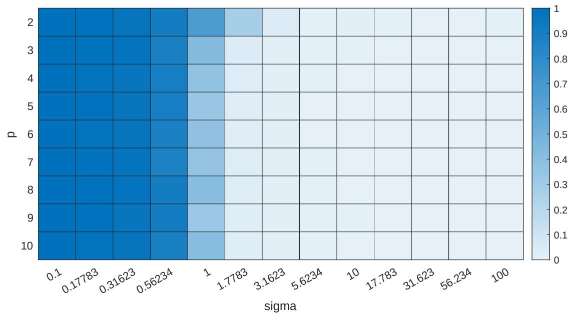
We can improve this result by making a slightly different choice of the $\dot{Y}_i$'s. Note that we can write our choice in the previous section as

$$\dot{Y}_i = \Gamma - Y_i\Gamma^*Y_i = \Gamma(I_p - Y_i^*Y_i) - (\Gamma Y^* - Y\Gamma^*)Y.$$

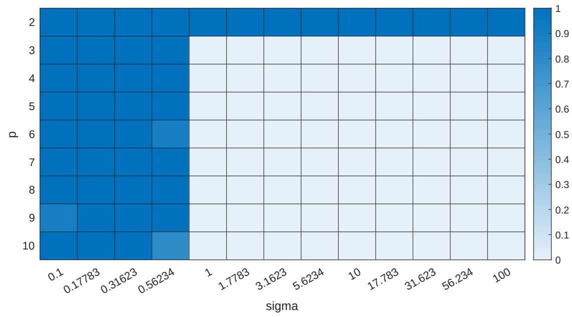
The first term has rows orthogonal to Y, while the second has a skew-symmetric matrix left-multiplying Y (rather than right-multiplying as before). We can rescale these terms arbitrarily:

$$\dot{Y}_i = a\Gamma(I_p - Y_i^*Y_i) + b(\Gamma Y_i^* - Y_i\Gamma^*)Y_i$$

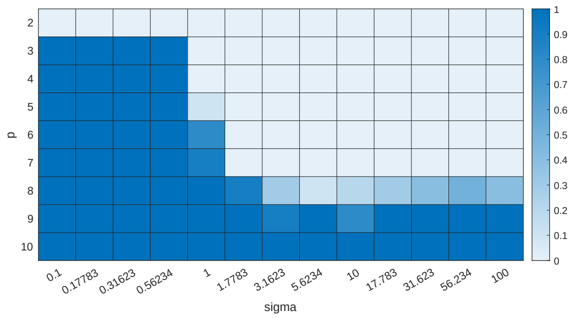
for any (potentially complex) numbers a and b. For $a, b > 0$ (which turns out to be the only sensible choice), this is closely related to our choice of metric on the Stiefel manifold. We will choose $a = 2$ and $b = 1$, which makes $\dot{Y}$ the (Euclidean) orthogonal projection of $2Z\Gamma$ onto T_Y. Intrinsically, this is quite similar to a complex adaptation of [12], and the result is quite similar.



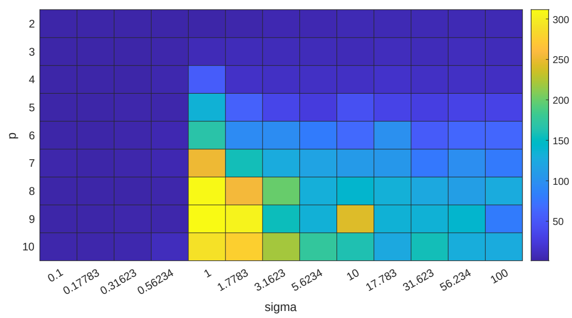
(a) Normalized correlation $\|Z^TY\|_{\mathrm{F}}^2/n^2r$



(b) Solution is rank-r



(c) Solution is rank-deficient



(d) Solve time (s)

Figure 2: Experiments ($r = 2$) with a circulant graph on $n = 400$ vertices, each vertex having degree 10. All values are the average of 10 experiments with random noise and initialization.

