Consensus seeking gradient descent flows on boundaries of convex sets

Johan Markdahl

arXiv:1912.08170v2 [math.OC] 1 Apr 2020

***Abstract*— Consensus on nonlinear spaces is of use in many control applications. This paper proposes a gradient descent flow algorithm for consensus on hypersurfaces. We show that if an inequality holds, then the system converges for almost all initial conditions and all connected graphs. The inequality involves the hypersurface Gauss map and the gradient and Hessian of the implicit equation. Moreover, for the inequality to hold, it is necessary that the manifold is the boundary of a convex set. The literature already contains an algorithm for consensus on hypersurfaces. That algorithm on any ellipsoid is equivalent to our algorithm on the unit sphere. In particular, that algorithm achieves almost global synchronization on ellipsoids. These findings suggest that strong convergence results for consensus seeking gradient descent flows may be established on manifolds that are the boundaries of convex sets.**

I. Introduction

Consensus on nonlinear spaces is of interest in many application areas including robotics [1], flocking [2], opinion dynamics [3], machine learning [4], and quantum synchronization [5]. The problem of almost global consensus on nonlinear spaces is interesting from an applied point of view since it makes the probability of reaching consensus from a random initial condition independent of the number of agents. It is also interesting from a theoretical perspective since the global geometry and topology is what differentiates a Riemannian manifold $\mathcal{M}$ from Euclidean space $\mathbb{R}^m$. This paper explores how a consensus seeking gradient descent flow algorithm being almost globally convergent depends on the geometry and topology of the manifold it evolves on.

Consider a consensus seeking gradient descent flow of a disagreement function on a manifold $\mathcal{M}$ [6]. Global convergence results are known for some special cases. For example, the consensus manifold $\mathcal{C}$ is almost globally asymptotically stable (AGAS) for all connected networks over spheres of dimension $n \geq 2$ [7]. There is hence at least one AGAS consensus protocol on the boundary of every compact, star-shaped set for $n \geq 2$, obtained by lifting the protocol on the corresponding n-sphere to $\mathcal{M}$. Another example; a necessary condition states that $\mathcal{C}$ cannot be AGAS if $\mathcal{M}$ is simply connected [8]. Since the boundary of any compact convex set of dimension n in $\mathbb{R}^m$ is homeomorphic to the n-sphere in $\mathbb{R}^{n+1}$, the boundaries of compact convex sets are simply connected for $n \geq 2$. Moreover, the boundary of any compact convex set can be described as a hypersurface in $\mathbb{R}^{n+1}$.

Working with arbitrary manifolds is difficult. Like spheres, hypersurfaces can be characterized by a single constraint, wherefore the methodology of [7], [9] can be applied. For technical reasons we limit consideration to closed analytic hypersurfaces, *i.e.*, hypersurfaces that are analytic, compact and without boundaries. This paper provides a sufficient condition for $\mathcal{C}$ to be AGAS for all connected networks over such hypersurfaces. The condition can only be satisfied if the hypersurface is the boundary of a convex set. However, because the condition is based on a quadratic Taylor expansion of the disagreement function, it cannot be used for manifolds on which the quadratic term vanishes. The question concerning the boundary of any convex set hence remains unresolved.

There is another algorithm for consensus on hypersurfaces in the literature [10]–[12]. Compared to our algorithm, it is more restricted in terms of the hypersurfaces it can be applied to. We show that this algorithm on any ellipsoid is equivalent to our algorithm on the unit sphere. Almost global convergence on ellipsoid have been established for graphs that are either complete or acyclic [11]. The result for our algorithm on networks over the sphere [7] also applies to the algorithm [11] on networks over ellipsoids, showing that it converges almost globally for all connected graphs.

