

Riemannian Consensus for Manifolds With Bounded Curvature

Roberto Tron, *Member, IEEE*, Bijan Afsari, *Member, IEEE*, and René Vidal, *Senior Member, IEEE*

Abstract—Consensus algorithms are popular distributed algorithms for computing aggregate quantities, such as averages, in ad-hoc wireless networks. However, existing algorithms mostly address the case where the measurements lie in Euclidean space. In this work we propose Riemannian consensus, a natural extension of existing averaging consensus algorithms to the case of Riemannian manifolds. Unlike previous generalizations, our algorithm is intrinsic and, in principle, can be applied to any complete Riemannian manifold. We give sufficient convergence conditions on Riemannian manifolds with bounded curvature and we analyze the differences with respect to the Euclidean case. We test the proposed algorithms on synthetic data sampled from the space of rotations, the sphere and the Grassmann manifold.

Index Terms—Grassmann manifold, Riemannian manifold.

I. INTRODUCTION

CONSIDER a set of low-power sensors, where each sensor can collect measurements from the surrounding environment and can communicate with a subset of neighboring nodes through wireless channels. We are interested in *distributed 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 a global quantity of interest (e.g., the average of the measurements). Natural candidates for this scenario are *consensus algorithms*, where each node maintains a local estimate of the global average, which is updated with the estimates from the local neighbors. The interesting characteristic of consensus algorithms is that they converge exponentially fast under very mild communication assumptions, even in the case of a time-varying network topology. However, traditional consensus algorithms have been mainly studied for the case where the measurements and the state of each node lie 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], [6]. However, these approaches are not easily generalizable to other manifolds. The work of [7] considers the problems of consensus and balancing on the more general class of compact homogeneous manifolds. However, the approach is in part *extrinsic*, i.e., it is based on specific embeddings of the manifolds in Euclidean space (where classical Euclidean consensus can be employed) and requires the ability to project the updates of Euclidean consensus onto tangent spaces. In this approach, 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 [8], 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 a totally *intrinsic* approach, which does not depend on specific embeddings of the manifold and does not require the definition of projection operations. Instead, it relies only on the intrinsic properties of the manifold, such as geodesic distances and exponential and logarithm maps. However, [1] focuses only on specific manifolds ($SE(3)$ and $SO(3)$) and does not provide a thorough convergence analysis. The work of [6] on the circle, while derived extrinsically, reduces to a special non-linear intrinsic protocol. It considers issues with local convergence and local minima on the circle, similarly to what we do for general manifolds in the present work. Other works on distributed algorithms for data lying in manifolds include [3], [9], which address the problem of coordination on Lie groups, and [2], which addresses the problem of camera localization. However, these works do not apply to general manifolds, as we consider in this paper.

Paper Contributions: In this paper, we propose a natural extension of consensus algorithms to measurements lying in a Riemannian manifold for the case where the network topology is fixed. We define a cost function which is the natural equivalent to the one for averaging consensus in the Euclidean case. We then obtain our Riemannian consensus algorithm as an application of Riemannian gradient descent to this cost function. This requires only the ability to compute the exponential and logarithm maps for the manifold of interest. We 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). We also point out analogies and differences with respect to the Euclidean case.

Our work makes 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, we provide sufficient conditions for the local and, in special cases,

Manuscript received December 30, 2011; revised July 18, 2012 and September 29, 2012; accepted September 29, 2012. Date of publication October 18, 2012; date of current version March 20, 2013. This work was supported by the grant NSF CNS-0834470. Recommended by Associate Editor L. Schenato.

The authors are with the Center for Imaging Science, Johns Hopkins University, Baltimore MD, 21202 USA (e-mail: tron@cis.jhu.edu; bijan@cis.jhu.edu; rvidal@cis.jhu.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TAC.2012.2225533

global convergence to the sub-manifold of consensus configurations. These conditions depend on the network connectivity, the geometric configuration of the measurements and the curvature of the manifold. We also provide stronger results that hold when additional assumptions on the manifold and network connectivity are made. Finally, we show that, while Euclidean consensus converges to the Euclidean mean of the initial measurements, the Riemannian extension does not converge to the Fréchet mean, which is the Riemannian equivalent of the Euclidean mean.

Paper Outline: In Section II we review Euclidean consensus and relevant notions from Riemannian geometry and optimization. In Section III we describe our extension of consensus algorithms to data in manifolds. Our main contributions are presented in Section IV and Section IV-F. We first give convergence results for the case of general manifolds. We then strengthen our results for the particular case of manifolds with constant, non-negative curvature. In Section V we test the proposed algorithm on manifolds such as the special orthogonal group, the sphere and the Grassmann manifold. In the Appendix we report all the additional derivations and proofs that support our claims.

II. MATHEMATICAL BACKGROUND

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

A. Review of Euclidean Consensus

Consider a network with N nodes. We represent the network as a connected, 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 $N_i = \{j \in V | \{i, j\} \in E\}$ and the number of neighbors or *degree* of node i as $|N_i|$. The maximum degree of the graph G is denoted as $\deg(G) = \max_i \{|N_i|\}$.

Assume that 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 $\bar{u} = (1/N) \sum_{i=1}^N u_i$. Note that this is a global quantity, in the sense that involves information from all the nodes. The well-known average consensus algorithm, to which we refer as Euclidean consensus, computes \bar{u} by iterating the difference equation

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

where $x_i(l)$ is the state of node i at iteration l and $\varepsilon \leq 1/\deg(G)$ is the step-size. It is easy to verify that the mean of the states is preserved at each iteration, i.e.,

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

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 N -tuples in which each element belongs to \mathbb{R} or another manifold \mathcal{M} . The cost (3) is convex and its global minima are achieved when the nodes reach a *consensus configuration*, i.e., when $x_i = y \in \mathbb{R}$ for all $i \in V$. 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., [10]). That is, all the states converge to a unique global minimizer which corresponds to the average of the initial measurements.

In addition, 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 [11].

B. Review of Concepts From Riemannian Geometry

In this section we present our notation for the Riemannian geometry concepts used throughout the paper. We refer the reader to [12]–[14] 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 assume that \mathcal{M} is *geodesically complete*, i.e., there always exists a minimal length geodesic between any two points in $x, y \in \mathcal{M}$. The length of this geodesic is said to be the *distance* between the two points, and is denoted as $d(x, y)$. The typical manifolds of practical interest (such as the one we mention in Section II-C) are all complete.

Let v be a unit-length tangent vector in $T_x \mathcal{M}$, i.e., $\|v\| = \langle v, v \rangle^{1/2} = 1$. We can 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 . Let $\tilde{\mathcal{I}}_x \subset T_x \mathcal{M}$ be the maximal open set on which \exp_x is a diffeomorphism and define the *interior set* [12, p. 216] as $\mathcal{I}_x = \exp_x \tilde{\mathcal{I}}_x$. The exponential map is invertible on \mathcal{I}_x and we can define the *logarithm map* as $\log_x y = \exp_x^{-1} y$ for $y \in \mathcal{I}_x$. We denote an *open geodesic ball* [14, p. 70] of radius $r > 0$ centered at $x \in \mathcal{M}$ as $\mathcal{B}_{\mathcal{M}}(x, r) \subset \mathcal{M}$. We also denote as $\text{inj}_x \mathcal{M}$ the *injectivity radius* of \mathcal{M} at $x \in \mathcal{M}$, i.e., 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 $x \in \mathcal{M}$.

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 $(d/dt)f(\gamma(t))|_{t=0}$, where $\gamma(t)$ is any curve such that $\gamma(0) = x$ and $\dot{\gamma}(0) = v$. The *gradient* of f 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 $v \in T_x \mathcal{M}$

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

Similarly, the Hessian is defined as the self-adjoint linear operator $\text{Hess}f(x)$ such that [14]

$$\left. \frac{d^2}{dt^2} f(\gamma(t)) \right|_{t=0} = \langle v, \text{Hess}\varphi(x)v \rangle. \quad (5)$$

Intuitively, as in the Euclidean case, $\text{grad}f$ indicates the direction along which f increases the most, while $\text{Hess}f$ indicates the local, second-order behavior of f . A point $x \in \mathcal{M}$ is called a *critical point* [15] of f if either $\text{grad}_x f(x) = 0$, i.e., it is a *stationary point*, or the gradient does not exist. In this paper, we mainly need the gradient of the squared distance function, which is given by

$$\frac{1}{2} \text{grad}_x d^2(x, y) = -\log_x(y). \quad (6)$$

Note that the gradient might be undefined if $d(x, y) \geq \text{inj}\mathcal{M}$.

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 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 . Intuitively, the curvature of a manifold indicates if two geodesics with tangents in σ spread slower (positive curvature) or faster (negative curvature) than in Euclidean space (which has zero curvature). See the Appendix A for a formal definition. For instance, on a sphere (which has constant positive curvature) geodesics eventually meet at antipodal points.

Related to the curvature and injectivity radius, we define the convexity radius $r^* > 0$ as

$$r^* = \frac{1}{2} \min \left\{ \text{inj}\mathcal{M}, \frac{\pi}{\sqrt{\Delta}} \right\} \quad (7)$$

where we use the convention that, if $\Delta \leq 0$, $1/\sqrt{\Delta} = +\infty$. Any open ball with radius $r \leq r^*$ is guaranteed to be convex [14]. The quantity r^* will play an important role in our convergence conditions.

In the following, we also make 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 use the notation $\mathbf{x} = (x_1, \dots, x_N)$ to indicate a point in \mathcal{M}^N and $\mathbf{v} = (v_1, \dots, v_N) \in T_{\mathbf{x}}\mathcal{M}^N$ to indicate a tangent vector. We 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 . We will use this notation to state results that involve the states of all the nodes.

C. Examples of Manifolds

We use the following manifolds as examples throughout this paper.

Euclidean Space: The space \mathbb{R}^n is a Riemannian manifold where the tangent space at 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 $\delta = \Delta = 0$ and $r^* = \text{inj}\mathbb{R}^n = +\infty$.

Orthogonal and Special Orthogonal Groups: The n -dimensional orthogonal group is defined as the set of $n \times n$ orthogonal matrices, $O(n) = \{R \in \mathbb{R}^{n \times n} : R^T R = I\}$. This group

has two connected components. One of them is the *special orthogonal group* $SO(n)$, i.e., the set of all possible rotations in \mathbb{R}^n , and has the additional property $\det(R) = 1$. The tangent space at a point R is $T_R SO(n) = \{Rv : v \in \mathfrak{so}(n)\}$, where $\mathfrak{so}(n)$ is the space of $n \times n$ skew-symmetric matrices. The standard bi-invariant Riemannian metric is given by $\langle v_1, v_2 \rangle = (1/2)\text{tr}(v_1^T v_2)$, $v_1, v_2 \in T_R SO(n)$. In this metric, the curvature bounds are $\Delta = 1/2$, and $\delta = 0$, except when $n = 3$, for which the curvature is constant and $\delta = \Delta = 1/4$. Also, $\text{inj}SO(n) = \pi$ and $r^* = \pi/2$.

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 $\pi/2$ and $r^* = \pi/4$.

The Sphere: The n -dimensional sphere is defined as $\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} : x^T x = 1\}$. The tangent space at a point x is defined as $T_x \mathbb{S}^n = \{y \in \mathbb{R}^{n+1} : y^T x = 0\}$. As metric, we use the standard inner product in \mathbb{R}^{n+1} . The geodesics follow great circles and the curvature is constant and $\delta = \Delta = 1$. Also, $\text{inj}\mathbb{S}^n = \pi$ and $r^* = \pi/2$.

More details about these manifolds and about the computation of the exp and log maps can be found in [16], [17].

D. Review of Riemannian Gradient Descent

Let $\varphi : \mathcal{M} \rightarrow \mathbb{R}$ be a twice differentiable, bounded below function defined on a Riemannian manifold \mathcal{M} . Given an initial point $x_0 \in \mathcal{M}$, one can define a Riemannian steepest descent algorithm, as shown by Alg. 1.

