Rotation Averaging and Strong Duality

Anders Eriksson[1], Carl Olsson[2,3], Fredrik Kahl[2,3] and Tat-Jun Chin[4]

[1]School of Electrical Engineering and Computer Science, Queensland University of Technology
[2]Department of Electrical Engineering, Chalmers University of Technology
[3]Centre for Mathematical Sciences, Lund University
[4]School of Computer Science, The University of Adelaide

Abstract

In this paper we explore the role of duality principles within the problem of rotation averaging, a fundamental task in a wide range of computer vision applications. In its conventional form, rotation averaging is stated as a minimization over multiple rotation constraints. As these constraints are non-convex, this problem is generally considered challenging to solve globally. We show how to circumvent this difficulty through the use of Lagrangian duality. While such an approach is well-known it is normally not guaranteed to provide a tight relaxation. Based on spectral graph theory, we analytically prove that in many cases there is no duality gap unless the noise levels are severe. This allows us to obtain certifiably global solutions to a class of important non-convex problems in polynomial time.

We also propose an efficient, scalable algorithm that outperforms general purpose numerical solvers and is able to handle the large problem instances commonly occurring in structure from motion settings. The potential of this proposed method is demonstrated on a number of different problems, consisting of both synthetic and real-world data.

1. Introduction

Rotation averaging appears as a subproblem in many important applications in computer vision, robotics, sensor networks and related areas. Given a number of relative rotation estimates between pairs of poses, the goal is to compute absolute camera orientations with respect to some common coordinate system. In computer vision, for instance, non-sequential structure from motion systems such as [21, 11, 22] rely on rotation averaging to initialize bundle adjustment. The overall idea is to consider as much data as possible in each step to avoid suboptimal reconstructions. In the context of rotation averaging this amounts to using as many camera pairs as possible.

The problem can be thought of as inference on the camera graph. An edge (i, j) in this undirected graph represents a relative rotation measurement $\tilde{R}_{ij}$ and the objective is to find the absolute orientation R_i for each vertex i such that $R_i\tilde{R}_{ij} = R_j$ holds (approximately in the presence of noise) for all edges. The problem is generally considered difficult due to the need to enforce non-convex rotation constraints. Indeed, both L_1 and L_2 formulations of rotation averaging can have local minima, see Fig. 1. Wilson *et al.* [28] studied local convexity of the problem and showed that instances with large loosely connected graphs are hard to solve with local, iterative optimization methods.

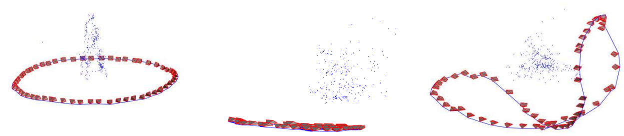

Figure 1: In many structure from motion pipelines, camera orientations are estimated with rotation averaging followed by recovery of camera centres (red) and 3D structure (blue). Here are three solutions corresponding to different local minima of the same rotation averaging problem.

In contrast, our focus is on global optimality. In this paper we show that convex relaxation methods can in fact overcome the difficulties with local minima in rotation averaging. We utilize Lagrangian duality to handle the quadratic non-convex rotation constraints. While such an approach is normally not guaranteed to provide a tight relaxation we give analytical error bounds that guarantee there will be no duality gap. For instance, it is sufficient that each angular residual is less than $42.9°$ to ensure optimality for complete camera graphs. Additionally, we develop a scalable and efficient algorithm, based on block coordinate descent, that outperforms standard semidefinite program (SDP) solvers for this problem.

Related work. Rotation averaging has been under intense study in recent years, see [19, 20, 21, 2, 25, 8]. Despite progress in practical algorithms, they largely come without guarantees. One of the earliest averaging methods was due to Govindu [15], who showed that when representing the rotations with quaternions the problem can be viewed as a linear homogeneous least squares problem. There is however a sign ambiguity in the quaternion representation that has to be resolved before the formulation can be applied. It was observed by Fredriksson and Olsson in [14] that since both the objective and the constraints are quadratic, the Lagrange dual can be computed in closed form. The resulting SDP was experimentally shown to have no duality gap for moderate noise levels.

A more straightforward rotation representation is 3×3 matrices. Martinec and Pajdla [21] approximately solve the problem by ignoring the orthogonality and determinant constraints. A similar relaxation was derived by Arie-Nachimson *et al.* in [1]. In addition, an SDP formulation was presented which is equivalent to the one we address here, but with no performance guarantees. The tightness of SDP relaxations for 2D rotation averaging is studied in [30].

A number of robust approaches have been developed to handle outlier measurements. A sampling scheme over spanning trees of the camera graph is developed by Govindu in [16]. Enqvist *et al.* [11] also start from a spanning tree and add relative rotations that are consistent with the solution. In [17] the Weiszfeld algorithm is applied to single rotation averaging with the L_1 norm. In [18] convexity properties of the single rotation averaging problem are given. To our knowledge these results do not generalize to the case of multiple rotations. In [9] a robust formulation is solved using IRLS and in [3] Cramér-Rao lower bounds are computed for maximum likelihood estimators, but neither with any optimality guarantees.

A closely related problem is that of pose graph estimation, where camera orientations and positions are jointly optimized. In this context Lagrangian duality has been applied [6, 7]. In [26] a consensus algorithm that allows for efficient distributed computations is presented. A fast verification technique for pose graph estimation was given in [5]. In a recent paper [23] an SDP relaxation for pose graph estimation with performance guarantees is analyzed. It is shown that there is a noise level β for which the relaxation is guaranteed to provide the optimal solution. However, the result only shows the existence of β. Its value which is dependent on the problem instance is not computed. In contrast our result for rotation averaging gives explicit noise bounds.

The main contributions of this paper are:

- We apply Lagrangian duality to the rotation averaging problem with the chordal error distance and study the properties of the obtained relaxations.
- We develop strong theoretical bounds on the noise level that guarantee exact global recovery based on spectral graph theory.
- We develop a conceptually simple and scalable algorithm which is able to handle large problem instances occurring in structure from motion problems.
- We present experimental results that confirm our theoretical findings.

