On the Distribution of Minima in Intrinsic-Metric Rotation Averaging

Kyle Wilson
Washington College
Department of Math and Computer Science
Chestertown, MD

kwilson24@washcoll.edu

David Bindel
Cornell University
Department of Computer Science
Ithaca, NY

bindel@cs.cornell.edu

Abstract

Rotation Averaging is a non-convex optimization problem that determines orientations of a collection of cameras from their images of a 3D scene. The problem has been studied using a variety of distances and robustifiers. The intrinsic (or geodesic) distance on $\mathrm{SO}(3)$ is geometrically meaningful; but while some extrinsic distance-based solvers admit (conditional) guarantees of correctness, no comparable results have been found under the intrinsic metric.

In this paper, we study the spatial distribution of local minima. First, we do a novel empirical study to demonstrate sharp transitions in qualitative behavior: as problems become noisier, they transition from a single (easy-to-find) dominant minimum to a cost surface filled with minima. In the second part of this paper we derive a theoretical bound for when this transition occurs. This is an extension of the results of [24], which used local convexity as a proxy to study the difficulty of problem. By recognizing the underlying quotient manifold geometry of the problem we achieve an n-fold improvement over prior work. Incidentally, our analysis also extends the prior l_2 work to general l_p costs. Our results suggest using algebraic connectivity as an indicator of problem difficulty.

1. Introduction

The rotation averaging problem arises in computer vision as part of estimating the 3-dimensional poses of a set of cameras. We are given many images of some scene, produced by cameras at unknown locations and with unknown orientations. When two images overlap, it is often possible to estimate the *relative orientation* of the associated cameras. By considering many overlapping camera pairs we obtain many such pairwise measurements.

Essentially, rotation averaging is an optimization problem on a graph in which vertices represent cameras and edges represent measurements of relative orientation. The goal is to pick an *absolute orientation* for each vertex in a way that

best agrees with the relative orientation measurements. Both the relative and absolute orientations must be in $\mathrm{SO}(3)$, the group of orientation-preserving rotations of 3D space.

We may choose the cost function used in rotation averaging for convenience of computation, for the existence of theoretical guarantees, or for modeling concerns. From a geometric perspective, costs based on intrinsic metrics—geodesic distances on $\mathrm{SO}(3)$—are a natural choice. However, no such method is presently known to have a guarantee of global convergence. In fact, the l_2 geodesic cost is known to yield a non-convex problem. Formulations based on extrinsic costs—non-manifold distances within representations—have been more tractable and sometimes give optimality guarantees. When solvers fail the junk solutions doom downstream problems, such as translations averaging and bundle adjustment.

This paper considers cost functions based on the intrinsic, geodesic distance. Two recent surveys [9, 21] propose a familiar pattern: use an extrinsic solver to generate a candidate solution, then refine with with a solver chosen for its modeling properties—even if the latter solver only guarantees local optimality. This is a common pattern in geometric computer vision, with a well-known exemplar being the "Gold-Standard" algorithm for estimating homographies [15]. The extrinsic-intrinsic solver pattern leads to some key questions: "Is this initial guess good enough for us to discover the best solution?" and "Why do some problems require much better initial guesses than others?" In this paper, we seek partial answers to these questions. (Note that we do not propose a new solver.) We make two main contributions:

1. We empirically investigate the cost surfaces of intrinsic rotation averaging problems. We demonstrate how properties of the problem instance can lead to challenging distributions of local minima.

2. We derive bounds on this behavior by improving the local convexity analysis in [24]. We show how rotation averaging's gauge symmetry leads to a quotient-manifold description of the optimization problem, which leads to an n-fold improvement in bounds.

2. Related Work

Relaxed Approaches. The seminal paper introducing the rotation averaging problem [13] also proposed a solver based on the unit quaternion representation of 3D rotations. This is a prototypical *relaxed* method: by ignoring the length-1 constraint and minimizing Euclidean distances between quaternions, the problem becomes linear. The solution may not satisfy the unit constraint (hence not actually represent rotations), but it can be normalized easily. Many later methods follow the same pattern of solving a relaxed problem, then "rounding" the solution to satisfy the constraints. Most works in this line use a rotation matrix representation and relax the orthnormality and determinant constraints. Martinec and Pajdla [17] relaxed to a least-squares problem. Arie-Nachimson *et al.* [3] give a spectral solution to a similar cost. Arrigoni *et al.* [4] solve their relaxed problem in a low rank + sparse framework for greater robustness to outliers. Wang and Singer [23] seek robustness by solving an l_1 relaxed problem. Unfortunately, even if the unconstrained problem can be solved exactly, the final rounded solution is generally not an optimal solution to any particular problem, and the "relaxation gaps" due to rounding are not well understood.

Manifold Approaches. Manifold methods use the tools of Riemannian geometry to optimize directly on the rotations manifold. In practice, this involves iteratively solving a series of Euclidean tangent space problems. Govindu [14] pioneered this approach, and Hartley *et al.* [16] and Chatterjee and Govindu [10] have proposed robustified methods. Tron *et al.* [20] give a distributed consensus solver for a specific "reshaped" cost. Unfortunately, while such solvers converge to a local minimum, that local minimum may not be globally optimal. Two survey papers [9, 21] propose using an initialize-and-refine pattern, where a globally solveable relaxed method generates an initial guess which a manifold-based method then refines.

Verification and Guarantees. More recently, there has been particular interest in validation and performance guarantees for solvers. Fredriksson and Olsson [12] return to quaternions, and are able to verify—via Lagrangian Duality—that their solution is optimal (if noise levels are low enough). Briales and Gonzales-Jimenez [7] give a verifier for camera poses, rather than rotations. Boumal *et al.* [5, 6] take a statistical approach and study the efficiency of maximum likelihood estimators via Cramér-Rao bounds. They develop visualizations to show how the problem's graph topology affects the uncertainty in the output. In contrast, this paper visualizes how topology affects the macro-scale shape of the cost surface, independent of a choice of solution.

