

A convex approach to consensus on $SO(n)$

Nikolai Matni and Matanya B Horowitz

Abstract

This paper introduces several new algorithms for consensus over the special orthogonal group. By relying on a convex relaxation of the space of rotation matrices, consensus over rotation elements is reduced to solving a convex problem with a unique global solution. The consensus protocol is then implemented as a distributed optimization using (i) dual decomposition, and (ii) both semi and fully distributed variants of the alternating direction method of multipliers technique – all with strong convergence guarantees. The convex relaxation is shown to be exact at all iterations of the dual decomposition based method, and exact once consensus is reached in the case of the alternating direction method of multipliers. Further, analytic and/or efficient solutions are provided for each iteration of these distributed computation schemes, allowing consensus to be reached without any online optimization. Examples in satellite attitude alignment with up to 100 agents, an estimation problem from computer vision, and a rotation averaging problem on $SO(6)$ validate the approach.

I. INTRODUCTION

Optimization, coordination, and consensus over the group of rotation matrices (i.e. over elements of $SO(n)$) is a problem of fundamental importance in a wide range of applications, from satellite attitude and spin estimation [1], [2], vehicle coordination [3], frequency synchronization [4], (distributed) visual pose estimation [5], [6], [7] and protein folding [8].

Traditional consensus in Euclidean space has a rich history [9], with results offering strong convergence guarantees under mild and realistic assumptions using *purely local* protocols. Further, with the resurfacing of distributed optimization methods such as dual decomposition [10] and the alternating direction method of multipliers [11], these methods have been successfully

N. Matni is with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA. nmatni@caltech.edu.

M. Horowitz is with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA. mhorowit@caltech.edu.

applied to achieve so-called “fast consensus” in a distributed [12] and semi-distributed (i.e. sensor fusion) [11] setting. Unfortunately, generalizing these approaches to manifolds, even those with as much structure as the group of rotation matrices, has proven non-trivial.

Fortunately, although the space of rotation matrices is highly non-linear, it is a Lie Group. This structure has allowed for sophisticated consensus methods to be created with so-called “almost-global” convergence – that is to say the only *stable* stationary points of the protocol are global minimizers of the underlying optimization. These consensus schemes fall under two broad categories, namely *intrinsic* and *extrinsic* approaches.

In the former, the consensus and/or optimization is performed directly on the manifold with respect to its geodesic distance: although many such methods exist in the literature¹, those found in [13] appear to be the most general and offer the strongest convergence guarantees. This technique is able to achieve almost-global convergence, despite the inherent non-convexity of the underlying optimization, through an appropriate re-weighting of the consensus problem’s objective function.

The category of extrinsic protocols² rely on embedding the underlying manifold into a Euclidean space, applying standard consensus protocols in this space, and then projecting the resulting states back onto the manifold [14]. These methods can also be constructed to provide almost-global convergence, but are highly dependent on the choice of embedding and projection methods, and do not easily generalize.

A common theme in all of the above methods is the use of sophisticated tools from differential geometry and Lie Group theory – although elegant and appropriate for an applied mathematics community, such mathematical overhead may prove to be a barrier for the adoption of such methods by engineering practitioners.

Contributions: This paper addresses all of these issues by exploiting a recent characterization of the convex hull of $SO(n)$ [1]. In particular, we reduce the consensus optimization to a *convex* one with a *unique global minimizer* (under a mild technical assumption), and leverage distributed optimization techniques such as dual decomposition [10] and the alternating direction method of multipliers [11] to implement a consensus protocol. We further show that this global optimizer

¹The interested reader should refer to the references of [13] for a more detailed overview of prior work.

²Once again, a more exhaustive overview of the state of the art of extrinsic methods can be found in [14], and the references therein.

lies in $SO(n)$, rather than $\text{conv}\{SO(n)\}$, and thus the convex relaxation is indeed *exact* once consensus is achieved.

We then prove that for a dual decomposition based method, the consensus path (i.e. all of the iterates of the protocol) are in fact in $SO(n)$, whereas for the ADMM protocol, we provide empirical evidence that an appropriate choice of algorithm parameters lead to iterates that are in or near $SO(n)$. Finally, we show that our algorithms can be efficiently implemented, with individual agents only needing to perform an eigen-decomposition of a symmetric matrix, and/or a projection of a vector onto the probability simplex.

We illustrate the usefulness of our approach on a rotation averaging problem (motivated by satellite formation control) for $N = 8$, $N = 50$ and $N = 100$ agents arranged in a *ring topology*, as well as on pose estimation problems based on observations given by several cameras, and on a consensus problem in $SO(6)$. Finally, it is worth noting that if the results on the spectrahedral representation of the convex hull of $SO(n)$ from [1] and the well established distributed computation techniques we rely upon are taken as fact, the derivations require only elementary linear algebra and basic convex analysis.

Paper outline: In Section II, we begin by fixing notation and terminology, and then formulate the consensus problem over $SO(n)$ as a convex optimization over the convex hull of the group of rotations. In particular, we show that the problem reduces to optimizing a linear functional defined by the average of the agents' data over this space. In Section III, we provide an overview of the required results on the spectrahedral representation of the convex hull of $SO(n)$ [1], and on distributed optimization techniques [10], [11]. In Section IV we derive our consensus algorithms, and in Section V we illustrate their effectiveness on several challenging consensus problems. We end in Section VI with conclusions and directions for future work.

II. PROBLEM FORMULATION

A. Notation and Terminology

We denote the Euclidean norm of a vector $x \in \mathbb{R}^n$ by $\|\cdot\|$. The space of $n \times n$ symmetric matrices is denoted by \mathbb{S}^n , and is equipped with the natural inner product $\langle A, B \rangle = \text{trace}(A^\top B) = \text{trace}(AB)$ which induces the Frobenius norm of an element $M \in \mathbb{S}^n$ as $\|M\|_F^2 = \text{trace}(M^2)$. The set of n dimensional rotation matrices, $SO(n)$, is defined as

$$SO(n) := \{R \in \mathbb{R}^{n \times n} \mid R^\top R = I, \det R = 1\}. \quad (1)$$

The convex hull operator is denoted by $\text{conv}\{\cdot\}$, and we use $M \succeq 0$ to indicate that $M \in \mathbb{S}^n$ is positive semi-definite. As is standard, we refer to the set of unit trace positive semi definite matrices, that is to say those $M \in \mathbb{S}^n$ satisfying $M \succeq 0$, $\text{trace}M = 1$, as the *free spectrahedron* – this is the generalization of the probability simplex to \mathbb{S}^n .

