Why Rotation Averaging is Easy

Anders Eriksson, Carl Olsson, Fredrik Kahl, Olof Enqvist and Tat-Jun Chin

1School of Electrical Engineering and Computer Science, Queensland University of Technology
2Department of Electrical Engineering, Chalmers University of Technology
3Centre for Mathematical Sciences, Lund University
4School 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 very challenging to solve globally.

In this work we show how to surpass 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. We analytically prove that unless the noise levels are severe, there will be no duality gap. 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, with convincing results.

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 [18, 10, 19] rely on rotation averaging to initialize bundle adjustment. While sequential approaches only consider subsets of the data in each reconstruction step, the overall idea of the non-sequential methods is to consider as much data as possible in each step to avoid suboptimal reconstructions. In the context of rotation averaging this amounts to use as many camera pairs as possible.

The problem can be thought of as inference on the so-called camera graph. An edge \((i, j)\) in this graph represents a relative rotation measurement \(R_{ij}\) and the objective is to find the absolute orientation \(R_i\) for each node \(i\) such that 

\[ R_i R_{ij} = R_j \]

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. [24] studied local convexity of the problem and showed that instances with large loosely connected graphs are hard to solve with local, iterative optimization methods.

In contrast, in this paper we show that convex relaxation methods can in fact overcome these difficulties with a minimum of extra computational cost. 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 analytically prove that unless the noise level is severe, there will be no duality gap. Specifically, if the worst error of the optimal solution has an angular residual less than 50.7°, then our method will recover the optimal solution. 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 [8]. Despite progress in practical algorithms, they largely come without guarantees. One of the earliest averaging methods was due to Govindu [13], 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.
Additionally, since a quaternion only represents a rotation if its length is one, a norm constraint has to be incorporated for each rotation. It was observed by Fredriksson and Ols-son in [12] 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. In order to avoid solving a costly SDP for large scale problems, it was proposed that given a candidate primal solution to compute the dual solution in order to evaluate the duality gap. In this way it is possible to test and verify optimality of a candidate solution obtained with any feasible algorithm.

A more straightforward rotation representation is $3 \times 3$ matrices. Martinec and Pajdla [18] approximately solve the problem by ignoring the orthogonality and determinant constraints by computing eigenvectors of the resulting homogeneous system of equations. A similar spectral relaxation was derived by Arie-Nachimson et al. in [2]. In addition, an SDP formulation was presented which is equivalent to the one we address in this paper. However, [2] does not give any performance bounds which is our main objective.

To address outlier measurements in the camera graph, a number of robust approaches have been developed. A sampling scheme over spanning trees of the camera graph is developed by Govindu in [14]. Enqvist et al. [10] also start from a spanning tree and add relative rotations that are consistent with the solution. In [15] the Weiszfeld algorithm is applied to single rotation averaging with the $L_1$ norm. In [16] 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. Here camera orientations and positions are jointly optimized. In this context Lagrangian duality has been applied [6, 7] both for optimization and for verification similar to [12]. In [22] 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 very recent paper [20] an SDP relaxation for pose graph estimation with performance guarantees is analyzed. It is shown that there is a noise level $\beta$ for which the relaxation is guaranteed to provide the exact globally optimal solution. However, The result only shows the existence of $\beta$. Its value which is dependent on the problem instance is not computed. In contrast our result for multiple rotation averaging gives an explicit noise bound which is valid for any problem instance.

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 are independent of the problem instance and guarantee exact global recovery.
- We show that the standard SDP relaxation can be reduced and 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.

2. Problem Statement

Here we consider the problem of rotation averaging to be defined as the task of determining a set of $n$ absolute rotations $R_1, \ldots, R_n$ given distinct estimated relative rotations $\tilde{R}_{ij}$. Available relative rotations are represented by the index set $\mathcal{N}$. Under ideal conditions this amounts to finding the $n$ rotations compatible with the $|\mathcal{N}|$ linear relations,

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

for all $(i, j) \in \mathcal{N}$. However, in the presence of noise, a solution to (1) is not guaranteed to exist. Instead, it is typically solved in a least-metric sense,

$$\min_{R_i, \ldots, R_n} \sum_{(i, j) \in \mathcal{N}} d(R_i \tilde{R}_{ij}, R_j)^p,$$

where $d$ is a distance metric.
where \( p \geq 1 \) and \( d \) is a distance function on the space of rotations.

