Semidefinite relaxations for optimization problems over rotation matrices

James Saunderson   Pablo A. Parrilo   Alan S. Willsky

***Abstract*— Optimization problems with variables constrained to be in $SO(d)$—orthogonal matrices with determinant one—arise in attitude estimation, molecular imaging, and computer vision applications, among others. Recently it has been shown that the convex hull of $SO(d)$ can be described in terms of linear matrix inequalities. This allows us to devise new semidefinite programming-based reformulations and relaxations of problems involving rotation matrices.**

In this paper we illustrate the use of these techniques for two different types of optimization problems over $SO(d)$. The first type of problem arises in jointly estimating the attitude and spin-rate of a spin-stabilized satellite. We show how to exactly reformulate such problems as semidefinite programs. The second type of problem arises when estimating the orientations of a network of objects (such as cameras, satellites or molecules) from noisy *relative orientation* measurements. For this class of problems we formulate new semidefinite relaxations that are tighter than those existing in the literature, and show that they are exact when the underlying graph is a tree.

I. INTRODUCTION

Optimization problems with variables constrained to be in the set of rotation matrices

$$SO(d) := \{X \in \mathbb{R}^{d\times d} : X^T X = I_d,\ \det(X) = 1\}$$

arise in numerous applications, such as attitude estimation [1], computer vision [2], and molecular imaging [3]. Since $SO(d)$ is non-convex, optimization problems with decision variables constrained to be in $SO(d)$ are non-convex and are often approached using local optimization methods. Of particular interest are those exploiting the manifold structure of $SO(d)$ [4].

In this paper we take a *global* approach to problems with rotation-matrix constraints by attempting to *convexify* such problems. The basic idea is a standard one: reformulate the problem as the maximization of a linear functional over a complicated constraint set S, and then observe that it is equivalent to maximize the same linear functional over the convex hull $\operatorname{conv} S$ of that constraint set. If we have a tractable representation of $\operatorname{conv} S$ we obtain a tractable convex reformulation of the problem. It may be the case that $\operatorname{conv} S$ is very complicated, nevertheless we can obtain convex *relaxations* of the original problem using tractable outer approximations of $\operatorname{conv} S$. Such relaxations may be exact for certain instances (i.e. produce optimal solutions for the original non-convex problem). In any case the optimal value of a convex relaxation always provides global bounds on the optimal cost that can be used, for instance, to assess the quality of stationary points produced by local optimization methods.

It has recently be shown by the authors [5] that the convex hull of $SO(d)$ can be expressed as the feasible region of a semidefinite program, allowing the possibility of devising improved semidefinite programming-based relaxations and even exact reformulations of certain optimization problems over $SO(d)$. In this paper we apply these new semidefinite representations to two particular problem classes described in the following two subsections.

A. Joint attitude and spin-rate estimation

The first problem class consists of optimization problems we call *generalized joint attitude and spin-rate estimation* problems. These are optimization problems of the form

$$\max_{\substack{R\in SO(d)\\ \omega\in[0,2\pi)}} \langle A_0, R\rangle + \sum_{t=1}^{T}\langle A_t\cos(\omega t) + B_t \sin(\omega t), R\rangle. \quad (1)$$

As shown by Psiaki [1], problems of this form (with $d = 3$) generalize the classical satellite attitude (i.e. orientation) estimation problem to a situation where the aim is to jointly estimate the spin-rate and attitude of a spinning satellite. We discuss how (1) arises in this attitude estimation context in Section III.

The first main contribution of this paper is to show that problems of the form in (1) can be *exactly reformulated* as semidefinite programs of size $(T+1)2^{d-1}$. In the physically relevant case $d = 3$ these semidefinite programs have size $4(T+1)$.

B. Optimization over relative rotations

The second problem class we consider consists of optimization problems where the decision variables $R_i \in SO(d)$ are indexed by the vertices V of a graph $G = (V,E)$ and the cost function depends only on the *relative rotation* $R_i^{-1}R_j = R_i^T R_j$ over the edges $\{i,j\} \in E$, i.e.

$$\min_{R_1,\ldots,R_n\in SO(d)} \sum_{\{i,j\}\in E} f_{ij}(R_i^T R_j) \quad (2)$$

where each of the $f_{ij} : \mathbb{R}^{d\times d} \to \mathbb{R}$ are convex functions. This problem class includes a natural discrete-time filtering problem on $SO(d)$ (see Section IV-A) as well the problem of synchronization over rotations discussed, for instance, in [3], [6], [7] (see Section IV-B).

This research was supported by AFOSR under Grants FA9550-12-1-0287 and FA9550-11-1-0305.

The authors are with the Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge MA 02139, USA {jamess, parrilo, willsky}@mit.edu

The second main contribution of this paper is to give new semidefinite relaxations for optimization problems over relative rotations (2) that are tighter than the relaxations considered, for instance, in [6], [8]. We prove, in Theorem 4, that our relaxations are exact when the f_{ij} are linear and the underlying graph is a tree.

C. Notation

We briefly summarize notation not explicitly defined elsewhere. We denote by $\mathcal{S}^m$ the space of $m\times m$ real symmetric matrices and by $\mathcal{S}^m_+ \subset \mathcal{S}^m$ the positive semidefinite matrices. If $X \in \mathcal{S}^m$ we write $X \succeq 0$ to mean $X \in \mathcal{S}^m_+$. We denote $d_1 \times d_2$ real matrices by $\mathbb{R}^{d_1\times d_2}$ and for any $X, Y \in \mathbb{R}^{d_1\times d_2}$ we define $\langle X, Y\rangle = \text{tr}(X^TY)$. If $X \in \mathbb{R}^{d_1\times d_2}$ its Frobenius norm is $\|X\|_F := \langle X, X\rangle^{1/2}$. If n is a positive integer we use the shorthand $[n]$ for $\{1, 2, \ldots, n\}$. If $G = (V, E)$ is an undirected graph we write the edge joining vertices i and j as $\{i, j\}$ whereas if G is directed, we write the directed edge from vertex i to vertex j as (i, j).