Most recently, Ericsson *et al.* [11] propose a new solver based on Lagrangian duality. Unusually, this method uses an extrinsic cost but makes no relaxations (and hence suffers no rounding gap). When strong duality holds, this method gives a global solution. For complete graphs, they show that it is sufficient that residuals are less than $42.9°$. We find it interesting that two very different approaches (theirs: chordal distance and Lagrangian duality and ours: geodesic distance and local convexity) lead to statements with similar form. Briales and Gonzales-Jimenez [8] and Rosen *et al.*'s [18] fast relaxation methods also give verifiers.

This present paper is most closely related to Wilson *et al.* [24] which studied l_2 intrinsic costs via a local convexity analysis. It concluded that some rotation averaging problem instances exhibit large areas of convex behavior, but their extent is related to problem size, noisiness, and connectivity. However, the local convexity is not apparent until an underlying gauge ambiguity is removed. This paper returns to the analysis in [24] with a more geometrically natural description of gauge which yields greatly improved bounds. The present analysis is also more general, as it considers an l_p cost rather than the l_2 cost of [24]. Also, unlike [11] and [24], we empirically investigate what happens outside the optimality bounds.

Outline of Remaining Sections. The remainder of this paper proceeds as follows: Section 3 gives notation and background on representations of rotations. It also states the rotation averaging problem in the form we will use. Section 4 is an empirical investigation into the shape of the cost surfaces of our problems. We use a spectral embedding to visualize the locations and strengths of local minima. These experiments clearly demonstrate transitions from easy problems dominated by a single minimum to difficult problems with a preponderance of bad local minima. In Section 5 we bound this behavior by improving the local convexity analysis from [24], and finally we suggest applications and conclude in Section 6.

3. Preliminaries

In this section we introduce notation and formally define the rotation averaging problem. We will also note some properties which are readily apparent.

3.1. 3D Rotations and Their Representations

The special orthogonal group in 3 dimensions—SO(3)—is the group of rotations of $\mathbb{R}^3$. SO(3) is a smooth 3-dimensional manifold, so it is in fact a Lie group—a notion of continuous symmetry. While this suffices as a geometric definition, for computational purposes we require coordinatized representations. Euler's rotation theorem states that every rotation can be decomposed (not uniquely) into an angle θ and an axis $\hat{\mathbf{v}}$. Alternately, every element of $SO(3)$ is uniquely represented by a 3×3 orthogonal matrix with determinant 1. As orthogonal matrices, rotation matrices have the property that $R^{-1} = R^\top$. The angle-axis representation $\theta\hat{\mathbf{v}}$ is closely related to the tangent space to SO(3), which is 3-by-3 skew symmetric matrices. We write a tangent space

element:

$$[\theta\hat{\mathbf{v}}]_\times = \begin{bmatrix} 0 & -\theta v_z & \theta v_y \\ \theta v_z & 0 & -\theta v_x \\ -\theta v_y & \theta v_x & 0 \end{bmatrix}, \text{ where } \hat{\mathbf{v}} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}.$$

Rotation matrices form a 3-dimensional submanifold embedded in $\mathbb{R}^{3\times3}$, giving rise to at least two common metrics:

$$d_{\text{intrinsic}}(\mathtt{R}, \mathtt{S}) = \angle(\mathtt{R}^\top \mathtt{S}) \tag{1}$$

$$d_{\text{extrinsic}}(\mathtt{R}, \mathtt{S}) = \|\mathtt{R} - \mathtt{S}\|_F. \tag{2}$$

where $\angle(\cdot)$ is the angle of a given rotation (*a la* Euler's rotation theorem) and $\|\cdot\|_F$ denotes the Frobenius matrix norm. The intrinsic metric can be shown to be a geodesic distance—that is, the length of the shortest path *that follows the curvature of the rotation manifold*. The extrinsic metric cuts through non-manifold space: it is a distance in $\mathbb{R}^{3\times3}$, and so is generally shorter than the $d_{\text{intrinsic}}$. The two are related by $d_{\text{extrinsic}} = 2\sqrt{2}\sin\left(\frac{d_{\text{intrinsic}}}{2}\right)$. This paper is about rotation averaging under the intrinsic metric, so we will omit the subscripts and simply write $d_{\text{intrinsic}}$ as d.

3.2. Problem Definition

A rotation averaging problem instance consists of a graph $G = (V, E)$ with measurements $\widetilde{\mathcal{R}} = \{\widetilde{\mathtt{R}}_{ij} | (i,j) \in E\}$ of relative orientation on each edge. Writing $n = |V|$, our goal is to choose an absolute orientation $\mathcal{R} \in \text{SO}(3)^n$ for each vertex in a way that best respects those measurements:

$$\min_{\mathcal{R} \in \text{SO}(3)^n} \phi_p(\mathcal{R}) = \sum_{(i,j)\in E} d(\widetilde{\mathtt{R}}_{ij}, \mathtt{R}_i \mathtt{R}_j^\top)^p. \tag{3}$$

We will refer to ϕ_p as the l_p cost function for this problem. Our results in Section 5 will apply for the $p > 1$ cases.

3.3. Local Minima

The l_p problems in Equation (3) are non-convex, manifold-valued optimization problems. Analogous problems in Euclidean space are convex; the nonconvexity of our problem arises entirely from the manifold geometry. Iterative solvers of the type we consider start from an initial guess, then iterate toward a stationary point (which we hope is a local minimum). This local minimum need not be a global minimizer for the problem, and Section 4 shows that in some situations the problem may have many bad local minima. The guiding concern of the analysis in this paper is improving our understanding of this failure mode. Note that the multiplicity and spatial distribution of local minima are properties of a problem instance itself, not of a solver. In principle, this difficulty can be largely mitigated with a sufficiently high quality guess. Thus, it would be helpful to have some theoretical guidance as to what quality of guess is required.

