

Average Consensus on Riemannian Manifolds with Bounded Curvature

Roberto Tron, Bijan Afsari and René Vidal

Abstract—Consensus algorithms are a popular choice for computing averages and other similar quantities in ad-hoc wireless networks. However, existing algorithms mostly address the case where the measurements live in a Euclidean space. In this paper, we propose distributed algorithms for averaging measurements lying in a Riemannian manifold. We first propose a direct extension of the classical average consensus algorithm and derive sufficient conditions for its convergence to a consensus configuration. Such conditions depend on the network connectivity, the geometric configuration of the measurements and the curvature of the manifold. However, the consensus configuration to which the algorithm converges may not coincide with the Fréchet mean of the measurements. We thus propose a second algorithm that performs consensus in the tangent space. This algorithm is guaranteed to converge to the Fréchet mean of the measurements, but needs to be initialized at a consensus configuration. By combining these two methods, we obtain a distributed algorithm that converges to the Fréchet mean of the measurements. We test the proposed algorithms on synthetic data sampled from manifolds such as the space of rotations, the sphere and the Grassmann manifold.

I. INTRODUCTION

Consider a set of low-power sensors that can communicate with each other through a wireless network. Assume that each sensor measures some common quantity, e.g., the temperature in a field or the pose of an object, and that the goal is to compute the average of all these measurements. However, due to power constraints, each node in the network has a limited communication range. We are, therefore, interested in *distributed estimation algorithms* in which each node performs some local computation via communication with a few neighboring nodes and all the nodes collaborate to reach an agreement on the global quantity of interest (e.g., the average of the measurements). Natural candidates for this scenario are iterative *consensus algorithms*, where each node maintains a local estimate of the global average, which is updated with a weighted average of the estimates from the local neighbors. The main advantage of consensus algorithms is that they converge exponentially to the centralized solution under very mild communication assumptions, even in the case of a time-varying network topology. However, traditional consensus algorithms have been mainly studied for averaging quantities that live in Euclidean spaces.

Prior work. In the last few years, there has been an increasing interest in extending consensus algorithms to data lying on manifolds. This problem arises in a number of applications, including distributed pose estimation [1], camera sensor network localization [2] and satellite attitude synchronization

[3]. Early works consider specific manifolds such as the sphere [4] or the N -torus [5]. However, these approaches are not easily generalizable to other manifolds. The work of [6] considers the problems of consensus and balancing on the more general class of compact homogeneous manifolds. The approach is *extrinsic*, i.e., it is based on specific embeddings of the manifolds in Euclidean space (where classical consensus can be employed) and requires the ability to project the Euclidean consensus configuration onto the manifold. Since the approach is extrinsic, convergence properties for both fixed and time-varying network topologies follow directly from existing results in the Euclidean case. A similar approach is taken in [7], where the extrinsic approach is extended to the case where the mean is time-varying.

To the best of our knowledge, the work of [1] is the first one to propose an *intrinsic* approach, which does not depend on specific embeddings of the manifold and does not require the definition of a projection operation. Instead, it relies only on the intrinsic properties of the manifold, such as geodesic distances and exponential and logarithm maps. However, the work of [1] focuses only on a specific manifold ($SO(3)$) and does not provide a thorough convergence analysis. Other works on distributed algorithms for data lying in manifolds include [3], [8], which address the problem of coordination on Lie groups, and [2], which addresses the problem of localizing a camera sensor network. However, these works cannot be applied to the problem of computing distributed averages in general manifolds, which we consider in this paper.

Paper contributions. In this paper, we propose distributed algorithms for averaging measurements lying in a Riemannian manifold. We assume a fixed network topology and aim to compute the Fréchet mean of the measurements (the point in the manifold that minimizes the sum of squared geodesic distances to the measurements) in a distributed way. For this purpose, we define two energy functions on the manifold and minimize them using Riemannian gradient descent. This requires computing the exponential and logarithm maps for the manifold of interest. We also derive sufficient conditions for the convergence of the proposed algorithms to a consensus configuration (i.e., where all the nodes converge to the same estimate). Moreover, we show that some existing results for Euclidean consensus (such as the choice of the step-size) are valid also for spaces of constant, non-negative curvature.

Our work has several important contributions with respect to the state of the art. First, our formulation is completely intrinsic, in the sense that it is not tied to a specific embedding of the manifold. Second, we consider more general (complete and not necessarily compact) Riemannian manifolds. Third,

This work was supported by the grant NSF CNS-0834470

The authors are with the Center for Imaging Science, Johns Hopkins University, Baltimore MD, 21202, USA

we provide sufficient conditions for the convergence of the algorithm to a consensus configuration, which depend on the network connectivity, the geometric configuration of the measurements and the sectional curvature of the manifold.

Paper outline. In §II we review classical consensus and some relevant notions from Riemannian geometry. In §III and §IV we describe two extensions of consensus algorithms to data in manifolds, each one with its advantages and disadvantages, and we show how they can be combined to overcome the limitations of both. In §VI we test the proposed algorithms on manifolds such as the special orthogonal group, the circle and the Grassmann manifold.

II. MATHEMATICAL BACKGROUND

In this section, we review some basic concepts related to Euclidean consensus algorithms and differential geometry that are relevant to our development in the rest of the paper.

A. Review of Euclidean consensus algorithms

Consider a sensor network with N nodes. In the basic setup of average consensus the network is represented using an undirected graph $G = (V, E)$. The vertices $i \in V = \{1, \dots, N\}$ represent the nodes of the network while the edges $(i, j) \in E \subseteq V \times V$ represent the communication links between nodes i and j . The set of neighbors of node i is denoted as $N_i = \{j \in V \mid (i, j) \in E\}$. The number of neighbors or degree of node i is denoted as $d_i = |N_i|$ and the maximum degree of the graph G as $\text{Deg}(G) = \max_i \{d_i\}$.