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

\[
\text{SO}(3) \in \{ R \in \mathbb{R}^{3\times 3} : R^T R = I, \ \det(R) = 1 \},
\]

with the ordinary matrix multiplication as the operation. Rotations can also be parametrised using the axis-angle representation predicated by Euler’s rotation theorem. This theorem states that any rotation in three-dimensional space is equivalent to a single rotation about a fixed axis \( v \) by a magnitude of \( \alpha \) degrees. Both these representations of rotations in three dimensions will be used interchangeably.

A number of distinct choices of metrics on \( \text{SO}(3) \) exist, see Hartley et al. [17] for a comprehensive discussion. In this work we restrict ourselves to the chordal distance, the most commonly used metric when analysing Lagrangian duality in rotation averaging. It has proven to be a convenient choice of distance function 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,

\[
d(R, S) = ||R - S||_F.
\]

It can be shown [17] that the chordal distance can also be written as \( d(R, S) = 2\sqrt{2}\sin \frac{||\alpha||}{2} \), where \( \alpha \) is the rotation angle of \( RS^{-1} \). Letting the rotations be embedded in \( \mathbb{R}^{3\times 3} \),

\[
\text{SO}(3) \ni R = [u \ v \ w]\begin{bmatrix}
 u^T v^T w^T \\
 u^T w^T v^T \\
 v^T w^T u^T
\end{bmatrix},
\]

the orthogonality constraints in (3) become,

\[
\begin{aligned}
 u^T u &= v^T v = w^T w = 1, \\
 u^T v &= u^T w = v^T w = 0.
\end{aligned}
\]

The positivity constraint on the determinant can then be expressed as the bilinear equalities,

\[
 u \times v = w \iff \begin{cases}
 w^uv^z - w^xv^y = w^x, \\
 w^ux^z - w^xv^y = w^y, \\
 w^ux^y - w^vy^x = w^z.
\end{cases}
\]

with the above choice of metric, the rotation averaging problem is defined as

\[
\arg \min_{R_1, \ldots, R_n \in \text{SO}(3)} \sum_{(i,j) \in \mathcal{N}} \| R_i \bar{R}_{ij} - R_j \|^2_F,
\]

which, with \( \langle A, B \rangle = \text{trace} A^T B \), can be simplified to

\[
\arg \min_{R_1, \ldots, R_n \in \text{SO}(3)} \sum_{(i,j) \in \mathcal{N}} \langle R_i \bar{R}_{ij}, R_j \rangle.
\]

Together with our orthogonality constraints (6)-(8), this constitutes our primal problem.

Later, when deriving the dual formulation, it will be convenient with a compact matrix formulation. Let \( r \) be a vector with all unknowns,

\[
r = [u_1^T \ v_1^T \ w_1^T \ u_2^T \ v_2^T \ w_2^T \ u_3^T \ v_3^T \ w_3^T \ u_n^T \ v_n^T \ w_n^T]^T.
\]

Then, each term \( \langle R_i \bar{R}_{ij}, R_j \rangle \) in the objective function can be written as a quadratic form \( r^T Q_{ij} r \) for some symmetric matrix \( Q_{ij} \). Introducing the index variables

\[
W_{ij} = \begin{cases}
 0 & (i, j) \notin \mathcal{N}, \\
 1 & (i, j) \in \mathcal{N},
\end{cases}
\]

for all \( i, j \leq n \), allows us to write the objective function as \( r^T Q r \) for \( Q = -\sum_{i,j} W_{ij} Q_{ij} \). Similarly the orthogonality constraints can be written as quadratic constraints. This leads to the following equivalent quadratic program,

\[
(P) \quad \min_{r \in \mathbb{R}^n} r^T Q r \quad \text{s.t.} \quad r^T E_{ik} r = 1 \quad i = 1, ..., n, \quad k = 1, ..., 3,
\]

with \( r^T E_{ik} r = 1, \ r^T F_{ik} r = 0 \) and \( r^T G_{ik} r + 2g_{ik}^T r = 0 \) encoding (6)-(8), respectively. Note that without loss of generality, all matrices can be chosen such that they are symmetric. We can also assume that if \( (i, j) \in \mathcal{N} \), then \( (j, i) \) also belongs to \( \mathcal{N} \) and that \( W_{ij} = W_{ji} = \bar{R}_{ij} \). The exact form of the matrices is of less importance right now.