3.4. Some Simple Observations

Edge Directions. If $\widetilde{\mathtt{R}}_{ij}$ is a measurement of $\mathtt{R}_i \mathtt{R}_j^\top$—the relative rotation on edge (i, j)—then $\widetilde{\mathtt{R}}_{ij}^\top$ is a measurement of $\mathtt{R}_j \mathtt{R}_i^\top$. We will treat G as an undirected graph, and supply either $\widetilde{\mathtt{R}}_{ij}$ or $\widetilde{\mathtt{R}}_{ij}^\top$ as appropriate in context.

Minimally Constrained Problems. We assume throughout that G is connected. When G is a tree there always exists a solution with zero cost. Any additional edges overconstrain the problem, but this redundancy improves accuracy in the face of measurement noise.

Gauge Ambiguity. Any rotation averaging cost function of the form above is right-invariant to a multiplication of the form $\mathcal{R} \to \mathcal{R}\mathtt{S}$ which maps rotations $\mathtt{R}_i \mapsto \mathtt{R}_i\mathtt{S}, \ i \in V$. Given any point $\mathcal{R} \in \text{SO}(3)^n$, the cost function is constant on the entire *gauge-equivalent* set $\{\mathcal{R}\mathtt{S}, \mathtt{S} \in \text{SO}(3)\}$. The $3n$-dimensional problem domain can be considered to be partitioned into 3-dimensional gauge orbits, and solutions are only unique up to the choice of orbit representative. An extra *gauge-fixing* constraint can be used to pick the orbit representative uniquely. The simplest method of gauge-fixing is to pick one vertex arbitrarily (say, vertex k), and set $\mathtt{R}_k = \mathbf{I}_3$. Later in the paper, we will consider an alternate approach that factors out the symmetry. Interestingly, Briales and Gonzalez-Jimenez [7] have observed faster solver performance when moving from simple vertex fixing to a scheme similar to ours.

4. Some Empirical Observations

In this section, we demonstrate a way to visualize the spatial distribution of local minima in rotation averaging problems. It shows us that structural factors (such as edge density and connectivity) can greatly affect the distribution. We will describe how this relates to problem tractability.

Motivation for Empirical Mapping. A key failure mode of iterative solvers is finding a local minimum of poor quality. This problem can be largely avoided through good initial guesses (perhaps via a relaxation of the problem). Even then, it is helpful to have some guidance as to what quality of guess is necessary. Indeed, for some optimization problems, sufficiently good initial guesses are rare. In this section we study the problem experimentally, and then in Section 5 we attempt to explain our observations theoretically.

We begin by fixing a problem instance. Then we collect many local minima: we choose N_{lm} initial guesses sampled uniformly at random in $\text{SO}(3)^n$ and then run a Levenberg-Marquardt solver [2] to efficiently iteratively minimize an l_p cost. This gives us N_{lm} (possibly not distinct) local minima.

Now that we have a large, randomly sampled collection of local minima, we visualize their distribution. Because the local minima live in the search space $\text{SO}(3)^n$, a $3n$-dimensional manifold, we are unable to plot them directly.

Instead, we first compute (gauge-aligned) pairwise distances between each of the N_{lm} minima. This effectively gives us a graph with labeled edge lengths. We use a spectral projection to embed this graph in a 2D plane. These 2D embeddings are show in Figure 1 and described in greater detail below.

Embedding Experimental Details. Each row in Figure 1 is a series of experiments on the same graph; see Table 1 for details. To create each problem instance, we select a ground truth solution for the given graph by choosing elements of SO(3) uniformly at random. We then generate noisy edge measurements by synthetically adding noise (zero-mean and variance σ_n^2 Gaussian in the tangent space) to each edge's ground truth relative rotation. The columns of Figure 1 are increasing noise levels. Within a row, only σ_n varies. The graph and ground truth solution are fixed.

We use $N_{lm} = 200$ random starts for each experiment. For the spectral graph embedding, we use the diffusion kernel $d \mapsto e^{-d^2/\sigma_d}$ to transform distances to similarities, where d is the distance between two local minima. We used $\sigma_d = \frac{\pi}{4}$ for all of our experiments. The synthetic noise σ_n is labeled on each experiment in Figure 1.

row	graph parameters	connectivity	$\alpha_{\max}$ [11]
	Watts-Strogatz small world graphs		
(1)	$n = 40, k = 16, p = 0.0$	$\lambda_2(G) = 4.6$	14.7°
(2)	$n = 40, k = 16, p = 0.2$	$\lambda_2(G) = 7.7$	19.0°
(3)	$n = 40, k = 16, p = 0.5$	$\lambda_2(G) = 8.2$	19.2°
(4*)	$n = 40, k = 16, p = 1.0$	$\lambda_2(G) = 9.5$	20.2°
	G_{nm} random graphs		
(5)	$n = 40, m = 200$	$\lambda_2(G) = 4.2$	13.6 °
(4*)	$n = 40, m = 240$	$\lambda_2(G) = 9.5$	20.2 °
(6)	$n = 40, m = 400$	$\lambda_2(G) = 12.0$	12.0 °

Table 1. Descriptions of the graphs in Figure 1. The algebraic connectivity $\lambda_2(G)$ features in the results in Section 5. The quantity $\alpha_{\max}$ is from [11]: their solution is certifiably optimal if all residuals are less than $\alpha_{\max}$. *(*) Note that graph (4) is an instance of both Watts-Strogatz and G_{nm} random graphs when $p = 1$.*

How to Interpret the Embeddings. Each small colored circle in the plots is a local minimum. Within each plot, absolute distances are not meaningful, but relative distances within each plot correspond (imperfectly) to distances between minima. Gauge is quotiented out when computing distances, as described in Section 5. The colors and radii of the points are both proportionate to the objective function of the minima (cool colors are lower cost and warm colors are higher).