II. Preliminaries

The boundary $\partial\mathcal{S}$ of any compact convex set $\mathcal{S} \subset \mathbb{R}^m$, $\dim \partial\mathcal{S} = n < m$, can be transformed into a hypersurface in $\mathbb{R}^{n+1}$ by a change of coordinates. For technical reasons we focus on closed analytic hypersurfaces. A closed analytic hypersurface $\mathcal{M} \subset \mathbb{R}^{n+1}$ can without loss of generality be characterized as a set on the form

$$\mathcal{M} = \{\mathbf{y} \in \mathbb{R}^{n+1} \,|\, c(\mathbf{y}) = 0\},$$

where $c : \mathbb{R}^{n+1} \to \mathbb{R}$ is an analytic function. The Jordan-Brouwer theorem implies that $\mathcal{M}$ separates the set on which c is positive from the set on which c is negative [13]. One of the sets is bounded while the other is unbounded. If the gradient $\nabla c(\mathbf{y})$ is considered as a vector located at $\mathbf{y}$, then it points towards the region on which c is positive. This set can be assumed to be unbounded without loss of generality.

A hypersurface $\mathcal{M}$ is called nonsingular if $\nabla c(\mathbf{y}) \neq \mathbf{0}$ for all $\mathbf{y} \in \mathcal{M}$. Assume that $\mathcal{M}$ is nonsingular. Let

$$\mathbf{n}(\mathbf{y}) = \tfrac{\nabla c(\mathbf{y})}{\|\nabla c(\mathbf{y})\|}$$

denote the unit normal obtained from the Gauss map $\mathbf{n} : \mathcal{M} \to \mathcal{S}^n$. The projection $\Pi : \mathbb{R}^n \to \mathsf{T}_{\mathbf{y}}\mathcal{M}$ on the tangent space of $\mathcal{M}$ at $\mathbf{y}$ is given by

$$\mathbf{z} \mapsto (\mathbf{I}_{n+1} - \mathbf{n}(\mathbf{y})\mathbf{n}^\top(\mathbf{y}))\,\mathbf{z},$$

This work is supported by the UL project OptBioSys.

J. Markdahl is with the Luxembourg Centre for Systems Biomedicine at the University of Luxembourg. Email: markdahl@kth.se

where the Gram-Schmidt rule $\mathbf{n} \perp \mathbf{z} - \langle \mathbf{z}, \mathbf{n} \rangle \mathbf{n}$ is used to cancel the normal component of $\mathbf{z}$. This expression allows us to calculate the gradient $\Pi \nabla f(\mathbf{y})$ of $f(\mathbf{y})$ on $\mathcal{M}$ as

$$\Pi \nabla f(\mathbf{y}) = (\mathbf{I}_{n+1} - \tfrac{\nabla c(\mathbf{y})}{\|\nabla c(\mathbf{y})\|} (\tfrac{\nabla c(\mathbf{y})}{\|\nabla c(\mathbf{y})\|})^\top) \nabla f(\mathbf{y}),$$

where $\nabla f(\mathbf{y})$ is the Euclidean gradient of $f(\mathbf{y})$ in $\mathbb{R}^{n+1}$.

A set is almost globally asymptotically stable if almost all system trajectories converge to it:

*Definition 1 (*AGAS*):* A Lyapunov stable equilibrium set $\mathcal{S}$ of a dynamical system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ on a Riemannian manifold $(\mathcal{M}, g)$, where $\mathcal{M} \subset \mathbb{R}^{n+1}$, is said to be AGAS if $\lim_{\mathbf{x} \to \infty} \mathbf{x}(t) \in \mathcal{S}$ for all $\mathbf{x}(0) \in \mathcal{M} \backslash \mathcal{N}$, where $\mathcal{N}$ has Riemannian measure zero.