Algorithm 1 A Riemannian steepest gradient descent algorithm

Input: An initial element $x_0 \in \mathcal{M}$, a sequence of step sizes $\varepsilon(l) \in \mathbb{R}$

- 1) **Initialize** $x(0) = x_0$
 - 2) **For** $l \in \mathbb{N}$, **repeat**
 - a) $w = -\text{grad}_x \varphi(x(l))$
 - b) $x(l+1) = \exp_{x(l)}(\varepsilon(l)w)$
-

At each iteration l , the algorithm moves from the current estimate $x(l)$ to a new estimate $x(l+1)$ along the geodesic in the direction of the negative gradient with a step size $\varepsilon(l)$.

Alg. 1 gives only a basic version of a gradient-based descent algorithm on Riemannian manifolds. Many variations are possible, e.g., in the computation of the descent direction and of the step size or in the choice of the curve used to search for $x(l+1)$ (which does not need to be a geodesic). We refer to [15] and [16] for some examples of such variations.

Choice of a Fixed Step Size and Convergence: Ideally, one could compute the step size $\varepsilon(l)$ at each iteration by employing methods based on a line search. However, it might be more efficient or necessary to employ a pre-determined fixed step size, which is maintained constant throughout all the iterations, i.e., $\varepsilon(l) \equiv \varepsilon$. This happens, for instance, when the evaluation of the cost function is computationally expensive, as it is the case

for the distributed optimization problem in Section III. It is well known that the choice of ε affects the convergence of Alg. 1. For small step sizes, the algorithm will exhibit slow convergence. On the other hand, if the step size is too large, the algorithm might fail to converge at all.

In this section we review methods for choosing a fixed step size for Alg. 1. These methods mostly rely on bounds for the maximum eigenvalue of the Hessian of the cost function. The results will be instrumental in our convergence proofs. The ideas in this section are fairly standard for the case $\mathcal{M} = \mathbb{R}^n$ (see, for instance, [18, p. 46]), but here we review the case where \mathcal{M} is a general manifold (see also [19]).

We give the following definition of a bound on the Hessian.

Definition 1: Given a twice differentiable function $\varphi(x)$ defined on an open subset \mathcal{X} of a manifold \mathcal{M} , we say that the Hessian $\text{Hess}\varphi(x)$ is uniformly bounded on \mathcal{X} if there exists a finite, non-negative constant μ_{max} such that, for any $x_0 \in \mathcal{X}$ and any $v \in T_{x_0}\mathcal{M}$, the second derivative of φ along $\gamma_{x_0}(t) = \exp_{x_0}(tv)$ satisfies

$$\left. \frac{d^2}{dt^2} \varphi(\gamma_{x_0}(t)) \right|_{t=0} = \langle v, \text{Hess}\varphi(x_0)v \rangle \leq \mu_{max} \|v\|^2. \quad (8)$$

We can use this bound to obtain results on the step size ε and on the convergence of Alg. 1.

Proposition 2: Let μ_{max} be a uniform bound on $\text{Hess}(\varphi)$ as in Def. 1. Assume $\gamma_{x_0}(t) = \exp_{x_0}(-t\text{grad}_x\varphi(x_0)) \in \mathcal{X}$ for all $t \in (0, 2\mu_{max}^{-1})$ and let $\tilde{\varphi}(t) = \varphi(\gamma_{x_0}(t))$. Then $\tilde{\varphi}(t) \leq \tilde{\varphi}(0)$ for $t \in (0, 2\mu_{max}^{-1})$, with equality if and only if $x_0 \in \mathcal{X}$ is a stationary point of φ . In this case we say that t is an *admissible* step size.

Proof: Note that $\dot{\tilde{\varphi}}(0) = -\|\text{grad}\varphi(x_0)\|^2$ and $\ddot{\tilde{\varphi}}(0) \leq \mu_{max}\|\text{grad}\varphi(x_0)\|^2$. Using a second order Taylor expansion of $\tilde{\varphi}$ around $t = 0$ and using (8), one can show that

$$\tilde{\varphi}(t) \leq \tilde{\varphi}(0) - \|\text{grad}\varphi(x_0)\|^2 \varepsilon + \frac{\mu_{max} \|\text{grad}\varphi(x_0)\|^2}{2} \varepsilon^2 \quad (9)$$

for all $t \in (0, 2\mu_{max}^{-1})$. From this we can derive

$$\tilde{\varphi}(0) - \tilde{\varphi}(t) \geq \|\text{grad}\varphi(x_0)\|^2 t \left(1 - \frac{\mu_{max}t}{2} \right). \quad (10)$$

Note that the RHS of the inequality is strictly positive, because $t \in (0, 2\mu_{max}^{-1})$. Also, we have equality throughout if and only if $\|\text{grad}\varphi(x_0)\|^2 = 0$. ■

Proposition 3: In Alg. 1, let $\varepsilon(l) = \varepsilon \in (0, 2\mu_{max}^{-1})$, and assume $\exp_{x(l)}(-t\text{grad}_x\varphi(x(l))) \in \mathcal{X}$ for all $t \in [0, \varepsilon]$, and all $l \in \mathbb{N}$. Then any cluster point of the sequence $\{x(l)\}$ is a stationary point of \mathcal{X} .

Proof: We use a fairly standard argument. Since φ is bounded below, from (10) we have

$$\begin{aligned} & \varepsilon \left(1 - \frac{\mu_{max}\varepsilon}{2} \right) \sum_{l=0}^L \|\text{grad}_x\varphi(x(l))\|^2 \\ & \leq \sum_{l=0}^L \varphi(x(l)) - \varphi(x(l+1)) \\ & = \varphi(x(0)) - \varphi(x(L+1)) < \infty \end{aligned} \quad (11)$$

for all $L \in \mathbb{N}$. Therefore, the series $\sum_{l=0}^{\infty} \|\text{grad}_x\varphi(x(l))\|^2$ converges, $\lim_{l \rightarrow \infty} \|\text{grad}_x\varphi(x(l))\|^2 = 0$ and the gradient vanishes, i.e., $\lim_{l \rightarrow \infty} \text{grad}_x\varphi(x(l)) = 0$. Since $\text{grad}\varphi$ is continuous, any cluster point of $\{x(l)\}$ is a stationary point of φ . ■

In the context of Alg. 1, Prop. 2 implies that, as long as $\varepsilon \in (0, 2\mu_{max}^{-1})$ and $x(l+1) = \gamma_{x(l)}(\varepsilon) \in \mathcal{X}$, the cost function decreases at every iteration, except at stationary points where it remains constant. However, we stress here the fact that neither Prop. 2, nor Alg. 1, imply that each new iterate $x(l+1)$ belongs to \mathcal{X} when $x(l) \in \mathcal{X}$. Therefore, additional considerations are needed in order to derive complete results for the convergence of Alg. 1 to a stationary point (see Section IV-C and Section IV-F).

E. Fréchet Mean

In order to compare the consensus algorithm that we will propose to Euclidean consensus, we need a generalization of the concept of empirical mean for Riemannian manifolds. Let $\{u_i\}_{i=1}^N$ be a set of points in a Riemannian manifold \mathcal{M} . Similar to the geometric definition of mean in the Euclidean case, the *Fréchet mean* \bar{u} of the points is defined as the global minimizer of the sum of squared geodesic distances, i.e.

$$\bar{u} = \arg \min_{u \in \mathcal{M}} \sum_{i=1}^N d^2(u_i, u). \quad (12)$$

If the points lie in a ball of radius smaller than r^* , the global minimizer is unique and belongs to the same ball [20]. Moreover, for spaces of constant, non-negative curvature the Fréchet mean belongs to the closed convex hull of the measurements [20].

Note that Alg. 1 can be used for the computation of the Fréchet mean \bar{u} . In this case, the negative gradient is $w = (1/N) \sum_{i=1}^N \log_{\bar{u}}(u_i)$, which is essentially a mean in $T_{\bar{u}}\mathcal{M}$.

The conditions for the convergence to \bar{u} (as opposed to other critical points) are, in general, only partially known [21]. These conditions depend on the spread of the points $\{u_i\}$, the step size ε and the initialization x_0 of the algorithm.

III. RIEMANNIAN CONSENSUS

In this section we present our proposed algorithm, which we call *Riemannian consensus*. This algorithm can be considered as a direct extension of the Euclidean consensus to the Riemannian case. The basic idea is to use the formulation of consensus as an optimization problem and define a potential function on the Riemannian manifold of interest which is equivalent to the cost in (3). Riemannian gradient descent is then applied to obtain the update rules for each node.

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

$$\varphi(\mathbf{x}) = \frac{1}{2} \sum_{\{i,j\} \in E} d^2(x_i, x_j). \quad (13)$$

The gradient of φ with respect to x_i can be computed as

$$\text{grad}_{x_i} \varphi = \frac{1}{2} \text{grad}_{x_i} \sum_{j \in N_i} d^2(x_i, x_j) = - \sum_{j \in N_i} \log_{x_i}(x_j) \quad (14)$$

where we used the facts that the graph is undirected, $d(\cdot, \cdot)$ is symmetric and $d(x_i, x_i) = 0$.

Note that φ is differentiable, and (14) well defined, at least on $\mathcal{E}_{\mathcal{M}^N}(\text{inj}\mathcal{M})$ where $\text{inj}\mathcal{M}$ is the injectivity radius and

$$\mathcal{E}_{\mathcal{M}^N}(r) = \{\mathbf{x} \in \mathcal{M}^N : d(x_i, x_j) < r, \forall \{i, j\} \in E\}. \quad (15)$$

Alg. 2 is our proposed consensus protocol on \mathcal{M} and is obtained by applying Alg. 1 (Riemannian gradient descent) to the cost φ . As mentioned before, this protocol is a natural extension of the Euclidean case. In fact, when $\mathcal{M} = \mathbb{R}$ with the standard metric, the updates (16) reduce to the standard Euclidean updates (1). On the one hand, the convergence analysis for Euclidean consensus is simple: the cost (3) is a simple quadratic function, and simple tools from optimization theory and linear algebra are sufficient. On the other hand, a similar analysis for Riemannian consensus is not trivial, because we need to take into account the geometry of the manifold. In particular, Alg. 2 does not always converge to a consensus configuration, and the algorithm is well defined only when $\mathbf{x}(l) \in \mathcal{E}_{\mathcal{M}^N}(\text{inj}\mathcal{M})$ for all $l \in \mathbb{N}$.

Algorithm 2 Riemannian consensus

Input: The measurements u_i at each node $i \in \{1, \dots, N\}$

- 1) **For** each node $i \in \{1, \dots, N\}$ **in parallel**
 - a) **Initialize** the state with the local measurement,

$$x_i(0) = u_i$$
 - b) **For** $l \in \mathbb{N}$, **compute** the update

$$x_i(l+1) = \exp_{x_i(l)}(-\varepsilon \text{grad}_{x_i} \varphi(\mathbf{x}(l))). \quad (16)$$

IV. CONVERGENCE TO THE CONSENSUS SUB-MANIFOLD

In this section we analyze the convergence properties of Alg.

2. We divide our treatment in three parts:

Section IV-A) We notice that the cost can have multiple local minima and we define a non-trivial subset $\mathcal{S} \subset \mathcal{M}^N$ that contains all *global* minimizers but no other critical point.

Section IV-B) We give a distributed method to choose a step-size ε for such that the cost is non-increasing at every iteration.

Section IV-C) We derive local, sufficient conditions under which the algorithm is guaranteed to converge to the *consensus sub-manifold*, i.e., the set of consensus configurations.

These results pertain to general manifolds and general network topologies. However, for special cases and under additional assumptions, we also give results on:

Section IV-D) *Global* convergence to the *set* of global minimizer.

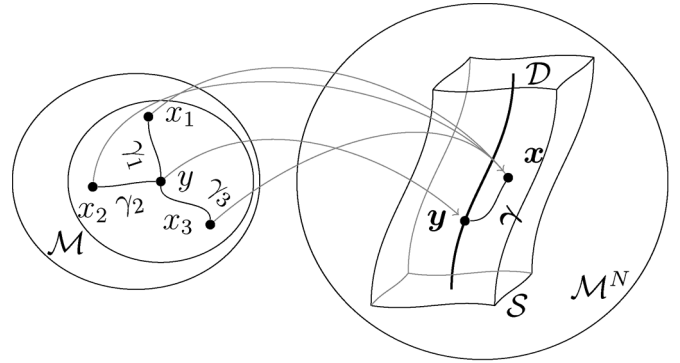


Fig. 1. Construction of the geodesic for testing if φ has a local minimum at (x_i, x_j) .