Jittered points are also plotted behind in light grey to reveal multiplicity. For example, the jitter reveals that of the four unique minima in the plot at row 3, column 1, the minimum with lowest cost is strongly dominant. The dominance of the largest minimum ($\%_{\max}$) is labeled on each experiment. This is the percent of the N_{lm} runs that reached the highest multiplicity minimum.

The qualitative results in Figure 1 are stable under repeated trials. Different random graphs and random synthetic noise always result in different 2D embeddings, but all of the trends discussed below are still present.

Observations and Trends. Rows (1) through (4) are Watts-Strogatz small world random graphs [19]. This family of graphs has a rewiring parameter p that smoothly interpolates between a k-connected cycle graph and G_{nm}, the random graph model that randomly samples from all graphs with exactly n vertices and m edges.

These graphs are interesting to us because G_{nm} graphs (and the closely related G_{np} Erdős Rényi graphs) are not that representative of real rotation averaging problems. For their size, they have small radius and high algebraic connectivity. By varying the Watts-Strogatz rewiring parameter p we keep the size of the graph fixed but vary the connectivity. Furthermore, the Watts-Strogatz model comes closer to capturing a qualitative property of image graphs: cameras which have similar pose are more likely to overlap, yet dissimilar cameras do occasionally see the same thing (perhaps on a tall or prominent structure).

Row (1) is the least-connected extreme of the Watts-Strogatz family, with the lowest algebraic connectivity (Table 1). Row (4) is the opposite: with 100% rewiring, this is just a G_{nm} random graph with 240 edges.

Some problem instances look very easy to solve. For instance, in row (3) column 1, 97.5% of the random initializations were sufficient to find the dominant minimum. It certainly seems plausible that this minimum is in fact the global optimum. If such a high percentage of random initial guesses suffice, then a non-random guess (such as from a relaxed solver) seems very likely to succeed.

In contrast, consider row (1) column 5. This is a much noisier problem. Now there are many distinct local minima, many of which are very close to each other. An iterative solver may have difficulty finding a true global optimum.

Finally, consider the trends in these graphs:

- At low noise levels there are very few distinct local minima. One of these is usually dominant and appears to be the global optimum.
- As noise levels increase, more distinct local minima appear, and they get closer to the dominant minimum.
- At fixed noise levels, the more connected graphs look easier. They have fewer distinct minima and these are better separated.
- Going from rows (1) to (2) to (3), the connectivity goes up and the problems appear to get easier. However, rows (3) and (4) look very similar. The extra connectivity does not appear to have much affect. Is there a critical level of connectivity?

Rows (4) - (6) vary a different graph parameter. Recall that the fully rewired Watts-Strogatz graph in (4) is also a
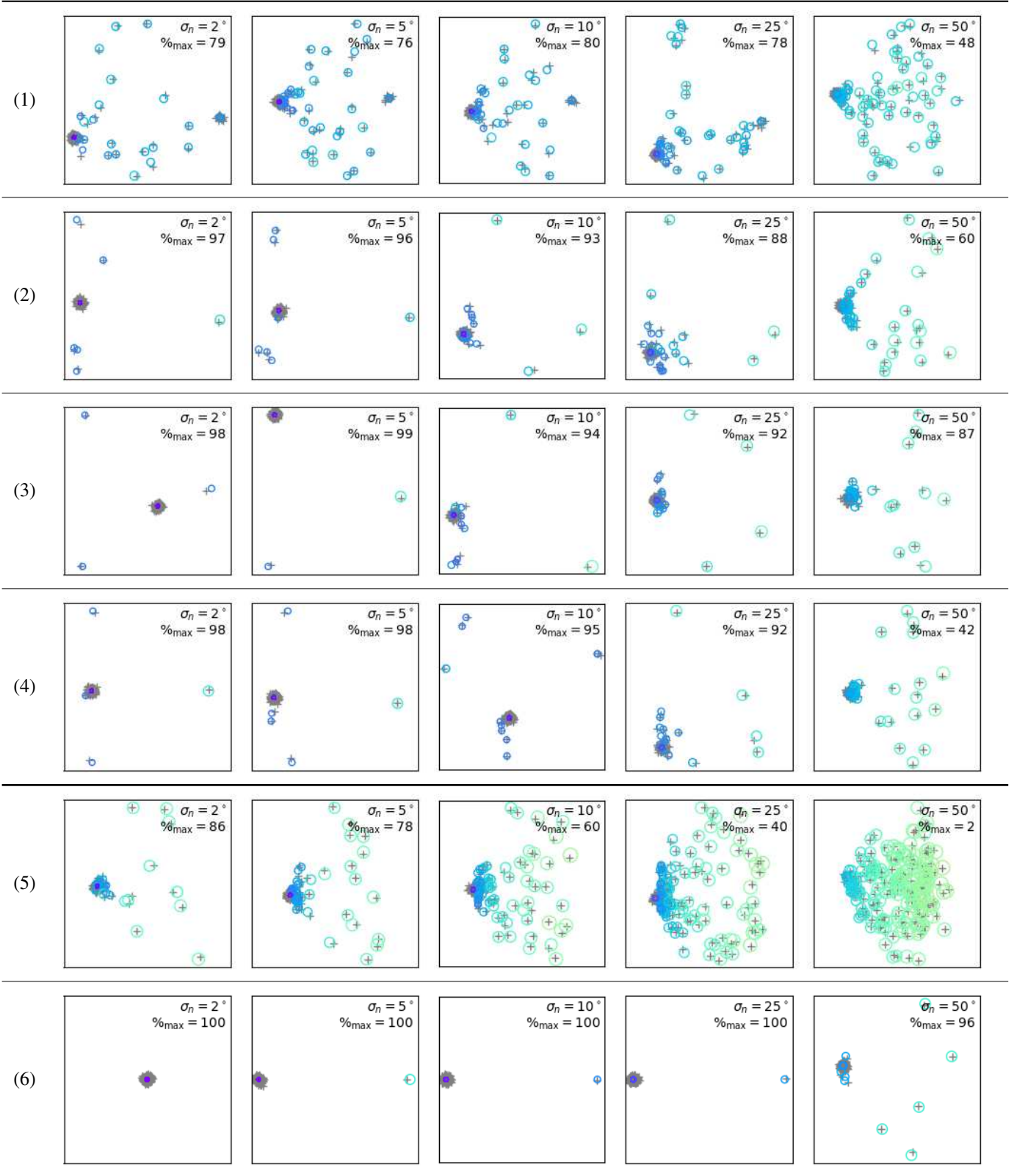
Figure 1. 2D embeddings of the local minima of rotation averaging problems. Each row is a fixed underlying graph, and the columns are increasing noise levels (σ_n). Both color and circle radius encode the objective function (cool colors and small radii are low cost). To reveal multiplicity, a small jittered + is plotted in grey behind each point. Additionally, $\%_{\max}$ gives the multiplicity of the most common minimum. Notice that some problem/noise level combinations have essentially one unique dominant local minimum, while others have many. Table 1 describes the graphs for each row.