B. The Generalized Average

Consider N agents, each with state $R_t^i \in SO(n)$, interacting according to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $(i, j) \in \mathcal{E} \subset 2^\mathcal{V}$ if and only if agent j can communicate its state to agent i (note we assume $(i, i) \in \mathcal{E}$ for all $i \in \mathcal{V}$). Further define the set of incoming neighbors at node i as $\mathcal{N}_i := \{j \mid (i, j) \in \mathcal{E}\}$ to be those nodes that can communicate their state directly to node i according to the graph topology.

In this paper, we consider the problem of implementing a (semi) distributed algorithm such that each state R_t^i converges to a consensus value \bar{R} , which we define as the generalized average rotation of the system given by

$$\bar{R} := \arg \min_{R \in SO(n)} \sum_{i=1}^N \|C^i - RB^i\|_F^2, \quad (2)$$

where each C^i and B^i are appropriately sized local data matrices.

Before justifying the use of this cost function and the generalized average terminology, we begin by providing a *convex* reformulation of optimization (2).

Lemma 1: For \bar{R} defined as in (2), we have that

$$\bar{R} = \arg \max_{R \in \text{conv}\{SO(n)\}} \left\langle \frac{1}{N} \sum_{i=1}^N C^i (B^i)^\top, R \right\rangle. \quad (3)$$

Proof: Expanding the objective of optimization (2), we may rewrite it as

$$\|C^i\|_F^2 - 2 \langle C^i, RB^i \rangle + \|RB^i\|_F^2. \quad (4)$$

Noting that elements of $SO(n)$ leave the Frobenius norm invariant, the only non-constant term is the middle inner-product. Multiplying by appropriate constants and taking the negative sign into account, we then obtain the following non-convex optimization

$$\bar{R} = \arg \max_{R \in SO(n)} \left\langle \frac{1}{N} \sum_{i=1}^N C^i (B^i)^\top, R \right\rangle. \quad (5)$$

However, noting that we are optimizing a linear functional over a basic semi-algebraic set, we may replace the feasible set with its convex hull, yielding the desired result. ■

Thus from the objective of (3), we see that the optimal solution is the element of $SO(n)$ that maximizes a linear functional defined by the average of a suitable function of the data – hence the generalized average moniker. We now present two examples to make these ideas concrete.

Example 1 (Rotation Averaging): Consider the case where $B^i = I$ and $C^i = R_0^i$, that is to say we wish to find the average value (with respect to the Frobenius norm) of the initial states of each node. Then optimization (3) reduces to

$$\bar{R} = \arg \max_{R \in \text{conv}\{SO(n)\}} \left\langle \frac{1}{N} \sum_{i=1}^N R_0^i, R \right\rangle \quad (6)$$

Example 2 (Multi-view pose estimation/localization): Let $n = 3$, and suppose that $B^i = X^i \in \mathbb{R}^{3 \times m}$ are m appropriately re-centered points obtained via either LIDAR or stereo camera observations, and that $C^i = Y^i \in \mathbb{R}^{3 \times m}$ are the corresponding points of the internal model of the object being observed. If all of the Y^i are identical, this corresponds to using multiple cameras to estimate the pose of a single object, whereas different Y^i could arise in, for example, a simultaneous localization and mapping (SLAM) scheme. The optimization then becomes

$$\bar{R} = \arg \max_{R \in \text{conv}\{SO(n)\}} \left\langle \frac{1}{N} \left(\sum_{i=1}^N Y^i (X^i)^\top \right), R \right\rangle \quad (7)$$

Based on the previous discussion, we let $D^i := C^i (B^i)^\top$ be a general data matrix available at node i , and consider solving the following optimization problem in a *distributed manner* such that each local iterate R_t^i converges to the desired generalized average

$$\bar{R} := \arg \max_{R \in \text{conv}\{SO(n)\}} \sum_{i=1}^n \langle D^i, R \rangle. \quad (8)$$

Our approach will be to exploit recent results on the spectrahedral representation of the *convex hull* of $SO(n)$ [1], and to leverage first order distributed optimization methods such as dual decomposition and alternating direction method of multipliers (ADMM) [11] to implement a consensus protocol that is guaranteed to converge to \bar{R} .

In particular, we will consider two types of consensus settings: a *completely distributed* setting in which information can only be exchanged between neighbors (i.e. node i has access to R^j if and only if $(i, j) \in \mathcal{E}$), and a *semi-distributed* setting, in which information can not only be exchanged between neighbors, but can also be routed along so that each agent has access to every other agent's information.

III. PRELIMINARIES

A. The convex hull of $SO(n)$

In the previous section, we showed that our consensus problem amounts to optimizing a linear functional over the convex hull of $SO(n)$. However, unless this latter object has a tractable representation, we have not made much progress. Fortunately, it has recently been shown that $\text{conv}\{SO(n)\}$ admits a *spectrahedral* representation of dimension $2^{n-1} \times 2^{n-1}$. Although this scaling is poor in n , most applications have low n , e.g., rigid body transformations have $n = 2, 3$.

To that end, we recall the characterization (12) of $\text{conv}\{SO(n)\}$ from Corollary 1.6 of [1]. In particular, for $1 \leq i, j \leq n$, we have that

$$A_{ij} = -P_{\text{even}}^\top \lambda_i \rho_j P_{\text{even}}, \quad (9)$$

where λ_i and ρ_j are given by

$$\begin{aligned} \lambda_i &= \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}}^{i-1} \otimes \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \otimes \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}^{n-i}, \\ \rho_i &= \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{i-1} \otimes \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \otimes \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}}_{n-i} \end{aligned} \quad (10)$$

and

$$P_{\text{even}} = \frac{1}{2} \begin{bmatrix} I + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ I - \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \end{bmatrix}. \quad (11)$$

The convex hull can then be written as

$$\text{conv}\{SO(n)\} = \left\{ \begin{bmatrix} \langle A_{11}, Z \rangle & \langle A_{12}, Z \rangle & \cdots & \langle A_{1n}, Z \rangle \\ \langle A_{21}, Z \rangle & \langle A_{22}, Z \rangle & \cdots & \langle A_{2n}, Z \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle A_{n1}, Z \rangle & \langle A_{n2}, Z \rangle & \cdots & \langle A_{nn}, Z \rangle \end{bmatrix} \mid Z \succeq 0, \text{trace} Z = 1 \right\}. \quad (12)$$

For our purposes, however, it will suffice to note that $\text{conv}\{SO(n)\}$ can be written as

$$\text{conv}\{SO(n)\} = \{\mathcal{A}(Z) \mid Z \succeq 0, \text{trace}(Z) = 1\}, \quad (13)$$

for the affine map $\mathcal{A} : \mathbb{S}^{2^{n-1}} \rightarrow \mathbb{R}^{n \times n}$ defined in terms of the matrices A_{ij} as given in (12). Further let $\mathcal{A}^\dagger : \mathbb{R}^{n \times n} \rightarrow \mathbb{S}^{2^{n-1}}$ denote its adjoint operator with respect to $\langle \cdot, \cdot \rangle$. Patient computation reveals that

$$\mathcal{A}^\dagger(Y) = \sum_{i,j} A_{ij} Y_{ij}. \quad (14)$$

1) *An eigenvalue problem:* We now exploit this characterization of $\text{conv}\{SO(n)\}$ to provide an analytic solution to the following optimization problem, which will be an essential element of our consensus algorithms.