Each node measures a scalar quantity $u_i \in \mathbb{R}$, $i \in V$. The goal is to obtain a distributed algorithm to compute the average of these measurements $\bar{u} = \frac{1}{N} \sum_{i=1}^N u_i$, which is a global quantity (in the sense that involves information from all the nodes). The well-known average consensus algorithm computes the average \bar{u} by iterating the difference equation

$$x_i(k+1) = x_i(k) + \varepsilon \sum_{j \in N_i} (x_j(k) - x_i(k)), \quad x_i(0) = u_i, \quad (1)$$

where $x_i(k)$ is the state of node i at iteration k and $\varepsilon \leq \frac{1}{\text{Deg}(G)}$ is the step-size. It is easy to verify that the mean of the states is preserved at each iteration, i.e.,

$$\frac{1}{N} \sum_{i=1}^N x_i(k) = \frac{1}{N} \sum_{i=1}^N x_i(k+1) = \bar{u}. \quad (2)$$

It is also easy to see that (1) is in fact a gradient descent algorithm that minimizes the function

$$\varphi(\mathbf{x}) = \frac{1}{2} \sum_{(i,j) \in E} (x_i - x_j)^2. \quad (3)$$

where $\mathbf{x} = (x_1, \dots, x_N)$ denotes the vectors of all states in the network. From now on, we will use bold letters to denote vectors (in which each element belongs to \mathbb{R} or another manifold \mathcal{M}). The cost (3) is convex and the global minimum is achieved when the nodes reach a *consensus configuration*, i.e., when $x_i = y$ for all $i \in V$ and for some $y \in \mathbb{R}$. It can be shown that with the initial conditions stated in (1) and when the graph G is connected, we have $\lim_{k \rightarrow \infty} x_i(k) = \bar{u}$,

for all $i \in V$ (see, e.g., [9]). Notice also that the average consensus algorithm can be easily extended to multivariate data $\mathbf{u}_i \in \mathbb{R}^D$ by applying the scalar algorithm to each coordinate of \mathbf{u}_i . It can also be extended to situations where the network topology changes over time (see [10] for the convergence analysis).

B. Riemannian geometry notation

In this section we present our notation for the Riemannian geometry concepts used throughout the paper. We refer the reader to [11], [12] and [13] for further details.

Let $(\mathcal{M}, \langle \cdot, \cdot \rangle)$ be a Riemannian manifold with metric $\langle \cdot, \cdot \rangle$. The *tangent space* of \mathcal{M} at a point $x \in \mathcal{M}$ is denoted as $T_x \mathcal{M}$. Using the metric it is possible to define *geodesic* curves, which are the generalization of straight lines in \mathcal{M} . For the remainder of the paper, we will always assume that \mathcal{M} is *geodesically complete*, i.e., there always exists a length-minimizing geodesic between any two points in \mathcal{M} . Most of the manifolds of practical interest are geodesically complete. We define the distance $d(x, y)$ between two points $x, y \in \mathcal{M}$ as the length of the shortest geodesic curve between them.

Let v be a unit-length tangent vector in $T_x \mathcal{M}$, i.e., $\|v\| = \langle v, v \rangle^{\frac{1}{2}} = 1$. We can then define the *exponential map* $\exp_x : T_x \mathcal{M} \rightarrow \mathcal{M}$, which maps each tangent vector $tv \in T_x \mathcal{M}$ to the point in $\gamma(t) \in \mathcal{M}$ obtained by following the geodesic $\gamma(t)$ passing through x with direction v for a distance t . We denote as $\tilde{\mathcal{I}}_x \subset T_x \mathcal{M}$ the maximal open set for which \exp_x is a diffeomorphism and we define the *interior set* [11, p.216] as $\mathcal{I}_x = \exp_x \tilde{\mathcal{I}}_x$. The exponential map is therefore invertible for all the points in \mathcal{I}_x and we can define the *logarithm map* $\log_x : \mathcal{I}_x \rightarrow T_x \mathcal{M}$ as $\log_x = \exp_x^{-1}$. We also denote as $\text{inj}_x \mathcal{M}$, $x \in \mathcal{M}$ the radius of the maximal geodesic ball centered at x entirely contained in \mathcal{I}_x and as $\text{inj} \mathcal{M}$ the infimum of $\text{inj}_x \mathcal{M}$ over all the points $x \in \mathcal{M}$.

Given a smooth function $f : \mathcal{M} \rightarrow \mathbb{R}$, and a tangent vector $v \in T_x \mathcal{M}$, one can define the *directional derivative* of f in the direction v at x as $\left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0}$, where $\gamma(t)$ is any curve such that $\gamma(0) = x$ and $\dot{\gamma}(0) = v$. The *gradient* of f on $(\mathcal{M}, \langle \cdot, \cdot \rangle)$ at $x \in \mathcal{M}$ is defined as the unique tangent vector $\text{grad}_x f(x) \in T_x \mathcal{M}$ such that, for all $v \in T_x \mathcal{M}$,

$$\langle \text{grad}_x f(x), v \rangle_x = \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0}. \quad (4)$$

Intuitively, as in the Euclidean case, the gradient indicates the direction along which f increases the most. A point $x \in \mathcal{M}$ is called a *critical point* of f if either $\text{grad}_x f(x) = 0$, i.e., it is a *stationary point*, or the gradient does not exist [14]. In this paper, we will mainly need the gradient of the squared distance function, which is given by:

$$\frac{1}{2} \text{grad}_x d^2(x, y) = -\log_x(y). \quad (5)$$

Given a point $x \in \mathcal{M}$, we denote the sectional curvature of σ , a two-dimensional subspace in $T_x \mathcal{M}$, as $K_\sigma(x)$. From now on we will assume that the sectional curvature of the manifold \mathcal{M} is bounded above by Δ and below by δ . In other words, $\delta \leq K_\sigma(x) \leq \Delta$ for any point $x \in \mathcal{M}$ and

any two-dimensional subspace $\sigma \subset T_x \mathcal{M}$. If $\delta = \Delta = \kappa$, then \mathcal{M} is said to be of constant curvature κ . Related to the curvature, we define the radius $r^* > 0$ as

$$r^* = \frac{1}{2} \min\{\text{inj } \mathcal{M}, \frac{\pi}{\sqrt{\Delta}}\}, \quad (6)$$

where we use the convention that, if $\Delta \leq 0$, $\frac{1}{\sqrt{\Delta}} = +\infty$. Note that any ball with radius $r \leq r^*$ is guaranteed to be convex. In addition, for the sake of brevity, we define the functions

$$S_\kappa(t) = \begin{cases} \frac{\sin(\sqrt{\kappa}t)}{\sqrt{\kappa}} & \kappa > 0 \\ t & \kappa = 0 \\ \frac{\sinh(\sqrt{|\kappa|}t)}{\sqrt{|\kappa|}} & \kappa < 0 \end{cases}, \quad C_\kappa(t) = \begin{cases} \cos(\sqrt{\kappa}t) & \kappa > 0 \\ 1 & \kappa = 0 \\ \cosh(\sqrt{|\kappa|}t) & \kappa < 0 \end{cases}. \quad (7)$$

In the following, we will make also use of the product manifold $\mathcal{M}^N = \mathcal{M} \times \dots \times \mathcal{M}$, which is the N -fold cartesian product of \mathcal{M} with itself. We will use the notation $\mathbf{x} = (x_1, \dots, x_N)$ to indicate a point in \mathcal{M}^N and $\mathbf{v} = (v_1, \dots, v_N) \in T_x \mathcal{M}$ to indicate a tangent vector. We will use the natural metric $\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^N \langle v_i, w_i \rangle$. As a consequence, geodesics, exponential maps, and gradients can be easily obtained by using the respective definitions on each copy of \mathcal{M} in \mathcal{M}^N . This notation will be useful when stating results that involve the states of all the nodes at the same time.