G_{nm} random graph with 240 edges. Row (5) is a G_{nm} graph with a few fewer edges, and (6) with a few more. Notice how sharply these changes affect the plots of minima! Row (5), which has fewer edges, appears with only $2°$ of noise to look as messy as row (4) does with $25°$. Row (6) is still well-behaved with a remarkable $25°$ of noise.

Practical Takeaways. These visualizations indicate that some instances of the l_p rotation averaging problem are easy—in the sense that a single minimum dominates the cost surface. This is presumably the global optimum, and almost any initialization is adequate to find it.

On the other hand, some problems are qualitatively different. They have many local minima, and these minima are close together. Furthermore, there may even be minima that are very similar in cost but not close to each other. This raises important questions about whether the problem is well-enough posed for solutions (even the global solution, were it available) to be useful.

In these experiments, for the same number of nodes and edges, more connected graphs were better. Unsurprisingly, for all else equal, lower noise was also better. See the supplemental for a discussion of outlier edges.

Looking at the last column of Table 1, we see that the Eriksson *et al.*'s certification [11] is sufficient, but not necessary. The bound behaves as expected on the Watts-Strogatz graphs, but not on the G_{nm} problems. This may be due to fragile dependence on the maximum degree of the graph.

5. Improved Local Convexity Bounds

Now we turn to seeking a theoretical description of the effects shown in Section 4. We proceed by studying the extent of *local convexity* in a rotation averaging problem instance. We generally follow the calculations in [24], but we add two pieces: (1) the calculations are generalized from l_2 to l_p costs, and (2) we describe the gauge ambiguity in a geometrically natural way, leading to a large improvement in the bounds.

Convex problems enjoy a prominent place in optimization because they provide a guarantee: all local minima are global minima. For rotation averaging, this helpful property does not hold. However, we can proceed with an analysis using *local convexity*, a far weaker, but closely related property.

Definition 1 *A rotation averaging problem $(G, \widetilde{\mathcal{R}})$ is locally convex at a solution $\mathcal{R}$ if there exists an open ball around $\mathcal{R}$ on which the problem is convex.*

Practically speaking, for functions that are at least twice differentiable, local convexity is a statement that the Hessian matrix of the problem is positive semi-definite.[1]

---

[1] This will serve for most of our analysis, but when $1 \leq p < 2$ and at a point where a residual is zero we lack differentiability. Instead, we look at the cost function along geodesics parameterized at constant speed within a
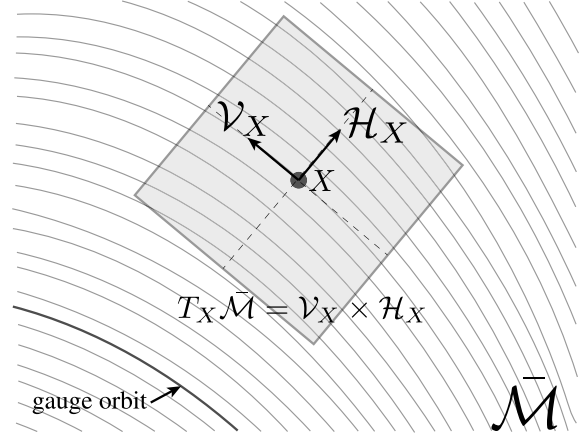

Figure 2. A picture of a 2D manifold $\bar{\mathcal{M}}$ partitioned into 1D orbits. The tangent space $T_X\bar{\mathcal{M}}$ at any point $X \in \bar{\mathcal{M}}$ can be decomposed as the product of two orthogonal vector spaces: a vertical space (which follows the orbits) and a horizontal space (which is orthogonal to the orbits).

In the context of rotations averaging, local convexity limits local minima, because if we have local convexity on a convex region, there cannot be distinct local minima there.

In [24], it was shown that the l_2 geodesic variant of the rotations averaging cost function frequently exhibits local convexity, and that the extent of that local convexity is bounded by properties of the problem instance. The crux of that argument was an appreciation of the *gauge ambiguity*—the local convexity in the problem was only visible after the gauge had been treated. In this section, we revisit that treatment with a more natural description of the gauge.