III. Distributed control design

We use a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to model interactions between agents. Each node $i \in \mathcal{V}$ corresponds to an agent and each edge $\{i, j\} \in \mathcal{E}$ corresponds to a pair of communicating agents. The graph is assumed to be connected. Items associated with agent i carry the subindex i; we denote the state of agent i by $\mathbf{x}_i \in \mathcal{M}$, the normal of $\mathcal{M}$ at $\mathbf{x}_i$ by $\mathbf{n}_i$, the projection onto the tangent space of $\mathcal{M}$ at $\mathbf{x}_i$ by Π_i, the neighbor set of agent i by $\mathcal{N}_i = \{j \in \mathcal{V} \,|\, \{i, j\} \in \mathcal{E}\}$, the Euclidean gradient of V with respect to $\mathbf{x}_i$ by $\nabla_i V$ *etc.* We call $\mathbf{x} = (\mathbf{x}_i)_{i=1}^N \in \mathcal{M}^N$ a configuration of agents.

Consider a dynamical system defined on $\mathcal{M}^N$. The dynamics of agent i could *e.g.*, be

$$\dot{\mathbf{x}}_i = \mathbf{u}_i \tag{1}$$

where $\mathbf{u}_i \in \mathsf{T}_i \mathcal{M}$ is the control signal. Another option is

$$\dot{\mathbf{x}}_i = \Pi_i \mathbf{u}_i = (\mathbf{I}_{n+1} - \mathbf{n}_i \mathbf{n}_i^\top) \mathbf{u}_i \tag{2}$$

where $\mathbf{u}_i \in \mathbb{R}^{n+1}$ and $\mathbf{n}_i = \mathbf{n}(\mathbf{x}_i)$ is introduced for the sake of notational convenience. Note that the right-hand sides of (1) and (2) belong to $\mathsf{T}_{\mathbf{x}_i} \mathcal{M}$. Suppose that $\mathbf{x}_i(0) \in \mathcal{M}$, that $\mathbf{u}_i$ is Lipschitz, and that $\mathcal{M}$ is a C^2 manifold. Then $\mathbf{x}_i(t) \in \mathcal{M}$ for all $t \in \mathbb{R}$ by the Bony-Brezis theorem since $\langle \mathbf{v}, \nabla_i c(\mathbf{x}_i) \rangle = 0$ for all $\mathbf{v} \in \mathsf{T}_{\mathbf{x}_i} \mathcal{M}$ and all $\mathbf{x}_i \in \mathcal{M}$. By confining $\dot{\mathbf{x}}_i$ to $\mathsf{T}_i \mathcal{M}$, we confine $\mathbf{x}_i$ to $\mathcal{M}$.

The input model (1) corresponds to a situation where the constraint $\mathbf{x}_i \in \mathcal{M}$ is adopted to accomplish a task whereas the model (2) refers to the case where the mechanical design of a systems constrains it to only be actuated in a certain fashion. An example of (1) is a team of satellites in orbit; they could leave the orbit if so desired. Examples of (2) include camera sensor networks where each camera is mounted on a spherical joint. The orientation of camera i is always some $\mathbf{x}_i \in \mathcal{S}^2$ regardless of the control input.

The goal of consensus seeking systems is for the agents to asymptotically approach the consensus manifold

$$\mathcal{C} = \{(\mathbf{x})_{i=1}^N \in \mathcal{M}^N\}. \tag{3}$$

The set $\mathcal{C}$ is a manifold $\mathcal{C} \simeq \mathcal{M}$ by the diffeomorphism $\mathcal{M}^N \to \mathcal{C} : (\mathbf{x})_{i=1}^N \mapsto \mathbf{x}$. If the agents are satellites in orbit that satisfy $\mathbf{x}_i \in \mathcal{S}^2$, then this would be interpreted as all N agents meeting up at one point. If the agents are rigid bodies whose pointing direction (reduced attitude) is modelled as $\mathbf{x}_i \in \mathcal{S}^2$, then a consensus implies that all N bodies are pointing in the same direction.