C. Examples of manifolds

We will use the following manifolds as examples throughout the paper.

Euclidean space. The usual Euclidean space \mathbb{R}^n can be interpreted as the simplest Riemannian manifold, where the tangent space of a point is a copy of \mathbb{R}^n , the metric is the usual inner product, and geodesics are straight lines. It has constant curvature $\kappa = 0$ and injectivity radius $+\infty$.

The orthogonal and special orthogonal groups. The n -dimensional orthogonal group is defined as $O(n) = \{R \in \mathbb{R}^{n \times n} : R^T R = I\}$. This is the group of orthogonal $n \times n$ matrices. This group has two connected components. One of them is the *special orthogonal group* $SO(n)$, which has the additional property $\det(R) = 1$, and essentially describes all possible rotations in the n -dimensional Euclidean space. The Lie algebra for the group is $\mathfrak{so}(n)$, the space of $n \times n$ skew-symmetric matrices. The Riemannian metric at the identity is given by $\langle v_1, v_2 \rangle = \frac{1}{2} \text{tr}(v_1^T v_2)$, $v_1, v_2 \in \mathfrak{so}(n)$. In this metric, the curvature bounds are $\Delta = \frac{1}{2}$, and $\delta = 0$, except when $n = 3$, for which $\delta = \Delta$. Also, the injectivity radius is π and $r^* = \frac{\pi}{2}$.

The Grassmann manifold. The (n, p) Grassmann manifold $\text{Grass}(n, p)$ is the space of p -dimensional subspaces in \mathbb{R}^n . It can also be viewed as a quotient space $O(n)/(O(p) \times O(n-p))$, which provides a Riemannian structure for it through immersion in $O(n)$ [16]. The curvature bounds are $\Delta = 2$, and $\delta = 0$. The injectivity radius is $\frac{\pi}{2}$ and $r^* = \frac{\pi}{4}$.

The sphere. The n -dimensional sphere is defined as $\mathbb{S}^n = \{Y \in \mathbb{R}^{n+1} : Y^T Y = 1\}$. The tangent space at a point Y is

defined as $T_Y \mathbb{S}^n = \{Z \in \mathbb{R}^{n+1} : Z^T Y = 0\}$. As metric, we use the standard inner product between vectors in \mathbb{R}^{n+1} . The geodesics follow great circles and the curvature is constant $\delta = \Delta = 1$.

More details about these manifolds and about the computation of the exp and log maps can be found in [15], [16].

D. Fréchet mean

Let $\{u_i\}_{i=1}^N$ be a set of points in a smooth Riemannian manifold \mathcal{M} . The Fréchet mean \bar{u} of this set is defined as the global minimizer of the sum of squared geodesic distances, i.e.,

$$\bar{u} = \underset{u \in \mathcal{M}}{\text{argmin}} \sum_{i=1}^N d^2(u_i, u). \quad (8)$$

The minimizer is unique if the points lie in a ball of radius smaller than r^* [18]. It can be shown that a necessary condition for \bar{u} to be the Fréchet mean is that

$$\sum_{i=1}^N \log_{\bar{u}}(u_i) = 0. \quad (9)$$

The condition (9) leads to the following iterative algorithm for computing the Fréchet mean

- 1) Set the initial mean in \mathcal{M} as $\bar{u} = u_1$.
- 2) Compute the mean in $T_{\bar{u}} \mathcal{M}$ as $w = \frac{1}{N} \sum_{i=1}^N \log_{\bar{u}}(u_i)$.
- 3) While $\|w\| < \delta$, update \bar{u} to $\bar{u} = \exp_{\bar{u}}(\varepsilon w)$, for $\varepsilon \leq 1$ and go to step 2.

The most general conditions for convergence of this algorithm are only partially known, see [17] for instance.

III. CONSENSUS IN THE MANIFOLD

Our first algorithm, which we refer to as *consensus in the manifold*, can be considered as a direct extension of the classical consensus algorithm to the non-Euclidean case.

Following the notation introduced in Section II-A, let us denote the measurement and the state at node i as $u_i \in \mathcal{M}$ and $x_i \in \mathcal{M}$, respectively. By a straightforward generalization of the Euclidean case in (3), we define the potential function φ

$$\varphi(\mathbf{x}) = \frac{1}{2} \sum_{(i,j) \in E} d^2(x_i, x_j). \quad (10)$$

Notice that the gradient of φ with respect to the i -th element can be explicitly calculated as

$$\text{grad}_{x_i} \varphi = \frac{1}{2} \text{grad}_{x_i} \sum_{j \in N_i} d^2(x_i, x_j) = - \sum_{j \in N_i} \log_{x_i}(x_j). \quad (11)$$

While simplifying the above equation, we used the facts that the graph is undirected and that $d(\cdot, \cdot)$ is symmetric.

Our proposed consensus protocol on \mathcal{M} corresponds to using a Riemannian gradient descent search for minimizing the cost function φ . Essentially, $x_i(k)$ is initialized with the measurements u_i and then it is updated along the geodesic corresponding to the gradient $-\text{grad}_{x_i} \varphi$ with a step size ε , yielding the protocol

$$x_i(k+1) = \exp_{x_i(k)}(-\varepsilon \text{grad}_{x_i(k)} \varphi), \quad x_i(0) = u_i. \quad (12)$$

Note that this protocol is an extension of the Euclidean case, because when $\mathcal{M} = \mathbb{R}$ with the standard metric, the iterations in (12) reduce to those in (1).

In the remainder of this section, we analyze the convergence properties of the consensus in the manifold algorithm. Specifically, we want to give conditions for convergence to a consensus configuration $x_i = x_j, \forall i, j \in V$. We divide our treatment in three parts. First, we notice that the function (10) can have multiple local minima and we characterize a subset of \mathcal{M}^N that contains only *global* minima. Second, we give conditions on the choice of the step-size ε for which the algorithm is guaranteed to reduce the cost (10) at each iteration. Finally, we give sufficient conditions under which the algorithm is guaranteed to converge to the set of global minima, which correspond to consensus configurations.

Global minima of the cost function φ . We first show that a global minimizer of φ corresponds to a consensus configuration. Define the diagonal space of \mathcal{M}^N as

$$\mathcal{D} = \{(y, \dots, y) \in \mathcal{M}^N : y \in \mathcal{M}\}. \quad (13)$$

This space represents all the possible consensus configurations of the network, where all the nodes agree on a state. We have the following propositions.

Proposition 1: If G is connected, then $y \in \mathcal{D}$ if and only if y is a global minimizer of φ .