Lemma 2: The optimal solution $R^* \in SO(n)$ to the optimization

$$\max_{R \in \text{conv}\{SO(n)\}} \langle D, R \rangle \quad (15)$$

is given by

$$R^* = \mathcal{A}(\mu\mu^\top), \quad (16)$$

where μ is the orthonormal eigenvector corresponding to the largest eigenvalue ν of $\mathcal{A}^\dagger(D)$.

Proof: Using the characterization of $\text{conv}\{SO(n)\}$ (13), optimization (15) is equivalent to

$$\max_{Z \succeq 0, \text{trace} Z = 1} \langle \mathcal{A}^\dagger(D), Z \rangle. \quad (17)$$

It is then straightforward to verify that (i) Slater's condition is satisfied, and thus strong duality holds, and that (ii) $(\mu\mu^\top, \nu_{\max})$ form a Karush-Kuhn-Tucker pair for the optimization. Finally, as $Z^* := \mu\mu^\top$ is of rank 1, $\mathcal{A}(Z^*)$ is an extreme point of $\text{conv}\{SO(n)\}$, and thus an element of $SO(n)$. ■

Corollary 1: Suppose that the maximal eigenvalue ν of $\mathcal{A}^\dagger(D)$ has multiplicity 1. Then the unique optimal solution Z^* to the optimization

$$\max_{Z \succeq 0, \text{trace} Z = 1} \langle \mathcal{A}^\dagger(D), Z \rangle \quad (18)$$

is given by $Z = \mu\mu^\top$.

Proof: As we are optimizing a linear functional over the free spectrahedron, the optimal solution must be an extreme point satisfying $Z^* = zz^\top$, $\|z\| = 1$. Thus, to show uniqueness it suffices to note that $\langle zz^\top, \mathcal{A}^\dagger(D) \rangle = z^\top \mathcal{A}^\dagger(D) z < \nu$ for all $z \neq \mu$ satisfying $\|z\| = 1$. ■

Remark 1: The hypothesis that the maximal eigenvalue ν of $\mathcal{A}^\dagger(D)$ has multiplicity 1 will be satisfied with probability one for randomly distributed data matrices.

B. Dual Decomposition

Dual decomposition is a general technique to solve optimization problems with separable objectives, and has the advantage of yielding a distributed algorithm. Specifically, given the problem

$$\begin{aligned} \min_{x_i \in \mathbb{R}^{n_i}} \quad & f(x) = \sum_{i=1}^N f^i(x^i) \\ \text{s.t.} \quad & Ax = b \end{aligned} \tag{19}$$

where the optimization variable x is partitioned according to the objective as $x = [x_1^T, \dots, x_N^T]^T$, and the constraint matrix A can be also be partitioned according to this structure, with $A = [A_1, \dots, A_N]$. The Lagrangian for (19) is then

$$\begin{aligned} L(x, \lambda) &= \sum_{i=1}^N f^i(x^i) + \lambda^T A_i x^i - \frac{1}{N} \lambda^T b \\ &= \sum_{i=1}^N L^i(x^i, \lambda) \end{aligned}$$

with Lagrange multipliers introduced as λ . This gives rise to the dual function

$$g(\lambda) = \inf_x L(x, \lambda).$$

This dual function may be optimized via gradient ascent. Due to the separability of the Lagrangian, this yields the following procedure

$$\begin{aligned} x_{t+1}^i &= \operatorname{argmin}_{x^i} L^i(x^i, \lambda_t) \\ \lambda_{t+1} &= \lambda_t + \alpha_t (Ax_{t+1} - b) \end{aligned}$$

where it is seen that the optimization may be performed in parallel over the primal variables x_i at each iteration k . At each iteration, these updated primal variables are gathered for the update of the dual variable, which may then be distributed to the computation units. These dual variable updates may also be performed in parallel according to the separability of A . The result is a distributed algorithm for solving (19), which can be shown to converge under certain conditions (which our problem will be shown to satisfy).

C. Alternating Direction Method of Multipliers

The Alternating Direction Method of Multipliers (ADMM) [11], [15] provides a principled method for parallelization of convex problems. It is adopted here due to its particularly strong and general convergence guarantees, allowing us to inherit these guarantees in our consensus algorithm.

ADMM is a “meta”-optimization scheme, where each step is carried out by solving a convex optimization problem. Consider the optimization

$$\begin{aligned} & \text{minimize} && f(x) + g(z) \\ & \text{subject to} && Ax + Bz = c \end{aligned} \tag{20}$$

over the variables x and z and convex functions f and g . Define an augmented Lagrangian

$$L_\alpha = f(x) + g(z) + y^T (Ax + Bz - c) + \frac{\alpha}{2} \|Ax + Bz - c\|_2^2,$$

where $\alpha > 0$ is an algorithm parameter, and y is the dual variable associated with the equality constraint. The constrained optimization is solved through alternately minimizing the augmented Lagrangian over the primal variables x , z , and updating the dual variable y ,

$$\begin{aligned} x_{t+1} &:= \operatorname{argmin}_x L_\alpha(x, z_t, y_t) \\ z_{t+1} &:= \operatorname{argmin}_z L_\alpha(x_{t+1}, z, y_t) \\ y_{t+1} &:= y_t + \alpha (Ax_{t+1} + Bz_{t+1} - c). \end{aligned}$$

We will be concerned with the following two assumptions:

Assumption 1: The (extended real valued) functions $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are closed, proper, and convex.

Assumption 2: The unaugmented Lagrangian has a saddle point.