As a measure of the distance to consensus, consider the disagreement function $V : \mathcal{M}^N \to \mathbb{R}$ given by

$$V(\mathbf{x}) = \tfrac{1}{2} \sum_{i,j \in \mathcal{E}} a_{ij} \|\mathbf{x}_j - \mathbf{x}_i\|^2, \tag{4}$$

where $a_{ij} \in [0, \infty)$. Clearly, $V = 0$ if and only if $\mathbf{x} = (\mathbf{x}_i)_{i=1}^N \in \mathcal{C}$, *i.e.*, no disagreement. The consensus seeking algorithm that we study in this paper is the gradient descent flow of (4). The gradient of V on $\mathcal{M}$ is given by

$$\Pi_i \nabla_i V = (\mathbf{I}_n - \mathbf{n}_i \mathbf{n}_i^\top) \sum_{j \in \mathcal{N}_i} a_{ij} (\mathbf{x}_i - \mathbf{x}_j).$$

We are now ready to state the main algorithm of this paper. This algorithm previously appears in [6], although their work is limited to the case when the norm of the states are constant, $\|\mathbf{x}_i\| = k$, *i.e.*, the case when $\mathcal{M}$ is a sphere. Moreover, they only show local stability results.

Algorithm 2: The consensus seeking gradient descent flow on $\mathcal{M}$ is given by

$$\dot{\mathbf{x}}_i = -\Pi_i \nabla_i V = (\mathbf{I} - \mathbf{n}_i \mathbf{n}_i^\top) \sum_{j \in \mathcal{N}_i} a_{ij} (\mathbf{x}_j - \mathbf{x}_i). \tag{5}$$

Suppose that $\mathcal{M}$ is closed, then the solution $\mathbf{x}(t) = (\mathbf{x}_i)_{i=1}^N$ to (5) is unique and exists for all $t \in \mathbb{R}$ [14].

There is another algorithm for consensus on hypersurfaces in the literature:

Algorithm 3 (Zhu [11]): The consensus seeking algorithm on $\mathcal{M}$ is given by

$$\begin{aligned} \dot{\mathbf{x}}_i &= \left(\mathbf{I} - \tfrac{\mathbf{x}_i \nabla c(\mathbf{x}_i)^\top}{\langle \mathbf{x}_i, \nabla c(\mathbf{x}_i) \rangle}\right) \sum_{j \in \mathcal{N}_i} a_{ij} (\mathbf{x}_j - \mathbf{x}_i) \\ &= \left(\mathbf{I} - \tfrac{\mathbf{x}_i \nabla c(\mathbf{x}_i)^\top}{\langle \mathbf{x}_i, \nabla c(\mathbf{x}_i) \rangle}\right) \sum_{j \in \mathcal{N}_i} a_{ij} \mathbf{x}_j \end{aligned} \tag{6}$$

To briefly compare Algorithm 2 and 3, note that Algorithm 2 requires that $\nabla c(\mathbf{x}) \neq 0$ on $\mathcal{M}$ whereas Algorithm 3 also requires $\langle \mathbf{x}, \nabla c(\mathbf{x}) \rangle \neq 0$. The two algorithms are identical when $\nabla c(\mathbf{x}) = k\mathbf{x}$ for some $k \in \mathbb{R}$, *i.e.*, when $\mathcal{M}$ is $\mathcal{S}^n$. Indeed, both algorithms are conceived of as generalizations of a consensus algorithm on the n-sphere [15], [16]. In general, it may be more difficult to establish convergence of Algorithm 3 since it is not a gradient descent flow. We provide the following result about Algorithm 3:

Proposition 4: The system (6) on an ellipsoid is equivalent to the system (5) on the unit sphere.