D. Outline

In Section II we summarize some of the main results from [5] describing two different semidefinite representations of conv $SO(d)$. In Section III we discuss a natural formulation (due to Psiaki [1]) of the joint spin-rate and attitude estimation problem for a spinning satellite. We establish that the associated optimization problem can be exactly reformulated as a semidefinite program. In Section IV we consider new semidefinite relaxations for optimization problems over relative rotations, and prove that our relaxations are exact when the underlying graph is a tree. Finally in Section V we present numerical results showing the difference between the semidefinite relaxations for optimization over relative rotations proposed in this paper, and a standard existing relaxation.

II. Semidefinite representations of conv $SO(d)$

In this section we briefly review some of the main results of [5] where two different semidefinite descriptions of conv $SO(d)$ are given. Both descriptions are expressed in terms of a collection $(A_{ij})_{i,j=1}^d$ of $2^{d-1} \times 2^{d-1}$ symmetric matrices. We now describe these matrices. Let

$$\sigma_0 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix},\ \ \sigma_1 = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix},\ \ \sigma_2 = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix},$$

$$\lambda_i = \overbrace{\sigma_1\otimes\cdots\otimes\sigma_1}^{i-1}\otimes\sigma_2\otimes\overbrace{\sigma_0\otimes\cdots\otimes\sigma_0}^{d-i},$$

$$\rho_i = \underbrace{\sigma_0\otimes\cdots\otimes\sigma_0}_{i-1}\otimes\sigma_2\otimes\underbrace{\sigma_1\otimes\cdots\otimes\sigma_1}_{d-i}\ \text{ and}$$

$$P_{\text{even}} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}\otimes\overbrace{\sigma_0\otimes\cdots\otimes\sigma_0}^{d-1} + \frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix}\otimes\overbrace{\sigma_1\otimes\cdots\otimes\sigma_1}^{d-1}$$

where P_{even} is a $2^d \times 2^{d-1}$ zero-one matrix with exactly one 1 per column and at most one 1 per row. Then define

$$A_{ij} = -P_{\text{even}}^T\lambda_i\rho_jP_{\text{even}}.$$

for $1 \leq i, j \leq d$. With the $(A_{ij})_{i,j=1}^d$ defined, we now state the representations of conv $SO(d)$ given in [5]. The first expresses conv $SO(d)$ as the intersection of the positive semidefinite cone with an affine subspace, showing conv $SO(d)$ is a *spectrahedron* and implying that it has many nice geometric and algebraic properties.

Theorem 1: Let $R := \text{diag}(1, 1, \ldots, 1, -1)$ and $d \geq 4$. Then $X \in \text{conv}\, SO(d)$ if and only if

$$\begin{bmatrix}0 & X\\X^T & 0\end{bmatrix} \preceq I_{2d},\ \text{ and }\ \sum_{i,j=1}^{d} A_{ij}[RX]_{ij} \preceq (d-2)I_{2^{d-1}}.$$

In the case $d = 2$ we have

$$\text{conv}\, SO(2) = \left\{\begin{bmatrix}c & -s\\s & c\end{bmatrix} : \begin{bmatrix}1-c & s\\s & 1+c\end{bmatrix} \succeq 0\right\}.$$

When $d = 3$, $X \in \text{conv}\, SO(3)$ if and only if

$$\begin{bmatrix}1-X_{11}-X_{22}+X_{33} & X_{13}+X_{31} & X_{12}-X_{21} & X_{23}+X_{32}\\ X_{13}+X_{31} & 1+X_{11}-X_{22}-X_{33} & X_{23}-X_{32} & X_{12}+X_{21}\\ X_{12}-X_{21} & X_{23}-X_{32} & 1+X_{11}+X_{22}+X_{33} & X_{31}-X_{13}\\ X_{23}+X_{32} & X_{12}+X_{21} & X_{31}-X_{13} & 1-X_{11}+X_{22}-X_{33}\end{bmatrix} \succeq 0.$$

The second representation of conv $SO(d)$ given in [5] expresses conv $SO(d)$ as a projection of $2^{d-1} \times 2^{d-1}$ unit trace positive semidefinite matrices.

Theorem 2: $X \in \text{conv}\, SO(d)$ if and only if there is $Z \succeq 0$ such that $\text{tr}(Z) = 1$ and

$$X = \mathcal{A}_d(X) := \begin{bmatrix}\langle A_{11}, Z\rangle & \langle A_{12}, Z\rangle & \cdots & \langle A_{1d}, Z\rangle\\ \langle A_{21}, Z\rangle & \langle A_{22}, Z\rangle & \cdots & \langle A_{2d}, Z\rangle\\ \vdots & \vdots & \ddots & \vdots\\ \langle A_{d1}, Z\rangle & \langle A_{d2}, Z\rangle & \cdots & \langle A_{dd}, Z\rangle\end{bmatrix}. \quad (3)$$

In Section IV-B the constraint $X \in \text{conv}\, SO(d)$ can be expanded in terms of linear matrix inequalities according to either Theorem 1 or Theorem 2. In contrast, for the reformulation in Section III to work, it is crucial that we use the representation in Theorem 2 rather than the representation in Theorem 1.

III. Joint spin-rate and attitude estimation

A. Wahba's problem

The *attitude* of a satellite is the rotation $R \in SO(3)$ that transforms a reference coordinate system (the sun-fixed frame) to a coordinate system fixed with respect to the satellite (the body-fixed frame). Attitude estimation problems involve estimating the satellite attitude given noisy measurements. One family of these are called 'vector measurements' and consist of measurements in the body-fixed frame of reference directions that are known in the sun-fixed frame (e.g. the directions of stars, the earth's magnetic field orientation, etc.).