If it can be demonstrated that the optimization problem obeys these assumptions, then the following general theorem is available:

Theorem 1: (See [11]) Given Assumptions 1, 2 then the ADMM iterates satisfy the following:

- **Residual convergence:** $r_t \rightarrow 0$ as $t \rightarrow \infty$, i.e. the iterates approach feasibility
- **Objective convergence:** $f(x_t) + g(z_t) \rightarrow p^*$ as $t \rightarrow \infty$, i.e. the objective function of the iterates approaches the optimal value
- **Dual variable convergence:** $y_t \rightarrow y^*$ as $t \rightarrow \infty$, where y^* is a dual optimal point

IV. CONSENSUS ALGORITHMS

A. General Approach

As is standard in the use of distributed optimization algorithms for consensus, we will rewrite optimization (8) in a manner more amenable to our purposes by introducing local variables at each node, and enforcing consistency according to the connectivity of the graph. In particular, we now look to solve the following optimization in a distributed manner:

$$\begin{aligned} & \underset{R^i, i \in \mathcal{V}}{\text{maximize}} && \sum_{i=1}^N \langle D^i, R^i \rangle \\ & \text{s.t.} && R^i \in \text{conv} \{SO(n)\} \quad \forall i \in \mathcal{V} \\ & && R^i = R^j \quad \forall (i, j) \in \mathcal{E}, \end{aligned} \tag{21}$$

or in its parameterized form,

$$\begin{aligned} & \underset{Z^i, i \in \mathcal{V}}{\text{maximize}} && \sum_{i=1}^N \langle \mathcal{A}^\dagger(D^i), Z^i \rangle \\ & \text{s.t.} && Z^i \succeq 0, \quad \text{trace} Z^i = 1 \quad \forall i \in \mathcal{V} \\ & && Z^i = Z^j \quad \forall (i, j) \in \mathcal{E}. \end{aligned} \tag{22}$$

It is evident that so long as the graph is strongly connected, that the solution to this optimization is identical to that of (8), and thus we work with this distributed optimization for the remainder of the paper.

B. Completely Distributed Consensus via Dual Decomposition

We write the Lagrangian of optimization (22) as

$$L(Z^i, Y^{ij}) = \sum_{i=1}^N \langle \mathcal{A}^\dagger(D^i), Z^i \rangle - \sum_{(i,j) \in \mathcal{E}} \langle Y^{ij}, Z^i - Z^j \rangle \tag{23}$$

where each Z^i is constrained to be positive semi-definite and of unit trace, and the Y^{ij} are symmetric. The gradient of the dual objective function with respect to Y^{ij} is then easily verified to be $Z^j - Z^i$, allowing us to write the following dual-ascent algorithm for the distributed optimization³:

$$\begin{aligned} Z_{t+1}^i &= \underset{Z^i \succeq 0, \text{trace} Z^i = 1}{\text{argmax}} \quad \left\langle \mathcal{A}^\dagger(D^i) - \sum_{j: (i,j) \in \mathcal{E}} Y_t^{ij}, Z^i \right\rangle \\ Y_{t+1}^{ij} &= Y_t^{ij} - \alpha_t (Z_{t+1}^j - Z_{t+1}^i) \end{aligned} \tag{24}$$

³Note that our presentation is slightly non-standard as our primal problem is a maximization, and our dual a minimization.

Leveraging Lemma 2, we can solve for the Z^i update analytically through an eigenvalue decomposition of a $2^{n-1} \times 2^{n-1}$ symmetric matrix, making this method computationally appealing for small to moderate n .

1) *Convergence Guarantees:* Assuming that the hypothesis of Corollary 1 holds at each iteration (which will be true for generic data), then the iterates are guaranteed to converge to the global optimizer Z^* of (22) (c.f. Prop. 3.9 of [16]) for appropriately chosen step size α_t .

2) *Iterate properties:* As the objective of the Z_{t+1}^i optimization step in (24) is linear in Z^i , by Lemma 2, each iterate is an element of $SO(n)$. Thus, this approach is well suited for use in “true consensus” protocols in which each iterate is required to be a valid rotation element.

C. Completely Distributed Consensus via ADMM

Following the approach of [12], we introduce additional auxiliary variables X^j in optimization (22) and rewrite it as

$$\begin{aligned} \underset{Z^i, X^j, i, j \in \mathcal{V}}{\text{maximize}} \quad & \sum_{i=1}^N \langle \mathcal{A}^\dagger(D^i), Z^i \rangle \\ \text{s.t.} \quad & Z^i \succeq 0, \text{ trace} Z^i = 1 \quad \forall i \in \mathcal{V} \\ & X^j \succeq 0, \text{ trace} X^j = 1 \quad \forall j \in \mathcal{V} \\ & Z^i = X^j \quad \forall (i, j) \in \mathcal{E}. \end{aligned} \quad (25)$$

We then write the augmented Lagrangian of optimization (25) as

$$\begin{aligned} L(Z^i, X^j, Y^{ij}) = & \sum_{i=1}^N \langle \mathcal{A}^\dagger(D^i), Z^i \rangle - \dots \\ & \sum_{(i,j) \in \mathcal{E}} \left[\langle Y^{ij}, Z^i - X^j \rangle + \frac{\alpha}{2} \|Z^i - X^j\|_F^2 \right] \end{aligned} \quad (26)$$

where each Z^i and X^j are constrained to be positive semi-definite and of unit trace, and the Y^{ij} are symmetric. Exploiting the identity $\sum_{(i,j) \in \mathcal{E}} = \sum_{i=1}^N \sum_{j:(i,j) \in \mathcal{E}} = \sum_{j=1}^N \sum_{i:(j,i) \in \mathcal{E}}$, the ADMM algorithm then becomes

$$\begin{aligned} Z_{t+1}^i &= \underset{Z^i \succeq 0, \text{trace} Z^i = 1}{\text{argmax}} \quad \langle M_t^i, Z^i \rangle - \frac{|\mathcal{N}_i| \alpha}{2} \|Z^i\|_F^2 \\ X_{t+1}^j &= \underset{X^j \succeq 0, \text{trace} X^j = 1}{\text{argmax}} \quad \langle N_t^j, X^j \rangle - \frac{|\mathcal{N}_j| \alpha}{2} \|X^j\|_F^2 \\ Y_{t+1}^{ij} &= Y_t^{ij} - \alpha (Z_{t+1}^j - Z_{t+1}^i) \end{aligned} \quad (27)$$

where

$$\begin{aligned} M_t^i &= \mathcal{A}^\dagger(D^i) - \sum_{j:(i,j) \in \mathcal{E}} [Y_t^{ij} - \alpha X_t^j] \\ N_t^j &= \sum_{i:(j,i) \in \mathcal{E}} [Y_t^{ij} + \alpha Z_{t+1}^i]. \end{aligned} \quad (28)$$