Now, for Γ chosen randomly as before,

$$\begin{aligned}\mathbf{E}\,\dot{Y}_i\dot{Y}_j^* &= 4\,\mathbf{E}\,\Gamma(I_p - Y_i^*Y_i)(I_p - Y_j^*Y_j)\Gamma^* + \mathbf{E}(\Gamma Y_i^* - Y_i\Gamma^*)Y_iY_j^*(Y_j\Gamma^* - \Gamma Y_j^*)\\ &\quad + 2\,\mathbf{E}\,\Gamma(I_p - Y_i^*Y_i)Y_j^*(Y_j\Gamma^* - \Gamma Y_j^*) + 2\,\mathbf{E}(\Gamma Y_i^* - Y_i\Gamma^*)Y_i(I_p - Y_j^*Y_j)\Gamma^*\\ &= 4\langle I_p - Y_i^*Y_i, I_p - Y_j^*Y_j\rangle I_r + \langle Y_i^*Y_i, Y_j^*Y_j\rangle I_r + \langle Y_i, Y_j\rangle Y_iY_j^*\\ &\quad + 2\langle I_p - Y_i^*Y_i, Y_j^*Y_j\rangle I_r + 2\langle Y_i^*Y_i, I_p - Y_j^*Y_j\rangle I_r\\ &= \big(4(p-r) + \|Y_iY_j^*\|_{\mathrm{F}}^2\big)I_r + \langle Y_i, Y_j\rangle Y_iY_j^*.\end{aligned}$$

The usual arguments and the previous bound on the imaginary part yield

$$\begin{aligned}(4p-2r)\sum L_{ij}\langle Y_i, Y_j\rangle &\le \sum L_{ij}(r\|Y_iY_j^*\|_{\mathrm{F}}^2 + \langle Y_i, Y_j\rangle^2)\\ &\le \sum L_{ij}\left(r\|Y_iY_j\|_{\mathrm{F}}^2 + \frac{1}{2}\operatorname{tr}^2(Y_iY_j^* + Y_jY_i^*) - r\|Y_iY_j^*\|_{\mathrm{F}}^2\right)\\ &= \frac{1}{2}\sum L_{ij}\operatorname{tr}^2(Y_iY_j^* + Y_jY_i^*)\\ &= 4r\sum L_{ij}\langle Y_i, Y_j\rangle - \frac{1}{2}\sum A_{ij}\|Y_i - Y_j\|_{\mathrm{F}}^4.\end{aligned}$$

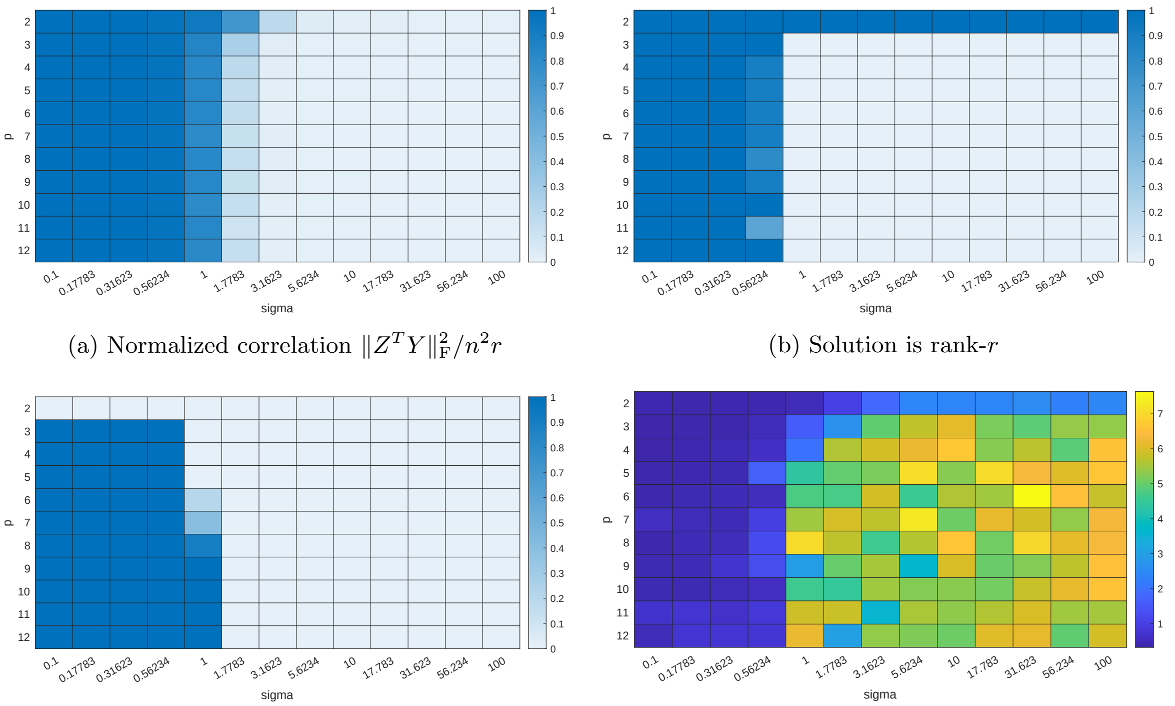
Thus, if G is connected, $2p \ge 3r$ implies all of the Y_i's are the same, or, equivalently, $YY^T = ZZ^T$.