The most basic attitude estimation problem from vector measurements is known as *Wahba's problem* [9]. Suppose $y_1, y_2, \ldots, y_k$ are noisy measurements in the body-fixed frame of known directions $x_1, x_2, \ldots, x_k$ in the sun-fixed frame. We think of the attitude R and the directions x_i as fixed, with R being unknown.[1] Assume that the y_i are independent and have a von Mises-Fisher distribution on the sphere [10] with mean Rx_i and concentration parameter κ_i. Hence the y_i have joint distribution

$$p(y_1, \ldots, y_k; R) \propto \exp(\textstyle\sum_{i=1}^{k} \kappa_i \langle y_i, Rx_i \rangle)$$

where the constant of proportionality does not depend on R. Then if $X = \begin{bmatrix} x_1 & \cdots & x_k \end{bmatrix}$ and $Y = \begin{bmatrix} y_1 & \cdots & y_k \end{bmatrix}$, the maximum likelihood estimate of the attitude R can be found by solving

$$\begin{aligned} \max_{R \in SO(3)} \sum_{i=1}^{k} \kappa_i \langle y_i, Rx_i \rangle &= \max_{R \in SO(3)} \langle YKX^T, R \rangle \\ &= \max_{R \in \operatorname{conv} SO(3)} \langle YKX^T, R \rangle \end{aligned} \quad (4)$$

where $K = \operatorname{diag}(\kappa_1, \kappa_2, \ldots, \kappa_k)$. Using the semidefinite representations of the convex hull of $SO(3)$ given either in Theorem 1 or Theorem 2 the problem in (4) can be expressed as a semidefinite program.

There are other ways to solve (4), the most common being the q-method [11], which involves using the quaternion parameterization of $SO(3)$ to rewrite (4) as a symmetric eigenvalue problem. The benefit of having a semidefinite representation of $\operatorname{conv} SO(3)$ is that it allows us to reformulate more complex related problems in the framework of semidefinite programming.

B. Joint attitude and spin-rate estimation

A generalization of Wahba's problem, proposed recently by Psiaki [1], involves jointly estimating the attitude and spin-rate of a spinning satellite.

We now assume the satellite is spinning around a *known* axis (say its major axis) at an *unknown rate* ω rad/sample which we assume lies in the interval $[0, 2\pi)$.[2]

By choosing the body-fixed coordinate system appropriately we see that the attitude at time t is given by

$$R[t] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega t) & -\sin(\omega t) \\ 0 & \sin(\omega t) & \cos(\omega t) \end{bmatrix} R$$

where $R := R[0]$ is the initial attitude. At each time $t = 0, 1, 2 \ldots, T$ we obtain a batch of noisy vector measurements $y_1[t], \ldots, y_k[t]$ in the body-fixed frame of known reference directions $x_1[t], \ldots, x_k[t]$ in the sun-fixed frame. As for Wahba's problem, assume that the $y_i[t]$ are independent (for different t and i) and have von Mises-Fisher distribution with mean $R[t]x_i[t]$ and concentration parameter $\kappa_i[t]$. Let $X[t] = \begin{bmatrix} x_1[t] & \cdots & x_k[t] \end{bmatrix}$, $Y[t] = \begin{bmatrix} y_1[t] & \cdots & y_k[t] \end{bmatrix}$, and $K[t] = \operatorname{diag}(\kappa_1[t], \kappa_2[t], \ldots, \kappa_k[t])$. Under these assumptions on the $y_i[t]$, the joint maximum likelihood estimate of the initial attitude R and the spin-rate ω can be found by solving

$$\max_{\substack{R \in SO(3) \\ \omega \in [0, 2\pi)}} \langle A_0, R \rangle + \sum_{t=1}^{T} \langle A_t \cos(\omega t) + B_t \sin(\omega t), R \rangle \quad (5)$$

where

$$A_0 = Y[0]K[0]X[0]^T + \sum_{t=1}^{T} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} Y[t]K[t]X[t]^T,$$

$$A_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} Y[t]K[t]X[t]^T, \text{ and}$$

$$B_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} Y[t]K[t]X[t]^T.$$

This is clearly of the general form (1) stated in the introduction. For the remainder of the section we work with the general form in (1).

C. An exact SDP reformulation of the general joint attitude and spin-rate estimation problem

If we define $\mathcal{M}_{d,T} \subset (\mathbb{R}^{d \times d})^{2T+1}$ to be

$$\{(R, R\cos(\omega), R\sin(\omega), \ldots, R\cos(T\omega), R\sin(T\omega)) : \\ R \in SO(d), \omega \in [0, 2\pi)\}$$

then (1) can be reformulated as the maximization of a linear functional over $\mathcal{M}_{d,T}$ which is equivalent to the maximization of the same linear functional over $\operatorname{conv} \mathcal{M}_{d,T}$:

$$\begin{aligned} &\max_{(X_t)_{t=0}^T, (Y_t)_{t=1}^T} \langle A_0, X_0 \rangle + \sum_{t=1}^{T} [\langle A_t, X_t \rangle + \langle B_t, Y_t \rangle] \\ &\text{subject to} \quad (X_0, X_1, Y_1, \ldots, X_T, Y_T) \in \operatorname{conv} \mathcal{M}_{d,T}. \end{aligned}$$

To reformulate the problem (1) as a semidefinite program, it remains to show that the convex hull of $\mathcal{M}_{d,T}$ has a semidefinite representation. This is the case because

1) $\operatorname{conv} \mathcal{M}_{d,T}$ is the image of another convex body, $\operatorname{conv} \widetilde{\mathcal{M}}_{2^{d-1},T}$, under a linear map (Proposition 1 to follow). The proof of this requires the description of $\operatorname{conv} SO(d)$ in Theorem 2.
2) For any positive integers m and T, $\operatorname{conv} \widetilde{\mathcal{M}}_{m,T}$ is the feasible region of a linear matrix inequality involving symmetric matrices of size $(T+1)m$.