D. Semi-Distributed Consensus via ADMM

We first modify optimization (22) by introducing a fusion variable Z^0 :

$$\begin{aligned} \underset{Z^0, Z^i, i \in \mathcal{V}}{\text{maximize}} \quad & \sum_{i=1}^N \langle \mathcal{A}^\dagger(D^i), Z^i \rangle \\ \text{s.t.} \quad & Z^i \succeq 0, \text{ trace } Z^i = 1 \quad \forall i \in \mathcal{V} \cup \{0\} \\ & Z^i = Z^0 \quad \forall i \in \mathcal{V}. \end{aligned} \quad (29)$$

We write the augmented Lagrangian of optimization (22) as

$$\begin{aligned} L_\alpha(Z^i, Z, Y^i) = & \sum_{i=1}^N \langle \mathcal{A}^\dagger(D^i), Z^i \rangle - \dots \\ & \sum_{i=1}^N \left(\langle Y^i, Z^i - Z^0 \rangle + \frac{\alpha}{2} \|Z^i - Z^0\|_F^2 \right), \end{aligned} \quad (30)$$

where once again each Z^i is constrained to be positive semi-definite and of unit trace, and the Y^i are symmetric. The ADMM algorithm then becomes

$$\begin{aligned} Z_{t+1}^i &= \underset{Z^i \succeq 0, \text{ trace } Z^i = 1}{\text{argmax}} \quad \langle \mathcal{A}^\dagger(D^i) - Y_t^i + \alpha Z_t^0, Z^i \rangle \dots \\ &\quad - \frac{\alpha}{2} \|Z^i\|_F^2 \\ Z_{t+1}^0 &= \underset{Z^0 \succeq 0, \text{ trace } Z^0 = 1}{\text{argmax}} \quad \left\langle \sum_{i=1}^N (Y_t^i + \alpha Z_{t+1}^i), Z^0 \right\rangle \dots \\ &\quad - \frac{\alpha}{2} \|Z^0\|_F^2 \\ Y_{t+1}^i &= Y_t^i - \alpha (Z_{t+1}^0 - Z_{t+1}^i). \end{aligned} \quad (31)$$

E. Convergence Guarantees and Efficient Updates

Assumption 1 is satisfied as the objectives are linear, and Assumption 2 is satisfied as strong duality is easily verified to hold: thus this algorithm is guaranteed to converge for appropriately chosen regularization parameter α – further as the global optimizer is unique and the objective functions continuous, convergence to an optimal cost implies convergence of the iterates to the optimal solution.

We note that as written, the Z^i and X^j updates do not admit an obvious efficient solution. The next subsection will show how this problem can be reduced to a projection onto the probability

simplex, which admits efficient algorithms that are of linear complexity in the dimension of the vector being projected [17].

1) Z^i and X^j updates: We note that the Z^i and X^j updates in equations (27) and (31) are all of the form

$$\operatorname{argmax}_{W \succeq 0, \operatorname{trace} W = 1} \langle \kappa \alpha T, W \rangle - \frac{\kappa \alpha}{2} \|W\|_F^2 \quad (32)$$

for appropriate constant matrix T and scalar κ . We begin by observing that the optimal solution W^* to the above program is the projection of T onto the free spectrahedron.

Lemma 3: The optimal solution W^* to optimization (32) is given by the projection of T onto the free spectrahedron.

Proof: Multiplying the objective by $-\frac{2}{\kappa \alpha}$ and accordingly replacing the maximization with a minimization, we obtain

$$W^* = \operatorname{argmin}_{W \succeq 0, \operatorname{trace} W = 1} -2 \langle T, W \rangle + \|W\|_F^2. \quad (33)$$

It suffices to note that adding $\|T\|_F^2$ to the objective does not change the optimal update's value, thus reducing the optimization to

$$W^* = \operatorname{argmin}_{W \succeq 0, \operatorname{trace} W = 1} \|T - W\|_F^2, \quad (34)$$

which is none other than the projection of T onto the free spectrahedron. ■

From this Lemma, the following two corollaries are immediate.

Corollary 2: The Z_{t+1}^i update as given in equation (31) is given by the projection of Q_t^i onto the free spectrahedron, where

$$Q_t^i := \begin{cases} \sum_{i=1}^N \frac{1}{\alpha} Y_t^i + Z_{t+1}^i & \text{for } i = 0 \\ \frac{1}{\alpha} (\mathcal{A}^\dagger(D^i) - Y_t^i) + Z_t^0 & \text{for } i = 1, \dots, N \end{cases} \quad (35)$$

Corollary 3: The Z_{t+1}^i and X_{t+1}^j update as given in equation (27) are given by the projection of $\frac{1}{|\mathcal{N}_i| \alpha} M_t^i$ and $\frac{1}{|\mathcal{N}_j| \alpha} N_t^j$, respectively, onto the free spectrahedron.

Finally, since in all cases, T is symmetric, the aforementioned projection can be further reduced to a projection of the eigenvalues of T onto the probability simplex.

Lemma 4: Let $T = U_t^i \Lambda_t^i (U_t^i)^\top$ be a diagonalization of T . Then the projection of T onto the free spectrahedron is given by $U_t^i [\Lambda_t^i]^+ (U_t^i)^\top$, where $[\cdot]^+$ denotes a projection of the diagonal elements of a matrix onto the probability simplex.

Proof: Follows immediately from diagonalizing T and the invariance of the Frobenius norm under orthogonal transformations. ■

2) *Iterate properties:* Although we are able to guarantee convergence to an element of $SO(n)$, the iterates may lie in the interior of $\text{conv}\{SO(n)\}$ – the iterate optimizations for Z_{t+1}^i and X_{t+1}^j have a quadratic objective, and thus extreme point solutions are not guaranteed. However, these ADMM schemes are still well suited for distributed computation on the group of rotations, in which only the consensus point is actually used.

In practice however, we have noticed that through an appropriate choice of α , all iterates are in, or very near to, $SO(n)$. In particular, by choosing a sufficiently low value of α (but still large enough to maintain good convergence), the objectives are dominated by their linear term, thus leading to (near) boundary solutions – current work is aimed at formalizing this intuition. Indeed this heuristic was used in the following examples to generate valid consensus paths within $SO(n)$.