Later, we will use that \( I_3 \otimes \bar{R} = \begin{bmatrix}
 \bar{R} & 0 & 0 \\
 0 & \bar{R} & 0 \\
 0 & 0 & \bar{R}
\end{bmatrix} \),

\[
(\otimes \text{ denoting the Kronecker product}) \text{ where } \bar{R} \text{ is the symmetric matrix}
\]

\[
\bar{R} = \begin{bmatrix}
 0 & W_{12} & \cdots & W_{1n} & \bar{R}_{12} & \cdots & \bar{R}_{1n} \\
 W_{21} & 0 & \cdots & W_{2n} & \bar{R}_{21} & \cdots & \bar{R}_{2n} \\
 \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
 W_{n1} & \cdots & \cdots & 0 & \bar{R}_{n1} & \cdots & \bar{R}_{nn}
\end{bmatrix}.
\]

The Karush-Kuhn-Tucker (KKT) conditions for (13) provide first-order necessary conditions for optimality. If
$r^*$ is a local optimizer, then $\exists \lambda_{ik}, \mu_{ik}, \nu_{ik}, i = 1, ..., n, k = 1, ..., 3$, such that

\begin{equation}
(\text{Stationarity})\quad Q r^* - \sum_{i,k} \left[ (\lambda_{ik} E_{ik} + \mu_{ik} F_{ik} + \nu_{ik} G_{ik}) r^* + \nu_{ik} g_{ik} \right] = 0, \tag{16a}
\end{equation}

\begin{equation}
(\text{Primal feasibility})
\begin{align*}
& r^T E_{ik} r^* = 1, \quad (16b) \\
& r^T F_{ik} r^* = 0, \quad (16c) \\
& r^T G_{ik} r^* + 2 g^T_{ik} r^* = 0, \quad (16d) \\
& i = 1, ..., n, \quad k = 1, ..., 3. \quad (16e)
\end{align*}
\end{equation}

### 3. A Dual Formulation of Rotation Averaging

Next we introduce two dual formulations associated with our primal problem ($P$). For the purposes of this derivation it will prove convenient to first homogenize our problem. Letting,

\begin{equation}
\bar{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \bar{E}_{ik} = \begin{bmatrix} E_{ik} & 0 \\ 0 & 0 \end{bmatrix}, \quad (17)
\end{equation}

\begin{equation}
\bar{F}_{ik} = \begin{bmatrix} F_{ik} & 0 \\ 0 & 0 \end{bmatrix}, \quad \bar{G}_{ik} = \begin{bmatrix} G_{ik} & g_{ik} \\ g^T_{ik} & 0 \end{bmatrix}, \quad (18)
\end{equation}

allows us to rewrite ($P$) as the equivalent, homogenized primal problem

\begin{equation}
(\bar{P}) \quad \min_{\bar{r} \in \mathbb{R}^{2n+1}} \bar{r}^T \bar{Q} \bar{r}
\end{equation}

\begin{equation}
s.t. \quad \bar{r}^T \bar{E}_{ik} \bar{r} = 1, \quad (19a)
\end{equation}

\begin{equation}
\bar{r}^T \bar{F}_{ik} \bar{r} = 0, \quad (19b)
\end{equation}

\begin{equation}
\bar{r}^T \bar{G}_{ik} \bar{r} = 0, \quad (19c)
\end{equation}

\begin{equation}
\bar{r}^T \bar{r} = 1, \quad (19d)
\end{equation}

\begin{equation}
\bar{e}_{2n+1} = 1, \quad (19e)
\end{equation}

with $i$ and $k$ as in (13e). The associated Lagrangian function is given by

\begin{equation}
L(\bar{r}; \lambda, \mu, \nu, \tau) = \bar{r}^T \bar{Q} \bar{r} + \sum_{i,k} \lambda_{ik} + \tau - \bar{r}^T \left( \sum_{i,k} \left[ \lambda_{ik} \bar{E}_{ik} + \mu_{ik} \bar{F}_{ik} + \nu_{ik} \bar{G}_{ik} \right] + \tau \bar{H} \right) \bar{r}, \quad (20)
\end{equation}

with $\bar{H} = e_{2n+1} e_{2n+1}^T$, $e_k$ denotes the k-th vector in the Euclidean standard basis. It is then straightforward from $L(\bar{r}; \lambda, \mu, \nu, \tau)$ to derive the dual problem as

\begin{equation}
(D) \quad \max_{\lambda, \mu, \nu, \tau} \sum_{i,k} \lambda_{ik} + \tau \quad (21a)
\end{equation}

\begin{equation}
s.t. \quad Q - \sum_{i,k} \left[ \lambda_{ik} E_{ik} + \mu_{ik} F_{ik} + \nu_{ik} G_{ik} \right] - \tau \bar{H} \geq 0. \quad (21b)
\end{equation}