1.1. Notation and Conventions

Let $G = (V, E)$ denote an undirected graph with vertex set V and edge set E and let $n = |V|$. The adjacency matrix A is by definition the $n \times n$ matrix with elements

$$a_{ij} = \begin{cases} 0 & (i,j) \notin E \\ 1 & (i,j) \in E \end{cases} \quad \text{for } i,j = 1,\dots,n. \tag{1}$$

The degree d_i is the number of edges that touch vertex i, and the degree matrix D is the diagonal matrix $D = \operatorname{diag}(d_1, \dots, d_n)$. The Laplacian L_G of G is defined by

$$L_G = D - A. \tag{2}$$

It is well-known that L_G has a zero eigenvalue with multiplicity 1. The second smallest eigenvalue λ_2 of L_G, also known as the *Fiedler value*, reflects the connectivity of G. For a connected graph G, which is the only case of interest to us, we always have $\lambda_2 > 0$.

The group of all rotations about the origin in three dimensional Euclidean space is the *Special Orthogonal Group*, denoted $\mathrm{SO}(3)$. This group is commonly represented by rotation matrices, orthogonal 3×3 real-valued matrices with positive determinant, i.e.,

$$\mathrm{SO}(3) \in \{R \in \mathbb{R}^{3\times 3} \mid R^T R = I,\ \det(R) = 1\}. \tag{3}$$

If we omit $\det(R) = 1$, we get the *Orthogonal Group*, $\mathrm{O}(3)$.

We will use the convention that $\lambda_i(A)$ is the i:th smallest eigenvalue of the symmetric matrix A. The trace of matrix A is denoted by $\operatorname{tr}(A)$ and the Kronecker product of matrices A and B by $A \otimes B$. The norm $\|A\|$ is the standard operator 2-norm and $\|A\|_F$ the Frobenius norm.

2. Problem Statement

The problem of rotation averaging is defined as the task of determining a set of n absolute rotations $R_1, \dots, R_n$ given distinct estimated relative rotations $\tilde{R}_{ij}$. Available relative rotations are represented by the edge set E of the camera graph V. Under ideal conditions this amounts to finding the n rotations compatible with the linear relations,

$$R_i \tilde{R}_{ij} = R_j, \tag{4}$$

for all $(i,j) \in E$. However, in the presence of noise, a solution to (4) is not guaranteed to exist. Instead, it is typically solved in a least-metric sense,

$$\min_{R_1,\dots,R_n} \sum_{(i,j)\in E} d(R_i \tilde{R}_{ij}, R_j)^p, \tag{5}$$

where $p \geq 1$ and $d(\cdot,\cdot)$ is a distance function.

A number of distinct choices of metrics on $\mathrm{SO}(3)$ exist, see Hartley *et al.* [19] for a comprehensive discussion. In this work we restrict ourselves to the chordal distance, the most commonly used metric when analyzing Lagrangian duality in rotation averaging. It has proven to be a convenient choice as it is quadratic in its entries leading to a particularly simple derivation and form of the associated dual problem.

The chordal distance between two rotations R and S is defined as their Euclidean distance in the embedding space,

$$d(R,S) = \|R - S\|_F. \tag{6}$$

It can be shown [19] that the chordal distance can also be written as $d(R,S) = 2\sqrt{2}\sin\frac{|\alpha|}{2}$, where α is the rotation angle of RS^{-1}. With this choice of metric, the rotation averaging problem is defined as

$$\underset{R_1,\dots,R_n \in \mathrm{SO}(3)}{\arg\min} \sum_{(i,j)\in E} \|R_i \tilde{R}_{ij} - R_j\|_F^2, \tag{7}$$

which, with trace notation, can be simplified to

$$\underset{R_1,\dots,R_n \in \mathrm{SO}(3)}{\arg\min} -\sum_{(i,j)\in E} \mathrm{tr}\left(R_i \tilde{R}_{ij} R_j^T\right), \tag{8}$$

which constitutes our *primal problem*.

It will be convenient with a compact matrix formulation. Let

$$\tilde{R} = \begin{bmatrix} 0 & a_{12}\tilde{R}_{12} & \dots & a_{1n}\tilde{R}_{1n} \\ a_{21}\tilde{R}_{21} & 0 & \dots & a_{2n}\tilde{R}_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1}\tilde{R}_{n1} & a_{n2}\tilde{R}_{n2} & \dots & 0 \end{bmatrix}, \tag{9}$$

where $\tilde{R}_{ij} = \tilde{R}_{ji}^T$ and a_{ij} are the elements of the adjacency matrix A of the camera graph G and let

$$R = \begin{bmatrix} R_1 & R_2 & \dots & R_n \end{bmatrix}. \tag{10}$$

We may now write the primal problem as

$$\begin{aligned} (P) \quad & \min && -\mathrm{tr}\left(R\tilde{R}R^T\right) \\ & \text{s.t.} && R \in \mathrm{SO}(3)^n. \end{aligned} \tag{11}$$

3. Optimality Conditions

3.1. Necessary Local Optimality Conditions

We now turn to the KKT conditions of our primal problem (P). The constraint set $R \in \mathrm{SO}(3)^n$ consists of two types of constraints; the orthogonality constraints $R_i^T R_i = I$ and the determinant constraints $\det(R_i) = 1$.