Proof: Note that each term of φ in (10) is non-negative, hence $\varphi \geq 0$. In addition, note that $\varphi = 0$ implies that for each pair $\{i, j\} \in E$, we have $d^2(x_i, x_j) = 0$. By definition, $d(x_i, x_j) = 0$ if and only if the points in the manifold are equal, i.e., $x_i = x_j$. Since G is connected, we can conclude that φ achieves its global minimum, i.e., $\varphi = 0$, if and only if $x_i = x_j = y$ for all i and j and some $y \in \mathcal{M}$. ■

We define the set $\mathcal{S} \subset \mathcal{M}^N$ as

$$\mathcal{S} = \{(x_1, \dots, x_N) \in \mathcal{M}^N : \exists y \in \mathcal{M} \text{ for which } \max_{i \in V} d(x_i, y) < r^*\}. \quad (14)$$

Intuitively, \mathcal{S} is a tube in \mathcal{M}^N centered around the diagonal space \mathcal{D} and having a “square” section (see Fig. 1). It is known that a sufficient condition for the uniqueness of the Fréchet mean is that $\mathbf{u} = (u_1, \dots, u_n) \in \mathcal{S}$ [18].

The following result represents our first contribution.

Theorem 1: A point $x_0 \in \mathcal{S}$ is a critical point for φ if and only if $x_0 \in \mathcal{D}$. In other words, the set \mathcal{S} contains all the global minima and no other critical points of φ .

To prove this theorem we need the following Lemma [15].

Lemma 1: Let x_1, x_2, y be three points in \mathcal{M} such that $d(x_i, y) < r^*, i = 1, 2$. Define the unique minimal geodesics $\gamma_i(t)$ such that $\gamma_i(0) = y$ and $\gamma_i(1) = x_i, i = 1, 2$. Define also $\phi_{12}(t) = d(\gamma_1(t), \gamma_2(t))$. Then $\frac{d}{dt}\phi_{12}^2 \geq 0$ for $t \in (0, 1]$, with equality if and only if $x_1 = x_2$.

Proof: [of Thm. 1] If $x_0 \in \mathcal{D}$, then x_0 is obviously a global minimizer of φ , and hence a critical point. On the other hand, $x_0 \notin \mathcal{D}$ cannot be a critical point of φ because, as we will show now, there exists a geodesic $\gamma : [0, 1] \rightarrow \mathcal{M}^N$ such that $\gamma(1) = x_0$ and along which the derivative of

$\varphi(\gamma(t))$ at $t = 1$ is not zero. Notice that since $x_0 \in \mathcal{S}$, there exists a $y \in \mathcal{M}$ such that $\max_{i \in V} d(x_{0i}, y) < r^*$. Define the unique minimal geodesics $\gamma_i(t)$ such that $\gamma_i(0) = y$ and $\gamma_i(1) = x_{0i}$. Then the curve $\gamma(t) = (\gamma_1(t), \dots, \gamma_N(t))$ is a minimal geodesic in \mathcal{M}^N (see also Fig. 1). It follows that

$$\left. \frac{d}{dt}\varphi(\gamma(t)) \right|_{t=1} = \frac{1}{2} \sum_{(i,j) \in E} \left. \frac{d}{dt}d^2(\gamma_i(t), \gamma_j(t)) \right|_{t=1}. \quad (15)$$

Since $d(x_{0i}, y) < r^*$, from Lemma 1 we deduce that all the terms on the right hand side of (15) are not negative and at least one is strictly positive (otherwise, since G is connected, we would have $x_0 \in \mathcal{D}$). From the definition of gradient, $\text{grad}_x \varphi(x_0) = 0$ if and only if the directional derivative $\left. \frac{d}{dt}\varphi(\gamma(t)) \right|_{t=1} = 0$ for any curve $\gamma(t)$ passing through x_0 , i.e., $\gamma(1) = x_0$. Since $\left. \frac{d}{dt}\varphi(\gamma(t)) \right|_{t=1} > 0$, x_0 is not a critical point. ■

We remark that the bound given by r^* in the proof above might be quite conservative for some specific geodesics γ . In practice, there might be a set containing \mathcal{D} and no other critical points which is larger than \mathcal{S} . In fact, if the graph G is a tree, we can show the following stronger result.

Theorem 2: If G is a tree, any stationary point x_0 of φ on \mathcal{M}^N is a global minimizer, i.e., $x_0 \in \mathcal{D}$.

Proof: [Sketch] From (11), we can deduce that $\text{grad}_x \varphi = 0$ implies $x_i = x_j$ whenever node i is a leaf of the tree and node j is its parent. Using induction from the leafs to the root of the tree, the claim follows. ■

Choice of the stepsize ε . In this section we provide bounds on the value of ε that guarantee the convergence of the consensus protocol (12) to a critical point of φ . As it will turn out, these bounds on ε depend on bounds on the Hessian of the cost function $\text{Hess } \varphi(x)$ [13, p. 142]. We will use the following definition.

Definition 1: Given a twice differentiable function defined on an open subset \mathcal{X} of a manifold \mathcal{N} , we say that the Hessian $\text{Hess } \varphi(x)$ is uniformly upper bounded on \mathcal{X} if there exists a finite, strictly positive constant μ_{max} such that, for any $x_0 \in \mathcal{X}$ and any $v \in T_{x_0}\mathcal{N}$, the second derivative of φ along

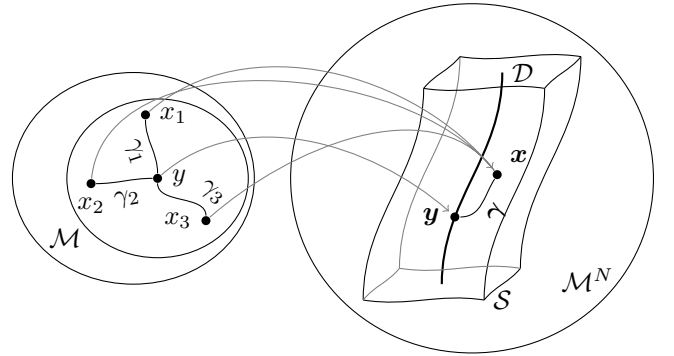


Fig. 1: Construction of the geodesic for testing if φ has a local minimum at (x_i, x_j) .

$\tilde{\mathbf{x}}(t) = \exp_{\mathbf{x}_0} t\mathbf{v}$ can be bounded as:

$$\left. \frac{d^2}{dt^2} \varphi(\tilde{\mathbf{x}}(t)) \right|_{t=0} = \langle \mathbf{v}, \text{Hess } \varphi(\mathbf{x}_0) \mathbf{v} \rangle \leq \mu_{\max} \|\mathbf{v}\|^2. \quad (16)$$

These bounds on the Hessian play an important role because of the following theorem [15].

Theorem 3: Let μ_{\max} be a uniform bound on the Hessian of φ . If $\tilde{\mathbf{x}}_0(\varepsilon) = \exp_{\mathbf{x}_0} \varepsilon \text{grad}_{\mathbf{x}} \varphi(\mathbf{x}_0) \in \mathcal{X}$ for all $\varepsilon \in (0, 2\mu_{\max}^{-1})$, then $\varphi(\tilde{\mathbf{x}}_0(\varepsilon)) \leq \varphi(\tilde{\mathbf{x}}_0(0))$ for $\varepsilon \in (0, 2\mu_{\max}^{-1})$, with equality if and only if $\mathbf{x}_0 \in \mathcal{X}$ is a stationary point of φ .