Taking the dual yet again, we arrive at the dual of the dual

\begin{equation}
(DD) \quad \min_{\bar{X}} < \bar{Q}, \bar{X} > \quad (22a)
\end{equation}

s.t. \quad $< \bar{E}_{ik}, \bar{X} > = 1$, \quad (22b)

$< \bar{F}_{ik}, \bar{X} > = 0$, \quad (22c)

$< \bar{G}_{ik}, \bar{X} > = 0$, \quad (22d)

$< \bar{H}, \bar{X} > = 1$, \quad (22e)

$\bar{X} \geq 0$, \quad (22f)

with $i = 1, ..., n, k = 1, ..., 3$. Given the equivalence of problems ($P$) (13) and ($P^h$) (19), the two problems above ($D$) and ($DD$) can consequently be understood as dual problems of ($P$) as well as ($P^h$). From weak duality we have that $P^* \geq D^* = DD^*$, i.e., the solution to (22) will always be an underestimate of the solution to (13). The problem ($DD$) is also known as the semidefinite relaxation of ($P$). It can further be shown that this semidefinite relaxation is in fact the tightest possible convex relaxation to the rotation averaging problem (13), see [1]. The KKT conditions for (22) are given by

\begin{equation}
(\text{Stationarity}) \quad \bar{Q} - \sum_{i,k} \left[ \lambda_{ik} \bar{E}_{ik} + \mu_{ik} \bar{F}_{ik} + \nu_{ik} \bar{G}_{ik} \right] - \tau \bar{H} - \bar{Z} = 0, \quad (23a)
\end{equation}

\begin{equation}
(\text{Primal feasibility})
\begin{align*}
& < \bar{E}_{ik}, \bar{X} > = 1, \quad (23b) \\
& < \bar{F}_{ik}, \bar{X} > = 0, \quad (23c) \\
& < \bar{G}_{ik}, \bar{X} > = 0, \quad (23d) \\
& < \bar{H}, \bar{X} > = 1, \quad (23e) \\
& \bar{X} \geq 0, \quad (23f)
\end{align*}
\end{equation}

\begin{equation}
(\text{Dual feasibility}) \quad \bar{Z} \geq 0, \quad (23g)
\end{equation}

\begin{equation}
(\text{Complementary slackness}) \quad \bar{X} \bar{Z} = 0, \quad (23h)
\end{equation}

where $\bar{Z}$ are Lagrangian multipliers. The subsequent analysis will build considerably on the above conditions.

### 4. Main Result

In this section, we will state our main result which says that strong duality holds for our primal and dual problems under rather generous conditions. An outline of all the steps involved in the proof will also be given.

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 by itself. In Section 8, we will develop a new algorithm for efficiently computing the optimal solution.
4.1. Strong Duality Theorem

Returning to our initial, primal rotation averaging problem (9). The goal is to find rotations \( R_i \) and \( R_j \) such that the sum of the residuals \( \|R_i \tilde{R}_{ij} - R_j\|^2 \) is minimized. Let \( \alpha_{ij} \) denote the residual rotation angle between \( R_i \tilde{R}_{ij} \) and \( R_j \), i.e., \( \angle(R_i \tilde{R}_{ij}, R_j) \). Due to noise, the residual errors will typically not be zero for the optimal solution. For strong duality to hold, we need to bound the residual error.

**Theorem 4.1** (Strong Duality). Let \( R_i^\ast, i, \ldots, n \) denote the optimal solution to the primal problem \((P)\). Let \( \alpha_{ij} \) denote the angular residuals, i.e., \( \alpha_{ij} = \angle(R_i^\ast \tilde{R}_{ij}, R_j^\ast) \) for all \((i, j) \in \mathcal{N}\). Then strong duality will hold for \((P)\) if

\[
|\alpha_{ij}| < \alpha_{\max} \quad \forall (i, j) \in \mathcal{N}
\]

where \( \alpha_{\max} = 2 \arcsin \left( \frac{1}{2} \sqrt{\frac{\gamma}{1 - \gamma}} \right) \approx 0.8841 \text{ rad} \approx 50.7^\circ \).