6 Simulations

We implemented an algorithm for solving (4) and ran experiments on several graphs. We used Matlab with the Manopt toolbox [51] to optimize over a product of Stiefel manifolds with the default second-order trust-region algorithm.

We ran experiments on three graphs:

- A circulant graph on 400 vertices, each having degree 10 (results in Figure 2).
- A single realization of the Erdős–Rényi random graph on 400 vertices, each vertex having expected degree 10 (Figure 3).



(a) Normalized correlation $\|Z^T Y\|_F^2/n^2 r$

(b) Solution is rank-r

(c) Solution is rank-deficient

(d) Solve time (s)

Figure 3: Experiments ($r = 2$) on a graph on $n = 400$ vertices which is a single instance from an ER model with average degree 10. All values are the average of 10 experiments on the same graph with random noise and initialization.

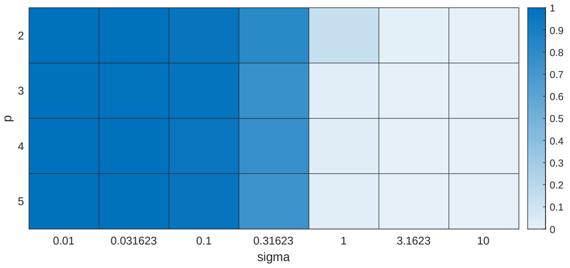
- The pose graph from the Freiburg Building 079 (`fr079`) SLAM dataset,[9] which has 989 vertices with average degree 3.4 (Figure 4).

All experiments are in the real number case and use $r = 2$. In the case $p = r = 2$, the random initial point is chosen to be in the same connected component as the ground truth. The noise matrix Δ is chosen with i.i.d. $\mathcal{N}(0, \sigma^2)$ entries in the nonzero blocks (with the constraint $\Delta = \Delta^T$). Figures 2 to 4).

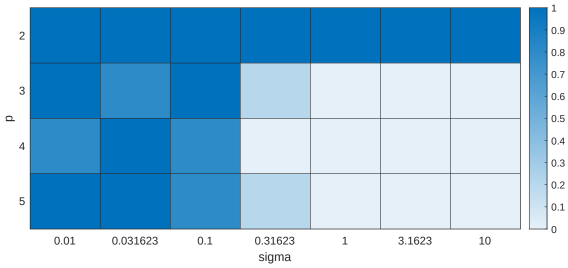
The results are summarized in Figures 2 to 4. A few points worth highlighting are the following:

- There is a clear phase transition as the noise standard deviation σ increases; the recovery performance, solution rank, and algorithm running time all jump dramatically in approximately the same place.
- Other properties show markedly different trends from one graph to another.
  - For the circulant graph (Figure 2), the algorithm running time is highest close to the phase transition and decreases for larger σ. Furthermore, the solution rank is almost never larger than 8 even at very high noise levels.
  - For the Erdős–Rényi and SLAM graphs (Figures 3 and 4), there is less clear structure in the running times, and the solution is never rank-deficient for noise levels past the phase transition.
- There is no apparent problem with ending in bad local optima when p is small (even in the case $p = r$, though in this case the experiments are artificially aided by initializing the solution to the correct connected component of $\mathrm{O}(r)$). This is likely due to the fact that, especially for such large graphs, the likelihood of a random initialization landing in the basin of attraction of a bad local optimum is small (though it is nonzero, at least in the case of the circulant graph).
- Even along the phase transition, the choice of p has no apparent effect on the solution quality (correlation with Z). The case $p = r = 2$ is a curious exception, as the correlation is *higher* (however, note the previous caveat on initialization in this case).

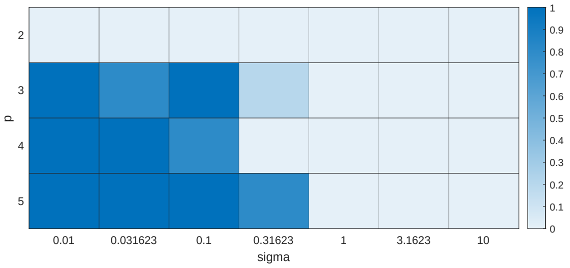
[9]Recorded by Cyrill Stachniss and available at https://www.ipb.uni-bonn.de/datasets/.



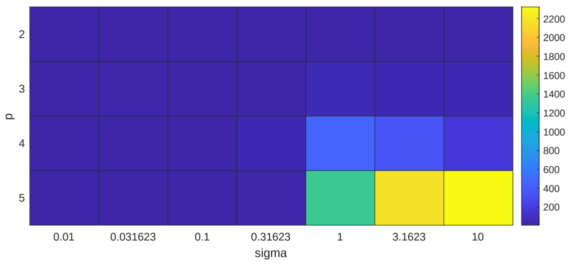
(a) Normalized correlation $\|Z^T Y\|_F^2/n^2 r$



(b) Solution is rank-r



(c) Solution is rank-deficient



(d) Solve time (s)

Figure 4: Experiments ($r = 2$) on the `fr079` dataset pose graph with $n = 989$ vertices of average degree 3.4. All values are the average of 5 experiments with random noise and initialization.

References

[1] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert, "Initialization techniques for 3d SLAM: A survey on rotation estimation and its use in pose graph optimization," in *Proc. IEEE Conf. Robot. Autom. (ICRA)*, Seattle, Washington, USA, May 2015.

[2] L. Wang, A. Singer, and Z. Wen, "Orientation determination of cryo-EM images using least unsquared deviations," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2450–2483, Jan. 2013.

[3] D. Martinec and T. Pajdla, "Robust rotation and translation estimation in multiview reconstruction," in *Proc. Conf. Comput. Vis. Pattern Recog. (CVPR)*, Minneapolis, Minnesota, USA, Jun. 2007.

[4] M. A. Iwen, B. Preskitt, R. Saab, and A. Viswanathan, "Phase retrieval from local measurements: Improved robustness via eigenvector-based angular synchronization," *Appl. Comput. Harmon. Anal.*, vol. 48, pp. 415–444, 2020.

[5] N. Boumal, "A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints," Tech. Rep., 2015. arXiv: `1506.00575 [math.OC]`.

[6] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "SE-sync: A certifiably correct algorithm for synchronization over the special Euclidean group," *Int. J. Robot. Res.*, vol. 38, no. 2-3, pp. 95–125, 2019.

[7] F. Dellaert, D. M. Rosen, J. Wu, R. Mahony, and L. Carlone, "Shonan rotation averaging: Global optimality by surfing $SO(p)^n$," in *Proc. European Conf. Comput. Vis. (ECCV)*, Virtual conference, Aug. 2020, pp. 292–308.

[8] N. Boumal, V. Voroninski, and A. S. Bandeira, "Deterministic guarantees for Burer-Monteiro factorizations of smooth semidefinite programs," *Commun. Pure Appl. Math.*, vol. 73, no. 3, pp. 581–608, 2019.