Section IV-F) *Local* convergence to a *single* global minimizer.

A. Global Minimizers of φ and the Set \mathcal{S}

We first show that the global minimizers of φ corresponds to consensus configurations. Let us define the consensus sub-manifold \mathcal{D} as the diagonal space of \mathcal{M}^N , i.e.

$$\mathcal{D} = \{(y, \dots, y) \in \mathcal{M}^N : y \in \mathcal{M}\}. \quad (17)$$

This set represents the manifold of all possible consensus configurations of the network, where all the nodes agree on a state. The following lemma shows that \mathcal{D} is exactly the set of global minimizers of φ . It follows easily from the non-negativity of the distance.

Lemma 4: If G is connected, then $\mathbf{x} \in \mathcal{D}$ if and only if \mathbf{x} is a global minimizer of φ .

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

$$\mathcal{S} = \{(x_1, \dots, x_N) \in \mathcal{M}^N : \exists y \in \mathcal{M} \text{ s.t. } x_i \in \mathcal{B}_{\mathcal{M}}(y, r^*) \forall i \in V\} \quad (18)$$

where r^* is the convexity radius given in (7). 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). Note that $\mathbf{x} \in \mathcal{S}$ is equivalent to saying that there exists $y \in \mathcal{M}$ such that, $\max_{i \in V} d(x_i, y) < r^*$ for all $i \in V$. As it happens, a sufficient condition for the uniqueness of the Fréchet mean is $\mathbf{u} = (u_1, \dots, u_n) \in \mathcal{S}$ [20].

We can now state our first main contribution.

Theorem 5: A point $\mathbf{x} \in \mathcal{S}$ is a critical point for φ if and only if $\mathbf{x} \in \mathcal{D}$. In other words, the set \mathcal{S} contains all the global minima and no other critical points of φ .

Note that, outside of \mathcal{S} , the cost φ might have local minima (e.g., see Section V, Fig. 3). However, as long as the iterates do not leave \mathcal{S} , Alg. 2 will behave as in the Euclidean case insofar as it converges toward the set of global minima. In order to prove Thm. 5, we need the following lemma, which is proven in Appendix C.

Lemma 6: 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 $(d/dt)\phi_{12}^2(t) \geq 0$ for $t \in (0, 1]$, with equality if and only if $x_1 = x_2$.

In Euclidean space, the result of Lemma 6 is trivial, because the distance between the two geodesics is always non-de-

creasing. However, the same is not true in general. On a sphere, the distance between two geodesics starting from the north pole with equal speed would start to decrease after passing the equator (which is exactly at distance r^*).

Proof of Theorem 5: If $\mathbf{x} \in \mathcal{D}$, then, from Prop. 4, \mathbf{x} is a global minimizer of φ and hence a critical point. On the other hand, we will now show that $\mathbf{x} \notin \mathcal{D}$ cannot be a critical point of φ because there exists a geodesic $\boldsymbol{\gamma} : [0, 1] \rightarrow \mathcal{M}^N$ such that $\boldsymbol{\gamma}(1) = \mathbf{x}$ and along which $(d/dt)\varphi(\boldsymbol{\gamma}(t))|_{t=1} \neq 0$. Then, from the definition in (4), $\text{grad}\varphi(\mathbf{x}) \neq 0$.

In order to construct such a geodesic, notice that, since $\mathbf{x} \in \mathcal{S}$, there exists $y \in \mathcal{M}$ such that $\max_{i \in V} d(x_i, y) < r^*$. Define unique minimal geodesics $\gamma_i(t)$ such that $\gamma_i(0) = y$ and $\gamma_i(1) = x_i$. Then $\boldsymbol{\gamma}(t) = (\gamma_1(t), \dots, \gamma_N(t))$ is a minimal geodesic in \mathcal{M}^N (see also Fig. 1). It follows that:

$$\left. \frac{d}{dt}\varphi(\boldsymbol{\gamma}(t)) \right|_{t=1} = \frac{1}{2} \sum_{\{i,j\} \in E} \left. \frac{d}{dt} d^2(\gamma_i(t), \gamma_j(t)) \right|_{t=1}. \quad (19)$$

Since $d(x_i, y) < r^*$, from Lemma 6 we know that each term in the sum in (19) (i.e., each derivative) is positive except for the case where $\gamma_i(1) = \gamma_j(1)$, i.e., $x_i = x_j$ and $\log_{x_i}(x_j) = 0$. If *all* the terms of the sum were zero, since G is connected, we would have that $x_i = x_j$ for all $\{i, j\} \in E$, i.e., $\mathbf{x} \in \mathcal{D}$. However, by assumption $\mathbf{x} \notin \mathcal{D}$, hence at least one of the terms in (19), and therefore the entire sum in (19) is greater than zero and $\boldsymbol{\gamma}$ is the desired geodesic. ■

Special Case Where G is a Tree: In general, \mathcal{S} is not maximal, because the result in Lemma 6 can be quite conservative: There might exist a set containing \mathcal{D} and no other critical points which is larger than \mathcal{S} . In fact, the following holds when G is a tree.

Proposition 7: If G is a tree, any stationary point \mathbf{x} of φ is a global minimizer, i.e., $\mathbf{x} \in \mathcal{D}$.

Proof: We will now introduce some new notation exclusively for the purposes of this proof. Pick an arbitrary node as the root of the tree and denote as $x_{h,i}$ the state of the i -th node among the ones at hop-distance h from the root (e.g., $x_{0,1}$ is the state at the root). Also, denote as $x_{P(h,i)}$ the single parent and $x_{C(h,i,j)}$, $j \in \{1, |N_{h,i}| - 1\}$ the j -th children of $x_{h,i}$. Using this notation we can rewrite (14) as

$$\text{grad}_{x_{h,i}} \varphi(\mathbf{x}) = -\log_{x_{h,i}} x_{P(h,i)} - \sum_{j=1}^{|N_{h,i}|-1} \log_{x_{h,i}} x_{C(h,i,j)} \quad (20)$$

with the appropriate modifications for the leaves and the root of G . Now assume $\text{grad}_{\mathbf{x}}\varphi = 0$. For a leaf node, (20) becomes $\log_{x_{h,i}} x_{P(h,i)} = 0$ (since leaves do not have any child) and therefore $x_{h,i} = x_{P(h,i)}$. Now assume that, for a given hop-distance h , we have $x_{h,i} = x_{C(h,i,j)}$ for all indices h, i and j . Then, according to (20), again $x_{h,i} = x_{P(h,i)}$. It is then simple to show, by induction, that $x_i = x_{0,1}$ for any $i \in \{1, \dots, N\}$. Therefore, $\text{grad}_{\mathbf{x}}\varphi = 0$ implies $\mathbf{x} \in \mathcal{D}$. ■

We will use Thm. 5 to show local convergence in general manifolds (Section IV-C) and manifolds of non-negative, constant curvature (Section IV-F), while we will use Prop. 7 for proving global convergence when G has linear topology (Section IV-D).

B. Choice of the Stepsize ε

In this section we provide results on the range of admissible step sizes ε which can be computed in a distributed way, and which guarantees convergence of Alg. 2. From Prop. 2, we already know that any $\varepsilon \in (0, 2\mu_{max}^{-1})$ is admissible, where μ_{max} is a bound on $\text{Hess}\varphi(\mathbf{x})$, as per Def. 1. However, we need to estimate a value for μ_{max} , and it should be possible to compute this value in a distributed way. The following proposition breaks the problem in two parts: one regarding the network topology, and the other regarding the distance function.

Proposition 8: Given a graph G , define $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ as

$$\varphi(\mathbf{x}) = \sum_{\{i,j\} \in E} \varphi_{ij}(x_i, x_j) \quad (21)$$

where, for all $i \in V$, $x_i \in \mathcal{X}_i \subseteq \mathcal{M}$ and, for all $\{i, j\} \in E$, $\varphi_{ij} : \mathcal{X}_i \times \mathcal{X}_j \rightarrow \mathbb{R}$ and $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_N$. Let μ_{max}^d be a bound for $\text{Hess}\varphi_{ij}$ on $\mathcal{X}_i \times \mathcal{X}_j$, $\{i, j\} \in E$. Then, a bound on the Hessian of the complete function φ on \mathcal{X} is given by

$$\mu_{max} = \mu_{max}^d \text{deg}(G) \quad (22)$$

where $\text{deg}(G)$ is the maximum node degree of the graph G .

Proof: The gradient of (21) at a point $\mathbf{x} = (x_1, \dots, x_N)$ is given by $\mathbf{v} = \text{grad}_{\mathbf{x}}\varphi(\mathbf{x})$ where $v_i = \text{grad}_{x_i}\varphi(\mathbf{x})$. Define the cost function restricted to the geodesic along the gradient descent direction as $\tilde{\varphi}(t) = \varphi(\exp_{\mathbf{x}_0}(-t\mathbf{v}))$. Similarly, define the restriction for each pairwise term $\tilde{\varphi}_{ij}(t) = \varphi_{ij}(\exp_{x_{0,i}}(-tv_i), \exp_{x_{0,j}}(-tv_j))$. Using Def. 1, we have

$$\left. \frac{d^2}{dt^2} \tilde{\varphi}_{ij}(t) \right|_{t=0} < \mu_{max}^d (\|v_i\|^2 + \|v_j\|^2). \quad (23)$$

The second derivative of $\tilde{\varphi}(t)$, and hence the Hessian of φ , can be uniformly bound as

$$\begin{aligned} \left. \frac{d^2}{dt^2} \tilde{\varphi}(t) \right|_{t=0} &= \sum_{\{i,j\} \in E} \left. \frac{d^2}{dt^2} \tilde{\varphi}_{ij}(t) \right|_{t=0} \\ &< \mu_{max}^d \sum_{\{i,j\} \in E} (\|v_i\|^2 + \|v_j\|^2) \\ &= \mu_{max}^d \sum_{i \in V} |N_i| \|v_i\|^2 \\ &\leq \mu_{max}^d \text{deg}(G) \sum_{i \in V} \|v_i\|^2 \\ &= \mu_{max}^d \text{deg}(G) \|\mathbf{v}\|^2. \end{aligned} \quad (24)$$

The claim of the theorem follows. ■

In our case, the global cost function φ is given by (13), and $\mathcal{X} \subset \mathcal{E}_{\mathcal{M}^N}(d_{max})$ where $\mathcal{E}_{\mathcal{M}^N}$ is given in (15) and we use $d_{max} \leq \text{inj}\mathcal{M}$ to represent the maximum allowed distance between the states of any two neighboring nodes. The bound μ_{max}^d in (22) is given by the following.

Proposition 9: The Hessian of the function $\varphi_{ij}(x_i, x_j) = (1/2)d^2(x_i, x_j)$ can be bounded on $\mathcal{E}_{\mathcal{M}^2}(d_{max})$ by

$$\mu_{max}^d(d_{max}) = \max \left\{ 2, d_{max} \left(\frac{C_\delta(d_{max})}{S_\delta(d_{max})} + \frac{1}{S_\Delta(d_{max})} \right) \right\} \quad (25)$$

where $C_\kappa(t)$, $S_\kappa(t)$ are defined as

$$\begin{aligned} S_\kappa(t) &= \frac{1}{\sqrt{\kappa}} \sin(\sqrt{\kappa}t), & C_\kappa(t) &= \cos(\sqrt{\kappa}t) & \text{for } \kappa > 0 \\ S_\kappa(t) &= t, & C_\kappa(t) &= 1 & \text{for } \kappa = 0 \\ S_\kappa(t) &= \frac{1}{\sqrt{|\kappa|}} \sinh(\sqrt{|\kappa|}t), \\ C_\kappa(t) &= \cosh(\sqrt{|\kappa|}t) & \text{for } \kappa < 0. \end{aligned} \quad (26)$$

The proof can be found in Appendix D. We remark that the bound on μ_{max}^d is sharp, in the sense that it can be achieved for manifolds with constant curvature [17]. In fact, for the Euclidean space and for spaces of non-negative constant curvature (e.g., the sphere or $SO(3)$) this bound is $\mu_{max}^d = 2$, and it is independent from d_{max} . However, in general, the bound depends on d_{max} . Still, we might be able to use a uniform upper bound, say, in terms of $\text{inj}\mathcal{M}$. In this case, for $d_{max} \leq \text{inj}\mathcal{M}$, we get $\mu_{max}^d \simeq 3.8$ for both the Grassmann manifold and $SO(n)$, $n \geq 4$.