For positive integers m, T, define $\widetilde{\mathcal{M}}_{m,T} \subset (\mathcal{S}^m)^{2T+1}$ by

$$\{(qq^T, qq^T\cos(\omega), qq^T\sin(\omega), \ldots, qq^T\cos(T\omega), qq^T\sin(T\omega)) : \\ q^Tq = 1,\ \omega \in [0, 2\pi)\}.$$

We now state the relationship between $\operatorname{conv} \mathcal{M}_{d,T}$ and $\operatorname{conv} \widetilde{\mathcal{M}}_{2^{d-1},T}$.

Proposition 1: $\operatorname{conv} \mathcal{M}_{d,T} = \widetilde{\mathcal{A}}(\operatorname{conv} \widetilde{\mathcal{M}}_{2^{d-1},T})$ where $\widetilde{\mathcal{A}}$ is the linear map that sends $((X_t)_{t=0}^T, (Y_t)_{t=1}^T)$ to

$$(\mathcal{A}_d(X_0), \mathcal{A}_d(X_1), \mathcal{A}_d(Y_1), \ldots, \mathcal{A}_d(X_T), \mathcal{A}_d(Y_T))$$

[1] We could alternatively take a Bayesian view, thinking of R as random with some prior distribution.

[2] In fact we need only assume that there is an interval $[a, a + 2\pi)$ containing ω. An assumption of this form is required due to aliasing—our formulation cannot distinguish between frequencies differing by multiples of 2π.

and $\mathcal{A}_d : \mathcal{S}^{2^{d-1}} \to \mathbb{R}^{d\times d}$ is defined in (3) of Section II.

Proof: See the appendix. ■

To complete the section we state a linear matrix inequality description of $\operatorname{conv} \widetilde{\mathcal{M}}_{m,T}$.

Proposition 2: $\operatorname{conv} \widetilde{\mathcal{M}}_{m,T}$ is the set of $(2T+1)$-tuples $(X_0, X_1, Y_1, \ldots, X_T, Y_T)$ of symmetric matrices s.t. $\operatorname{tr}(X_0) = 1$ and

$$\begin{bmatrix} X_0 & X_1 & X_2 & \cdots & X_T \\ X_1 & X_0 & X_1 & \ddots & \vdots \\ X_2 & X_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & X_1 \\ X_T & \cdots & \cdots & X_1 & X_0 \end{bmatrix} + \begin{bmatrix} -Y_T & -Y_{T-1} & \cdots & -Y_1 & 0 \\ -Y_{T-1} & \iddots & \iddots & 0 & Y_1 \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ -Y_1 & 0 & \iddots & \iddots & Y_{T-1} \\ 0 & Y_1 & \cdots & Y_{T-1} & Y_T \end{bmatrix} \succeq 0.$$

Proof: See [12, Appendix A]. The idea of the proof is to combine the matrix version of the Fejér-Riesz theorem (that Hermitian positive semidefinite-valued matrices with univariate trigonometric polynomial entries are Hermitian squares) with a symmetry reduction argument. ■

Combining Propositions 1 and 2 establishes our main result for this section.

Theorem 3: For any positive integer T, the joint attitude and spin-rate estimation problem in (5) can be reformulated as a semidefinite program of size $4(T+1)$.

IV. Optimization over relative rotations

We now turn to optimization problems over many rotation-matrix valued variables (e.g. satellites in formation or a network of cameras) indexed by the vertices of a graph. The edges in the graph may, for instance, correspond to satellites within low-power communication range, or cameras that can see each other. In particular we focus on cases where the cost function on edge (i, j) depends only on the *relative rotation* $R_i^T R_j$ of the variables corresponding to the endpoints of that edge, giving rise to problems of the form

$$\min_{R_1,\ldots,R_n \in SO(d)} \sum_{\{i,j\}\in E} f_{ij}(R_i^T R_j) \tag{6}$$

where each of the $f_{ij} : \mathbb{R}^{d\times d} \to \mathbb{R}$ are convex functions and $f_{ij}(X) = f_{ij}(X^T)$. By choosing different graph structures and cost functions f_{ij} we capture a natural discrete-time filtering problem on $SO(d)$ (see Section IV-A) as well as many of the formulations of 'synchronization over rotations' considered, for instance, in [6], [7], [8] (see Section IV-B).

At this stage we would like to emphasize that if $(R_1, \ldots, R_n)$ is an optimizer of (6) then so is $(RR_1, \ldots, RR_n)$ for any $R \in SO(d)$. As such we can 'dehomogenize' the problem by choosing a vertex in each connected component of G and setting $R_i = I_d$ for these vertices. This leads to an equivalent 'inhomogeneous' problem class of the form

$$\min_{R_1,\ldots,R_n \in SO(d)} \sum_{\{i,j\}\in E'} f_{ij}(R_i^T R_j) + \sum_{i\in V'} f_i(R_i) \tag{7}$$

which may be more natural in some settings. Importantly, though, the graphs corresponding to the homogeneous problem and the inhomogeneous problem are *different*. If $G' = (V', E')$ is the graph for the inhomogeneous problem (7) then the graph $G = (V, E)$ for the corresponding homogeneous problem is the *suspension graph* of G' obtained by adding a new vertex and connecting it to all existing vertices. For example if G' is a chain of length 6, G is the graph shown in Figure 1, which is not a chain.

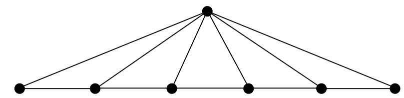

Fig. 1. Graph corresponding to discrete-time filtering problem written in homogeneous form.