[9] L. O'Carroll, V. Srinivas, and A. Vijayaraghavan, "The Burer-Monteiro SDP method can fail even above the Barvinok-Pataki bound," Nov. 22, 2022. arXiv: `2211.12389 [math.OC]`.

[10] I. Waldspurger and A. Waters, "Rank optimality for the Burer–Monteiro factorization," *SIAM J. Optim.*, vol. 30, no. 3, pp. 2577–2602, 2020.

[11] J. Markdahl, J. Thunberg, and J. Goncalves, "Towards almost global synchronization on the Stiefel manifold," in *Proc. IEEE Conf. Decision Control (CDC)*, Miami, Florida, USA, Dec. 2018.

[12] J. Markdahl, J. Thunberg, and J. Goncalves, "High-dimensional Kuramoto models on Stiefel manifolds synchronize complex networks almost globally," *Automatica*, vol. 113, 2020.

[13] N. Boumal, *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023, ISBN: 9781009166157.

[14] A. Townsend, M. Stillman, and S. H. Strogatz, "Dense networks that do not synchronize and sparse ones that do," *Chaos*, vol. 30, no. 8, 2020.

[15] J. Markdahl, J. Thunberg, and J. Goncalves, "Almost global consensus on the n-sphere," *IEEE Trans. Autom. Control*, vol. 63, no. 6, pp. 1664–1675, 2018.

[16] J. Markdahl, "Synchronization on Riemannian manifolds: Multiply connected implies multistable," *IEEE Trans. Autom. Control*, vol. 66, no. 9, pp. 4311–4318, 2021.

[17] R. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation averaging," *Int. J. Comput. Vis.*, vol. 103, pp. 267–305, 2013.

[18] D. M. Rosen, K. J. Doherty, A. T. Espinoza, and J. J. Leonard, "Advances in inference and representation for simultaneous localization and mapping," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 215–242, May 2021.

[19] S. Burer, R. D. C. Monteiro, and Y. Zhang, "Rank-two relaxation heuristics for MAX-CUT and other binary quadratic programs," *SIAM J. Optim.*, vol. 12, no. 2, pp. 503–521, 2002.

[20] S. Burer and R. D. C. Monteiro, "Local minima and convergence in low-rank semidefinite programming," *Math. Program.*, vol. 103, pp. 427–444, 2005.

[21] A. S. Bandeira, N. Boumal, and V. Voroninski, "On the low-rank approach for semidefinite programs arising in synchronization and community detection," in *Proc. Conf. Learn. Theory (COLT)*, New York, Jun. 2016.

[22] S. Ling, "Solving orthogonal group synchronization via convex and low-rank optimization: Tightness and landscape analysis," *Math. Program.*, 2022, forthcoming.

[23] S. Mei, T. Misiakiewicz, A. Montanari, and R. I. Oliveira, "Solving SDPs for synchronization and MaxCut problems via the Grothendieck inequality," in *Proc. Conf. Learn. Theory (COLT)*, Amsterdam, Netherlands, Jul. 2017.

[24] R. Y. Zhang, "Improved global guarantees for the nonconvex burer–monteiro factorization via rank overparameterization," 2022. arXiv: 2207.01789 [math.OC].

[25] A. Singer and H.-t. Wu, "Vector diffusion maps and the connection Laplacian," *Commun. Pure Appl. Math.*, vol. 65, no. 8, pp. 1067–1144, 2012.

[26] A. Singer, "Angular synchronization by eigenvectors and semidefinite programming," *Appl. Comput. Harmon. Anal.*, vol. 30, pp. 20–36, 2011.

[27] A. S. Bandeira, A. Singer, and D. A. Spielman, "A Cheeger inequality for the graph connection Laplacian," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 4, pp. 1611–1630, 2013.

[28] A. S. Bandeira, "Random Laplacian matrices and convex relaxations," *Found. Comput. Math.*, vol. 18, pp. 345–379, 2018.

[29] L. Wang and A. Singer, "Exact and stable recovery of rotations for robust synchronization," *Inform. Inference.*, vol. 2, no. 2, pp. 145–193, 2013.