In our case, $\mathcal{N} = \mathcal{M}^N$ and the set $\mathcal{X} = \mathcal{E}$ is defined as

$$\mathcal{E} = \{\mathbf{x} \in \mathcal{M}^N : d(x_i, x_j) < 2r^* \forall (i, j) \in E\}. \quad (17)$$

and we also use the specific notation $\mathcal{E}_{\mathcal{M}^2}$ for the case $N = 2$. Given the special structure of our cost function φ in (10), the bound μ_{\max} can be obtained from the bound on the Hessian of each pairwise distance in (10), which is given by the following:

Theorem 4: The Hessian of the function $(x_i, x_j) \mapsto \frac{1}{2}d^2(x_i, x_j)$ can be bounded on $\mathcal{E}_{\mathcal{M}^2}$ by

$$\mu_{\max}^d(t) = \max \left\{ 2, t \left(\frac{C_\delta(t)}{S_\delta(t)} + \frac{1}{S_\Delta(t)} \right) \right\}, \quad (18)$$

where $t = d(x_i, x_j)$, $C_\kappa(t)$ and $S_\kappa(t)$ are defined in (7), and δ and Δ are, respectively, lower and upper bounds on the sectional curvature of \mathcal{M} .

The proof can be found in [15]. We remark that the bound μ_{\max}^d is sharp, in the sense that it becomes exact for manifolds with constant curvature (i.e., $\delta = \Delta$, see [15]). In fact, for the Euclidean space and for spaces of non-negative constant curvature, e.g., the sphere or $SO(3)$, this bound is 2, independent of the distance between the points. In general, while the bound depends on the distance between the points, we might be able to find a uniform upper bound on $\mu_{\max}^d(t)$. For instance, if we take $d_{\max} = 2r^*$, then $\mu_{\max}^d \simeq 3.792$ for both $SO(n)$ with $n \geq 4$ and the Grassmann manifold.

Given μ_{\max}^d , the following proposition gives a relation with the bound μ_{\max} on the Hessian of the entire cost φ .

Proposition 2: Assume that there exists $d_{\max} < 2r^*$ such that $d(x_i, x_j) < d_{\max}$ for all $(i, j) \in E$. The Hessian of the function φ can be uniformly bounded on \mathcal{E} by the constant $\mu_{\max}(d_{\max}) = \text{Deg}(G) \mu_{\max}^d(d_{\max})$, where, again, $\text{Deg}(G)$ is the maximum graph degree.

This proposition is just a particular case of the more general Thm. 13 in [15].

We are now ready to state our second main contribution, whose proof follows easily from Thm. 3 and Prop. 2.

Theorem 5: Assume that the curvature of \mathcal{M} is bounded below by δ and above by Δ . Assume also that, for a given k , $d_{\max} < 2r^*$, $d(x_i(k), x_j(k)) < d_{\max}$ for all $(i, j) \in E$. In addition, assume $\exp_{\mathbf{x}(k)}(-\varepsilon \text{grad}_{\mathbf{x}} \varphi(\mathbf{x}(k))) \in \mathcal{E}$ for all $\varepsilon \in (0, 2\mu_{\max}^{-1}(d_{\max}))$. If $\mathbf{x}(k+1)$ is given by the consensus protocol in (12) with $\varepsilon \in (0, 2\mu_{\max}^{-1}(d_{\max}))$, then $\varphi(\mathbf{x}(k+1)) \leq \varphi(\mathbf{x}(k))$, with equality if and only if $\mathbf{x}(k)$ is a stationary point of φ .

Notice that $\mathcal{S} \subseteq \mathcal{E}$. However, in general, \mathcal{E} might be much larger than \mathcal{S} , especially when each node has a small number of neighbors.

From Thm. 5, we can deduce a simple corollary.

Corollary 1: For spaces of constant, non-negative curvature $\delta = \Delta \geq 0$, we have $\mu_{\max}^d \equiv 2$ and $\varepsilon \in (0, \text{Deg}(G)^{-1})$.

This tells us that the bound for the Euclidean case can be applied also for the case of manifolds with positive constant curvature. In other cases (e.g., for manifolds of negative curvature) we need to reduce ε according to the maximum distance between the states of two neighboring nodes.

To compute the same ε at each node, one needs distributed algorithms to compute $\text{Deg}(G)$ and d_{\max} . The maximum degree $\text{Deg}(G)$ can be computed using a consensus-like algorithm where each node initializes its state with its own degree and repeatedly updates its estimate by taking the maximum of the estimates in the local neighborhood. If the manifold \mathcal{M} is compact, the bounds on the maximum distance d_{\max} between the states of two neighboring nodes can be precomputed. If \mathcal{M} is not compact, d_{\max} can be bounded by first using consensus to compute the value of the cost function evaluated at the measurements, $\varphi(\mathbf{u})$, and then using ideas similar to the ones we will see in the proof of Thm. 7.

Using the results above, we can establish our first theorem on the convergence of our consensus protocol.

Theorem 6: If the assumptions of Thm. 5 hold for any iteration k , then any cluster point of the sequence of iterates $\mathbf{x}(k) \in \mathcal{E}$ generated by (12) is a critical point of φ in \mathcal{E} . The proof, which can be found in [15], uses a fairly standard argument to show that $\lim_{k \rightarrow \infty} \text{grad}_{\mathbf{x}} \varphi(\mathbf{x}(k)) = 0$.

Convergence to the set of global minimizers. In this section we show that there exists a set $\mathcal{S}_{\text{conv}} \subset \mathcal{S} \subset \mathcal{M}^N$ such that the algorithm will converge to the set of global minimizers of φ for any initialization in $\mathcal{S}_{\text{conv}}$.

Theorem 7: Let $D = \text{diam}(G)$ denote the diameter of the network graph G and define $\mathcal{S}_{\text{conv}} = \{\mathbf{x} \in \mathcal{M}^N : \varphi(\mathbf{x}) < \frac{(r^*)^2}{2D}\}$. Then, $\mathcal{S}_{\text{conv}} \subseteq \mathcal{S}$ and if the consensus protocol (12) is initialized with measurements $\mathbf{u} \in \mathcal{S}_{\text{conv}}$, then $\mathbf{x}(k)$ converges to \mathcal{D} .