V. EXPERIMENTAL RESULTS

We tested our methods on a number of estimation and consensus problems. The dual decomposition and ADMM methods of §IV-B and §IV-D were employed to synchronize the attitude for a formation of spacecraft. This is then followed by a pose estimation problem from computer vision, where the semi-distributed algorithm of §IV-D allows for a consistent estimate of an object’s orientation to be achieved from a collection of disparate measurements. We finally end with an application of the dual decomposition method to a rotation averaging problem on $SO(6)$ to illustrate that our methods generalize seamlessly to higher dimensions.

A. Satellite Attitude Synchronization via Dual Decomposition

Similar to [13] we propose to synchronize the attitude of a $N = 8$ satellite formation. Each satellite is initialized with a random orientation, with the task being to find the average rotation, i.e. Example 1 of §II. The communication topology of the satellites is limited to a ring, and the problem solved via dual decomposition. The results are demonstrated in Figure 1, where convergent behavior is indeed observed. Computational time for each iteration for all satellites is typically under $1ms$ for unoptimized, interpreted (Matlab) code.

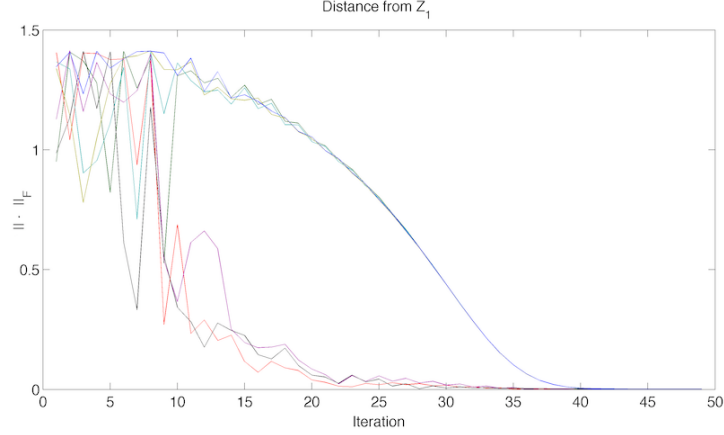


Fig. 1: Error between average rotation estimates for a satellite formation of $N = 8$ satellites with a ring communication topology. The gradient descent weight is set as $\alpha = 5$.

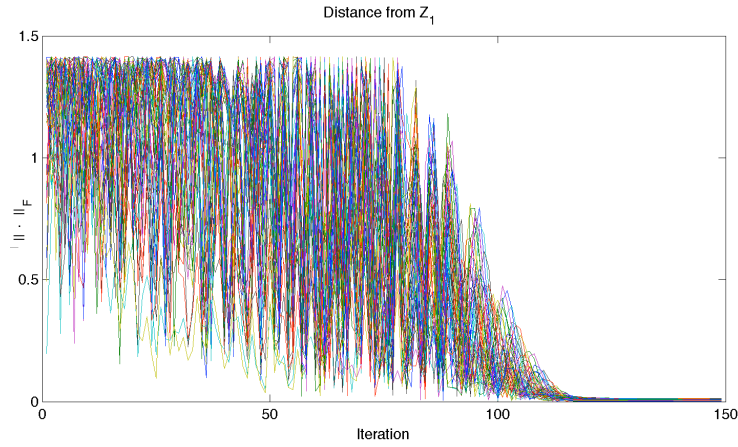


Fig. 2: Error between average rotation estimates for a satellite formation of $N = 50$ with nearest neighbor and second nearest neighbor communication. The gradient weight is set as $\alpha = 5$.

A larger example is then tested, where $n = 50$ and the communication topology is augmented to allow communication between both neighbors and second-neighbors. The resulting error trace is shown in Figure 2, where convergence is again observed. In practice, convergence is obtained for all examples for a wide range values $\alpha \in (.01 - 20)$.

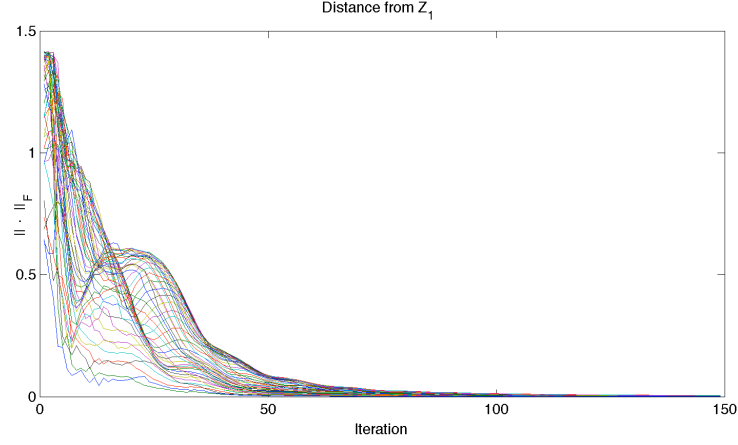


Fig. 3: Error between average rotation estimates for a satellite formation of $N = 50$ satellites with nearest- and second- neighbor communication using the distributed ADMM algorithm. The gradient weight is set as $\alpha = 0.2$.

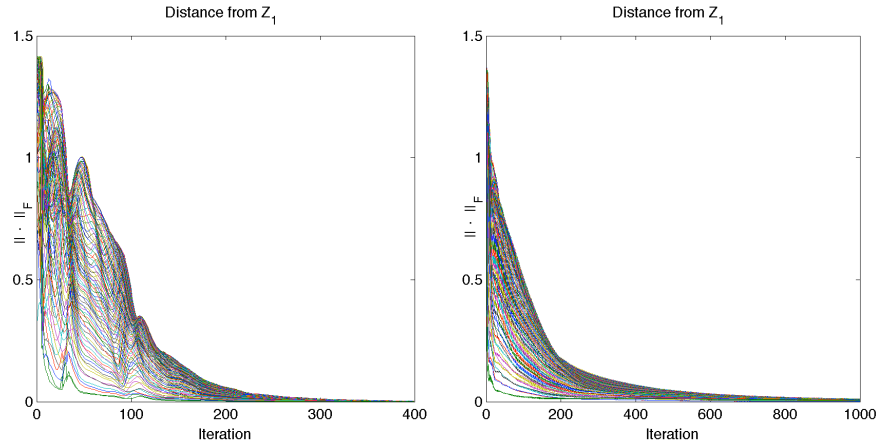


Fig. 4: Error between average rotation estimates for a satellite formation of $N = 100$ satellites arranged in a ring topology with nearest neighbor (left) and second nearest neighbor (right) communication using the decentralized ADMM algorithm. The gradient descent weight is set as $\alpha = 0.2$.