5.1. A more natural perspective on gauge

By gauge ambiguity, we mean that our cost function is invariant to right-multiplication by any element of $\mathrm{SO}(3)$. This is a free group action of $\mathrm{SO}(3)$ on $\mathrm{SO}(3)^n$, leading us to consider our optimization problem over the quotient manifold $\mathrm{SO}(3)^n / (\mathcal{R} \sim \mathcal{R}\mathbf{S})$. (The natural metric on this manifold is $d_{\mathrm{quot}}(\mathcal{R}_1, \mathcal{R}_2) = \min_{\mathbf{S} \in \mathrm{SO}(3)} d(\mathcal{R}_1, \mathcal{R}_2\mathbf{S})$.) This quotient no longer has a gauge ambiguity; we have exactly removed the symmetry in question. We would like to analyze the local convexity of this new problem, which raises the question, "How do derivatives of a function on a manifold relate to derivatives of that function on a quotient manifold?"

Quotients of Riemannian manifolds have a particularly nice structure [1, Ch. 3]. Points in the quotient space correspond to gauge-equivalent orbits. A tangent space $T_X\mathcal{M}$ on the quotient manifold is related to the tangent space $T_X\bar{\mathcal{M}}$ on $\bar{\mathcal{M}} = \mathrm{SO}(3)^n$ as shown in Figure 2. In particular, note

---

small ball. We only need consider geodesics which pass exactly through the point. Finally, we note that along such a geodesic the cost function is a scaled version of $f(\theta) = |\theta|^p$, which is known to be convex for $p \geq 1$.

that the original tangent space $T_X\bar{\mathcal{M}}$ is a direct product of two orthogonal vector spaces. The vertical space is the subspace consisting of motions that remain in the same gauge orbit, and the horizontal space is the space of motions orthogonal to that. Notably, the horizontal space is identified with the quotient manifold's tangent space: $\mathcal{H}_X = T_X\mathcal{M}$. Furthermore, through this identification, derivatives of functions on quotient manifolds are simply the projections onto $\mathcal{H}$ of derivatives in the original manifold.

We can explicitly state what these spaces are for our problem. A perturbation in a tangent space to $\mathrm{SO}(3)^n$ can be expressed as an n-tuple $(\mathbf{x_1}, \ldots, \mathbf{x_n})$, where the $\mathbf{x}_i$'s are all in $\mathbb{R}^3$. (Properly speaking, then the tangent space elements are the skew symmetric matrices $[\mathbf{x}_i]_\times$.) Each such tangent space element is mapped through the $\exp$ map to a new element of $\mathrm{SO}(3)^n$:

$$\exp_{\mathcal{R}}(\mathbf{x_1}, \ldots, \mathbf{x_n}) = \begin{bmatrix} \mathrm{R_1}\exp[\mathbf{x_1}]_\times \\ \vdots \\ \mathrm{R_n}\exp[\mathbf{x_n}]_\times \end{bmatrix}.$$

If this motion is gauge-preserving we must have

$$\exp[\mathbf{x_1}]_\times = \ldots = \exp[\mathbf{x_n}]_\times = \mathtt{S}$$

for some $\mathtt{S} \in \mathrm{SO}(3)$. The inverse of the $\exp$ map is $\log$, which maps from the manifold into the tangent space. For small $\mathbf{x}$, $\log[\mathbf{x}]_\times$ is one-to-one, so $\mathbf{x_1} = \mathbf{x_2} = \ldots = \mathbf{x}_n$. In other words, in Kronecker product notation, the vertical space is this subspace of $\mathbb{R}^{3n}$:

$$\mathcal{V} = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes \mathbf{1}_n, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \otimes \mathbf{1}_n, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \otimes \mathbf{1}_n. \right\} \tag{4}$$

It follows that the horizontal space is $\mathcal{H} = \mathcal{V}^\perp$. Interestingly, $\mathcal{V}$ and $\mathcal{H}$ depend only of n. The tangent spaces at every point $\mathcal{R}$ on the manifold share the same description of the vertical and horizontal spaces.

Returning to the original question, we wish to analyze the local convexity of the quotient-manifold version of the l_p rotation averaging problem. Recall that (given sufficient differentiability) local convexity holds when all eigenvalues of the Hessian matrix are non-negative. This will hold exactly when the projection of the Hessian matrix onto $\mathcal{V}^\perp$ is positive semi-definite:

Theorem 1 *A rotation averaging problem is locally convex with respect to the l_p geodesic cost function when*

$$\min_{\mathbf{x}\in\mathbf{V}^\perp,\|\mathbf{x}\|=1} \mathbf{x}^\top\mathbf{H}\mathbf{x} > 0, \tag{5}$$

where $\mathbf{H}$ is the Hessian of the l_p cost function (on $\mathrm{SO}(3)^n$; see below).

5.2. Some Lower-Bound Approximations

Now that we have a better way to factor the gauge ambiguity out of our local convexity analysis, we begin with the Hessian matrix computed in [24] and follow it through two approximations. The smallest eigenvalue of the (gauge-projected) Hessian matrix describes how the graph structure of the problem interacts with the magnitude and directions of residuals, determining local convexity. Unfortunately, this description is complicated. In this section we aim to get insight through sacrificing some accuracy for interpretability. In each case, our new gauge approach yields superior bounds over analogous prior results [24].