Our second main contribution follows by combining Prop. 8, 9 and 2:

Theorem 10: Let $\mu_{max} = \deg(G)\mu_{max}^d(d_{max})$. Then $\varepsilon \in (0, 2\mu_{max}^{-1})$ is an admissible step size.

As a corollary, we have that the same bounds apply to both the Euclidean case and the case of manifolds with positive constant curvature (such as the sphere and $SO(3)$).

Corollary 11: For spaces of constant curvature $\delta = \Delta \geq 0$, we can choose $\varepsilon \in (0, \deg(G)^{-1})$.

However, in general (e.g., in manifolds with negative curvature) we need to reduce ε according to d_{max} .

Computing ε at Every Node: We can devise a distributed method to compute the same ε at every node. The maximum degree $\deg(G)$ 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 [22]. Bounds on the maximum distance can be precomputed (in the case of compact manifolds), or estimated by using consensus to bound the value of the cost function for the measurements $\varphi(\mathbf{u})$ together with ideas similar to the ones we will see in the proof of Thm. 12 (see [17] for details) before the stop.

C. Local Convergence to the Consensus Sub-Manifold

Thm. 10 together with Prop. 3, implies convergence of the consensus algorithm to the set of critical points of φ , which includes local minima. However, we are interested in convergence to the set of global minimizers \mathcal{D} , which is contained in \mathcal{S} . If we could ensure that the iterates $\{\mathbf{x}(l)\}$ stay in \mathcal{S} , then we could deduce convergence to \mathcal{D} . With this goal in mind, our next main contribution uses \mathcal{S}_{conv} , a sub-level set of φ , to give local sufficient conditions for convergence.

Theorem 12: Let $D = \text{diam}(G)$ denote the diameter of the network graph G and define $\mathcal{S}_{conv} = \{\mathbf{x} \in \mathcal{M}^N : \varphi(\mathbf{x}) < (r^*)^2/2D\}$. Then, $\mathcal{S}_{conv} \subseteq \mathcal{S}$. Moreover, if the consensus protocol (16) is initialized with measurements $\mathbf{u} \in \mathcal{S}_{conv}$ and ε is admissible, then $\mathbf{x}(l)$ 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 geodesic distance between states x_p and x_q with the cost φ . Using the triangular and Jensen's inequalities, and the fact that $K \leq D$, we have

$$\begin{aligned} d^2(x_p, x_q) &\leq \left(\sum_{k=0}^{K-1} d(x_{i_k}, x_{i_{k+1}}) \right)^2 \leq K \sum_{k=0}^{K-1} d^2(x_{i_k}, x_{i_{k+1}}) \\ &\leq K \sum_{\{i,j\} \in E} d^2(x_i, x_j) \leq 2D\varphi(\mathbf{x}). \end{aligned} \quad (27)$$

This shows that if $\mathbf{x} \in \mathcal{S}_{conv}$, then $\varphi(\mathbf{x}) < r^*/2D$ and $d(x_p, x_q) < r^*$, for any $p, q \in V$. This implies $\mathbf{x} \in \mathcal{S}$. Next, since \mathcal{S}_{conv} is a sub-level set of φ and ε is admissible, one can show by continuity that the geodesic $\gamma_{\mathbf{x}(l)}(t) = \exp_{\mathbf{x}(l)}(-t \text{grad}_{\mathbf{x}} \varphi(\mathbf{x}(l)))$ is always in \mathcal{S}_{conv} for all $t \in (0, 2\mu_{max}^{-1}(d_{max}))$. Finally, since $\mathbf{x}(0) \in \mathcal{S}_{conv}$ and $\varphi(\mathbf{x}(l))$ is non-increasing, the sequence $\{\mathbf{x}(l)\}$ generated by the protocol will stay in $\mathcal{S}_{conv} \subseteq \mathcal{S}$. From this fact and from Prop. 3, any cluster point of the sequence $\mathbf{x}(l)$ will be a stationary point in \mathcal{S} , i.e., a global minimizer. ■

Note that we have shown convergence to a set and not to a single point. In principle, the proof does not exclude cases where the sequence $\{\mathbf{x}(l)\}$ has multiple cluster points in \mathcal{D} . In practice, we have convergence to a single global minimizer under much more relaxed conditions. We can strengthen Thm. 12 by making additional assumptions.

D. Cases of Global Convergence to the Consensus Sub-Manifold

In general, the basin of attraction $\mathcal{S}_{conv} \subset \mathcal{S}$ given by Thm. 12 can be quite small, because it depends on the diameter of the network, which might be large. Moreover, $\mathcal{S} \subseteq \mathcal{E}_{\mathcal{M}^N}(\text{inj}\mathcal{M}) \subset \mathcal{M}^N$ and, in general, $\mathcal{E}_{\mathcal{M}^N}(\text{inj}\mathcal{M})$ might be much larger than \mathcal{S} , especially when each node has a small number of neighbors. In this section we show that the basin of attraction of the global minimizers can be enlarged for particular manifolds and network topologies.

For instance, the following is a special case for Thm. 12.

Corollary 13: If $r^* = \infty$ and ε is admissible, then the iterates $\mathbf{x}(l)$ from the consensus protocol (16) converge to \mathcal{D} for any set of initial measurements $\mathbf{u} \in \mathcal{M}^N$.

This corollary can be used for \mathbb{R}^n and some other manifold with non-positive curvature (such as the hyperbolic space or the space of positive definite matrices), and for any graph G . On the other hand, if G has linear topology (i.e., it is a tree with a single branch), the following holds for any manifold \mathcal{M} .

Proposition 14: Assume G has linear topology, and the consensus protocol (16) is initialized with measurements $\mathbf{u} \in \mathcal{E}_{\mathcal{M}^N}(\text{inj}\mathcal{M})$. Then $\mathbf{x}(l)$ converges to \mathcal{D} .

Proof: First, notice that the edges of the network are exactly $\{i, i+1\}$ for $i \in \{1, \dots, N-1\}$. Then the assumptions imply $d(u_i, u_{i+1}) < \text{inj}\mathcal{M}$. We will now show that this same property is also satisfied by all the iterates $\mathbf{x}(l)$, i.e., $\mathbf{x}(l) \in \mathcal{E}_{\mathcal{M}^N}(\text{inj}\mathcal{M})$ for all $l \in \mathbb{N}$. For the sake of brevity, we will use the notation $d_i(l) = d(x_i(l), x_{i+1}(l))$, with the convention $d_0 = d_N = 0$,

and $w_i(l) = (\varepsilon/2)(\log_{x_i(l)} x_{i-1} + \log_{x_i(l)} x_{i+1})$. By using the triangular inequality twice, notice that

$$\begin{aligned} d_i(l+1) &= d\left(\exp_{x_i(l)} w_i(l), \exp_{x_{i+1}(l)} w_{i+1}(l)\right) \\ &\leq d_i(l) + \|w_i(l)\| + \|w_{i+1}(l)\| \end{aligned} \quad (28)$$

with equality if and only if $\{x_j(l)\}_{j=i-1}^{i+2}$ all lie in order on the same geodesic. In such case, $\|w_i\| = |d_i(l) - d_{i+1}(l)|$ and

$$\begin{aligned} d_i(l+1) &< d_i(l) + \|w_i(l)\| + \|w_{i+1}(l)\| \\ &= (1 - \varepsilon)d_i(l) - \frac{\varepsilon}{2}(d_{i-1}(l) + d_{i+1}(l)) \leq \text{inj } \mathcal{M}. \end{aligned} \quad (29)$$

This shows that, at any iteration, the distance between any two neighbors will be always less than $\text{inj } \mathcal{M}$, i.e., φ will always be differentiable at $\mathbf{x}(l)$. Combining this fact with Prop. 3 and Prop. 7, the claim follows. ■

E. Lack of Convergence to the Fréchet Mean

As we mentioned in Section II-A, when we minimize φ in Euclidean 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 simulations in Section V. 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. 12), this need not correspond to the desired Fréchet mean.

For computing the exact Fréchet mean of the measurements, one could extend the *consensus in the tangent space* algorithm from [1] to general manifolds. However, the convergence analysis of such algorithm is out of the scope of this paper.

F. Cases of Convergence to a Single Consensus Configuration

In this section we prove local convergence for the specific case of spaces with constant, non-negative curvature. We can strengthen Thm. 12 under three aspects.

- 1) The set of initializations for which convergence is guaranteed is enlarged from \mathcal{S}_{conv} to \mathcal{S} .
- 2) Convergence is to a single consensus state instead of the entire consensus set.
- 3) The consensus state is shown to lie in the convex hull of the initial measurements.

In the following, we define the convex hull of \mathcal{U} , $\text{hull}(\mathcal{U})$, as the minimal convex subset of \mathcal{M} containing \mathcal{U} (when it exists). A sufficient condition for $\text{hull}(\mathcal{U})$ to exist is that \mathcal{U} is contained in a convex ball. In this section, this will always be the case. Our strategy will be to show that the convex hull of the states shrinks at every iteration. This implies that the states stay in a compact set which, together with results from previous sections, implies convergence to single point. Before starting, we need a lemma which follows easily from the definition of $\text{hull}(\cdot)$.

Lemma 15: Let $\mathcal{U}, \mathcal{V} \subset \mathcal{M}$ be two sets such that $\mathcal{U} \subseteq \mathcal{V}$. Then $\text{hull}(\mathcal{U}) \subseteq \text{hull}(\mathcal{V})$.

We can then prove the following.

Proposition 16: Assume that \mathcal{M} has constant, non-negative curvature, $\mathbf{x}(l) \in \mathcal{S}$, and $x_i(l+1)$ is computed according to (16) with $\varepsilon \in (0, \mu_{max}^{-1}]$. Then $x_i(l+1) \in \text{hull}(\{x_i(l)\}_{i \in V})$. Moreover, this implies $x_i(l+1) \in \text{hull}(\{u_i\}_{i \in V})$, where $\{u_i\}_{i \in V}$ are the initial measurements (see Alg. 2).

Proof: From [21, Thm. 5] and Lemma 15, we have $x_i(l+1) \in \text{hull}(\{x_j(l)\}_{j \in N_i \cup \{i\}}) \subseteq \text{hull}(\{x_i(l)\}_{i \in V})$. Then, we also have $\text{hull}(\{x_i(l+1)\}_{i \in V}) \subseteq \text{hull}(\{x_i(l)\}_{i \in V})$ and, iteratively, $\text{hull}(\{x_i(l+1)\}_{i \in V}) \subseteq \text{hull}(\{x_i(0)\}_{i \in V}) = \text{hull}(\{u_i\}_{i \in V})$. The claim is proven. ■

We are now ready to show an improved version of Thm. 12.

Theorem 17: Assume that \mathcal{M} has constant, non-negative curvature and $\mathbf{u} \in \mathcal{S}$. Then the iterates given by protocol (16) with $\varepsilon \in (0, \mu_{max}^{-1}]$ satisfy, for all $j \in V$, $\lim_{l \rightarrow \infty} x_j(l) = y^*$, where $y^* \in \text{hull}(\{u_i\}_{i \in V})$.