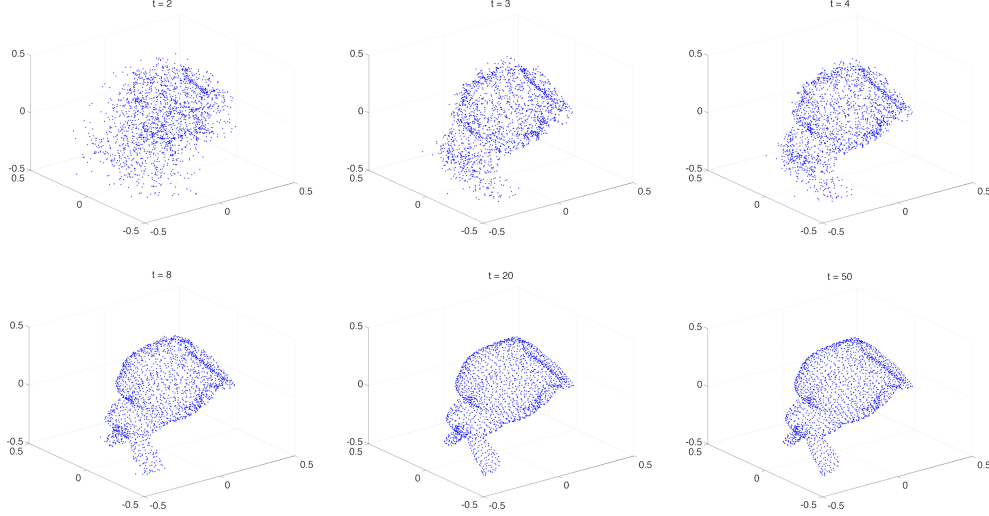


Fig. 5: Visualization of the estimated Stanford Bunny model for $N = 8$ different sensors. The inconsistency of the local estimates cause the local models (groups of vertices) to appear un-aligned. As consensus is achieved, the model becomes consistent.

B. Attitude Synchronization via Distributed ADMM

The problem of synchronizing the attitude of $N = 50$ satellites was then repeated using the distributed ADMM approach of §IV-C. Again, a ring topology was enforced, with Figure 3 showing the convergence of the algorithm when only allowing nearest neighbor and also second nearest neighbor communication. It is seen that the distributed ADMM solution is quicker and appears more regular than the dual ascent method in practice, although it is slightly more computationally demanding requiring both an eigen-decomposition and a projection onto the probability simplex.

Finally, to demonstrate the scalability properties of the algorithm, a $N = 100$ satellite example with nearest- and second nearest-neighbor connectivity was simulated, with results shown in Figures 4.

C. Parallel Pose Estimation

The semi-distributed ADMM scheme is applied to the problem of pose estimation from computer vision. This problem is to identify a transformation that, when applied to a known

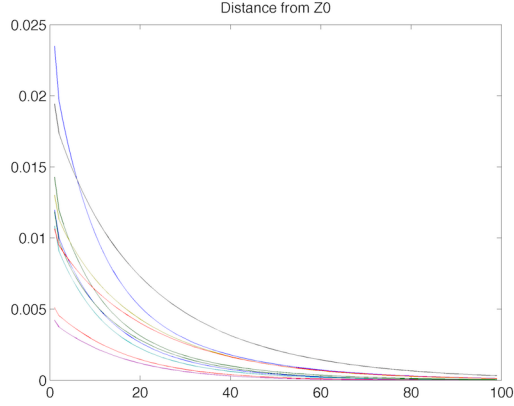


Fig. 6: Convergence of Algorithm (31) for $n = 8$ disparate sensors for the Stanford Bunny model.

object model, yields the object as perceived through the system's sensors. The convex relaxation has been applied previously in this context [7]. However, the approach presented in §IV-D provides computationally appealing benefits, and arises in a number of ways. The first is to obtain a consistent estimate from a number of disparate sensors, for instance from a collection of cameras viewing a common object. This may be further extended for use in sensor fusion [18], wherein the disparate sensors may provide different estimation modalities, and weights of the data matrices B^i , C^i in (2) may correspond to priors on data strength. Alternatively, this consensus problem may arise when the pose estimation problem is split in order to take advantage of parallel processing capabilities, as follows.

Pose estimation, i.e. Example 2 of §II, is solved for a 3D model of the Stanford Bunny, which has $s = 1889$ vertices, and which is normalized to have unit dimensions. A random orientation is applied to the model, and the data is corrupted by Gaussian noise with standard deviation $\sigma = 0.05$. The data is split between $N = 10$ processing nodes and solved using Algorithm (31) where the dual variable weighting is set to $\alpha = 2$.

A visualization of the consensus progress is shown in Figure 5, where it is seen that the differing estimates of the rotation cause the estimated model points to appear inconsistent. As a common estimate is obtained, it is seen that the points indeed do consistently match the model. Figure 6 demonstrates the convergence of the error measure.

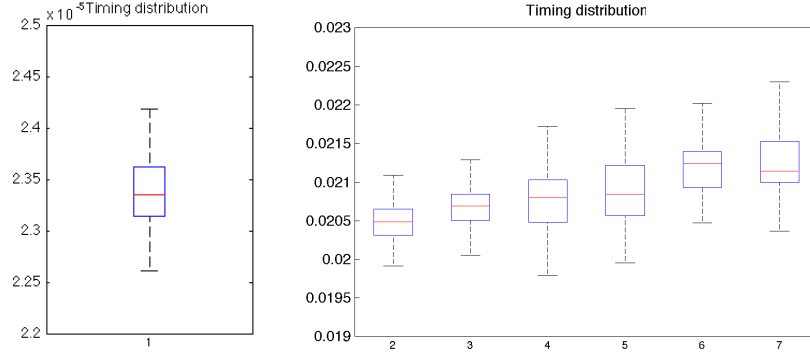


Fig. 7: Timing results when varying the number of active cameras for the pose estimation problem. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

The problem was then examined by varying the number of virtual cameras available to collect observation data. The model was down sampled to $s = 728$ vertices, and the standard deviation of the corrupting noise increased to $\sigma = 0.10$. The observed data was then divided among $N = 7$ sensors, and the estimation process repeated for $t = 100$ trials for varying numbers of these sensors being active. The error and timing results are shown in Figures 7 & 8. We again emphasize that the timing results are for unoptimized, interpreted (Matlab) code running in *serial* – significant speed up is expected with compiled code and parallelized implementations.