Consider relaxing the rotation averaging problem by removing the determinant constraint,

$$\begin{aligned} (P') \quad & \min && -\mathrm{tr}\left(R\tilde{R}R^T\right) \\ & \text{s.t.} && R \in \mathrm{O}(3)^n. \end{aligned} \tag{12}$$

The constraint $R \in \mathrm{O}(3)^n$ still requires the R_i's to be orthogonal. The orthogonal matrices consist of two disjoint, non-connected sets, with determinants 1 and -1 respectively. Hence, any local minimizer to the problem (P) also has to be a local minimizer, and therefore a KKT point, to (P'). We note that orthogonality can be enforced by restricting the 3×3 diagonal blocks of the symmetric matrix $R^T R$ to be identity matrices. If

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 & 0 & \dots \\ 0 & \Lambda_2 & 0 & \dots \\ 0 & 0 & \Lambda_3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{13}$$

is a symmetric matrix then the Lagrangian can be written

$$\begin{aligned} L(R,\Lambda) &= -\mathrm{tr}\left(R\tilde{R}R^T\right) - \mathrm{tr}\left(\Lambda(I - R^T R)\right) \\ &= \mathrm{tr}\left(R(\Lambda - \tilde{R})R^T\right) - \mathrm{tr}\left(\Lambda\right). \end{aligned} \tag{14}$$

Taking derivatives gives the KKT equations

$$\textit{(Stationarity)} \qquad (\Lambda^* - \tilde{R}){R^*}^T = 0 \tag{15a}$$

$$\textit{(Primal feasibility)} \qquad R^* \in \mathrm{SO}(3)^n. \tag{15b}$$

Equation (15a) states that the rows of a local minimizer R^* will be eigenvectors of the matrix $\Lambda^* - \tilde{R}$ with eigenvalue zero. This allows us to compute the optimal Lagrange multiplier Λ^* from a given minimizer R^*. By (15a) we see that

$$\Lambda_i^* R_i^{*T} = \sum_{j \neq i} a_{ij}\tilde{R}_{ij}R_j^{*T} \iff \Lambda_i^* = \sum_{j \neq i} a_{ij}\tilde{R}_{ij}R_j^{*T}R_i^* \tag{16}$$

for $i = 1, \dots, n$.

Lemma 3.1. *For a stationary point R^* to the primal problem (P), we can compute the corresponding Lagrangian multiplier Λ^* in closed form via (16).*

3.2. Sufficient Global Optimality Conditions

We begin this section by deriving the Lagrange dual of (P) which is a semidefinite program that we will use for optimization in later sections. The dual problem is defined by

$$\max_{\Lambda - \tilde{R} \succeq 0} \min_R L(R,\Lambda). \tag{17}$$

Since the (unrestricted) optimum of $\min_R L(R,\Lambda)$ is either $-\mathrm{tr}\left(\Lambda\right)$, when $\Lambda - \tilde{R} \succeq 0$, or $-\infty$ otherwise, we get

$$(D) \qquad \max_{\Lambda - \tilde{R} \succeq 0} -\mathrm{tr}\left(\Lambda\right). \tag{18}$$

It is clear (through standard duality arguments) that (D) gives a lower bound on (P). Furthermore, if R^* is a stationary point with corresponding Lagrangian multiplier Λ^* that satisfies $\Lambda^* - \tilde{R} \succeq 0$ then Λ^* is feasible in (D) and by (16), $-\mathrm{tr}\left(\Lambda^*\right) = -\mathrm{tr}\left(R^*\tilde{R}{R^*}^T\right)$, which shows that there is no duality gap between (P) and (D). Thus, the convex program (D) provides a way of solving the non-convex (P) when $\Lambda^* - \tilde{R} \succeq 0$.

It also follows that for the stationary point R^* we have $\operatorname{tr}\left(R^*\Lambda^*R^{*T}\right) = \operatorname{tr}\left(R^*\tilde{R}R^{*T}\right)$ due to (15a). We further note that if $\Lambda^* - \tilde{R} \succeq 0$ then by definition it is true that

$$x^T\left(\Lambda^* - \tilde{R}\right)x \geq 0, \tag{19}$$

for any $3n$-vector x. In particular, for any $R \in \mathrm{O}(3)^n$,

$$\begin{aligned} 0 &\leq \operatorname{tr}\left(R(\Lambda^* - \tilde{R})R^T\right) = \operatorname{tr}(\Lambda^*) - \operatorname{tr}\left(R\tilde{R}R^T\right) \\ &= \operatorname{tr}\left(R^*\Lambda^*R^{*T}\right) - \operatorname{tr}\left(R\tilde{R}R^T\right), \end{aligned} \tag{20}$$

which shows that $-\operatorname{tr}\left(R^*\tilde{R}R^{*T}\right) \leq -\operatorname{tr}\left(R\tilde{R}R^T\right)$ for all $R \in \mathrm{O}(3)^n$, that is, R^* is the global optimum.

Lemma 3.2. *If a stationary point R^* with corresponding Lagrangian multiplier Λ^* fulfills $\Lambda^* - \tilde{R} \succeq 0$ then:*

1. *There is no duality gap between (P) and (D).*
2. *R^* is a global minimum for (P).*

In the remainder of this paper we will study under which conditions $\Lambda^* - \tilde{R} \succeq 0$ holds and derive an efficient implementation for solving (D).

4. Main Result

In this section, we will state our main result which gives error bounds that guarentee that that strong duality holds for our primal and dual problems. From a practical point of view, the result means that it is possible to solve a convex semidefinite program and obtain the globally optimal solution to our non-convex problem, which is quite remarkable.

4.1. Strong Duality Theorem

Returning to our initial, primal rotation averaging problem (7). The goal is to find rotations R_i and R_j such that the sum of the residuals $\|R_i\tilde{R}_{ij} - R_j\|_F^2$ is minimized. For strong duality to hold, we need to bound the residual error.

Theorem 4.1 (Strong Duality)**.** *Let R_i^*, $i = 1, \ldots, n$ denote a stationary point to the primal problem (P) for a connected camera graph G with Laplacian L_G. Let α_{ij} denote the angular residuals, i.e., $\alpha_{ij} = \angle(R_i^*\tilde{R}_{ij}, R_j^*)$. Then R_i^*, $i = 1, \ldots, n$ will be globally optimal and strong duality will hold for (P) if*

$$|\alpha_{ij}| \leq \alpha_{\max} \quad \forall (i,j) \in E, \tag{21}$$

where

$$\alpha_{\max} = 2\arcsin\left(\sqrt{\frac{1}{4} + \frac{\lambda_2(L_G)}{2d_{\max}}} - \frac{1}{2}\right), \tag{22}$$

and $d_{\max}$ is the maximal vertex degree.