Proof: For the sake of brevity, let $\mathcal{U} = \text{hull}(\{u_i\}_{i \in V})$. We show the claim in three steps. The first step is to show that $\mathcal{U}^N \subseteq \mathcal{S}$. By definition of \mathcal{S} in (18), $\mathbf{u} \in \mathcal{S}$ implies that there exists $y \in \mathcal{M}$ such that $u_i \in \mathcal{B}_{\mathcal{M}}(y, r^*)$ for all $i \in \{1, \dots, N\}$. Hence $\mathcal{U} \subset \mathcal{B}_{\mathcal{M}}(y, r^*)$. It follows that for any point $\mathbf{v} \in \mathcal{U}^N$, we also have $v_i \in \mathcal{B}_{\mathcal{M}}(y, r^*)$, which means $\mathbf{v} \in \mathcal{S}$. Hence $\mathcal{U}^N \subseteq \mathcal{S}$. The second step of the proof is to show that the iterates $\{\mathbf{x}(l)\}$ converge to a specific, bounded subset of \mathcal{D} . From Prop. 16, we have that, for all $i \in V$, the sequence of iterates $\{x_i(l)\}$ remains in \mathcal{U} . Equivalently, we have that $\mathbf{x}(l) \in \mathcal{U}^N \subseteq \mathcal{S}$ for all $l \in \mathbb{N}$. From this fact and Prop. 3 we have therefore that the iterates $\{\mathbf{x}(l)\}$ converge to the set $\mathcal{D}_{\mathcal{U}} = \mathcal{D} \cap \mathcal{U}^N$. The third and final step of the proof is to show convergence to a single point. Since $\mathcal{D}_{\mathcal{U}}$ is compact, there exists an infinite subsequence of indices $l_k \subset \mathbb{N}$ such that $\lim_{k \rightarrow \infty} \mathbf{x}(l_k) = \mathbf{y}^* \in \mathcal{D}_{\mathcal{U}}$, i.e., the subsequence of iterates $\{\mathbf{x}(l_k)\}$ converges to a single point in $\mathcal{D}_{\mathcal{U}}$ of the form $\mathbf{y}^* = (y^*, \dots, y^*)$ where $y^* \in \text{hull}(\{u_i\}_{i \in V})$. This implies that for any arbitrarily small $\xi \geq 0$ there exists an $L \in \mathbb{N}$ large enough such that $\mathbf{x}(L) \in \mathcal{B}_{\mathcal{M}^N}(\mathbf{y}^*, \xi)$. This in turn implies that $x_i(L) \in \mathcal{B}_{\mathcal{M}}(y^*, \xi)$. Using Prop. 16 we therefore get that, for all $l \geq L$, $l \in \mathbb{N}$, we have $x_j(l) \in \text{hull}(\{x_i(l)\}) \subseteq \text{hull}(\{x_i(L)\}) \subset \mathcal{B}_{\mathcal{M}}(y^*, \xi)$. To summarize we have that $\forall \xi \geq 0, \exists L \in \mathbb{N}: \forall l \geq L, x_i(l) \in \mathcal{B}_{\mathcal{M}}(y^*, \xi)$, which, by definition, means $\lim_{l \rightarrow \infty} x_j(l) = y^*$. ■

Note that Thm. 17 requires $\varepsilon \in (0, \mu_{max}^{-1}]$ instead of $\varepsilon \in (0, 2\mu_{max}^{-1})$, as we used in Thm. 12. This is because we rely on the fact that the iterates $\{x_i(l)\}$ never leave $\text{hull}(\{u_i\}_{i \in V})$, which might not be true if $\varepsilon \in (\mu_{max}^{-1}, 2\mu_{max}^{-1})$. Finally, by combining Thm. 17 with Cor. 13, we obtain the known result that Euclidean consensus with $\mathcal{M} = \mathbb{R}^n$ has global convergence to a single consensus configuration.

While we conjecture that it should be possible to extend the results of this section to manifolds with non-constant curvature, extending Prop. 16 is not trivial. The problem is that, with non-constant curvature, the convex hull of a set of point becomes much more complex. For instance, in a sphere, the convex hull of three points is a two-dimensional triangle. However, with non-constant curvature, the convex hull is no more two-dimensional. See [21] for details. Therefore, the strategy adopted here cannot easily replace the one in Section IV-C.

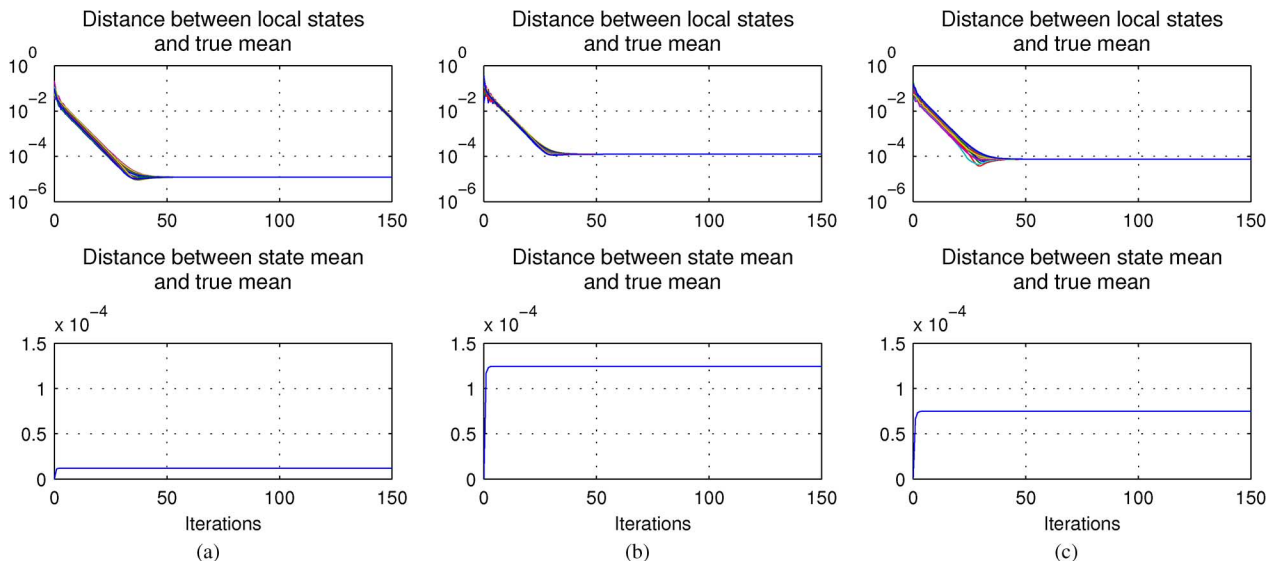


Fig. 2. Results for the algorithm applied to data in $SO(7)$, \mathbb{S}^6 and $Grass(7, 3)$. Top row: distances between each state and the Fréchet mean of the measurements for the Riemannian consensus algorithm. Bottom row: distance between Fréchet mean of the states and the true Fréchet mean. (a) Consensus in $SO(7)$; (b) consensus in \mathbb{S}^6 ; (c) consensus in $Grass(7, 3)$.

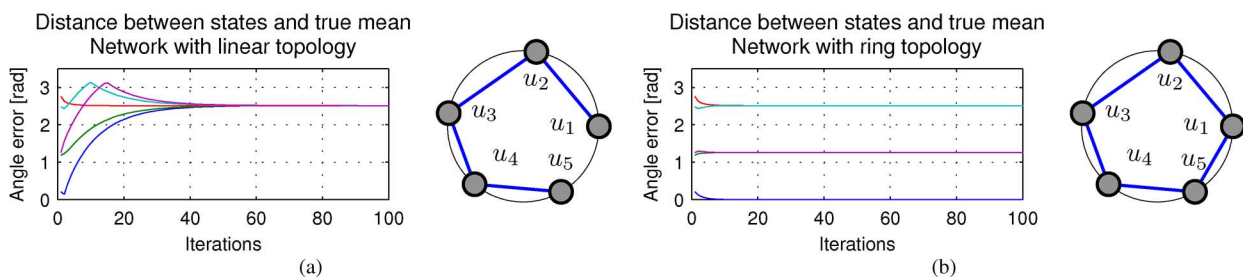


Fig. 3. Example where Riemannian consensus converges (a) or fails to converge (b) to a consensus configuration depending on the topology. These plots correspond to initial configurations around a closed geodesic of $SO(3)$, as portrayed on the right.

V. SIMULATIONS

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

The simulations 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 $i \in V$ is then defined as $u_i = \exp_{x_0}(v_{0i})$. We then run our Riemannian consensus algorithm for 150 iterations. We use step sizes compatible with the bounds found in Section IV-B. After each iteration, we compute the distance between each state and the Fréchet mean \bar{u} of the initial measurements (Fig. 2, top row). We also record the distance between the Fréchet mean of the states at each iteration and \bar{u} (Fig. 2, bottom row). We have selected $SO(7)$, \mathbb{S}^6 and $Grass(7,3)$ as particular examples. While we show that our algorithm can be applied to non-typical manifolds, similar results are obtained on manifolds such as $SO(3)$, $SO(2)$ or \mathbb{S}^2 .

A number of points can be made on the simulations. First, in this experiment, the measurements that we have generated are not too far one from the other, and the states converge to a consensus configuration. Even if we were able to theoretically show only convergence to a set, we can see here

that, in practice, we have convergence to a single point, even for non-constant curvature spaces. Second, the 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 true Fréchet mean. This error might be negligible in practical applications, but it is much greater than the achievable machine precision.

We include also two simulations (Fig. 3) for which the measurements are taken around a closed geodesic in $SO(3)$ and are far apart, i.e., $u \notin \mathcal{S}$ (see Thm. 5). With a linear network, the algorithm converges to a consensus configuration, as expected from Prop. 14. On the other hand, with a ring network, the algorithm gets trapped in a local minima and fails. The work [6] proposes strategies to avoid similar situations on the circle. It would be interesting to study if they could be extended to our general case.

VI. CONCLUSION

We proposed Riemannian consensus, a natural generalization of classical consensus algorithms to Riemannian manifolds. Our main contribution is finding sufficient conditions that guarantee convergence of the algorithm to a consensus configuration. These conditions depend on the curvature and topology of the manifold as well as the connectivity of the network.

APPENDIX

This appendix contains all the additional derivations and proofs for the claims in the paper. An expanded version of these results can be found in [17].

A. Additional Notation

In this section we review additional concepts and notation from Riemannian geometry. We focus only on those definitions and properties that are going to be applied in the remainder of this Appendix. We refer the reader to standard texts (e.g., [12], [14]) for the complete and precise definitions.

Following the notation introduced in Section II-B, let $(\mathcal{M}, \langle \cdot, \cdot \rangle)$ be a Riemannian manifold with its Riemannian metric. We denote the length of a curve $\gamma : [a, b] \rightarrow \mathcal{M}$ between two points $\gamma(a)$ and $\gamma(b)$ as $L(\gamma) = \int_a^b \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle^{1/2} dt$. We denote as ∇ the Levi-Civita connection on \mathcal{M} . In the following, we use the symbols $X(t)$, $Y(t)$ and $Z(t)$ to denote vector fields defined along a curve $\gamma(t)$. Unless necessary, we omit the dependency of curves and vector fields on the parameter t . Given X and Y , the metric compatibility property of ∇ implies $(d/dt)\langle X, Y \rangle = \langle \nabla X, Y \rangle + \langle X, \nabla Y \rangle$, where we use the notational convention $\nabla X = \nabla_{\dot{\gamma}} X$. The field X is said to be *parallel* if $\nabla X = 0$. In this case $X(t)$ is said to be the *parallel transport* of $X(0)$ from $\gamma(0)$ to $\gamma(t)$ along the curve, and we use the notation $X(t) = \tau_0^t X(0)$. The curve $\gamma(t)$ is said to be *geodesic* if it parallel transports its own tangent, i.e., $\nabla \dot{\gamma} = 0$.

The Riemannian curvature tensor R is defined as $R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z$, where X, Y and Z are smooth vector fields on \mathcal{M} . For brevity, we use the notational convention $R(X, Y, Z, W) = \langle R(X, Y)Z, W \rangle$. The curvature tensor has many symmetry properties. In particular, $R(X, Y, Z, W) = -R(Y, X, Z, W) = R(Z, W, X, Y)$. Therefore, $R(X, Y, Z, W) = 0$ whenever $X = Y$ or $Z = W$. Given a point $x \in \mathcal{M}$ and two linearly independent vectors $v, w \in T_x \mathcal{M}$ spanning a two-dimensional subspace $\sigma \subseteq T_x \mathcal{M}$, from the Riemannian curvature tensor one can define the sectional curvature for σ as $K_\sigma(x) = R(v, w, v, w) / (\|w\|^2 \|v\|^2 - \langle w, v \rangle^2)$.