Remark 2: Note that for the single camera case, the problem is no longer distributed. The computation of Lemma 2 is instead used, explaining the significant difference in computational time.

D. $SO(6)$ Consensus

Although applications of the Special Orthogonal group are most common for $SO(2)$ and $SO(3)$, efficient characterizations of $SO(n)$ for $n > 3$ are necessary for computation on high dimensional datasets [19]. In this application, rigid body transformations are used to change the perspective of a projection onto a lower dimensional structure. With this in mind, we apply Algorithm (24) to the rotation averaging problem where $n = 6$, and number of nodes $N = 12$,

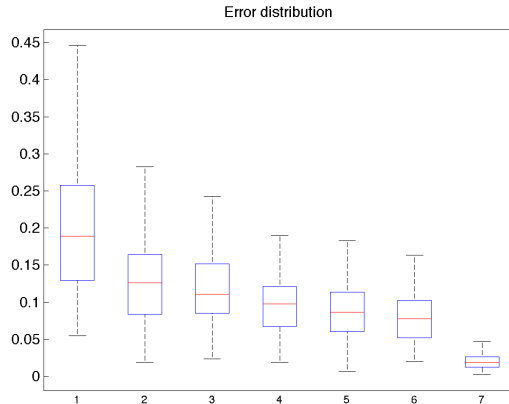


Fig. 8: Estimation error results when varying the number of active cameras for the pose estimation problem. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

with the results shown in Figure 9.

VI. CONCLUSION

In this paper we have presented a convex approach to consensus on $SO(n)$ by leveraging recent results on the spectrahedral representation of $\text{conv}\{SO(n)\}$ and techniques from distributed optimization. In particular, we show how meaningful consensus problems can be formulated as optimizations of linear functionals over the convex hull of the group of rotation matrices, and that these optimizations can be solved in a distributed manner using dual decomposition and ADMM techniques – these distributed computation methods are then used to implement a consensus protocol among the agents. The benefits of our approach include rapid and guaranteed convergence to globally optimal consensus points, scalability and mathematical transparency. Future work will focus on understanding how the choice of α in the ADMM algorithm affects boundary solution guarantees.

REFERENCES

- [1] J. Saunderson, P. A. Parrilo, and A. S. Willsky, “Semidefinite descriptions of the convex hull of rotation matrices,” *arXiv preprint arXiv:1403.4914*, 2014.

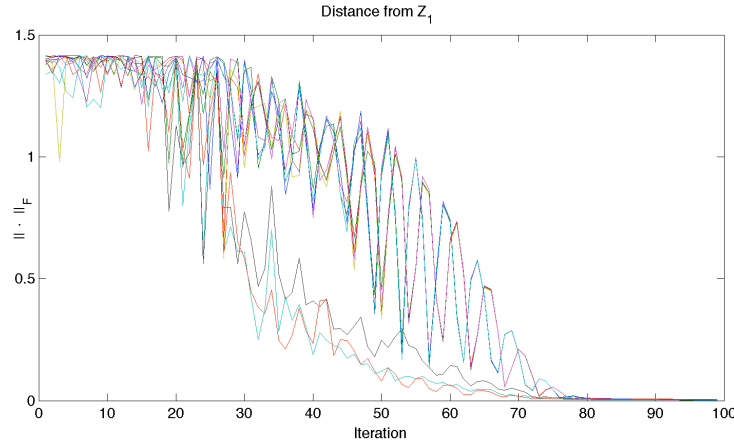


Fig. 9: Convergence of Algorithm (24) in calculating the rotational average for $n = 6$ and number of nodes $N = 12$. The gradient weight was set as $\alpha = 2$.

- [2] M. L. Psiaki, “Generalized Wahba problems for spinning spacecraft attitude and rate determination,” *The Journal of the Astronautical Sciences*, vol. 57, no. 1-2, pp. 73–92, 2009.
- [3] F. Bullo, J. Cortés, and S. Martinez, *Distributed control of robotic networks: a mathematical approach to motion coordination algorithms*. Princeton University Press, 2009.
- [4] S. H. Strogatz, “From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators,” *Physica D: Nonlinear Phenomena*, vol. 143, no. 1, pp. 1–20, 2000.
- [5] R. Tron, R. Vidal, and A. Terzis, “Distributed pose averaging in camera networks via consensus on SE(3),” in *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*. IEEE, 2008, pp. 1–10.
- [6] R. Tron and R. Vidal, “Distributed image-based 3-D localization of camera sensor networks,” in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*. IEEE, 2009, pp. 901–908.
- [7] M. B. Horowitz, N. Matni, and J. W. Burdick, “Convex relaxations of SE(2) and SE(3) for visual pose estimation,” in *International Conference on Robotics and Automation (ICRA)*, 2014, arXiv:1401.3700.
- [8] M. Longinetti, L. Sgheri, and F. Sottile, “Convex hulls of orbits and orientations of a moving protein domain,” *Discrete & Computational Geometry*, vol. 43, no. 1, pp. 54–77, 2010.
- [9] R. Olfati-Saber, J. A. Fax, and R. M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.
- [10] D. P. Palomar and M. Chiang, “A tutorial on decomposition methods for network utility maximization,” *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 8, pp. 1439–1451, 2006.
- [11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [12] T. Erseghe, D. Zennaro, E. Dall’Anese, and L. Vangelista, “Fast consensus by the alternating direction multipliers method,” *Signal Processing, IEEE Transactions on*, vol. 59, no. 11, pp. 5523–5537, 2011.

- [13] R. Tron, B. Afsari, and R. Vidal, “Intrinsic consensus on $SO(3)$ with almost-global convergence.” in *CDC*, 2012, pp. 2052–2058.
- [14] A. Sarlette and R. Sepulchre, “Consensus optimization on manifolds,” *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 56–76, 2009.
- [15] J. Eckstein, “Splitting methods for monotone operators with applications to parallel optimization,” Ph.D. dissertation, Massachusetts Institute of Technology, 1989.
- [16] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice-Hall, Inc., 1989.
- [17] N. Maculan *et al.*, “A linear-time median-finding algorithm for projecting a vector on the simplex of \mathbb{R}^n ,” *Operations research letters*, vol. 8, no. 4, pp. 219–222, 1989.
- [18] L. Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based on average consensus,” in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*. IEEE, 2005, pp. 63–70.
- [19] S. Thakur, “Framework for visualizing and exploring high-dimensional geometry,” Ph.D. dissertation, Indiana University, 2008.