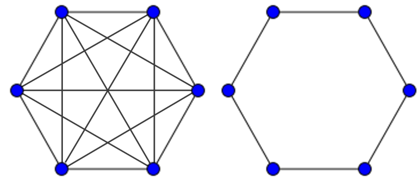

Figure 2: A complete graph (left) and a cycle graph (right), both with 6 vertices.

Note that any local minimizer that fulfills this error bound will be global, and conversely there are no non-global minimizers with error residuals fulfilling (21). It is clear that (22) will give a positive bound $\alpha_{\max}$ for any graph. Thus for any given problem instance, $\alpha_{\max}$ gives an explicit bound on the error residuals for which strong duality is guaranteed to hold. The strength of the bound will depend on the particular graph connectivity encapsulated by the Fiedler value $\lambda_2(L_G)$ and the maximal vertex degree $d_{\max}$. We will see that for tightly connected graphs the bound ensures strong duality under surprisingly generous noise levels. In [28] it was observed that local convexity at a point holds under similar circumstances.

Example. Consider a graph with $n = 3$ vertices that are connected, and all degrees are equal, $d_{max} = 2$. Now from the Laplacian matrix L_G, one easily finds that $\lambda_2 = 3$. This gives $\alpha_{\max} = \frac{\pi}{3}\mathrm{rad} = 60^\circ$. So, any local minimizer which has angular residuals less than 60° is also a global solution.

Complete graphs. Let us turn to a more general class of graphs, namely complete graphs with n vertices, see Fig. 2. As every pair of vertices is connected, it follows that $d_{\max} = n - 1$. Further, it is well-known (and easy to show) that $\lambda_2(L_G) = n$, see [13]. Again, for $n = 3$, we retrieve $\alpha_{\max} = \frac{\pi}{3}\mathrm{rad}$. As n becomes larger, we get a decreasing series of upper bounds which in the limit tends to $2\arcsin(\frac{\sqrt{3}-1}{2}) \approx 0.749\mathrm{rad} = 42.9^\circ$. Hence, as long as the residual angular errors are less than 42.9° - which is quite generous from a practical point of view - we can compute the optimal solution via a convex program. Also note that this bound holds independently of n.

Corollary 4.1. *For a complete graph G with n vertices, the residual upper bound $\alpha_{\max} = 2\arcsin(\frac{\sqrt{3}-1}{2}) \approx 0.749\mathrm{rad} = 42.9^\circ$ ensures global optimality and strong duality for any n.*

Cycle graphs. Now consider the other spectrum in terms of graph connectivity, namely cycle graphs. A cycle graph has a single cycle, or in other words, every vertex in the camera graph has degree two ($d_{\max} = 2$) and the vertices form a closed chain (Fig. 2). From the literature, we have that the Fiedler value $\lambda_2 = 2(1 -$

$\cos\frac{2\pi}{n}$). Inserting into (22) and simplifying, we get $\alpha_{\max} = 2\arcsin\left(\sqrt{\frac{1}{4}+\sin^2(\frac{\pi}{n})}-\frac{1}{2}\right)$. Again, for $n = 3$, we retrieve $\alpha_{\max} = \frac{\pi}{3}$rad. For larger values of n, the upper bound decreases rapidly. In fact, the upper bound is quite conservative and it is possible to show a much stronger upper bound using a different analysis. In the appendix, we prove the following theorem.

Theorem 4.2. *Let R_i^*, $i = 1,\dots,n$ denote a stationary point to the primal problem (P) for a cycle graph with n vertices. Let α_{ij} denote the angular residuals, i.e., $\alpha_{ij} = \angle(R_i^*\tilde{R}_{ij}, R_j^*)$. Then, R_i^*, $i = 1,\dots,n$ will be globally optimal and strong duality will hold for (P) if $|\alpha_{ij}| \le \frac{\pi}{n}$ for all $(i,j) \in E$.*

Requiring that the angular residuals $|\alpha_{ij}|$ must be less than π/n for the global solution may seem like a restriction, but it is actually not. To see this, note that a non-optimal solution to the rotation averaging problem can be obtained by choosing R_1 such that the first residual α_{12} is zero, and then continuing in the same fashion such that all but the last residual α_{1n} in the cycle is zero. In the worst case, $\alpha_{1n} = \pi$. However, this is (obviously) non-optimal. A better solution is obtained if we distribute the angular residual error evenly so that $\alpha_{ij} = \alpha = \frac{\alpha_{1n}}{n}$ (which is always possible, see Theorem 23 in [10]). In conclusion, the angular residuals $|\alpha_{ij}|$ of the globally optimal solution for a cycle graph is always less than or equal to $\frac{\pi}{n}$, and conversely, if the angular residual is larger than $\frac{\pi}{n}$ for a local minimizer, then it does not correspond to the global solution.

In Fig. 1, we have a real example of an orbital camera motion which is close to a cycle. It may seem hard to determine if the camera motion consists of one or more loops around the object - we give three different local minima for this example. Still, applying formula (22) for this instance gives $\alpha_{\max} = 8.89°$ which is typically sufficient in practice to ensure that the optimal solution can be obtained by solving a convex program. Before developing an actual algorithm, we shall prove our main result on strong duality.

4.2. Proof of Theorem 4.1

Recall that a sufficient condition for strong duality to hold is that $\Lambda^* - \tilde{R} \succeq 0$ (Lemma 3.2). To prove Theorem 4.1 we will show that this is true under the conditions of the theorem.