The Structure of the Hessian. We begin by giving the exact Hessian for the unquotiented l_p cost function, as derived by following the chain rule calculations in [24] for $p > 1$. Recall that the cost function (Equation (3)) is a sum of terms on each edge in the graph. Likewise, the Hessian is a sum of terms for each edge in the cost function:

$$\mathbf{H} = \sum_{(i,j)\in E} \mathbf{H}_{ij}. \tag{6}$$

where the terms are $3n$-by-$3n$ matrices composed of 3-by-3 blocks. For each term, exactly four 3-by-3 blocks (corresponding to the ith and jth block rows and columns) are non-zero. These blocks are expressed in terms of the residuals $\mathtt{R}_i^\top\widetilde{\mathtt{R}}_{ij}\mathtt{R}_j$ on each edge. We write these residuals in their tangent spaces as $[\theta_{ij}\hat{\mathbf{w}}_{ij}]_\times = \log\left(\mathtt{R}_i^\top\widetilde{\mathtt{R}}_{ij}\mathtt{R}_j\right)$. So, where θ_{ij} is the magnitude of the residual rotation on edge (i, j), and $\mathbf{w}_{ij}$ is its axis of rotation.

$$\mathbf{H}_{ij} = \begin{array}{c} \\ i \\ \\ j \\ \\ \end{array} \begin{bmatrix} & \vdots & & \vdots & \\ \ldots & \mathbf{S}_{ij} & \ldots & -\mathbf{S}_{ij}+\mathbf{A}_{ij}^\top & \ldots \\ & \vdots & & \vdots & \\ \ldots & -\mathbf{S}_{ij}+\mathbf{A}_{ij} & \ldots & \mathbf{S}_{ij} & \ldots \\ & \vdots & & \vdots & \end{bmatrix}, \tag{7}$$

where the 3-by-3 blocks making up each $\mathbf{H}_{ij}$ are

$$\mathbf{S}_{ij} = p(p-1)\theta_{ij}^{p-2}\mathbf{w}_{ij}\mathbf{w}_{ij}^\top + \frac{p}{2}\theta_{ij}^{p-1}(\mathbf{I}_3 - \mathbf{w}_{ij}\mathbf{w}_{ij}^\top),$$

$$\mathbf{A}_{ij} = \frac{p}{2}\theta_{ij}^{p-1}[\mathbf{w}_{ij}]_\times.$$

Notice that the symmetric $\mathbf{S}_{ij}$ terms of the Hessian are arranged in a graph-Laplacian like structure, but the antisymmetric $\mathbf{A}_{ij}$ terms are not.

An Isotropic Bound. A first approach to taming this Hessian is to find a bound that depends only on residual magnitudes. The best isotropic (direction independent) lower bound for the $\mathbf{S}_{ij}$ terms is

$$\mathbf{S}_{ij} \succ \min\left(\frac{p}{2}\theta_{ij}^{p-1}, p(p-1)\theta_{ij}^{p-2} - \frac{p}{2}\theta_{ij}^{p-1}\right)\mathbf{I}_3. \tag{8}$$

Let's call those values α_{ij}, so that we have $\mathbf{S}_{ij} \succ \alpha_{ij}\mathbf{I}_3$. Notice that α_{ij} is purely a convenience function of θ_{ij}. Now, the indefinite skew contribution $\mathbf{S}_{ij}$ can be isotropically bounded above as a block:

$$\begin{bmatrix} \mathbf{0} & \mathbf{A}_{ij}^\top \\ \mathbf{A}_{ij} & \mathbf{0} \end{bmatrix} \prec \frac{p}{2}\theta_{ij}^{p-1} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \end{bmatrix}. \qquad (9)$$

Continuing as in [24], we see that for a single Hessian term,

$$\mathbf{H}_{ij} \succ \left(\underbrace{\begin{bmatrix} \alpha_{ij} & -\alpha_{ij} \\ -\alpha_{ij} & \alpha_{ij} \end{bmatrix}}_{\text{PSD Laplacian term}} - \underbrace{\frac{p}{2}\theta_{ij}^{p-1} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \end{bmatrix}}_{\text{non-PSD curvature term}} \right) \otimes \mathbf{I}_3. \qquad (10)$$

Considering all the residual terms together, we will write this as $\mathbf{H} \succ (\mathbf{L}(\alpha_{ij}) - \mathbf{D}(\theta_{ij})) \otimes \mathbf{I}_3$. $\mathbf{L}$ is a standard graph Laplacian matrix [22] with edge weights α_{ij}, and $\mathbf{D}$ is a diagonal degree matrix with elements:

$$\mathbf{D}_{ii} = \sum_{j|(i,j)\in E} \frac{p}{2}\theta_{ij}^{p-1}. \qquad (11)$$

Multiplying by $\mathbf{D}^{-\frac{1}{2}}$ on both sides, we see that this is sufficient for $\mathbf{H} \succ \mathbf{0}$:

$$\mathbf{L}_{\text{norm}} = \left(\mathbf{D}(\theta_{ij})^{-\frac{1}{2}} \mathbf{L}(\alpha_{ij}) \mathbf{D}(\theta_{ij})^{-\frac{1}{2}} \right) \succ \mathbf{I}_n. \qquad (12)$$

The smallest eigenvalue of a graph Laplacian is 0, so this condition appears to be unsatisfiable. We must use what we know about gauge to proceed. Recall that considering the problem from a quotient manifold perspective effectively meant projecting the tangent space off of the vertical space and onto the horizontal space. This means that (allowing for the Kronecker product), we require:

$$\min_{\mathbf{x}\perp\mathbf{1}_n, \|\mathbf{x}\|=1} \mathbf{x}^\top \mathbf{L}_{\text{norm}} \mathbf{x} > 1. \qquad (13)$$

That is, $\mathbf{L}_{\text{norm}}$ is an particular weighted, normalized graph Laplacian. It represents the interaction of graph topology and residual noisiness (but not residual directions, which we approximated away), it bounds local convexity. Projecting this matrix off of the gauge vector $\mathbf{1}_n$ represents considering the problem as being solved on a quotient manifold.