Proof: Let $\mathcal{M}$ be an ellipsoid, *i.e.*,

$$\mathcal{M} = \{\mathbf{y} \in \mathbb{R}^n \,|\, c(\mathbf{y}) = \tfrac{1}{2} \langle \mathbf{y}, \mathbf{A}\mathbf{y} \rangle - 1 = 0\},$$

where $\mathbf{A}$ is a positive definite matrix. The dynamics (6) of the consensus seeking system on $\mathcal{M}$ under Algorithm 3 is

$$\dot{\mathbf{y}}_i = (\mathbf{I} - \tfrac{1}{\langle \mathbf{y}_i, \mathbf{A}\mathbf{y}_i \rangle} \mathbf{y}_i \mathbf{y}_i^\top \mathbf{A}) \sum_{\{i,j\} \in \mathcal{E}} a_{ij} (\mathbf{y}_j - \mathbf{y}_i).$$

Let $\mathbf{L}$ denote the Cholesky factor of $\mathbf{A}$, *i.e.*, $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$. Introduce $\mathbf{z}_i = \mathbf{L}^\top \mathbf{y}_i$ and note that $\|\mathbf{z}_i\|^2 = \langle \mathbf{y}_i, \mathbf{A}\mathbf{y}_i \rangle = 1$, *i.e.*, $\mathbf{z}_i \in \mathcal{S}^n$. Calculate

$$\begin{aligned}\dot{\mathbf{z}}_i &= \mathbf{L}^\top (\mathbf{I} - \tfrac{1}{\langle \mathbf{y}_i, \mathbf{L}\mathbf{L}^\top \mathbf{y}_i \rangle} \mathbf{y}_i \mathbf{y}_i^\top \mathbf{L}\mathbf{L}^\top) \sum_{j \in \mathcal{N}_i} a_{ij}(\mathbf{y}_j - \mathbf{y}_i) \\ &= (\mathbf{I} - \tfrac{\mathbf{z}_i}{\|\mathbf{z}_i\|}(\tfrac{\mathbf{z}_i}{\|\mathbf{z}_i\|})^\top) \sum_{j \in \mathcal{N}_i} a_{ij}(\mathbf{z}_j - \mathbf{z}_i) \\ &= (\mathbf{I} - \mathbf{z}_i \mathbf{z}_i^\top) \sum_{j \in \mathcal{N}_i} a_{ij} \mathbf{z}_j,\end{aligned}$$

which is the system (5) on the unit sphere. ■

IV. Almost global asymptotic stability

The main result of this paper states that for any closed, analytic manifold that satisfies a geometric condition, the consensus manifold $\mathcal{C}$ is an AGAS equilibrium manifold of the gradient descent flow (5). In the derivation of the main result, the condition appears as an expression which relates the relative information $\mathbf{x}_j - \mathbf{x}_i$ for any $\{i, j\} \in \mathcal{E}$ at an equilibria of the system to some geometric quantities evaluated at $\mathbf{x}_i$ and $\mathbf{x}_j$. However, it is difficult to say which pairs of points are part of an equilibrium and which are not. As such, we make the conservative requirement that the condition is satisfied at any pair of points $\mathbf{y}$, $\mathbf{z} \in \mathcal{M}$.

Let $\mathcal{Q}$ denote the set of all equilibria of the gradient descent flow (5) that does not belong to the consensus manifold $\mathcal{C}$ given by (3). Most of this section is concerned with establishing that each equilibria in $\mathcal{Q}$ is unstable; a result which is summarized in Proposition 10. This leads us to sufficient conditions for $\mathcal{C}$ to be an AGAS set of equilibria of the gradient descent flow (5). Before that we establish Proposition 5 which shows that the consensus manifold $\mathcal{C}$ given by (3) is asymptotically stable as a set. Note that Proposition 5 only requires $\mathcal{M}$ to be a closed analytic manifold, *i.e.*, a compact analytic manifold without boundary.

A. Local stability

Proposition 5: Let $\mathcal{M} \subset \mathbb{R}^m$ be a closed, analytic, embedded Riemannian manifold. The consensus manifold $\mathcal{C} = \{(\mathbf{x})_{i=1}^N \in \mathcal{M}^N\}$ is an asymptotically stable equilibrium set of the gradient descent flow $\dot{\mathbf{x}} = -(\Pi_i \nabla_i V)_{i=1}^N$ and

$$V(\mathbf{x}) = \tfrac{1}{2} \sum_{\{i,j\} \in \mathcal{E}} \|\mathbf{x}_j - \mathbf{x}_i\|^2.$$

Proof: The potential function of a gradient descent flow decreases with time,

$$\dot{V} = \langle \Pi_i \nabla V, \dot{\mathbf{x}} \rangle = -\|\Pi_i \nabla V\|^2. \tag{7}$$

Since $V \geq 0$ with $V = 0$ if and only if $\mathbf{x} \in \mathcal{C}$, we can take V as a Lyapunov function and conclude that $\mathcal{C}$ is stable.