To simplify the presentation we denote the residual rotations $\mathcal{E}_{ij} = R_i^*\tilde{R}_{ij}R_j^{*T}$ and define

$$D_{R^*} = \begin{bmatrix} R_1^* & 0 & 0 & \dots \\ 0 & R_2^* & 0 & \dots \\ 0 & 0 & R_3^* & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{23}$$

Then $D_{R^*}(\Lambda^* - \tilde{R})D_{R^*}^T =$

$$\begin{bmatrix} \sum_{j\neq 1} a_{1j}\mathcal{E}_{1j} & -a_{12}\mathcal{E}_{12} & -a_{13}\mathcal{E}_{13} & \dots \\ -a_{12}\mathcal{E}_{12}^T & \sum_{j\neq 2} a_{2j}\mathcal{E}_{2j} & -a_{23}\mathcal{E}_{23} & \dots \\ -a_{13}\mathcal{E}_{13}^T & -a_{23}\mathcal{E}_{23}^T & \sum_{j\neq 3} a_{3j}\mathcal{E}_{3j} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{24}$$

Note that $\sum_{j\neq i} a_{ij}\mathcal{E}_{ij} = \frac{1}{2}\sum_{j\neq i} a_{ij}(\mathcal{E}_{ij}+\mathcal{E}_{ij}^T)$ by symmetry of Λ^*. Since D_{R^*} is orthogonal, the matrix $\Lambda^* - \tilde{R}$ is positive semidefinite if and only if $D_{R^*}(\Lambda^* - \tilde{R})D_{R^*}^T$ is.

In the noise free case we note that the residual rotations will fulfill $\mathcal{E}_{ij} = I$ and therefore

$$D_{R^*}(\Lambda^* - \tilde{R})D_{R^*}^T = L_G \otimes I_3. \tag{25}$$

In the general noise case our strategy will therefore be to bound the eigenvalues of $D_{R^*}(\Lambda^* - \tilde{R})D_{R^*}^T$ by those of L_G for which well-known estimates exist. Thus, we will analyze the difference and define the matrix

$$\Delta = D_{R^*}(\Lambda^* - \tilde{R})D_{R^*}^T - L_G \otimes I_3. \tag{26}$$

The following results characterize the eigenvalues of Δ.

Lemma 4.1. *Let Δ_{ij}, $i = 1,...,n$, $j = 1,...,n$ be the 3×3 sub-blocks of Δ. If λ is an eigenvalue of Δ then*

$$|\lambda| \le \sum_{j=1}^{n} \|\Delta_{ij}\| \quad \textit{for some } i = 1,\dots,n. \tag{27}$$

Proof. The proof is similar to that of Gerschgorin's theorem [12]. Let $\Delta x = \lambda x$, with $\|x\| = 1$. Then $\lambda x_i = \sum_j \Delta_{ij}x_j$. Now pick i such that $\|x_i\| \ge \|x_j\|$ for all j. Then

$$|\lambda| = \left\|\lambda\frac{x_i}{\|x_i\|}\right\| = \left\|\sum_{j=1}^{n}\Delta_{ij}\frac{x_j}{\|x_i\|}\right\| \le \sum_{j=1}^{n}\|\Delta_{ij}\|. \tag{28}$$

□

Lemma 4.2. *Denote $\alpha_{\max}$ the largest (absolute) residual angle of all $\mathcal{E}_{ij}$ and assume $0 \le \alpha_{\max} \le \frac{\pi}{2}$. Then*

$$\|\Delta_{ii}\| \le 2d_i\sin^2(\frac{\alpha_{\max}}{2}) \quad \forall i = 1,\dots n, \tag{29}$$

where d_i is the degree of vertex i.

Proof. It is easy to see that by applying a change of coordinates $\mathcal{E}_{ij}$ can be written

$$\mathcal{E}_{ij} = V_{ij}\begin{bmatrix} \cos(\alpha_{ij}) & -\sin(\alpha_{ij}) & 0 \\ \sin(\alpha_{ij}) & \cos(\alpha_{ij}) & 0 \\ 0 & 0 & 1 \end{bmatrix} V_{ij}^T, \tag{30}$$

and therefore

$$\frac{1}{2}(\mathcal{E}_{ij}+\mathcal{E}_{ij}^T) = V_{ij}\begin{bmatrix} \cos(\alpha_{ij}) & 0 & 0 \\ 0 & \cos(\alpha_{ij}) & 0 \\ 0 & 0 & 1 \end{bmatrix} V_{ij}^T. \tag{31}$$

This gives

$$(\cos(\alpha_{ij})-1)I \preceq \frac{1}{2}(\mathcal{E}_{ij}+\mathcal{E}_{ij}^T)-I \preceq 0, \tag{32}$$

and since $\Delta_{ii}=\sum_{j\neq i} a_{ij}\left(\frac{1}{2}(\mathcal{E}_{ij}+\mathcal{E}_{ij}^T)-I\right)$ we get

$$d_i(\cos(\alpha_{\max})-1)I \preceq \Delta_{ii} \preceq 0. \tag{33}$$

Thus $\|\Delta_{ii}\| \leq d_i(1-\cos(\alpha_{\max})) = 2d_i\sin^2(\frac{\alpha_{\max}}{2})$. □

Lemma 4.3. *If $0 \leq \alpha_{max} \leq \frac{\pi}{2}$ and $i \neq j$ then*

$$\|\Delta_{ij}\| \leq 2a_{ij}\sin(\frac{\alpha_{\max}}{2}). \tag{34}$$

Proof. To estimate the off-diagonal blocks $\|\Delta_{ij}\| = a_{ij}\|I-\mathcal{E}_{ij}\|$ we note that for a unit vector v we have

$$\begin{aligned}\sqrt{\|v-\mathcal{E}_{ij}v\|^2} &= \sqrt{\|v\|^2-2\cos\angle(v,\mathcal{E}_{ij}v)+\|\mathcal{E}_{ij}v\|^2}\\ &\leq \sqrt{2(1-\cos(\alpha_{ij}))},\end{aligned} \tag{35}$$

where $\angle(v,\mathcal{E}_{ij}v)$ is the angle between v and $\mathcal{E}_{ij}v$. Furthermore, we will have equality if v is perpendicular to the rotation axis of $\mathcal{E}_{ij}$. Therefore

$$\|\Delta_{ij}\| = a_{ij}\sqrt{2(1-\cos(\alpha_{ij}))} \leq 2a_{ij}\sin(\frac{\alpha_{\max}}{2}). \tag{36}$$