A. Example 1: Discrete-time filtering on $SO(d)$

Suppose $R[0], R[1], \ldots, R[T]$ is a sequence of rotations which we wish to estimate from noisy observations of their action on vectors in $\mathbb{R}^d$. Such a situation could be used to model a dynamic attitude estimation problem such as the one described in Section III-B. We consider a basic probabilistic model that models the dynamics of the underlying sequence of rotations as random, in contrast with the deterministic model in the joint spin-rate and attitude estimation problem discussed in Section III-B.

In particular assume that at time t we observe $y_1[t], \ldots, y_k[t]$, noisy observations of the rotations $R[t]x_1[t], \ldots, R[t]x_k[t]$ of a collection $x_1[t], \ldots, x_k[t]$ of known vectors. Assume the $R[t]$ and $y_i[t]$ follow a hidden Markov model with state space $SO(d)$, observations $y_1[t], \ldots, y_k[t]$ that (conditioned on the state) are i.i.d. von Mises-Fisher distributed with mean $R[t]x_i[t]$ and concentration parameter κ, and transition kernel $\Pr[R[t+1]|R[t]]$ proportional to $\exp(w_t \operatorname{tr}(R[t]^T R[t+1]))$ with $w_t > 0$. One could think of the interaction between $R[t]$ and $R[t+1]$ as putting higher weight on sequences where consecutive rotations are similar (depending on the magnitude of w_t) and so putting higher weight on slowly varying sequences $R[t]$.

A solution of the following optimization problem gives a most probable sequence of rotations given the observed data (where the notation $Y[t]$ and $X[t]$ are from Section III-B)

$$\max \sum_{t=0}^{T} w_t \operatorname{tr}(R[t]^T R[t+1]) + \kappa \sum_{t=0}^{T} \langle R[t], Y[t]X[t]^T \rangle \tag{8}$$

where the maximization is over $R[0], \ldots, R[T] \in SO(d)$. We can rewrite (8) in the form of (6) by homogenizing as discussed in Section IV. The resulting graph G is of the form shown in Figure 1, and so is not a tree.

B. Example 2: Synchronization over rotations

Another family of optimization problems over relative rotations has received significant attention recently is that of *synchronization problems over rotations*. In this case there are multiple variables $R_1, \ldots, R_n \in SO(d)$ and we are given (noisy) measurements $\hat{R}_{ij}$ of the relative rotations $R_i^T R_j$ for some subset E of pairs of the variables. The aim is to recover the underlying rotations $R_1, \ldots, R_d$ (up to a global ambiguity caused by the fact that we only have measurements of the relative rotations). Many formulations have recently been proposed for this problem. For example, by choosing

$$f_{ij}(R_i^T R_j) = \|\hat{R}_{ij} - R_i^T R_j\|_F^2 = 2 - 2\langle \hat{R}_{ij}, R_i^T R_j \rangle$$

we obtain (up to an additive constant) a formulation with the f_{ij} being linear functionals. With

$$f_{ij}(R_i^T R_j) = \|\hat{R}_{ij} - R_i^T R_j\|_F$$

we obtain the least unsquared deviation formulation in [6]. Taking $f_{ij}(R_i^T R_j)$ to be the log-likelihood function of a certain mixture model on $SO(3)$ we obtain the robust formulation in [7].

C. Semidefinite relaxations

The $O(d)$-based relaxation: The standard convex relaxation of (6) proposed in the literature is a fairly direct generalization of the usual semidefinite relaxation for binary quadratic optimization. The idea is that if $R_1, R_2, \ldots, R_n \in SO(d)$ then $R_i^T R_i = I_d$. Hence the matrix of relative rotations

$$M := [R_i^T R_j]_{i,j=1}^n = \begin{bmatrix} I_d & R_1^T R_2 & \cdots & R_1^T R_n \\ R_2^T R_1 & I_d & \cdots & R_2^T R_n \\ \vdots & \vdots & \ddots & \vdots \\ R_n^T R_1 & R_n^T R_2 & \cdots & I_d \end{bmatrix}$$

satisfies $M_{ii} = I_d$ and $M \succeq 0$. (Note that throughout this section we think of $nd \times nd$ symmetric matrices of $n \times n$ block matrices with each block being $d \times d$ and use indices $1 \leq i, j \leq n$ to index the $d \times d$ blocks.)

These observations give rise to the following convex relaxation of (6) proposed in the literature (see, e.g., [6]):

$$\min_{M \in \mathcal{S}^{nd}} \sum_{\{i,j\} \in E} f_{ij}(M_{ij}) \text{ s.t. } \begin{cases} M \succeq 0, \\ M_{ii} = I_d, \ \forall i \in [n]. \end{cases} \quad (9)$$

We call it the *$O(d)$-based relaxation* because (9) only uses the fact that if $R_i \in SO(d)$ then $R_i^T R_i = I$, i.e. that

$$SO(d) \subset O(d) = \{X \in \mathbb{R}^{d \times d} : X^T X = I_d\}.$$

It does not use the additional constraint that $\det(R_i) = 1$.[3]

The $SO(d)$-based relaxation: We now propose a new convex relaxation for problems of the form (6) that makes use of the semidefinite representations of $\operatorname{conv} SO(d)$ from [5] described in Section II. Our relaxation is based on the following characterization of matrices of relative rotations.

Lemma 1: Suppose $R_1, R_2, \ldots, R_n \in SO(d)$ and $M = [R_i^T R_j]_{i=1,j}^n$. Then

- $M_{ii} = I_d$
- $M \succeq 0$
- $M_{ij} \in \operatorname{conv} SO(d)$ for $i \neq j$.
- $\operatorname{rank}(M) = d$.

[3]We note that in [8] a finite set of valid inequalities for $\operatorname{conv} SO(d)$ are added to M_{ij} for $i \neq j$, giving a formulation tighter than (9) and not as tight as the $SO(d)$-based relaxation.