Since $\mathcal{M}$ is closed, the gradient descent flow converges to a connected component of the set of critical points of V [17]. By (7), any sublevel set of V is forward invariant. Moreover, all sublevel sets contain $\mathcal{C}$. If there is an open sublevel set of V which does not intersect $\mathcal{Q}$, then there is an open neighborhood of $\mathcal{C}$ from which $\mathbf{x}$ converges to $\mathcal{C}$.

Since V is analytic it satisfies the Łojasiewicz inequality on Riemannian manifolds [18]. For every $\mathbf{x} \in \mathcal{C}$ there is an open ball $\mathcal{B}(\mathbf{x})$, an $\alpha < 1$, and a $k > 0$ such that

$$V(\mathbf{y})^\alpha \leq k \|\Pi \nabla V(\mathbf{y})\|$$

for all $\mathbf{y} \in \mathcal{B}(\mathbf{x})$. If $\mathbf{y} \in \mathcal{Q}$, then $\Pi \nabla V(\mathbf{y}) = \mathbf{0}$ whereby $V(\mathbf{y}) = 0$, which implies $\mathbf{y} \in \mathcal{C}$, a contradiction. Hence $\mathcal{Q} \cap \mathcal{B}(\mathbf{x}) = \emptyset$.

Consider the value of $q = \inf_{\mathbf{x} \in \mathcal{Q}} V(\mathbf{x})$. If $q = 0$, then there is a sequence $\{\mathbf{x}_k\}_{k=1}^\infty$ such that $\lim_{k \to \infty} V(\mathbf{x}_k) = 0$. Since $\mathcal{M}$ is a closed manifold, the sequence $\{\mathbf{x}_k\}_{k=1}^\infty$ has a subsequence which converges to some $\mathbf{y} \in \mathcal{M}$. Moreover, $V(\mathbf{y}) = 0$ whereby $\mathbf{y} \in \mathcal{C}$. For each $\varepsilon > 0$ there must be a $\mathbf{z}(\varepsilon) \in \mathcal{Q}$ (an element of the subsequence) such that $\|\mathbf{y} - \mathbf{z}(\varepsilon)\| < \varepsilon$. This contradicts $\mathcal{Q} \cap \mathcal{B}(\mathbf{y}) = \emptyset$. Hence $q > 0$ and all trajectories that start in the level set $\{\mathbf{x} \in \mathcal{M} \,|\, V(\mathbf{x}) < q\}$ converges to $\mathcal{C}$. ■

This result is similar to Proposition 7 in [6]. Note however that our proof of local stability only makes use of the properties of gradient descent flows of analytic function on closed manifolds. To show that $\mathcal{C}$ is AGAS we also need to consider the geometry and topology of $\mathcal{M}$. In particular, $\mathcal{C}$ is not an AGAS equilibrium manifold of (5) if $\mathcal{M}$ is a multiply connected hypersurface such as a circle or a torus [8]. The sufficient condition for AGAS established in this paper places requirements on $\mathcal{M}$ that exclude such cases.