□

Summarizing the results in Lemmas 4.1- 4.3 we get that the eigenvalues λ of Δ fulfill

$$\begin{aligned}|\lambda(\Delta)| &\leq 2d_i\sin^2(\frac{\alpha_{\max}}{2})+\sum_{j\neq i}2a_{ij}\sin(\frac{\alpha_{\max}}{2})\\ &\leq 2d_{\max}\sin(\frac{\alpha_{\max}}{2})\left(1+\sin(\frac{\alpha_{\max}}{2})\right),\end{aligned} \tag{37}$$

where $d_{\max}$ is the maximal vertex degree. Note that the same bound holds for all eigenvalues of Δ, in particular, the one with the largest magnitude $\lambda_{\max}(\Delta)$.

Now returning to our goal of showing that $D_{R^*}(\Lambda^*-\tilde{R})D_{R^*}^T \succeq 0$. Let $N=\begin{bmatrix} I & I & \dots \end{bmatrix}^T$. The columns of N will be in the nullspace of $D_{R^*}(\Lambda^*-\tilde{R})D_{R^*}^T$. Therefore $D_{R^*}(\Lambda^*-\tilde{R})D_{R^*}^T$ is positive semidefinite if $D_{R^*}(\Lambda^*-\tilde{R})D_{R^*}^T+\mu NN^T$ is, and hence it is enough to show that

$$\lambda_1\left(D_{R^*}(\Lambda^*-\tilde{R})D_{R^*}^T+\mu NN^T\right) \geq 0 \tag{38}$$

for sufficiently large μ. The Laplacian L_G is positive semidefinite with smallest eigenvalue $\lambda_1=0$ and corresponding eigenvector $v=\begin{pmatrix}1 & 1 & \dots & 1\end{pmatrix}^T$. Furthermore, as $N=v\otimes I_3$, it is clear that for sufficiently large μ we have $\lambda_1(L_G\otimes I_3+\mu NN^T)=\lambda_1(L_G+\mu vv^T)=\lambda_2(L_G)$. Since

$$D_{R^*}(\Lambda^*-\tilde{R})D_{R^*}^T+\mu NN^T = L_G\otimes I_3+\mu NN^T+\Delta, \tag{39}$$

we therefore get

$$\lambda_1(D_{R^*}(\Lambda^*-\tilde{R})D_{R^*}^T+\mu NN^T) \geq \lambda_2(L_G)-|\lambda_{\max}(\Delta)|. \tag{40}$$

If the right-hand side is positive, then so is the left-hand side. Using (37) for $\lambda_{\max}(\Delta)$ yields the following result.

Lemma 4.4. *The matrix $\Lambda^*-\tilde{R}$ is positive semidefinite if*

$$\lambda_2(L_G)-2d_{\max}\sin(\frac{\alpha_{\max}}{2})\left(1+\sin(\frac{\alpha_{\max}}{2})\right) \geq 0. \tag{41}$$

By completing squares, one obtains the equivalent condition

$$\left(\sin(\frac{\alpha_{\max}}{2})+\frac{1}{2}\right)^2 \leq \frac{\lambda_2(L_G)}{2d_{\max}}+\frac{1}{4}, \tag{42}$$

which proves Theorem 4.1.

What these results show, is that if there is a KKT point in (P), then it is also a KKT point to (P'). If this KKT point fulfills the prescribed error conditions it will be globally optimal in (P') and strong duality holds. But a solution that is globally optimal in (P') and feasible in (P) will also be globally optimal in (P) since the objective functions are the same. Thus, as long as there is a solution to (P) with with small enough errors the programs (P),(P') and (D) will all yield the same objective value.

5. Solving the Rotation Averaging Problem

The dual problem (D) is a convex semidefinite program, and although it is theoretically sound and provably solvable in polynomial time by interior point methods [4], in practice such problems quickly become intractable as the dimension of the entering variables grow.

In this section we present a first-order method for solving semidefinite programs with constant block diagonals. Our approach solves the dual of (D) and consists of two simple matrix operations only, matrix multiplication and square roots of 3×3 symmetric matrices, the latter which can be solved in closed form. Consequently, these two operations permit a simple and efficient implementation without the need for dedicated numerical libraries.

The dual of (D) is given by

$$\min_{Y\succeq 0}\max_{\Lambda} -\mathrm{tr}\,(\Lambda)+\mathrm{tr}\left(Y(\Lambda-\tilde{R})\right). \tag{43}$$

Let the matrix Y be partitioned as follows,

$$Y=\begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1n}\\ Y_{12}^T & Y_{22} & \dots & Y_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ Y_{1n}^T & \dots & \dots & Y_{nn}\end{bmatrix} \tag{44}$$

where each block $Y_{ij} \in \mathbb{R}^{3\times 3}$ for $i, j = 1, \ldots, n$. Since Λ is block-diagonal (13) it is clear that the inner maximization is unbounded when $Y_{ii} - I_{3\times 3} \neq 0$ and zero otherwise. We therefore get

$$(DD) \quad \begin{array}{ll} \min\limits_{Y} & -\text{tr}\left(\tilde{R}Y\right) \\ \text{s.t.} & Y_{ii} = I_3, \quad i = 1, ..., n, \\ & Y \succeq 0. \end{array} \tag{45}$$

Since $Y \succeq 0$ it is clear that

$$-\text{tr}\left(\Lambda\right) + \text{tr}\left(Y(\Lambda - R^*)\right) \geq -\text{tr}\left(\Lambda\right),$$

for all Λ of the form (13). Therefore $(DD) \geq (D)$ and assuming strong duality holds $(D) = (P)$. Furthermore if R^* is the global optimum of (P) then $Y = {R^*}^T R^*$ is feasible in (45) which shows that $(DD) = (P)$.