We denote by \mathcal{M}_Δ a complete simply connected Riemannian manifold with constant curvature Δ and with the same dimension as \mathcal{M} . Also, we define the shorthand notation $\sin(\sqrt{\kappa}t) = s_\kappa(t)$, $\cos(\sqrt{\kappa}t) = c_\kappa(t)$, $\sinh(\sqrt{|\kappa|}t) = \text{sh}_\kappa(t)$, $\cosh(\sqrt{|\kappa|}t) = \text{ch}_\kappa(t)$.

A geodesic triangle $\Delta(x_1, x_2, x_3)$ in a Riemannian manifold \mathcal{M} is a figure formed by three distinct points x_1, x_2 and x_3 , called the vertices, that are connected by three minimal geodesics, called the sides. We denote as $\gamma_i(t)$ the side opposite to the vertex x_i and we denote its length as $l_i = L(\gamma_i)$. We indicate as $\beta_i = \angle x_i = \angle(x_j, x_i, x_k)$ the oriented angle between the tangent vectors of the two geodesics emanating from x_i . A geodesic hinge $(y; \gamma_1, \gamma_2)$ in \mathcal{M} is a figure formed by a point y and two minimal geodesics segments γ_1, γ_2 emanating from y .

Given a vector field X along a normal (i.e., unit speed) geodesic γ , we define its tangential and perpendicular components as $X^\parallel = \langle X, \dot{\gamma} \rangle \dot{\gamma}$ and $X^\perp = X - X^\parallel$, respectively.

TABLE I

LAW OF COSINES FOR GEODESIC TRIANGLES
IN MANIFOLDS OF CONSTANT CURVATURE Δ

$$\begin{array}{l|l} \kappa = 0 & l_i^2 = l_{i+1}^2 + l_{i+2}^2 - 2l_{i+1}l_{i+2} \cos \beta_i \\ \kappa > 0 & (l_i)^2 = (l_{i+1})(l_{i+2}) + (l_{i+1})(l_{i+2}) \cos \beta_i \\ \kappa < 0 & (l_i)^2 = (l_{i+1})(l_{i+2}) - (l_{i+1})(l_{i+2}) \cos \beta_i \end{array}$$

A smooth vector field Y along a geodesic γ is said to be a Jacobi field if it satisfies the second order differential equation $\nabla \nabla Y + R(Y, \dot{\gamma})\dot{\gamma} = 0$. Intuitively, Jacobi fields represent a variation of γ under a perturbation of the endpoints. In fact, it is known [12, Chapter 2, Lemma 2.4] that a Jacobi field is uniquely determined by fixing the value of Y at the two endpoints of γ . Moreover, if Y_1 and Y_2 are two Jacobi field along γ , then also $Y = Y_1 + Y_2$ is a Jacobi field along γ .

B. General Results

In this section we collect useful results that can be easily obtained from the existing literature.

Laws of Cosines: In manifolds with constant curvature Δ , the angles and sides of geodesic triangles are related by the *laws of cosines* in Table I. Using these laws it is possible to show the following Lemma on geodesic triangles in manifolds with constant curvature [12, page 138].

Lemma 18: Let $T = \Delta(x_1, x_2, x_3)$, $T' = \Delta(x'_1, x'_2, x'_3)$ be two geodesic triangles in \mathcal{M}_Δ . The side lengths for T and T' are denoted as l_i and l'_i , respectively, $i = 1, 2, 3$ and let $l_i = l'_i$, $i = 1, 2$. If $\Delta > 0$, assume also $l_2, l_3 < \pi/\sqrt{\Delta}$. Then $\angle x'_i > \angle x_i$ if and only if $l'_i > l_i$.

Comparison Theorems for Geodesic Triangles and Hinges: We start by reporting a hinge version of the Alexander-Toponogov theorem [13, Exercise IX.1].

Theorem 19: Given a complete Riemannian manifold \mathcal{M} with curvature bounded above by Δ and a geodesic triangle $\Delta(x_1, x_2, x_3)$ in \mathcal{M} , assume $l_1 + l_2 + l_3 < 2 \min\{\text{inj} \mathcal{M}, \pi/\sqrt{\Delta}\}$. Consider the hinge $(x_3; \gamma_1, \gamma_2)$ and let $(\tilde{x}_3; \tilde{\gamma}_1, \tilde{\gamma}_2)$ be a geodesic hinge in \mathcal{M}_Δ such that $L(\gamma_1) = L(\tilde{\gamma}_1)$, $L(\gamma_2) = L(\tilde{\gamma}_2)$ and $\angle x_3 = \angle \tilde{x}_3$. Then $d(\gamma_1(l_1), \gamma_2(l_2)) \geq d(\tilde{\gamma}_1(l_1), \tilde{\gamma}_2(l_2))$.

We will need the following triangle version of Thm. 19.

Theorem 20: For a geodesic triangle $\Delta(x_1, x_2, x_3)$ in \mathcal{M} suppose that γ_1 and γ_2 are minimal and the perimeter $l = l_1 + l_2 + l_3 < 2\pi/\sqrt{\Delta}$. Then, there exist a geodesic triangle $\Delta(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$ in \mathcal{M}_Δ with the same side lengths $L(\gamma_i) = L(\tilde{\gamma}_i)$ and satisfying $\angle x_3 \leq \angle \tilde{x}_3$.

Proof: In addition to the triangles $\Delta(x_1, x_2, x_3)$ in \mathcal{M} and $\Delta(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$ in \mathcal{M}_Δ defined in the statement, define a hinge $(\tilde{x}_3; \tilde{\gamma}_2, \tilde{\gamma}'_1)$ such that $L(\gamma_1) = L(\tilde{\gamma}'_1)$ and $\angle x_3 = \angle \tilde{x}_3$. Define $\tilde{x}'_2 = \tilde{\gamma}'_1(l_1)$. From Theorem 19 just given above, $d(\tilde{x}_1, \tilde{x}'_2) \leq d(x_1, x_2)$ and, by assumption, $d(x_1, x_2) = d(\tilde{x}_1, \tilde{x}_2)$, hence $d(\tilde{x}_1, \tilde{x}'_2) \leq d(\tilde{x}_1, \tilde{x}_2)$. Using Lemma 18, it follows that $\angle \tilde{x}'_3 = \angle \tilde{x}_1 \tilde{x}_3 \tilde{x}'_2 \leq \angle \tilde{x}_1 \tilde{x}_3 \tilde{x}_2 = \angle \tilde{x}_3$ and hence $\angle x_3 = \angle \tilde{x}'_3 \leq \angle \tilde{x}_3$. A similar argument can be repeated for the other points x_1 and x_2 . ■

Orthogonal Decomposition of Jacobi Fields: Let Y be a Jacobi field along a normal geodesic γ . The following proposition shows that Y can be decomposed in two orthogonal parts.

Proposition 21: A Jacobi field Y along a geodesic γ can be decomposed as $Y = Y^\perp + Y^\parallel$, where Y^\perp and Y^\parallel are Jacobi fields which are, respectively, perpendicular and tangential to γ .

Proof: The projection of Y along $\dot{\gamma}$ is a function of the form $\langle Y, \dot{\gamma} \rangle = at + b$, because

$$\begin{aligned} \frac{d^2}{dt^2} \langle Y(t), \dot{\gamma}(t) \rangle &= \frac{d}{dt} \langle \nabla Y(t), \dot{\gamma}(t) \rangle \\ &= \langle \nabla \nabla Y(t), \dot{\gamma}(t) \rangle = R(Y, \dot{\gamma}, \dot{\gamma}, \dot{\gamma}) = 0. \end{aligned} \quad (30)$$

In the above we used, in succession, the metric compatibility property of ∇ , the definitions of geodesic and Jacobi field, and the properties of the curvature tensor. The constants a and b can be determined using boundary conditions. Similar calculations show that $Y^\parallel = \langle Y, \dot{\gamma} \rangle \dot{\gamma}$ is in fact a Jacobi field. Therefore, $Y^\perp = Y - Y^\parallel$ is also a Jacobi field. ■

Comparison Theorems for Jacobi Fields: We now review versions of the Rauch Comparison Theorems based on the presentation in [13, pages 388–389].

Theorem 22 [Rauch Comparison Theorem I]: Let $X(t)$, $X(0) = 0$ be a Jacobi field along and orthogonal to a normal geodesic γ . If the curvature is bounded above by Δ , then

$$\langle \nabla X, X \rangle \geq \frac{C_\Delta}{S_\Delta} \|X\|^2, \quad \|\nabla X(0)\| \leq \frac{\|X\|}{S_\Delta} \quad (31)$$

for all t between zero and the first conjugate value of $\gamma(0)$. The functions C_Δ and S_Δ are given in (26).

Note that here and in the following we omit, for the sake of clarity, the explicit dependence of C_Δ and S_Δ on the distance t . The proof uses the following Lemma [13, p. 387].

Lemma 23: Let $X(t)$ be a vector field along a geodesic $\gamma : t \in [0, \beta] \rightarrow \mathcal{M}$. If $X(0)$ and $\nabla X(0)$ are linearly dependent, or if $X(0) = 0$, then $(d/dt)\|X\|(0) = \|\nabla X(0)\|$.

Proof of Theorem 22: The proof is simply an adaptation of Thm. IX.2.1 in [13] to our goals, where we identify $\eta = X$, $\psi = \|\nabla X(0)\|S_\Delta$ and $\delta = \Delta$. In particular, that theorem states that $(d/dt)(\|\eta\|/\psi) = (1/\psi^2)((d\|\eta\|/dt)\psi - \eta(d\psi/dt)) \geq 0$, which implies

$$\frac{d\|\eta\|}{dt} \psi - \|\eta\| \frac{d\psi}{dt} \geq 0 \implies \frac{d\|\eta\|}{dt} \geq \frac{d\psi}{dt} \implies \frac{\langle \nabla \eta, \eta \rangle}{\|\eta\|^2} \geq \frac{C_\Delta}{S_\Delta}. \quad (32)$$

With the above, the first equality of (31) follows by Lemma 23.

The results in [13] also state that $\|\eta\| \geq \psi$, which is equivalent to the second part of (31). ■

Theorem 24 [Rauch Comparison Theorem II]: Let $X(t)$, $X(0) = 0$, be a Jacobi field along and orthogonal to a normal geodesic γ . If the curvature is bounded below by δ , then

$$\langle \nabla X, X \rangle \leq \frac{C_\delta}{S_\delta} \|X\|^2, \quad \|\nabla X(0)\| \geq \frac{\|X\|}{S_\delta} \quad (33)$$

for all t between zero and the first conjugate value of $\gamma(0)$. The functions C_δ and S_δ are given in (26).

Proof: This theorem is a restatement of Thm. IX.2.2 in [13] with $\eta = X$, $\psi = \|\nabla X(0)\|S_\delta$ and $\kappa = \delta$. ■

C. Derivative of the Distance and Proof of Lemma 6

This section is devoted to build results on the derivative of the distance between two points moving on the sides of a geodesic hinge, with the final goal of providing a proof for Lemma 6. We

first obtain expressions in terms of angles between geodesics for general manifolds.

Let $x_1, x_2 \neq x_1$, and y be three points in \mathcal{M} such that $d_i = d(x_i, y)$ satisfies $0 < d_i < r^*$, $i = 1, 2$, where r^* is defined in (7). Define the geodesic hinge $(y; \gamma_1, \gamma_2)$, where the sides are defined by the conditions $\gamma_1(0) = \gamma_2(0) = y$, $\gamma_1(1) = x_1$ and $\gamma_2(1) = x_2$. For each value of t , $0 < t \leq 1$, define the minimal geodesic segment $\gamma_{12,t}(s)$ joining $\gamma_1(t)$ to $\gamma_2(t)$. Note that, since $d_i < r^*$, $i = 1, 2$, by the triangular inequality we have that $d(x_1, x_2) < \text{inj}_{x_1} \mathcal{M}$, therefore $\gamma_{12,t}(s)$ is uniquely defined for $t \in (0, 1 + \epsilon)$, where ϵ is small enough (so that $(1 + \epsilon)d_i < r^*$, $i = 1, 2$). Denote the length of the geodesic segment $\gamma_{12,t}$ by $\phi_{12}(t) = L(\gamma_{12,t})$, which is nothing but the distance between $\gamma_1(t)$ and $\gamma_2(t)$ for a specific t . Our goal is to show that the derivative of ϕ_{12}^2 is strictly positive on $t \in (0, 1]$. Notice that $\gamma_{12,t}$ is defined for $t \in (0, 1 + \epsilon)$, hence the derivative is well defined for $t = 1$.