Proof: Consider any $p, q \in V$ and consider a shortest path in the graph $\{i_k\}_{k=0}^K$ from $i_0 = p$ to $i_K = q$. We will use this path to bound the real distance between states x_p and x_q with the cost function φ as follows:

$$\begin{aligned} d^2(x_p, x_q) &\leq \left(\sum_{k=0}^{K-1} d(x_{i_k}, x_{i_{k+1}}) \right)^2 \leq K \sum_{k=0}^{K-1} d^2(x_{i_k}, x_{i_{k+1}}) \\ &\leq K \sum_{(i,j) \in E} d^2(x_i, x_j) \leq 2D \varphi(\mathbf{x}), \quad \forall p, q \in V. \end{aligned} \quad (19)$$

In the equation above we used the triangular and Jensen's inequalities, and the fact that, by definition of diameter, $K \leq D$. This shows that if $\mathbf{x} \in \mathcal{S}_{\text{conv}}$, then $\varphi(\mathbf{x}) < \frac{r^*}{2D}$ and $d(x_p, x_q) < r^*$, for any $p, q \in V$. This means that $\mathbf{x} \in \mathcal{S}$.

Next, we show that if $\mathbf{x}(k) \in \mathcal{S}_{\text{conv}}$, then $\tilde{\mathbf{x}}(\varepsilon) = \exp_{\mathbf{x}(k)}(-\varepsilon \text{grad}_{\mathbf{x}} \varphi(\mathbf{x}(k))) \in \mathcal{S}_{\text{conv}}$ for all $\varepsilon \in$

$(0, 2\mu_{max}^{-1}(d_{max}))$. The idea is to show that \tilde{x} does not cross the boundary of \mathcal{S}_{conv} if $\varepsilon \in (0, 2\mu_{max}^{-1}(d_{max}))$. By way of contradiction, assume that there exist values of $\varepsilon \in (0, 2\mu_{max}^{-1}(d_{max}))$ such that $\varphi(\tilde{x}(\varepsilon)) = \frac{(r^*)^2}{2D}$ and denote as ε_0 the minimum of such values. From the continuity of $\varphi(\tilde{x}(\varepsilon))$, it follows that there must exist $\eta, \nu > 0$ arbitrarily small such that $\varphi(\tilde{x}(\varepsilon_0 - \nu)) \geq \frac{(r^*)^2}{2D} - \eta \geq \varphi(\tilde{x}(0))$. However, this is in contradiction with the result of Thm. 5, which implies that $\varphi(\tilde{x}(\varepsilon)) < \varphi(\tilde{x}(0)) \forall \varepsilon \in (0, \varepsilon_0)$, since $\tilde{x}(\varepsilon) \in \mathcal{S}_{conv} \forall \varepsilon \in (0, \varepsilon_0)$ and ε_0 is minimal.

Finally, we show that the algorithm converges to a global minimum. Since $x(0) \in \mathcal{S}_{conv}$ and $\varphi(x(k))$ decreases at each iteration, then the sequence of $x(k)$ generated by the algorithm will be guaranteed to be in $\mathcal{S}_{conv} \subset \mathcal{S}$. From this and Thm. 6, any cluster point of the sequence $x(k)$ will be a critical point in \mathcal{S} , which must be a global minimizer of φ . ■

Note that we have shown convergence to a set and not to a single point. In practice, the experiments in §VI show convergence to a single global minimizer under much more relaxed conditions. Nevertheless, as many other gradient-based algorithms, a rigorous proof of these properties would require a more delicate argument. In fact, we conjecture that, given the initial measurements u , there exists a compact set \mathcal{S}^* that the iterates $x(k)$ never leave. If this conjecture was valid, then it would be possible to prove convergence to a single point instead of to a set. Moreover, if $\mathcal{S}^* \subset \mathcal{S}$, then the convergence to a single global minimizer would be automatically ensured. In our future research we plan to investigate the possibility of obtaining these improved results.

Special cases with global convergence. In general, the conditions of Thm. 7 can be restrictive, because they depend on the diameter of the network, which might be large. Nevertheless, we can provide stronger results if we make additional assumptions on the manifold and network topologies. For instance, when $r^* = \infty$ (e.g., in \mathbb{R}^d or other manifolds with non-positive curvature), Thm. 7 implies global convergence on any graph G . On the other hand, if G has linear topology (i.e., it is a tree with a single branch), the following is true independently from the manifold topology.

Theorem 8: If G has linear topology, and the consensus protocol (12) is initialized with measurements u such that $d(u_i, u_j) < \text{inj } \mathcal{M}$, then $x(k)$ converges to \mathcal{D} .

The proof [15] relies on the fact that, for this particular choice of G , if $d(x_i(k), x_j(k)) < \text{inj } \mathcal{M}$ for all $(i, j) \in E$, then the same is true for $k + 1$, the gradient in (12) always exists and Thm. 2 holds. For some manifolds (e.g., the sphere), this result provides almost global convergence to \mathcal{D} (which is the best one can hope for in compact manifolds).

IV. CONSENSUS IN THE TANGENT SPACE

As we mentioned in §II-A, when we minimize φ in classical consensus, the states converge to a global minimizer which corresponds to the average of the initial measurements.

In the Riemannian case one would expect a similar behavior, where all the states converge to the Fréchet mean

of the measurements. However, in general this is not the case, as we will see in the experiments in §VI. Intuitively, this is due to the fact that the Fréchet mean of the states is not preserved after each iteration [1] and, even if the algorithm converges to a global minimizer (e.g., under the conditions of Thm. 7), this need not correspond to the desired Fréchet mean.

In this section, we propose an alternative method called *consensus in the tangent space* for computing the exact Fréchet mean in a distributed fashion. The algorithm follows the same iterations as the centralized Fréchet mean algorithm described in Section II-D, except that we perform consensus on tangent vectors (which are Euclidean quantities) in order to compute the updates.

More specifically, the consensus in the tangent space algorithm proceeds as follows. First, all the nodes are initialized with a common element of \mathcal{M} , say $x_i(0) = y$. The main requirement is that all the measurements u_i should not lie in the cut locus of y , i.e., $u_i \in \mathcal{I}_y$ for all $i \in V$. This step corresponds to Step 1 of the centralized algorithm in Section II-D. Here, however, each node maintains x_i as a local copy of the centralized estimate of the mean \bar{u} .

Then, at each iteration k , all the nodes compute w_i , the negative gradient of the distance between x_i and u_i

$$w_i(k) = -\text{grad}_{x_i} \frac{1}{2} d^2(x_i(k), u_i) = \log_{x_i(k)} u_i. \quad (20)$$

The nodes can employ Euclidean average consensus to obtain $w(k) = \frac{1}{N} \sum_{i=1}^N w_i(k)$ with arbitrary precision. This corresponds to Step 2 in the centralized Fréchet mean algorithm, which is now computed in a distributed fashion.

Each node finally updates its local estimate x_i using the tangent vector w as $x_i(k+1) = \exp_{x_i(k)}(w(k))$ and then repeats from the computation of the w_i (this corresponds to the final step of the centralized Fréchet mean algorithm).