Conversely any $nd \times nd$ symmetric matrix M satisfying these four conditions has a unique factorization as $M = [R_i^T R_j]_{i,j=1}^n$ with $R_i \in SO(d)$ and $R_1 = I_d$.

Proof: Whenever $R_1, R_2, \ldots, R_n \in SO(d)$ it is clear that $[R_i^T R_j]_{i,j=1}^n$ satisfies these four properties. Conversely if M is positive semidefinite and has rank d then there are matrices $R_1, \ldots, R_d$ such that $M = [R_i^T R_j]_{i,j=1}^n$. Since $M_{ii} = I_d$ for all $i \in [n]$ each of the R_i are orthogonal. Without loss of generality assume $R_1 = I_d$ (otherwise multiply all the matrices on the left by R_1^T). This removes any ambiguity in the factorization, showing it is unique.

It remains to show that $R_i \in SO(d)$ for $i \in [n]$. Since $M_{ij} \in \operatorname{conv} SO(d)$ for $i \neq j$ we have that $R_1^T R_j = R_j \in \operatorname{conv} SO(d)$ for $j \in [n]$. Since we know that $R_i \in O(d)$ for $i \in [n]$, to conclude the proof we need to show that

$$\operatorname{conv} SO(d) \cap O(d) = SO(d).$$

Since $SO(d) \subset \operatorname{conv} SO(d)$ and $SO(d) \subset O(d)$ we have that $\operatorname{conv} SO(d) \cap O(d) \supseteq SO(d)$. For the reverse inclusion note that $SO(d)$ and $O(d)$ are both subsets of the Frobenius norm sphere of radius $\sqrt{d}$, denoted $S \subset \mathbb{R}^{d \times d}$. It then follows that $\operatorname{conv} SO(d) \cap S = SO(d)$. Since, in addition, $O(d) \subset S$ we can conclude that $\operatorname{conv} SO(d) \cap O(d) \subseteq \operatorname{conv} SO(d) \cap S = SO(d)$ as required. ■

Note that Lemma 1 would also be true (but less interesting) if we replaced the condition $M_{ij} \in \operatorname{conv} SO(d)$ for $i \neq j$ with the stronger requirement that $M_{ij} \in SO(d)$ for $i \neq j$.

Omitting the rank constraint from the characterization in Lemma 1 we obtain our *$SO(d)$-based relaxation*:

$$\min_{M} \sum_{\{i,j\} \in E} f_{ij}(M_{ij}) \text{ s.t. } \begin{cases} M \succeq 0 & \\ M_{ii} = I_d & \forall i \in [n] \\ M_{ij} \in \operatorname{conv} SO(d) & \text{if } i \neq j. \end{cases} \quad (10)$$

We remark that using Theorem 1 to represent $\operatorname{conv} SO(d)$ shows that the feasible region of this semidefinite program is a spectrahedron.

Exactness for trees: We say that the $SO(d)$-based relaxation (10) is *exact* if it has an optimal solution $M^\star$ of rank d. In this case by Lemma 1 we can factorize $M^\star$ to obtain an optimal solution of the original non-convex problem (6).

One feature of the $SO(d)$-based relaxation (10) not enjoyed by the $O(d)$-based relaxation (9) is that it is exact when the underlying graph is a tree and the f_{ij} are linear functionals. We prove this in Theorem 4 (to follow). The key fact underlying the argument is the following.

Lemma 2: Let $G = (V, E)$ be a rooted tree with n vertices. Suppose vertex 1 is the root and the edges are oriented away from the root. Let $(R_{(i,j)})_{(i,j) \in E}$ be an arbitrary collection of $n-1$ elements of $SO(d)$ indexed by the oriented edges of the tree. Then there are $R_1, \ldots, R_n \in SO(d)$ with $R_1 = I_d$ such that $R_i^T R_j = R_{(i,j)}$ for every oriented edge $(i, j) \in E$.

Proof: Define $R_1 = I_d$. Since G is a tree with edges oriented away from vertex 1, for any $i \in V \setminus \{1\}$ there is a unique (oriented) path $1, v_1^i, \ldots, v_k^i = i$ from vertex 1 to vertex i. For $i \neq 1$ define

$$R_i = R_{(1,v_1^i)} R_{(v_1^i, v_2^i)} \cdots R_{(v_{k-1}^i, i)}.$$

Since each $R_{(i,j)} \in SO(d)$, it follows that $R_i \in SO(d)$ for each i. Furthermore, if $(i,j) \in E$ is an oriented edge then the oriented path from 1 to i is $1, v_1^i, \ldots, v_{k-1}^i, i$ and the oriented path from 1 to j is $1, v_1^i, \ldots, v_{k-1}^i, i, j$. Hence

$$\begin{aligned} R_i^T R_j &= R_{(v_{k-1}^i, i)}^T \cdots R_{(v_1^i, 1)}^T R_{(v_1^i, 1)} \cdots R_{(v_{k-1}^i, i)} R_{(i,j)} \\ &= R_{(i,j)} \end{aligned}$$

as required. ■

We now prove that the $SO(d)$-based relaxation (10) is exact when the underlying graph is a tree and the f_{ij} are linear.

Theorem 4: If each of the f_{ij} are linear and the underlying graph G is a tree then the convex relaxation (10) is exact.

Proof: For convenience of notation, root the tree at vertex 1, orient the edges away from the root, and assume the vertices are sorted so that if $\{i, j\}$ is an undirected edge (with $i < j$) then the corresponding oriented edge is (i, j).

We, in fact, show that the following optimization problem

$$\begin{aligned} &\min_M \sum_{(i,j)\in E} f_{ij}(M_{ij}) \\ &\text{s.t.} \quad M_{ij} \in \operatorname{conv} SO(d),\ \forall (i,j) \in E \end{aligned} \tag{11}$$

obtained from (10) by keeping the same cost function and keeping only the constraints corresponding to the edges of the graph, has an optimal solution $M^\star$ that satisfies $M^\star \succeq 0$, $M_{ii}^\star = I_d$, $M_{ij}^\star \in \operatorname{conv} SO(d)$ and $\operatorname{rank}(M^\star) = d$. This implies that the $SO(d)$-based relaxation (10) is also exact.