Hence, if we have a problem instance which has bounded noise, then we may instead solve the dual problem \((DD)\) or the dual of the dual \((DD)\), and obtain the optimal solution to our primal problem \((P)\) as \( P^\ast = D^\ast = DD^\ast \).

4.2. Proof Outline

The idea of the proof is straightforward. We will investigate under what conditions a solution to \((P)\) is also a solution to \((DD)\), because then \( P^\ast = DD^\ast \) and strong duality holds. In order to do so, we also need to involve the dual solution of \( D \).

It turns out that a sufficient condition for strong duality is that a certain matrix, which we denote \( \Omega \) shall be positive semidefinite, \( \Omega \succeq 0 \). In order to analyze when \( \Omega \) is positive semidefinite, we examine when \( \Omega \) is a block diagonally dominant matrix, since this implies that \( \Omega \succeq 0 \). Finally, we can show that if the bounded noise condition as stated in the theorem is satisfied, then \( \Omega \) is indeed block diagonally dominant.

In summary, the proof involves the following steps.

1. Given a solution to \((P)\), construct a dual solution \((D)\) and a candidate solution to \((DD)\).

2. Derive a sufficient condition such that the candidate solution is optimal to \((DD)\). The sufficient condition will be expressed as \( \Omega \succeq 0 \).

3. Show that the sufficient condition can also be expressed as an upper, fixed bound on the angular residuals by examining when \( \Omega \) is a block diagonally dominant matrix.

Before we start with the first step of the proof (Section 5), we will show some useful properties of the dual solution (Section 6) and introduce the concept of block diagonally dominant matrices (Section 6).

5. Properties of the Dual Solution

In this lemma, given a primal solution, we derive some important properties of the dual solution.

**Lemma 5.1.** Let \( r^\ast \) be a local minimizer of \((13)\) with \( \lambda^\ast, \mu^\ast, \nu^\ast \) the corresponding Lagrangian multipliers, then it holds that \( \nu^\ast = 0 \) and

\[
\sum_{i,k} \left[ \lambda^\ast_{ik} E_{ik} + \mu^\ast_{ik} F_{ik} \right] = - I_3 \otimes \Lambda^\ast = - \begin{bmatrix} \Lambda^\ast & 0 & 0 \\ 0 & \Lambda^\ast & 0 \\ 0 & 0 & \Lambda^\ast \end{bmatrix},
\]

where

\[
\Lambda^\ast = \begin{bmatrix} \Lambda^\ast_1 & 0 & \ldots & 0 \\ 0 & \Lambda^\ast_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Lambda^\ast_n \end{bmatrix}
\]

and

\[
\Lambda^\ast_i = \left( \sum_{j=1}^n W_{ij} \tilde{R}_{ij} R_{j}^{T} \right) R_{i}^\ast, \quad i = 1, \ldots, n,
\]

will be a solution to the stationarity condition \((16a)\).

**Proof.** We begin by proving that \((16a)\) permits a solution with \( \nu^\ast = 0 \). Consider relaxing the rotation averaging problem by removing \((8)\). Equations \((6)-(7)\) are still requiring \( \nu \)'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 with \((8)\) has to also be a local minimizer to the problem without \((8)\), thus satisfying the KKT conditions with \( \nu^\ast = 0 \).

To prove formula \((27)\) we first note that \( \sum_{i,k} \lambda_{ik} E_{ik} + \mu_{ik} F_{ik} \) is a block diagonal matrix with blocks

\[
\Lambda^\ast_i = \begin{bmatrix} \lambda_{i1} & \mu_{i1} & \mu_{i2} \\ \mu_{i1} & \lambda_{i2} & \mu_{i3} \\ \mu_{i2} & \mu_{i3} & \lambda_{i3} \end{bmatrix}.
\]

Computation of \((16a)\) with \( \nu^\ast = 0 \) now shows that \((16a)\) amounts to equations of the form

\[
\begin{bmatrix} u^\ast_i \\ v^\ast_i \\ w^\ast_i \end{bmatrix} = \sum_j W_{ij} \tilde{R}_{ij} \begin{bmatrix} u^\ast_j \\ v^\ast_j \\ w^\ast_j \end{bmatrix},
\]

for all \( i = 1, \ldots, n \) and similarly for the \( y \) and \( z \) coordinates. Summarizing we thus get

\[
\Lambda^\ast_i R_{i}^{T} = \sum_j W_{ij} \tilde{R}_{ij} R_{j}^{T},
\]

which is equivalent to \((27)\). □
6. Block Diagonally Dominant Matrices

To prove our main result, we will use the concept of block diagonally dominant matrices [11] and a theorem establishing an important property of this class of matrices. These are not new results but presented for completeness.