At a high level, the idea is to linearize the problem around the current estimates $x_i(k)$ for each iteration, and then use the standard consensus for linear spaces.

Remarks about convergence. Notice that all the states x_i are initialized with the same value and updated with the same sequence of vectors $w(k)$. Each state x_i will then follow the same sequence of updates as in the centralized Fréchet mean algorithm. Therefore, the same convergence properties hold.

It is important to notice that there are problems also with the consensus in the tangent space algorithm. First, it is relatively slow when compared to consensus in the manifold. In fact, it is necessary to run a full Euclidean consensus at each iteration of the Fréchet mean algorithm. This requires several communication rounds for computing \bar{w} with sufficient precision. Second, the nodes must be initialized with the same y , which may need to be chosen depending on the particular values of u_i .

V. COMBINED ALGORITHM

We have presented two Riemannian consensus algorithms with complementary properties. The consensus in the manifold algorithm is automatically initialized with the

measurements u_i at each node, as in the classical Euclidean consensus algorithm. Then, the states at each node are updated and the estimates of the different nodes converge to a common value. Unfortunately, due to the curvature of the manifold, this common value need not coincide with the true global Fréchet mean.

On the other hand, the consensus in the tangent space algorithm converges to the true Fréchet mean. However, it requires a common initialization y , which might need to be chosen depending on the data u_i .

To overcome this issue, one could initialize the method using $y = u_i$ for some $i \in V$. Alternatively, one can merge the two consensus algorithms as follows. First, we run the consensus in the manifold until it converges, thereby obtaining a common estimate for all the nodes which in practice, as we will see in the Experiment section, is close to the true Fréchet mean. This result is then used to initialize the consensus in the tangent space, which then converges to the Fréchet mean typically in a small number of iterations.

VI. EXPERIMENTS

In this section we evaluate the proposed algorithms on synthetic data drawn from the special orthogonal group, the sphere and the Grassmann manifold.

The experiments are performed using a synthetic network of $N = 15$ nodes with a 4-regular connectivity graph. To generate the measurements, we choose an arbitrary element $x_0 \in \mathcal{M}$ and compute N random tangent vectors v_{0i} in $T_{x_0}\mathcal{M}$ drawn from an isotropic Gaussian distribution with standard deviation $\sigma = 0.2$. The measurement at each node i is then defined as $u_i = \exp_{x_0}(v_{0i})$, $i = 1, \dots, N$.

We then run our consensus in the manifold algorithm for 150 iterations followed by consensus in the tangent space for 10 iterations (each one of which includes 200 iterations of standard Euclidean consensus). We use stepsizes compatible with the bounds found in §III. For each iteration of each algorithm we compute the distance between each state and the Fréchet mean \bar{u} of the initial measurements $\{u_i\}$. For the first algorithm we record also the distance between the Fréchet mean of the states at each iteration and \bar{u} .

As particular examples of manifolds we consider $SO(7)$, \mathbb{S}^6 and $\text{Grass}(7, 3)$. The collected results for each one of these manifolds are plotted in Fig. 2. As an implementation detail, notice that for $\text{Grass}(7, 3)$ the matrix representation in $\mathbb{R}^{7 \times 3}$ is not unique. In order to compute the average of the tangent vectors in the consensus in the tangent space algorithm, all the nodes must choose the same representative of the equivalence class. This can be done, for instance, by computing the QR factorization of the transposed matrix and then using the transpose of the R part as the representative.

A number of points can be made on the experiments. First, the consensus in the manifold clearly converges to a consensus configuration (because the measurements generated are not too far one from each other). Second, the same algorithm modifies the Fréchet mean of the states, especially in the first iterations. When this algorithm terminates, the estimated Fréchet mean is at a distance in the order of 10^{-4} from the

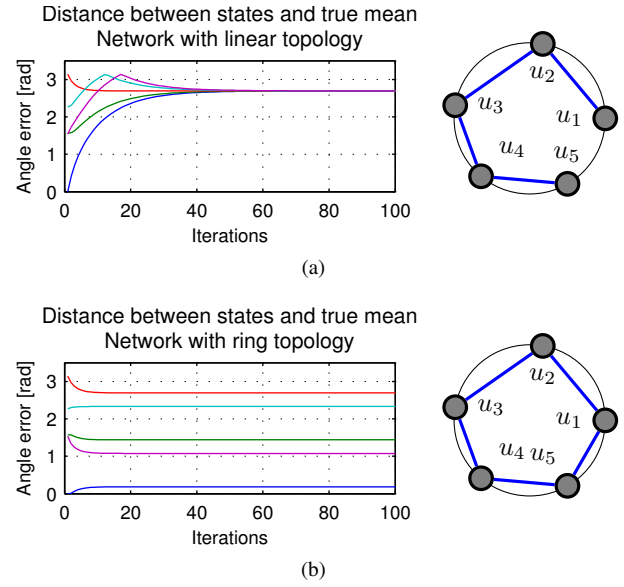


Fig. 3: An example where consensus in the manifold converges (a) or fails to converge (b) to a consensus configuration depending on the topology. These plots correspond to the initial configurations portrayed on the right.

true Fréchet mean. This error might be negligible in practical applications, but it is many order of magnitude greater than the achievable machine precision. The consensus in the tangent space reduces the errors in both cases down to an order of 10^{-15} .

We include also two experiments (Fig. 3) for which the measurements are taken around the circle and are far apart, i.e., $u \notin \mathcal{S}$ (see Thm. 1). With a linear network, the algorithm converges to a consensus configuration, as expected from Thm. 8. On the other hand, with a ring network, the algorithm gets trapped in a local minima and fails. These experiment suggests that the convergence of the algorithm depends on both the manifold and the network topologies. However, a complete investigation of this fact is out of the scope of this paper.

VII. CONCLUSION

In this paper, we proposed a framework that extends consensus algorithms for computing the Fréchet mean of data lying in Riemannian manifolds. We examined two algorithms: consensus in the manifold and consensus in the tangent space. The first one can be initialized with the measurements and converges to an approximate Fréchet mean, while the second one is slower and requires a careful initialization, but converges to the global Fréchet mean. Our main contribution is finding sufficient conditions that guarantee convergence of the algorithms to a consensus configuration. Experiments on various manifolds illustrated the applicability of our methods.

REFERENCES

- [1] R. Tron, R. Vidal, and A. Terzis, “Distributed pose averaging in camera networks via consensus on $SE(3)$,” in *International Conference on Distributed Smart Cameras*, 2008.