The optimization problem (11) is separable, so for any optimal $\hat{M}$ we have that when $(i,j) \in E$ the (i,j) block of $\hat{M}$, namely $\hat{M}_{ij}$, is optimal for

$$\min_{M_{ij}} f_{ij}(M_{ij}) \text{ s.t. } M_{ij} \in \operatorname{conv} SO(d). \tag{12}$$

Fix $(i,j) \in E$. Since f_{ij} is linear, the minimum of (12) is achieved at an extreme point of $\operatorname{conv} SO(d)$ and so is an element of $SO(d)$. So there is some optimal $\hat{M}$ for (11) satisfying $\hat{M}_{ij} \in SO(d)$ for $(i,j) \in E$. Apply Lemma 2 to the $\hat{M}_{ij}$ to conclude that there are $R_1 = I_d, R_2, \ldots, R_n$ such that $R_i^T R_j = \hat{M}_{ij}$. Then $M^\star = [R_i^T R_j]_{i,j=1}^n$ is also an optimal solution to (11) with the desired properties, completing the argument. ■

We remark that the same argument works when the underlying graph is a forest. In that case one can apply the argument separately to each connected component of the graph.

V. Numerical experiments

It is clear from the definitions that our SDP relaxation (10) is at least as tight as the standard SDP relaxation (9). In this section we describe some simple numerical experiments that illustrate the gap between these two relaxations for some natural families of problem instances. To solve the semidefinite programs described in this section we use the parser YALMIP [13] and the solver MOSEK.

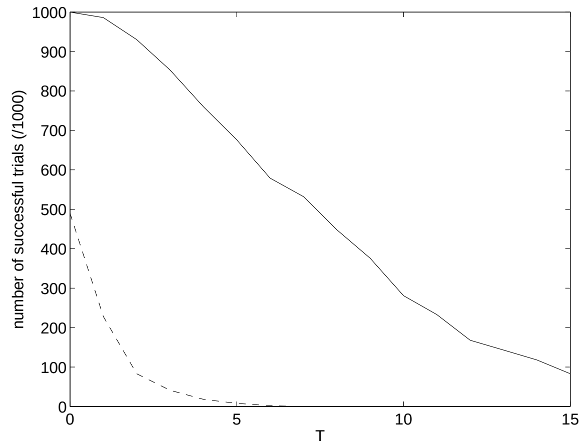

Fig. 2. We show the number of trials of randomly generated filtering problems (8) with $d = 3$ for which the $SO(d)$-based (solid) and $O(d)$-based (dashed) relaxations are, respectively, exact. For $T = 0$ the relaxations reduce to optimizing a random linear functional over the convex hull of $SO(3)$ and $O(3)$ respectively. In this case the $SO(d)$-based relaxation is always exact whereas the $O(d)$-based relaxation is exact when the linear functional has positive determinant—for our ensemble this occurs with probability $1/2$.

A. Discrete-time filtering on $SO(d)$

We consider problems with the discrete-time filtering structure from Section IV-A with $d = 3$ and $T = 1, 2, \ldots, 15$. For each of these values of T we sample 1000 independent random problem instances. For each instance we solve the $SO(d)$-based and $O(d)$-based relaxations. We record the number of instances for the two relaxations, respectively, solve the original non-convex optimization problem. Each random problem instance is specified by taking $w_t = 1$ for all $t = 0, 1, \ldots, T$ and sampling $T + 1$ matrices $A_0, A_1, \ldots, A_T \in \mathbb{R}^{d\times d}$ each having independent standard Gaussian entries.

The results of the experiment are shown in Figure 2. They indicate that the $SO(d)$-based relaxation is often exact, particularly for smaller problem sizes (i.e. small T), whereas the $O(d)$-based relaxation typically fails to be exact for problems of this type.

B. Optimization over relative rotations

We generate random 'planted' instances from an ensemble analyzed by Wang and Singer [6]. First fix $R_1, R_2, \ldots, R_n \in SO(3)$. For each $i < j$ sample $\hat{R}_{ij} \in SO(3)$ by

$$\hat{R}_{ij} = \begin{cases} R_i^T R_j & \text{with probability } p \\ R_{ij} & \text{with probability } 1 - p \end{cases}$$

where the R_{ij} are i.i.d. samples from the uniform distribution on $SO(3)$. Define $f_{ij}(X) = -\langle \hat{R}_{ij}, X\rangle$. This models a situation where there is a 'planted' consistent set of underlying relative rotations corrupted by noise.

For $n = 10$ and for $p = 0, 0.05, 0.1, \ldots, 0.95, 1$ we sample 1000 independent problem instances as described in the previous paragraph. For each instance we solve both the

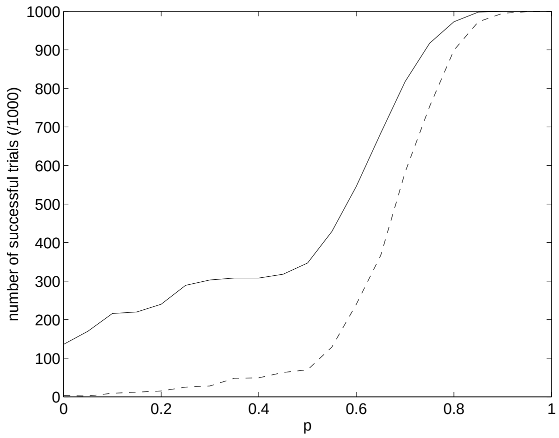

Fig. 3. Plot of the number of random instances of synchronization problems (see Section IV-B) of size $n = 10$ with probability p of obtaining a correct measurement on each edge, for which the $SO(d)$-based relaxation (solid) and the $O(d)$-based relaxation (dashed) are, respectively, exact.