An Even Simpler Bound. If we are a bit more destructive, we can reach a bound that cleanly separates structure and noise terms. Let's begin back with this sufficient condition for local convexity:

$$\mathbf{L}(\alpha_{ij}) \succ \mathbf{D}(\theta_{ij}). \qquad (14)$$

This next condition is still sufficient, but much weaker:

$$\lambda_{\min}(\mathbf{L}(\alpha_{ij})) > \lambda_{\max}(\mathbf{D}(\theta_{ij})). \qquad (15)$$

Now, the eigenvector corresponding to the zero eigenvalue of a graph Laplacian is $\mathbf{1}_n$, the vector of all ones. This means that (allowing for the Kronecker product) the gauge subspace corresponds perfectly with the 0-eigenvector of the graph Laplacian. Note that the largest value of a diagonal matrix is its largest entry, so we have a bound:

$$\lambda_2(\mathbf{L}(\alpha_{ij})) > \max_{i\in V} \sum_{j|(i,j)\in E} \frac{p}{2}\theta_{ij}^{p-1}. \qquad (16)$$

Note that in the $p=2$ case, the right hand side is simply the maximum weighted degree of the residual graph. If we wish to bound even more aggressively, we can bound by the smallest α, giving:

$$\underbrace{\lambda_2(\mathbf{L})}_{structure} > \underbrace{\frac{\max_{i\in V} \sum_{(i,j)\in E} \frac{p}{2}\theta_{ij}^{p-1}}{\min_{ij} \alpha_{ij}}}_{residuals}. \qquad (17)$$

This gives a perfect separation of the condition for local convexity into a graph structure term (the so-called *algebraic connectivity*) a term that depends on residual noise levels. Notice that this bound differs from the one given in [24] in that a factor of n has disappeared from the denominator of the structure term in Equation (17). This is solely a consequence of the improved gauge treatment substantially increases the practical utility of the bound.

6. Application and Conclusion

In this paper we studied the cost surface of the geodesic l_p rotation averaging problem. We gave empirical evidence that changes in problem structure and noise can cause sharp transitions between easy and challenging distributions of local minima. We then derived a bound on this behavior: connectivity must balance noise to get a single dominant minimum. These bounds are an n-fold improvement over prior work, achieved through treating the problem's gauge ambiguity in a geometrically natural way.

We see this analysis being useful for the development of new intrinsic-metric rotation averaging solvers. Our theoretical work shows how to remain in the easy regime. We hope that this analysis can provide guidance in pipeline design for forming rotation averaging problem instances.

There is a particular application to hierarchical / merging-based solver schemes. In these approaches (which are a current area of active study) small subproblems are selected, solved, and merged into a larger solution. Our results are especially applicable here because unlike in single-shot schemes where the problem is provided as-is, merging solvers have free choice of their subproblems. Schemes that can maintain elevated connectivity may do better at avoiding local minima traps.

References

[1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009. 6

[2] Sameer Agarwal, Keir Mierle, and Others. Ceres solver. http://ceres-solver.org. 3

[3] M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri. Global motion estimation from point matches. In *3DIMPVT*, 2012. 2

[4] F. Arrigoni, L. Magri, B. Rossi, P. Fragneto, and A. Fusiello. Robust absolute rotation estimation via low-rank and sparse matrix decomposition. In *3DV*, 2014. 2

[5] N. Boumal, A. Singer, and P-A. Absil. Robust estimation of rotations from relative measurements by maximum likelihood. In *Decision and Control*. IEEE, 2013. 2

[6] N. Boumal, A. Singer, P-A. Absil, and V. D. Blondel. Cramér–Rao bounds for synchronization of rotations. *Information and Inference*, 2014. 2

[7] J. Briales and J. Gonzales-Jimenez. Fast global optimality verification in 3D SLAM. *IROS*, 2016. 2, 3

[8] J. Briales and J. Gonzalez-Jimenez. Cartan-sync: Fast and global SE(d)-synchronization. *Robotics and Automation Letters*, 2017. 2

[9] 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. *ICRA*, 2015. 1, 2

[10] A. Chatterjee and V. M. Govindu. Robust relative rotation averaging. *PAMI*, 2017. 2

[11] A. Eriksson, C. Olsson, F. Kahl, and T. Chin. Rotation averaging and strong duality. *CVPR*, 2018. 2, 4, 6

[12] J. Fredriksson and C. Olsson. Simultaneous multiple rotation averaging using Lagrangian duality. In *ACCV*, 2012. 2

[13] V. M. Govindu. Combining two-view constraints for motion estimation. In *CVPR*, 2001. 2

[14] V. M. Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *CVPR*, 2004. 2

[15] G. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision (2nd Ed.)*. Cambridge University Press, 2004. 1

[16] R. Hartley, J. Trumpf, Y. Dai, and H.. Li. Rotation averaging. In *IJCV*, 2013. 2

[17] D. Martinec and T. Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *CVPR*, 2007. 2

[18] D. Rosen, L. Carlone, A. Bandiera, and J. Leonard. Se-sync: A certifiably correct algorithm for synchronization over the special euclidean group. *IJRR*, 2019. 2

[19] S. H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, March 2001. 4

[20] R. Tron, R. Vidal, and A. Terzis. Distributed pose averaging in camera networks via consensus in SE(3). In *ICDSC*, 2008. 2

[21] R. Tron, X. Zhou, and K. Daniilidis. A survey on rotation optimization in structure from motion. In *CVPR Workshops*, 2016. 1, 2

[22] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 2007. 8

[23] L. Wang and A. Singer. Exact and stable recovery of rotations for robust synchronization. In *Information and Inference*. IMA, 2013. 2

[24] K. Wilson, D. Bindel, and N. Snavely. When is rotations averaging hard? *ECCV*, 2016. 1, 2, 6, 7, 8