The first step is to obtain an expression for $d\phi_{12}/dt$.

Proposition 25: For a given $t_0 \in (0, 1]$, consider the geodesic triangle $\Delta(y, \gamma_1(t_0), \gamma_2(t_0))$ and let β_i be the angle at $\gamma_i(t_0)$. Then $(d\phi_{12}/dt)|_{t=t_0} = d_1 \cos \beta_1 + d_2 \cos \beta_2$.

Proof: Let $d(x_1, x_2)$ be the distance function on \mathcal{M} . By the definition of gradient we have

$$\begin{aligned} \left. \frac{d\phi_{12}}{dt} \right|_{t=t_0} &= \langle \text{grad}_{x_1} d(\gamma_1(t_0), \gamma_2(t_0)), \dot{\gamma}_1(t_0) \rangle \\ &\quad + \langle \text{grad}_{x_2} d(\gamma_1(t_0), \gamma_2(t_0)), \dot{\gamma}_2(t_0) \rangle \\ &= \left\langle \frac{-\log_{\gamma_1(t_0)} \gamma_2(t_0)}{\left\| \log_{\gamma_1(t_0)} \gamma_2(t_0) \right\|}, \dot{\gamma}_1(t_0) \right\rangle \\ &\quad + \left\langle \frac{-\log_{\gamma_2(t_0)} \gamma_1(t_0)}{\left\| \log_{\gamma_2(t_0)} \gamma_1(t_0) \right\|}, \dot{\gamma}_2(t_0) \right\rangle. \end{aligned} \quad (34)$$

Considering that $\|\dot{\gamma}_i(t)\| = d_i$, the claim follows. ■

The next step is to consider the particular case of manifolds with constant curvature $\Delta \geq 0$ (for our purposes, the case $\Delta < 0$ will be covered by the case $\Delta = 0$).

Proposition 26: Let \mathcal{M} be of constant curvature $\Delta \geq 0$. Using the same definitions given at the beginning of the section, we have $(d\phi_{12}/dt)(t) > 0$ for $t \in (0, 1]$.

Proof: Let $l_i(t) = d_i t$, $i = 1, 2$. In the case $\Delta = 0$, from the cosine law in Table I, we have $\phi_{12}(t) = |t| \sqrt{d_1^2 + d_2^2 - 2d_1 d_2 \cos \beta_3}$. The claim then easily follows. For the case $\Delta > 0$, as argued before, the triangular inequality implies $d(\gamma_1(t), \gamma_2(t)) < \pi/\sqrt{\Delta}$. In turn, this means that $\sin(\sqrt{\Delta}\phi_{12}) > 0$. Instead of the derivative of $\phi_{12}(t)$, it is more convenient to use the derivative $(d/dt) \cos(\sqrt{\Delta}\phi_{12}) = -\sin(\sqrt{\Delta}\phi_{12})(d\phi_{12}/dt)$. Note that $d\phi_{12}/dt > 0$ if and only if $-(d/dt) \cos(\sqrt{\Delta}\phi_{12}) > 0$, hence the two expressions are equivalent for our purposes.

Using the cosine law for $\Delta > 0$, we get

$$\begin{aligned} &-\frac{d}{dt} \cos(\sqrt{\Delta}\phi_{12}) \\ &= -\frac{d}{dt} (c_\Delta(l_1(t)) c_\Delta(l_2(t)) + s_\Delta(l_1(t)) s_\Delta(l_2(t)) \cos \beta_i) \\ &= \sqrt{\Delta} ((d_1 - d_2 \cos(\alpha)) s_\Delta(d_1 t) c_\Delta(d_2 t) \\ &\quad + (d_2 - d_1 \cos(\alpha)) c_\Delta(d_1 t) s_\Delta(d_2 t)). \end{aligned} \quad (35)$$

Assume, without loss of generality, $d_1 > d_2$ (if not, just swap the indexes throughout) and recall that $0 < t < \pi/2\sqrt{\Delta}d_1$. This implies that $s_\Delta(d_1t), c_\Delta(d_1t) > 0$ for $i = 1, 2$. Now, the condition $-(d/dt) \cos(\sqrt{\Delta}\phi_{12}) > 0$ can be written as

$$(d_1 - d_2 \cos(\alpha)) s_\Delta(d_1t) c_\Delta(d_2t) + (d_2 - d_1 \cos(\alpha)) c_\Delta(d_1t) s_\Delta(d_2t) > 0 \quad (36)$$

$$\frac{s_\Delta(d_1t) c_\Delta(d_2t)}{c_\Delta(d_1t) s_\Delta(d_2t)} > \frac{d_1 \cos(\alpha) - d_2}{d_1 - d_2 \cos(\alpha)}. \quad (37)$$

At this point, note that the RHS is always less or equal to one. Therefore, sufficient conditions for $-(d/dt) \cos(\sqrt{\Delta}\phi_{12}) > 0$ are

$$\begin{aligned} \frac{s_\Delta(d_1t) c_\Delta(d_2t)}{c_\Delta(d_1t) s_\Delta(d_2t)} > 1 &\implies \frac{s_\Delta(d_1t)}{c_\Delta(d_1t)} > \frac{s_\Delta(d_2t)}{c_\Delta(d_2t)} \\ &\implies \tan(\sqrt{\Delta}d_1t) > \tan(\sqrt{\Delta}d_2t). \end{aligned} \quad (38)$$

Due to the monotonicity properties of the tan function, this condition is always true under the assumptions above, i.e., $d_2 < d_1 < r^*$ and $t \in (0, 1]$. In other words, $-(d/dt) \cos(\sqrt{\Delta}\phi_{12}) > 0$, hence $d\phi_{12}/dt > 0$ and the claim follows. ■

We have now all the elements necessary to prove Lemma 6.

Proof of Lemma 6: We first consider the case where the three points are all distinct. Notice that showing $(d/dt)(\phi_{12}^2/2) = \phi_{12}(d\phi_{12}/dt) > 0$ is equivalent to showing $d\phi_{12}/dt > 0$. For any $t_0 \in (0, 1]$ consider the geodesic triangle $T = \Delta(y, \gamma_1(t_0), \gamma_2(t_0))$. Build a triangle $T_\Delta = \Delta(\tilde{y}, \tilde{x}_1, \tilde{x}_2)$ in \mathcal{M}_Δ having the same side lengths as T . Define the geodesics $\tilde{\gamma}_i(t) : t \rightarrow \mathcal{M}_\Delta$ such that $\tilde{\gamma}_i(0) = \tilde{y}$ and $\tilde{\gamma}_i(t_0) = \tilde{x}_i$, $i = 1, 2$. Define also $\tilde{\phi}_{12}(t) = d(\gamma_1(t), \gamma_2(t))$. Let $\beta_i = \angle \gamma_i(t_0)$ and $\tilde{\beta}_i = \angle \tilde{\gamma}_i(t_0)$, $i = 1, 2$. According to Thm. 20, $\beta_i \leq \tilde{\beta}_i$, $i = 1, 2$. Using Prop. 25 and Prop. 26 (if $\Delta < 0$, use $\Delta = 0$), this implies $d\phi_{12}/dt \geq d\tilde{\phi}_{12}/dt > 0$, and the claim is shown. Next, consider the case $x_2 = y$, $x_1 \neq x_2$. Then $\phi_{12} = d^2(x_1, y)t$ and the claim follows directly. The same applies by swapping the roles of x_1 and x_2 . Finally, if $x_1 = x_2 = y$, then trivially $d\phi_{12}/dt \equiv 0$. ■

D. Bounds on the Hessian of the Distance

In this section we compute and give bounds on the second derivative of the distance (and distance squared) between two points moving on geodesics. First, we derive a general expression that depends only on the relative velocities and angles between geodesics. Then, we compute concrete bounds for the case of manifolds with bounded sectional curvature. We refer to [17] for the case with constant curvature. From these bounds, we can then obtain the bound on the Hessian of the squared distance, which is used in Prop. 9.

The General Case: Define two geodesics $\gamma_1, \gamma_2 : (-\epsilon, \epsilon) \rightarrow \mathcal{M}$ such that there exist a minimal geodesic $\gamma_{12,t}(s)$ joining $\gamma_1(t)$ to $\gamma_2(t)$ for all $t \in (-\epsilon, \epsilon)$. Using the same notation as in Appendix C, we denote the length of the geodesic segment $\gamma_{12,t}$ by $\phi_{12}(t) = L(\gamma_{12,t})$, which is the distance between $\gamma_1(t)$ and $\gamma_2(t)$ for a specific t . In this section we will find bounds on the second derivative of $\phi_{12}(t)$ around $t_0 = 0$.

Define the geodesic variation $\alpha : [0, 1] \times [a, b] \rightarrow \mathcal{M}$, such that the map $s \mapsto \alpha(t_0, s)$ traces the geodesic $\gamma_{12,t_0}(s)$,

$\|\dot{\gamma}_{12,t}\| = 1$. Define $\partial_s \alpha = (D\alpha)\partial_s$ and $\partial_t \alpha = (D\alpha)\partial_t$, where ∂_s (resp., ∂_t) denotes the partial derivation operator with respect to the variable s (resp., t). Since $\alpha(t, s)$ traces geodesics, the vector field $X(s) = \partial_s \alpha|_{t=t_0}$ is a Jacobi field [12, page 36].

We then have the following theorem.

Theorem 27: Using the notation above, we have

$$\left. \frac{d^2 \phi_{12}}{dt^2} \right|_{t=t_0} = \langle \nabla X(s)^\perp, X(s)^\perp \rangle \Big|_0^l \quad (39)$$

$$\begin{aligned} \left. \frac{d^2 \phi_{12}^2}{dt^2} \right|_{t=t_0} &= \left(\left\langle X(s), \frac{\gamma_{12,t_0}(s)}{\|\gamma_{12,t_0}(s)\|} \right\rangle \Big|_0^l \right)^2 \\ &\quad + l \langle \nabla X(s)^\perp, X(s)^\perp \rangle \Big|_0^l \end{aligned} \quad (40)$$

where $l = \phi_{12}(0) < 2r^*$.

Proof: From [13, page 76] we get

$$\begin{aligned} \left. \frac{d^2 \phi_{12}}{dt^2} \right|_{t=t_0} &= \langle \nabla_{\partial_t} X, \dot{\gamma}_{12,t} \rangle \Big|_a^b \\ &\quad + \int_a^b \left(\|\nabla X^\perp\|^2 - R(\dot{\gamma}_{12,t}, X^\perp, \dot{\gamma}_{12,t}, X^\perp) \right) ds. \end{aligned} \quad (41)$$

Then, notice that since X^\perp is a Jacobi field (Prop. 21), $-R(\dot{\gamma}_{12,t}, X^\perp, \dot{\gamma}_{12,t}, X^\perp) = \langle \nabla \nabla X^\perp, X^\perp \rangle = (d/dt) \langle \nabla X^\perp, X^\perp \rangle - \langle \nabla X^\perp, \nabla X^\perp \rangle = (d/dt) \langle \nabla X^\perp, X^\perp \rangle - \|\nabla X^\perp\|^2$. Hence

$$\left. \frac{d^2 \phi_{12}}{dt^2} \right|_{t=t_0} = \int_a^b \frac{d}{dt} \langle \nabla X^\perp, X^\perp \rangle ds = \langle \nabla X(s)^\perp, X(s)^\perp \rangle \Big|_0^l \quad (42)$$

which is (39). Equation (40) follows from the chain rule. ■

Notice that the second derivative of ϕ_{12} depends only on the orthogonal component of the Jacobi field, X^\perp . Therefore, any two pairs of geodesics having the same $X^\perp(0) = \dot{\gamma}_1(t_0)^\perp$ and $X^\perp(l) = \dot{\gamma}_2(t_0)^\perp$, will have the same orthogonal Jacobi field and will yield the same second derivative of the distance ϕ_{12} . However, the tangential components of $\dot{\gamma}_1(t_0)$ and $\dot{\gamma}_2(t_0)$ play a role in the second derivative of the squared distance.