Thus, when strong duality holds, recovering a primal solution to (P) is then achieved by simply reading off the first three rows of Y^* and choosing their signs to ensure positive determinants of the resulting rotation matrices, see supplementary material for further details.

5.1. Block Coordinate Descent

In this section we present a block coordinate descent method for solving semidefinite programs with block diagonal constraints on the form (45). This method is a generalization of the row-by-row algorithms derived in [27].

Consider the following semidefinite program,

$$\begin{array}{ll} \min\limits_{S\in\mathbb{R}^{3n\times 3}} & \text{tr}\left(W^T S\right) \\ \text{s.t.} & \left[\begin{smallmatrix} I & S^T \\ S & B \end{smallmatrix}\right] \succeq 0. \end{array} \tag{46}$$

This is a subproblem that arises when attempting to solve (DD) in (45) using a block coordinate descent approach, i.e., by fixing all but one row and column of blocks in (44) and reordering as necessary. It turns out that this subproblem has a particularly simple, closed form solution, established by the following lemma.

Lemma 5.1. *Let B be a positive semidefinite matrix. Then, the solution to* (46) *is given by,*

$$S^* = -BW\left[\left(W^T BW\right)^{\frac{1}{2}}\right]^{\dagger}. \tag{47}$$

Here † denotes the Moore–Penrose pseudoinverse.

Proof. See supplementary material. □

6. Experimental Results

In this section we present an experimental study aimed at characterizing the performance and computational efficiency of the proposed algorithm compared to existing standard numerical solvers.

Algorithm 1 A block coordinate descent algorithm for the semidefinite relaxation (DD) in (45).

input: $\tilde{R}, Y^{(0)} \succeq 0,\ t = 0.$
repeat
- Select an integer $k \in [1, \ldots, n]$,
- B_k: the result of eliminating the k[th] row and column from Y^t.
- W_k: the result of eliminating the k[th] column and all but the k[th] row from $\tilde{R}$.
- $S_k^* = -B_k W_k \left[\left(W_k^T B_k W_k\right)^{\frac{1}{2}}\right]^{\dagger}$ as in (47).
- $Y^t = \left[\begin{smallmatrix} I & S_k^{*T} \\ S_k^* & B_k \end{smallmatrix}\right]$, (succeeded by the appropriate reordering).
- $t = t + 1$

until convergence

Synthetic data. In our first set of experiments we compared the computational efficiency of the Levenberg-Marquardt (LM) algorithm [29], a standard nonlinear optimization method, Algorithm 1 and that of SeDuMi [24], a publicly available software package for conic optimization.

We constructed a large number of synthetic problem instances of increasing size, perturbed by varying levels of noise. Each absolute rotation was obtained by rotation about the z-axis by $2\pi/n$ rad and by construction, forming a cycle graph. The relative rotations were perturbed by noise in the form of a random rotation about an axis sampled from a uniform distribution on the unit sphere with angles normally distributed with mean 0 and variance σ. The absolute rotations were initialized (if required) in a similar fashion but with the angles uniformly distributed over $[0, 2\pi]$ rad.

The results, averaged over 50 runs, can be seen in Table 1. As expected, the LM algorithm significantly outperforms our algorithm as well as SeDuMi, but it only manages to obtain the global optima in about $30 - 70\%$ of the time. As predicted by Theorem 4.2 and the discussion in Section 4.1 on cycle graphs, both Algorithm 1 and SeDuMi produce globally optimal solutions at every single problem instance, independent of the noise level and independent on the number of cameras. From this table we also observe that Algorithm 1 does appear to outperform SeDuMi quite significantly with respect to computational efficiency.

Real-world data. In our second set of experiments we compared the computational efficiency on a number of publicly available real-world datasets [11]. The results, again averaged over 50 runs, are presented in Table 2. Here, as in the previous experiment, both methods correctly produce the global optima at each instance. Algorithm 1 again significantly outperforms SeDuMi in computational cost, providing further evidence of the efficiency of the proposed algorithm. It can further be seen that Theorem 4.1 provides

n	σ [rad]	LM [29]		Alg. 1		SeDuMi [24]	
		$avg.error$ (%)	$time[s]$	$avg.error$	$time[s]$	$avg.error$	$time[s]$
20	0.2	1.49 (0.48)	0.012	9.34e-10	0.028	4.30e-09	0.501
	0.5	0.56 (0.73)	0.008	3.94e-08	0.023	3.72e-09	0.553
50	0.2	0.55 (0.50)	0.026	1.3e-09	0.17	6.85e-09	5.91
	0.5	0.17 (0.58)	0.017	1.83e-07	0.33	2.00e-09	6.32
100	0.2	0.15 (0.55)	0.042	1.46e-07	8.89	5.31e-09	47.0
	0.5	0.15 (0.45)	0.039	6.64e-08	7.97	7.41e-10	49.51
200	0.2	0.099 (0.40)	0.082	4.02e-08	17.01	4.15e-10	419.04
	0.5	0.031 (0.33)	0.071	6.79e-08	29.4	6.91e-10	391.23

Table 1: Comparison of running times and resulting errors on synthetic data. Here the errors are given with respect to the lowest feasible objective function value found. The fraction of the times the global optima was reached by the LM algorithm is indicated along side the average error.

Figure 3: Images and reconstructions of the datasets in Table 2.