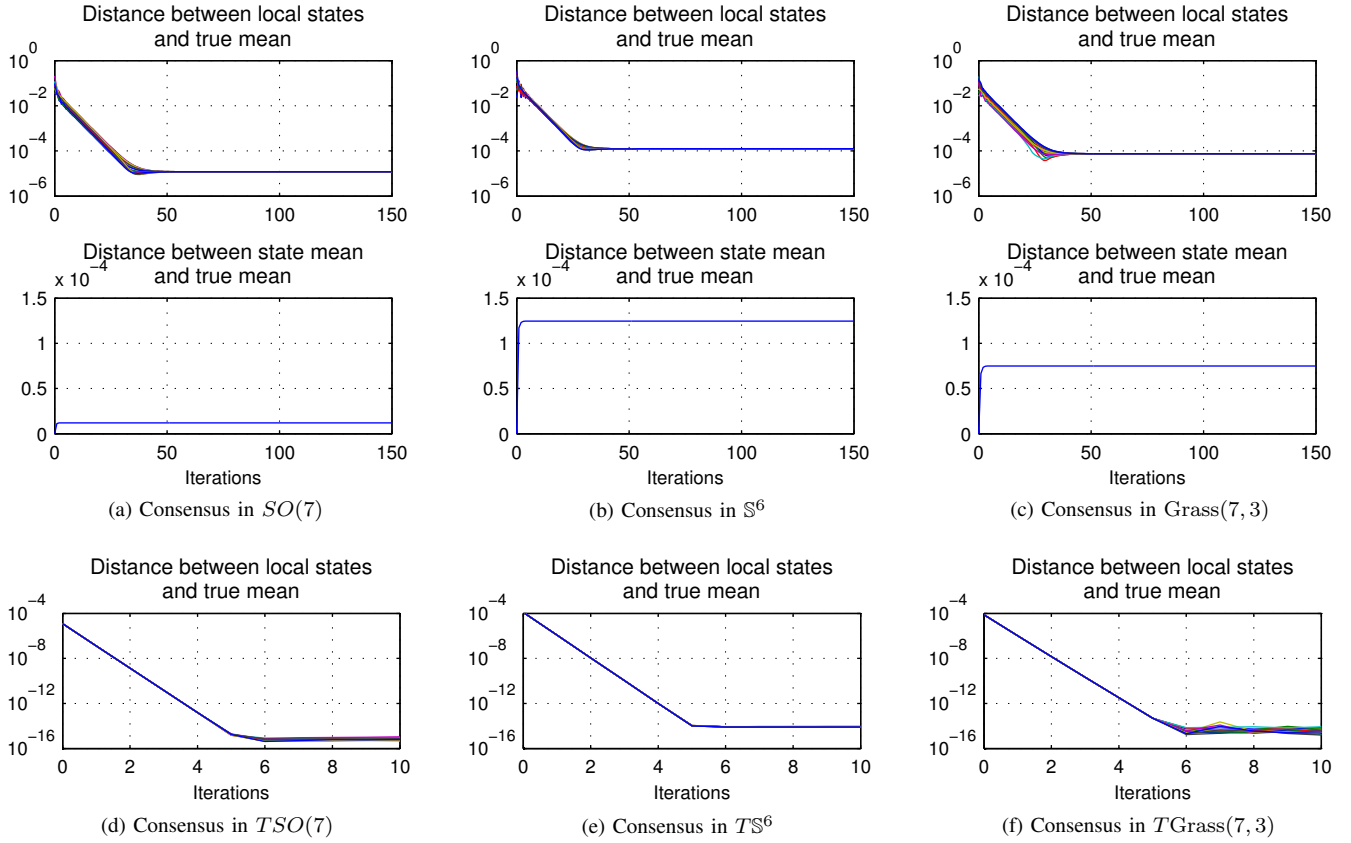


Fig. 2: Results for the algorithm applied to data in $SO(7)$, \mathbb{S}^6 and $\text{Grass}(7, 3)$. Top row: distances between each state and the Fréchet mean of the measurements for the consensus in the manifold algorithm. Middle row: distance between Fréchet mean of the states and the true Fréchet mean for the consensus in the manifold algorithm. Bottom row: distances between each state and the Fréchet mean of the measurements for the consensus in the tangent space algorithm.

- [2] R. Tron and R. Vidal, "Distributed image-based 3-D localization in camera sensor networks," in *Conference on Decision and Control*, 2009.
- [3] A. Sarlette, S. Bonnabel, and R. Sepulchre, "Coordinated motion design on Lie groups," *IEEE Transactions on Automatic Control*, vol. to be published, 2010.
- [4] R. Olfati-Saber, "Swarms on sphere: A programmable swarm with synchronous behaviors like oscillator networks," in *IEEE Conference on Decision and Control*, 2006, pp. 5060–5066.
- [5] L. Scardovi, A. Sarlette, and R. Sepulchre, "Synchronization and balancing on the N -torus," *Systems and Control Letters*, vol. 56, no. 5, pp. 335–341, 2007.
- [6] A. Sarlette and R. Sepulchre, "Consensus optimization on manifolds," *SIAM J. Control and Optimization*, vol. 48, no. 1, pp. 56–76, 2009.
- [7] T. Hatanaka, M. Fujita, and F. Bullo, "Vision-based cooperative estimation via multi-agent optimization," in *IEEE Conference on Decision and Control*, 2010.
- [8] Y. Igarashi, T. Hatanaka, M. Fujita, and M.W. Spong, "Passivity-based attitude synchronization in $SE(3)$," *IEEE Transactions on Control Systems Technology*, vol. 17, no. 5, pp. 1119–1134, 2009.
- [9] A. Olshevsky and J. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM Journal of Control and Optimization*, vol. 48, no. 1, pp. 33–55, 2007.
- [10] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 3, pp. 1520–1533, 2004.
- [11] T. Sakai, *Riemannian Geometry*, vol. 149 of *Translations of Mathematical Monographs*, American Mathematical Society, 1996.
- [12] I. Chavel, *Riemannian Geometry: a Modern Introduction*, vol. 98 of *Cambridge studies in advanced mathematics*, Cambridge University Press, 2 edition, 2006.
- [13] M. P. do Carmo, *Riemannian geometry*, Birkhäuser, Boston, MA, 1992.
- [14] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.
- [15] R. Tron, B. Afsari, and R. Vidal, "Average consensus on Riemannian manifolds with bounded curvature," Tech. Rep., Johns Hopkins University, 2011.
- [16] A. Edelman, T. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal of Matrix Analysis Applications*, vol. 20, no. 2, pp. 303–353, 1998.
- [17] H. Le, "Locating Fréchet means and with application to shape spaces," *Advances in Applied Probabilities*, vol. 33, no. 2, pp. 324–338, 2001.
- [18] B. Afsari, "Riemannian L^p center of mass: Existence, uniqueness, and convexity," *Proceedings of the AMS*, vol. 139, no. 2, pp. 655–673, 2011.
- [19] R. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [21] A. M. Bloch and A. Iserles, "Commutators of skew-symmetric matrices," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 15, no. 3, pp. 793 – 801, 2005.
- [22] B. O'Neill, "The fundamental equations of a submersion," *The Michigan Mathematical Journal*, vol. 13, pp. 459–469, 1966.