Manifolds With Bounded Curvature: In this section we give bounds on the second derivative of the squared distance in terms of the curvature bounds Δ and δ . In particular, we show the following.

Theorem 28: Define two geodesics $\gamma_1, \gamma_2 : (-\epsilon, \epsilon) \rightarrow \mathcal{M}$ such that $\gamma_2(t) \in \mathcal{B}_\mathcal{M}(\gamma_1(t), \text{inj}\mathcal{M})$ for all $t \in (-\epsilon, \epsilon)$. Let $\phi_{12}(t) = d(\gamma_1(t), \gamma_2(t))$ and define $l = \phi_{12}(0)$. Then

$$\left. \frac{d^2 \phi_{12}}{dt^2} \right|_{t=t_0} \leq \left(\frac{C_\delta(l)}{S_\delta(l)} + \frac{1}{S_\Delta(l)} \right) \left(\|\dot{\gamma}_1^\perp(0)\|^2 + \|\dot{\gamma}_2^\perp(0)\|^2 \right) \quad (43)$$

and

$$\left. \frac{d^2 \phi_{12}^2}{dt^2} \right|_{t=t_0} \leq \mu_{max}^d(l) \left(\|\dot{\gamma}_1(0)\|^2 + \|\dot{\gamma}_2(0)\|^2 \right) \quad (44)$$

where $\mu_{max}^d(l) = \max\{2, l((C_\delta(l)/S_\delta(l)) + (1/S_\Delta(l)))\}$.

By Def. 1, $\mu_{max}^d(l)$ is a bound on the Hessian of the squared distance evaluated at $(\gamma_1(0), \gamma_2(0))$. These bounds are sharp, in the sense that if $\delta = \Delta = \kappa$, we obtain the same bounds as in the constant curvature case [17].

Proof: We start from (39). Since it is not easy to express ∇X in terms of X , we will use bounds. We decompose the Jacobi field X^\perp in two components $X^\perp = X_1 + X_2$ where X_1 and X_2 are Jacobi fields satisfying the conditions $X_1(0) = 0$, $X_1(l) = X(l)^\perp$, $X_2(0) = X(0)^\perp$ and $X_2(l) = 0$. The fields X_1 and X_2 vanish at one of the endpoints, and we can exploit results from standard Riemannian geometry texts. Note that since $\gamma_2(t) \in \mathcal{B}_{\mathcal{M}}(\gamma_1(t), \text{inj}\mathcal{M})$, we have $l < 2r^*$ and, by the Morse-Schönberg Theorem [13, p. 86] there are no conjugate point pairs on γ_{12} . We can therefore apply the Rauch comparison theorems from Appendix B. More concretely, we can break down (39) as

$$\left. \frac{d^2 \phi_{12}}{dt^2} \right|_{t=t_0} = \langle \nabla X_1(l), X_1(l) \rangle + \langle \nabla X_2(l), X_1(l) \rangle - \langle \nabla X_1(0), X_2(0) \rangle - \langle \nabla X_2(0), X_2(0) \rangle \quad (45)$$

We now bound each term. Using Thm. 24, we have

$$\begin{aligned} \langle \nabla X_1(l), X_1(l) \rangle &\leq \frac{C_\delta(l)}{S_\delta(l)} \|X(l)^\perp\|^2, \\ \langle -\nabla X_2(0), X_2(0) \rangle &\leq \frac{C_\delta(l)}{S_\delta(l)} \|X(0)^\perp\|^2. \end{aligned} \quad (46)$$

Note that for X_2 , in order to apply Thm. 24, we need to reverse the parametrization of $\gamma_{12,t_0}(s)$ as $s' = l - s$. This has the effect that $\nabla X_2(s')|_{s'=l} = \nabla_{-\dot{\gamma}_{12,t_0}(s)} X_2(s)|_{s=0} = -\nabla X_2(0)$. This explains the negative sign in the second inequality of (46).

Using the Cauchy-Schwarz inequality, Thm. 22 and the inequality $ab \leq (a^2 + b^2)/2$, we have

$$\begin{aligned} \langle \nabla X_2, X_1^\perp(l) \rangle &\leq \frac{1}{S_\Delta(l)} \|X_2(0)\| \|X_1(l)\| \\ &\leq \frac{1}{2S_\Delta(l)} \left(\|X(0)^\perp\|^2 + \|X(l)^\perp\|^2 \right) \quad (47) \\ -\langle \nabla X_1(0), X_2(0) \rangle &\leq \frac{1}{S_\Delta(l)} \|X_1(l)\| \|X_2(0)\| \\ &\leq \frac{1}{2S_\Delta(l)} \left(\|X(0)^\perp\|^2 + \|X(l)^\perp\|^2 \right). \quad (48) \end{aligned}$$

Combining (46), (47) and (48) into (45), and recalling that $X(0) = \dot{\gamma}_1(0)$, $X(l) = \dot{\gamma}_2(0)$, we obtain (43). Combining this with (39), we have

$$\begin{aligned} \left. \frac{d^2 \phi_{12}^2}{dt^2} \right|_{t=t_0} &\leq \left(\|X(0)^\perp\| + \|X(l)^\perp\| \right)^2 \\ &+ l \left(\frac{C_\delta(l)}{S_\delta(l)} + \frac{1}{S_\Delta(l)} \right) \left(\|X(0)^\perp\|^2 + \|X(l)^\perp\|^2 \right) \quad (49) \end{aligned}$$

which, after using the inequality $ab \leq a^2 + b^2/2$, becomes (44). ■

REFERENCES

[1] R. Tron, R. Vidal, and A. Terzis, "Distributed pose averaging in camera networks via consensus on SE(3)," in *Proc., Int. Conf. Distrib. Smart Cameras*, 2008, pp. 1–10.

[2] R. Tron and R. Vidal, "Distributed image-based 3-D localization in camera sensor networks," in *Proc. Conf. Decision Control*, 2009, pp. 901–908.

[3] A. Sarlette, S. Bonnabel, and R. Sepulchre, "Coordinated motion design on Lie groups," *IEEE Trans. Autom. Control*, vol. 55, no. 5, pp. 1047–1058, May 2010.

[4] R. Olfati-Saber, "Swarms on sphere: A programmable swarm with synchronous behaviors like oscillator networks," in *Proc. IEEE Conf. Decision Control*, 2006, pp. 5060–5066.

[5] L. Scardovi, A. Sarlette, and R. Sepulchre, "Synchronization and balancing on the N -torus," *Syst. Control Lett.*, vol. 56, no. 5, pp. 335–341, 2007.

[6] A. Sarlette and R. Sepulchre, *The Complexity of Dynamical Systems: A Multi-Disciplinary Perspective, Chapter Synchronization on the Circle*. New York: Wiley, 2011.

[7] A. Sarlette and R. Sepulchre, "Consensus optimization on manifolds," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 56–76, 2009.

[8] T. Hatanaka, M. Fujita, and F. Bullo, "Vision-based cooperative estimation via multi-agent optimization," in *Proc. IEEE Conf. Decision Control*, 2010, pp. 2492–2497.

[9] Y. Igarashi, T. Hatanaka, M. Fujita, and M. W. Spong, "Passivity-based attitude synchronization in SE(3)," *IEEE Trans. Control Syst. Technol.*, vol. 17, no. 5, pp. 1119–1134, May 2009.

[10] A. Olshevsky and J. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 33–55, 2007.

[11] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 3, pp. 1520–1533, Mar. 2004.

[12] T. Sakai, *Riemannian Geometry*. New York: American Mathematical Society, 1996, vol. 149.

[13] I. Chavel, *Riemannian Geometry: A Modern Introduction*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2006, vol. 98.

[14] M. P. do Carmo, *Riemannian Geometry*. Boston, MA: Birkhäuser, 1992.

[15] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton Univ. Press, 2008.

[16] A. Edelman, T. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[17] R. Tron, "Distributed Optimization on Manifolds for Consensus Algorithms and Camera Network Localization," Ph.D. dissertation, The Johns Hopkins University, Baltimore, MD, 2012.

[18] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Boston, MA: Athena Scientific, 1999, vol. 2, Optimization and computation.

[19] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds*. Norwell, MA: Kluwer, 1994, vol. 297, Mathematics and Applications.

[20] B. Afsari, "Riemannian L^p center of mass: Existence, uniqueness, convexity," *Proc. AMS*, vol. 139, no. 2, pp. 655–673, 2011.

[21] B. Afsari, R. Tron, and R. Vidal, "On the convergence of gradient descent for locating the Riemannian center of mass," *SIAM J. Control Optim.*, to be published.

[22] J. Cortés, "Distributed algorithms for reaching consensus on general functions," *Automatica*, vol. 44, no. 3, pp. 726–737, 2008.



Roberto Tron (M'12) received the B.Sc. and M.Sc. degrees (with highest honors) from the Politecnico di Torino, Torino, Italy, in 2004 and 2007, respectively, the Diplôme d'Ingénieur from the Eurecom Institute, Biot, France, the DEA degree from the Université de Nice Sophia-Antipolis, in 2006, and is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD.

His research interests include motion segmentation and distributed algorithms on camera sensor networks.

Mr. Tron received the "General Chair's Interactive Presentation Recognition Award" (2009) and the "Best Student Paper Award Runner-up" (2011) from the IEEE Conference on Decision and Control.



Bijan Afsari (M'07) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1998 and the M.S. and Ph.D. degrees in electrical engineering and applied mathematics from the University of Maryland, College Park, in 2004 and 2009, respectively.

In the meantime, he experienced working in the financial industry. His main interest lies around geometrical methods for data analysis and computation, especially applications of Riemannian and non-Euclidean geometries. He is currently a research fellow in the Center for Imaging Science, Johns Hopkins University, Baltimore, MD, where he is primarily focused on geometrization of the spaces of linear dynamical systems for applications in the analysis of video sequence data.



René Vidal (SM'12) received the B.S. degree in electrical engineering (with highest honors) from the Pontificia Universidad Católica de Chile in 1997 and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California at Berkeley, in 2000 and 2003, respectively.

He was a Research Fellow at the National ICT Australia in the Fall of 2003 and has been a faculty member in the Department of Biomedical Engineering and the Center for Imaging Science of The Johns Hopkins University since 2004. He has held several visiting faculty positions at Stanford University, ENS Paris, the Catholic University of Chile, Université Henri Poincaré, Heriot Watt University, and the Australian National University. He was Associate Editor of the *SIAM Journal on Imaging Sciences* and the *Journal of Mathematical Imaging and Vision*. He was co-editor of the book *Dynamical Vision* and has co-authored more than 150 articles in computer vision, machine learning, biomedical image analysis, signal processing, hybrid systems, and robotics.

Dr. Vidal received the 2012 J. K. Aggarwal Prize “for outstanding contributions to the generalized principal component analysis (GPCA) and subspace clustering,” the 2011 Best Paper Award Finalist at the Conference on Decision and Control, the 2009 ONR Young Investigator Award, the 2009 Sloan Research Fellowship, the 2005 NFS CAREER Award, the 2004 Best Paper Award Honorable Mention at the European Conference on Computer Vision, the 2004 Sakrison Memorial Prize for “completing an exceptionally documented piece of research,” the 2003 Eli Jury award for “outstanding achievement in the area of Systems, Communications, Control, or Signal Processing,” the 2002 Student Continuation Award from NASA Ames, the 1998 Marcos Orrego Puelma Award from the Institute of Engineers of Chile, and the 1997 Award of the School of Engineering of the Pontificia Universidad Católica de Chile to the best graduating student of the school. He is Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI). He was or will be Program Chair for ICCV 2015, CVPR 2014, WMVC 2009 and PSIVT 2007, and area chair for ICCV 2013, CVPR 2013, ICCV 2011, ICCV 2007 and CVPR 2005. He is a member of the ACM.