Let

\[ A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \in \mathbb{R}^{n \times n} \]  

with \( A_{ij} \in \mathbb{R}^{m \times m} \) for all \( i, j = 1, \ldots, n \).

**Definition 6.1.** A block matrix \( A \) partitioned as in (31) is strictly block diagonally dominant if

\[ \sum_{j=1}^{n} ||A_{ij}|| > \sum_{j=1, j \neq i}^{n} ||A_{ij}||, \quad \forall i = 1, \ldots, n. \]  

**Theorem 6.1.** A symmetric strictly block diagonally dominant matrix with block diagonal entries,

\[ A_{ii} \succ 0, \quad i = 1, \ldots, n, \]  

is positive semidefinite.

**Proof.** The theorem follows from [11]. A simplified proof is also provided in the supplementary material. \( \square \)

7. Sufficient Conditions for Strong Duality

In this section we will derive sufficient conditions for strong duality to hold between our primal (13) and dual problem (22), beginning with the following lemma.

**Lemma 7.1.** Let \( r^* \) be the global minimizer of (13) with \( \Lambda^* \) as defined in Lemma 5.1 and \( \bar{R} \) as defined in (15). Then strong duality holds for the primal problem (P) if

\[ \Omega = \Lambda^* - \bar{R} \succeq 0. \]  

**Proof.** The lemma will follow if we can show that \( r^* \) also yields a solution to \((D\bar{D})\). Therefore, let

\[ \bar{X}^* = \left[ \begin{array}{c} r^* \\ \nu^* \end{array} \right], \quad \bar{Z}^* = \left[ \begin{array}{c} Q+I_3 \otimes \Lambda^* \\ 0 \end{array} \right]. \]  

We will prove the result by showing that the candidate solution satisfies all the conditions of (23) apart from (23g).

The stationary condition (23a) follows directly by Lemma 5.1,

\[ \bar{Q} - \sum_{i,k} \left[ \lambda^*_{ik} E_{ik} + \mu^*_{ik} F_{ik} + \nu^*_{ik} G_{ik} \right] - r^* H - \bar{Z}^* = \left[ Q - (-I_3 \otimes \Lambda^*) - (Q+I_3 \otimes \Lambda^*) \right] 0_1 = 0. \]

For the primal feasibility condition (23b) we have that

\[ < E_{ik}, \bar{X}^* >= r^* T E_{ik} r^* = 1, \quad \forall i, k, \]  

similarly for (23c)-(23e), where the equalities follow from \( r^* \) being a feasible solution to (13). Condition (23f) follows directly from (35).

Complementary slackness is also satisfied, as

\[ (\bar{X}^* \bar{Z}^*)^T = \left[ \begin{array}{c} Q r^* - \sum_{i,k} (\lambda^*_{ik} E_{ik} + \mu^*_{ik} F_{ik} r^* 0_1) \end{array} \right] 0_1 = 0. \]  

(38)

For dual feasibility (23g), we have that

\[ \bar{Z}^* = \left[ -I_3 \otimes \bar{R} + I_3 \otimes \Lambda^* \right] 0_1 \succeq 0 \iff \Lambda^* - \bar{R} \succeq 0. \]  

(39)

Note that we have used that \( Q \) can be written \( Q = -I_3 \otimes \bar{R} \), cf. (14).

Finally since \( r^* T Q r^* =< Q, r^* r^* T >=< \bar{Q}, \bar{X}^* > \) we have \( P^* = \bar{D} D^* = D^* \) and the lemma follows. \( \square \)

Building on Theorem 6.1 and Lemma 7.1, by replacing the sufficient condition \( \Omega \succeq 0 \) (34) by the stricter condition of \( \Omega \) being strictly block diagonally dominant, we can reformulate the condition and arrive at the following result.