$SO(d)$- and $O(d)$- based relaxations and record how many instances, for each p, are exact. The results are shown in Figure 3.

The experiment reveals two regimes among these problem instances. For large p both relaxations work well. This is because in these cases the linear functional we are optimizing is a slight perturbation of a linear functional pointing in the direction of a point in the constraint set for the underlying non-convex problem (which is a subset of the sphere). For small p, the $O(d)$-based relaxation typically fails, whereas the $SO(d)$-based relaxation can be exact even with independent random edge variables in $SO(3)$ (i.e. $p = 0$).

VI. Conclusion

In this paper we illustrated the use of semidefinite descriptions of the convex hull of rotation matrices for two classes of optimization problems over rotation matrices. We showed how to obtain an exact semidefinite programming reformulation of a joint satellite attitude and spin-rate estimation problem and described tighter semidefinite relaxations, exact for problems defined with respect to trees, for certain optimization problems over relative rotations.

References

[1] M. L. Psiaki, "Generalized Wahba problems for spinning spacecraft attitude and rate determination," *J. Astronautical Sciences*, vol. 57, no. 1-2, pp. 73–92, 2009.

[2] M. B. Horowitz, N. Matni, and J. W. Burdick, "Convex relaxations of SE(2) and SE(3) for visual pose estimation," 2014, arxiv:1401.3700.

[3] A. Singer and Y. Shkolnisky, "Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming," *SIAM J. Imaging Sci.*, vol. 4, no. 2, pp. 543–572, 2011.

[4] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[5] J. Saunderson, P. A. Parrilo, and A. S. Willsky, "Semidefinite descriptions of the convex hull of rotation matrices," 2014, arXiv:1403.4914.

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

[7] N. Boumal, A. Singer, and P.-A. Absil, "Robust estimation of rotations from relative measurements by maximum likelihood," in *Proceedings of the 52nd Conference on Decision and Control, CDC 2013*, 2013.

[8] M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri, "Global motion estimation from point matches," in *Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012*. IEEE, 2012, pp. 81–88.

[9] G. Wahba, "A least squares estimate of satellite attitude," *SIAM Rev.*, vol. 7, no. 3, pp. 409–409, 1965.

[10] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009, vol. 494.

[11] J. E. Keat, "Analysis of least-squares attitude determination routine DOAOP," Computer Sciences Corporation, Tech. Rep. CSC/TM-77/6034, February 1977.

[12] J. Saunderson, P. A. Parrilo, and A. S. Willsky, "A solution to Psiaki's first joint attitude and spin-rate estimation problem," 2014, in preparation.

[13] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip

Appendix

A. Proof of Proposition 1

Suppose

$$(R, R\cos(\omega), R\sin(\omega), \ldots, R\cos(T\omega), R\sin(T\omega)) \in \mathcal{M}_{d,T}.$$

Then since $R \in SO(d)$ by Theorem 2 there is $q \in \mathbb{R}^{2^{d-1}}$ such that $q^T q = 1$ and $\mathcal{A}_d(qq^T) = R$. (This is the case because R is an extreme point of $\operatorname{conv} SO(d)$ so its preimage under $\mathcal{A}_d$ must consist of extreme points of unit trace positive semidefinite matrices.) Hence

$$(R, R\cos(\omega), R\sin(\omega), \ldots, R\cos(T\omega), R\sin(T\omega)) = \\ \widetilde{\mathcal{A}}(qq^T, qq^T\cos(\omega), qq^T\sin(\omega), \ldots, qq^T\cos(T\omega), qq^T\sin(T\omega)).$$

This establishes that $\mathcal{M}_{d,T} \subseteq \widetilde{\mathcal{A}}(\widetilde{\mathcal{M}}_{2^{d-1},T})$ and so that

$$\operatorname{conv} \mathcal{M}_{d,T} \subseteq \operatorname{conv} \widetilde{\mathcal{A}}(\widetilde{\mathcal{M}}_{2^{d-1},T}) = \widetilde{\mathcal{A}}(\operatorname{conv} \widetilde{\mathcal{M}}_{2^{d-1},T}).$$

We now turn our attention to the reverse inclusion. Let

$$(qq^T, qq^T\cos(\omega), qq^T\sin(\omega), \ldots, qq^T\cos(T\omega), qq^T\sin(T\omega))$$

be an element of $\widetilde{\mathcal{M}}_{2^{d-1},T}$. Then since $\mathcal{A}_d(qq^T) \in \operatorname{conv} SO(d)$ we can express $\mathcal{A}_d(qq^T)$ as a convex combination of elements of $SO(d)$, i.e.,

$$\mathcal{A}_d(qq^T) = \sum_{i=1}^{\ell} \lambda_i R_i$$

where the $R_i \in SO(d)$ and the $\lambda_i \geq 0$ satisfy $\sum_{i=1}^{\ell} \lambda_i = 1$. Hence

$$\widetilde{\mathcal{A}}(qq^T, qq^T\cos(\omega), qq^T\sin(\omega), \ldots, qq^T\cos(T\omega), qq^T\sin(T\omega)) \\ = \sum_{i=1}^{\ell} \lambda_i (R_i, R_i\cos(\omega), R_i\sin(\omega), \ldots, \\ R_i\cos(T\omega), R_i\sin(T\omega))$$

and so $\widetilde{\mathcal{A}}(\widetilde{\mathcal{M}}_{2^{d-1},T}) \subseteq \operatorname{conv} \mathcal{M}_{d,T}$. This establishes the reverse inclusion that $\widetilde{\mathcal{A}}(\operatorname{conv} \widetilde{\mathcal{M}}_{2^{d-1},T}) \subseteq \operatorname{conv} \mathcal{M}_{d,T}$.