$Dataset$	n	$time[s]$ Alg. 1	$time[s]$ SeDuMi	$\lvert\alpha_{ij}\rvert$	$\alpha_{\max}$
Gustavus	57	3.25	8.28	6.33°	8.89°
Sphinx	70	3.87	14.40	6.14°	12.13°
Alcatraz	133	12.73	117.19	7.68°	43.15°
Pumpkin	209	9.23	688.65	8.63°	3.59°
Buddha	322	16.71	1765.72	7.29°	14.01°

Table 2: The average run time and largest resulting angular residual ($|\alpha_{ij}|$) and bound ($\alpha_{\max}$) on five different real-world datasets.

bounds sufficiently large to guarantee strong duality, and hence global optimality, in all the real-world instances except for one, the *Pumpkin* dataset. Although strong duality does indeed hold in this case, the resulting certificate is less than the largest angular residual obtained. The camera graph is comprised both of densely as well as sparsely connected cameras, resulting in a large value of $d_{\max}$ in combination with a small value of $d_{\min}$ (minimum degree). Since $\lambda_2 \leq d_{\min}$ a limited bound on $\alpha_{\max}$ follows directly from (22). This instance serves as a representative example of when the bounds of Theorem 4.1, although still valid and strictly positive, become too conservative in practice.

7. Conclusions

In this paper we have presented a theoretical analysis of Lagrangian duality in rotation averaging based on spectral graph theory. Our main result states that for this class of problems strong duality will provably hold between the primal and dual formulations if the noise levels are sufficiently restricted. In many cases the noise levels required for strong duality not to hold can be shown to be quite severe. To the best of our knowledge, this is the first time such practically useful sufficient conditions for strong duality have been established for optimization over multiple rotations.

A scalable first-order algorithm, a generalization of coordinate descent methods for semidefinite cone programming, was also presented. Our empirical validation demonstrates the potential of this proposed algorithm, significantly outperforming existing general purpose numerical solvers.

Acknowledgements. This work has been funded by the Australian Research Council through grants FT170100072 and DP160103490, the Swedish Research Council (no. 2016-04445), the Swedish Foundation for Strategic Research (Semantic Mapping and Visual Navigation for Smart Robots) and Vinnova / FFI (Perceptron, no. 2017-01942).

References

[1] M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri. Global motion estimation from point matches. In *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2012. 2

[2] F. Arrigoni, L. Magri, B. Rossi, P. Fragneto, and A. Fusiello. Robust absolute rotation estimation via low-rank and sparse matrix decomposition. In *International Conference on 3D Vision*, 2014. 1

[3] N. Boumal, A. Singer, P.-A. Absil, and V. Blondel. Cramér-Rao bounds for synchronization of rotations. *Information and Inference*, 3:1–39, 2014. 2

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. 6

[5] J. Briales and J. Gonzalez-Jimenez. Fast global optimality verification in 3D SLAM. In *International Conference on Intelligent Robots and Systems*, 2016. 2

[6] L. Carlone, G. C. Calafiore, C. Tommolillo, and F. Dellaert. Planar pose graph optimization: Duality, optimal solutions, and verification. *IEEE Transactions on Robotics*, 32(3):545–565, 2016. 2

[7] L. Carlone and F. Dellaert. Duality-based verification techniques for 2D SLAM. In *International Conference on Robotics and Automation*, 2015. 2

[8] 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 *International Conference on Robotics and Automation*, 2015. 1

[9] A. Chatterjee and V. Madhav Govindu. Efficient and robust large-scale rotation averaging. In *International Conference on Computer Vision*, 2013. 2

[10] O. Enqvist. *Robust Algorithms for Multiple View Geometry - Outliers and Optimality*. PhD thesis, Centre for Mathematical Sciences, Lund University, Sweden, 2011. 5

[11] O. Enqvist, F. Kahl, and C. Olsson. Non-sequential structure from motion. In *International Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, 2011. 1, 2, 7

[12] D. G. Feingold and R. S. Varga. Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem. *Pacific J. Math.*, 12(4):1241–1250, 1962. 5

[13] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973. 4

[14] J. Fredriksson and C. Olsson. Simultaneous multiple rotation averaging using Lagrangian duality. In *Asian Conference on Computer Vision*, 2012. 2

[15] V. Govindu. Combining two-view constraints for motion estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2001. 1

[16] V. Govindu. Robustness in motion averaging. In *European Conference on Computer Vision*, 2006. 2

[17] R. Hartley, K. Aftab, and J. Trumpf. L_1 rotation averaging using the Weiszfeld algorithm. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011. 2

[18] R. Hartley, J. Trumpf, and Y. Dai. Rotation averaging and weak convexity. In *International Symposium on Mathematical Theory of Networks and Systems*, 2010. 2

[19] R. Hartley, J. Trumpf, Y. Dai, and H. Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, 2013. 1, 2, 3

[20] F. Kahl and R. Hartley. Multiple-view geometry under the L_∞-norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1603–1617, 2008. 1

[21] D. Martinec and T. Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007. 1, 2

[22] P. Moulon, P. Monasse, and R. Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *International Conference on Computer Vision*, 2013. 1

[23] 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. *CoRR*, abs/1612.07386, 2016. 2

[24] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999. 7, 8

[25] R. Tron, B. Afsari, and R. Vidal. Intrinsic consensus on SO(3) with almost-global convergence. In *IEEE Conference on Decision and Control*, 2012. 1

[26] R. Tron and R. Vidal. Distributed 3-D localization of camera sensor networks from 2-D image measurements. *IEEE Transactions on Automatic Control*, 59(12):3325–3340, 2014. 2

[27] Z. Wen, D. Goldfarb, S. Ma, and K. Scheinberg. Row by row methods for semidefinite programming. Technical report, Columbia University, 2009. 7

[28] K. Wilson, D. Bindel, and N. Snavely. When is rotations averaging hard? In *European Conference on Computer Vision*, 2016. 1, 4

[29] S. Wright and J. Nocedal. Numerical optimization. *Springer Science*, 35:67–68, 1999. 7, 8

[30] Y. Zhong and N. Boumal. Near-optimal bounds for phase synchronization. *ArXiv e-prints*, Mar. 2017. 2