**Theorem 7.1** (Strong Duality). Let \( R^*_i, \quad i = 1, \ldots, n \) denote the solution to (P) in (13). Then strong duality will hold for \((P)\) if

\[ ||R^*_i \bar{R}_{ij} - R^*_j||_F^2 < 2(\sqrt{3} - 1), \]  

for all \((i, j) \in \mathcal{N} \).

**Proof.** See supplementary material. \( \square \)

As the chordal distance is computed as \( ||R^*_i \bar{R}_{ij} - R^*_j||_F = 2\sqrt{2} \sin\left(\frac{\pi}{2m}\right)\), see [17] for a proof, an equivalent condition is that all residual angles \( \alpha_{ij} \) satisfy

\[ |\alpha_{ij}| < 2 \arcsin\left(\frac{1}{2} \sqrt{\sqrt{3} - 1}\right), \]  

which is the alternative formulation of the Strong Duality Theorem in Section 4.1.
8. Solving the Rotation Averaging Problem

The dual of dual problem \((DD)\) in (22) 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 proposed method consists of two simple matrix operations only, matrix multiplication and square roots of semidefinite programs with constant block diagonals. Our methods \([4]\), in practice such problems quickly become intractable as the dimension of the entering variables grow.

When strong duality holds, recovering a primal solution to the KKT system (23). The existence of such a solution, which is both feasible and optimal with respect to (13) is guaranteed by standard results in duality theory. Here the solution becomes particularly simple. If \(\hat{Y}^*\) is a solution to (43) and \(v_1, v_2\) and \(v_3\) the three most significant eigenvectors. Then it is straightforward to show that

\[
\hat{r}^* = \begin{bmatrix} \pm v_1^T & \pm v_2^T & \pm v_3^T \end{bmatrix}^T ,
\]

and \(\hat{X}^* = \begin{bmatrix} r_1^* & r_2^* \end{bmatrix} \) will be a rank-1 solution to (23) and consequently an optimal solution to \((P)\) in (13). The signs of the vectors in (44) are chosen to ensure positive determinants of the resulting rotation matrices.

8.1. Block Coordinate Descent

In this section we present a block coordinate descent method for solving semidefinite programs with block diagonal constraints on the form (43). This method is a generalization of the row-by-row algorithms derived in [23].

Consider the following semidefinite program,

\[
\begin{align*}
\min_{S \in \mathbb{R}^{3 \times 3}} & < A, S > \\
\text{s.t.} & \begin{bmatrix} I & S \end{bmatrix} \succeq 0 .
\end{align*}
\]

This is a subproblem that arises when attempting to solve \((DD^*)\) in (43) using a block coordinate descent approach, i.e., by fixing all but one row and column of blocks in (42) and reordering as necessary. It turns out that this subproblem has a particularly simple, closed form solution, established by the following lemma.

Lemma 8.2. Let \(B \) be a positive semidefinite matrix. Then, the solution to (45) is given by;

\[
S^* = -BA \left( A^TBA \right)^{\frac{1}{2}} .
\]

Here \(^\dagger\) denotes the Moore–Penrose pseudoinverse.

Proof. See supplementary material.

\[\square\]

Algorithm 1 A block coordinate descent algorithm for the semidefinite relaxation \((DD^*)\) in (43).

| Algorithm 1 |
|-------------|
| **input:** \(\hat{R}, Y^{(0)} \succeq 0, \ t = 0.\) |
| **repeat** |
| · Select an integer \(k \in [1..n]\), |
| · \(B_k\): the result of eliminating the \(k^\text{th}\) row and column from \(Y^t\), |
| · \(A_k\): the result of eliminating the \(k^\text{th}\) column and all but the \(k^\text{th}\) row from \(\hat{R}\), |
| · \(S_k^* = -B_kA_k \left( A_k^T B_k A_k \right)^{\frac{1}{2}} \)^\dagger as in (46), |
| · \(Y^t = \begin{bmatrix} I & S_k^* \end{bmatrix} \), (succeeded by the appropriate reordering), |
| · \(t = t + 1\) |
| **until** convergence |

As the above algorithm permits initialization or warm-starting (a property not shared with most interior-point methods) we found that a significant speed up of our proposed method can be achieved by first running a standard local optimization methods, such as Levenberg-Marquardt [25], for a few iterations before applying Algorithm 1.
9. Experimental Results

In this section we present an experimental study aimed at characterising the performance and computational efficiency of the proposed algorithm. Here we were mainly interested in computational gain achieved by applying the approach presented in this paper over existing standard numerical solvers.