[30] F. Filbir, F. Krahmer, and O. Melnyk, "On recovery guarantees for angular synchronization," *J. Fourier Anal. Appl.*, vol. 27, 2021.

[31] K. J. Doherty, D. M. Rosen, and J. J. Leonard, "Performance guarantees for spectral initialization in rotation averaging and pose-graph SLAM," in *Proc. IEEE Conf. Robot. Autom. (ICRA)*, Philadelphia, PA, USA, May 2022.

[32] B. P. Preskitt, "Phase retrieval from locally supported measurements," Ph.D. thesis, UC San Diego, 2018, 216 pp.

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

[34] C. Gao and A. Y. Zhang, "SDP achieves exact minimax optimality in phase synchronization," *IEEE Transactions on Information Theory*, vol. 68, no. 8, pp. 5374–5390, Aug. 2022.

[35] A. Y. Zhang, "Exact minimax optimality of spectral methods in phase synchronization and orthogonal group synchronization," Sep. 12, 2022. arXiv: 2209.04962 [math.ST].

[36] A. Pikovsky and M. Rosenblum, "Introduction to focus issue: Dynamics of oscillator populations," *Chaos*, vol. 33, no. 1, 2023.

[37] R. Taylor, "There is no non-zero stable fixed point for dense networks in the homogeneous Kuramoto model," *Journal of Physics A: Mathematical and Theoretical*, vol. 45, 2012.

[38] S. Ling, R. Xu, and A. S. Bandeira, "On the landscape of synchronization networks: A perspective from nonconvex optimization," *SIAM J. Optim.*, vol. 29, no. 3, pp. 1879–1907, 2019.

[39] R. Yoneda, T. Tatsukawa, and J.-n. Teramae, "The lower bound of the network connectivity guaranteeing in-phase synchronization," *Chaos*, vol. 31, no. 6, 2021.

[40] J. Lu and S. Steinerberger, "Synchronization of Kuramoto oscillators in dense networks," *Nonlinearity*, vol. 33, no. 11, pp. 5905–5918, 2020.

[41] M. Kassabov, S. H. Strogatz, and A. Townsend, "Sufficiently dense Kuramoto networks are globally synchronizing," *Chaos*, vol. 31, no. 7, 2021.

[42] M. Kassabov, S. H. Strogatz, and A. Townsend, "A global synchronization theorem for oscillators on a random graph," Mar. 7, 2022. arXiv: 2203.03152 [math.DS].

[43] P. Abdalla, A. S. Bandeira, M. Kassabov, V. Souza, S. H. Strogatz, and A. Townsend, "Expander graphs are globally synchronising," Oct. 23, 2022. arXiv: 2210.12788 [math.CO].

[44] P. Abdalla, A. S. Bandeira, and C. Invernizzi, "Guarantees for spontaneous synchronization on random geometric graphs," Aug. 25, 2022. arXiv: 2208.12246 [math.PR].

[45] J. Markdahl, D. Proverbio, L. Mi, and J. Goncalves, "Almost global convergence to practical synchronization in the generalized Kuramoto model on networks over the n-sphere," *Commun. Phys.*, vol. 4, 2021.

[46] J. Markdahl, "A geometric obstruction to almost global synchronization on riemannian manifolds," Aug. 2, 2018. arXiv: 1808.00862 [math.OC].

[47] A. Hatcher, *Algebraic Topology*. Cambridge University Press, 2001, p. 556, ISBN: 9780521795401.

[48] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991, p. 607, ISBN: 052130587X.

[49] A. Lemon, A. M.-C. So, and Y. Ye, "Low-rank semidefinite programming: Theory and applications," *Foundations and Trends in Optimization*, vol. 2, no. 1-2, pp. 1–156, 2016.

[50] Z. Zhu, "A semidefinite programming method for graph realization and low rank matrix completion problem," Ph.D. dissertation, Stanford University, 2011.

[51] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *J. Mach. Learn. Res.*, vol. 15, no. 42, pp. 1455–1459, 2014.