**Synthetic Data.** In our first set of experiments we compared the computational efficiency of the Levenberg-Marquardt algorithm [25], a standard nonlinear optimization method, Algorithm 1 and that of SeDuMi [21], a publicly available software package for conic optimization.

For this purpose we constructed a large number of synthetic rotation averaging 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, the relative rotations were then perturbed by noise in the form of a random rotation about an axis sampled from a uniform distribution on the unit sphere. The magnitude of these rotations were normally distributed with mean 0 and variance $\sigma$. If required, the absolute rotations were initialized in a similar fashion but with the magnitudes uniformly distributed between 0 and $2\pi$ 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 established in the main theorem of this paper, both Algorithm 1 and SeDuMi produce negligible errors at every single problem instance. From this table we also observe that Algorithm 1 outperforms 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 [10]. 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 should also be noted that $\alpha_{\text{max}}$, the largest resulting angular residual, provided in Table 2, satisfy the bound of Theorem 4.1 with significant margin, in all the datasets. It could be argued that that these datasets to some extent can be regarded as representative of the typical noise levels occurring in real-world datasets. Consequently it would be expected that, in the absence of outliers, the angular residuals will be sufficiently small ($< 50.7^\circ$) and strong duality is guaranteed to hold, i.e., rotation averaging can be solved globally and efficiently in most instances. However, further experimental validation is needed before such a statement can be made with significant certainty.

10. Conclusions

In this paper we have presented a theoretical analysis of Lagrangian duality in rotation averaging. Our main result states that for this class of problems strong duality will provably hold between the primal and dual formulations under considerably generous conditions. 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.

References

[1] K. M. Anstreicher. On convex relaxations for quadratically constrained quadratic programming. *Mathematical programming*, 136(2):233–251, 2012.

[2] 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.

[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.

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

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

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

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

[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.

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

[10] 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.

[11] 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.
### 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 (in parenthesis) alongside the average error.

| n   | σ [rad] | avg.error (%) | time [s] | avg.error | time [s] | avg.error | time [s] |
|-----|---------|---------------|----------|-----------|----------|-----------|----------|
| 20  | 0.2     | 7.66 (0.84)   | 0.019    | 7.87e-09  | 0.0047   | 6.545e-11 | 0.48     |
|     | 0.5     | 6.98 (0.81)   | 0.022    | 4.03e-07  | 0.0049   | 8.915e-10 | 0.49     |
| 50  | 0.2     | 7.96 (0.57)   | 0.071    | 2.18e-07  | 0.082    | 7.872e-10 | 2.15     |
|     | 0.5     | 6.76 (0.53)   | 0.012    | 3.62e-06  | 0.078    | 4.898e-10 | 2.07     |
| 100 | 0.2     | 5.36 (0.43)   | 0.36     | 1.62e-06  | 0.93     | 3.662e-10 | 17.84    |
|     | 0.5     | 4.13 (0.46)   | 0.68     | 1.49e-05  | 0.93     | 1.919e-10 | 18.44    |
| 200 | 0.2     | 1.98 (0.48)   | 1.65     | 7.73e-06  | 9.22     | 4.34e-10  | 228.21   |
|     | 0.5     | 1.17 (0.52)   | 2.16     | 7.61e-06  | 9.07     | 3.80e-10  | 230.03   |

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

| Dataset  | n   | Alg. 1 | SeDuMi [21] | $\alpha_{\text{max}}$ |
|----------|-----|--------|--------------|------------------------|
| Gustavus | 57  | 0.034  | 3.97         | 6.33°                  |
| Sphinx   | 70  | 0.040  | 7.01         | 6.14°                  |
| Alcatraz | 133 | 0.066  | 58.91        | 7.68°                  |
| Pumpkin  | 209 | 0.25   | 286.85       | 8.63°                  |
| Buddha   | 322 | 6.83   | 1504.44      | 7.29°                  |

---

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

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

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

[15] 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

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

[17] R. Hartley, J. Trumpf, Y. Dai, and H. Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, 2013. 3, 6

[18] 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

[19] 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

[20] 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

[21] 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. 8, 9

[22] 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

[23] Z. Wen, D. Goldfarb, S. Ma, and K. Scheinberg. Row by row methods for semidefinite programming. Technical report, Columbia University, 2009. 7
[24] K. Wilson, D. Bindel, and N. Snavely. When is rotations averaging hard? In *European Conference on Computer Vision*, 2016.

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