B. Main result

Our main result is establishes almost global convergence to the consensus manifold if the following assumption holds:

Assumption 6: Suppose c satisfies

$$\langle \mathbf{n}(\mathbf{y}), \mathbf{n}(\mathbf{z}) \rangle^2 + \frac{\langle \mathbf{y} - \mathbf{z}, \nabla c(\mathbf{y}) \rangle (\Delta c(\mathbf{y}) - \langle \mathbf{n}(\mathbf{y}), \nabla^2 c(\mathbf{y}) \mathbf{n}(\mathbf{y}) \rangle)}{\|\nabla c(\mathbf{y})\|^2} \geq 1,$$

for all $\mathbf{y}, \mathbf{z} \in \mathcal{M}$ and with equality only if $\mathbf{y} = \mathbf{z}$, where $\mathbf{n} : \mathcal{M} \to \mathcal{S}^{n-1}$ is the Gauss map and Δ is the Laplace-Beltrami operator, $\Delta c(\mathbf{y}) = \operatorname{tr} \nabla^2 c(\mathbf{y})$.

Theorem 7: Let c be a real analytic function that satisfies Assumption 6. The consensus manifold is an AGAS equilibrium set of the gradient descent flow

$$\begin{aligned} V &= \tfrac{1}{2} \sum_{\{i,j\} \in \mathcal{E}} a_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2, \\ \dot{\mathbf{x}}_i &= -\nabla_i V \\ &= (\mathbf{I} - \tfrac{\nabla c(\mathbf{x}_i)}{\|\nabla c(\mathbf{x}_i)\|}(\tfrac{\nabla c(\mathbf{x}_i)}{\|\nabla c(\mathbf{x}_i)\|})^\top) \sum_{\{i,j\} \in \mathcal{E}} a_{ij}(\mathbf{x}_j - \mathbf{x}_i), \end{aligned}$$

on the N-fold product of $\mathcal{M} = \{\mathbf{y} \in \mathbb{R}^n \,|\, c(\mathbf{y}) = 0\}$.

Since the proof of the main result is somewhat long, we have broken it into parts. First, we need two definitions.

Definition 8: Let Σ be a dynamical system on $\mathcal{S} \subset \mathbb{R}^{n+1}$ whose solution $\mathbf{\Phi}(t; \mathbf{x})$, $\mathbf{\Phi}(0; \mathbf{x}) = \mathbf{x}$ exists for all $t \in \mathbb{R}$ and all $\mathbf{x} \in \mathcal{S}$. The system Σ is said to be pointwise convergent if for each $\mathbf{x} \in \mathcal{S}$ there is exactly one ω-limit point $\lim_{i \to \infty} \mathbf{\Phi}(t_i; \mathbf{x})$ for all $(t_i)_{i=1}^\infty$ such that $\lim_{i \to \infty} t_i = \infty$.

Definition 9: An equilibrium point $\mathbf{y} \in \mathbb{R}^{n+1}$ of a dynamical system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ is said to be exponentially unstable

if the Jacobian matrix of $\mathbf{f}(\mathbf{x})$ evaluated at $\mathbf{y}$ has a strictly positive eigenvalue.

For pointwise convergent systems, any set of exponentially unstable equilibria have a region of attraction with Riemannian measure zero [19]. The system (5) is pointwise convergent due to being a gradient descent flow of an analytic function on an analytic manifold [20]. The problem of establishing almost global convergence has hence been reduced to showing that all equilibria besides those belonging to the consensus manifold are exponentially unstable.

C. Positive eigenvalues

Let $\mathbf{L}(\mathbf{x}) \in \mathbb{R}^{N(n+1)\times N(n+1)}$ denote the linearization matrix of the gradient descent flow (5) at the point $\mathbf{x} \in \mathcal{M}^N$. Our aim is to show that the eigenvalues of $\mathbf{L}(\mathbf{x})$ are positive for all equilibria $\mathbf{x} \notin \mathcal{C}$. Note that $\mathbf{L}(\mathbf{x})$ is related to the Hessian matrix $\mathbf{H}(\mathbf{x}) \in \mathbb{R}^{Nn\times Nn}$ of V as $\mathbf{L}(\mathbf{x}) = -\mathbf{H}(\mathbf{x})$ [17].

Proposition 10: Let $\mathcal{M} \subset \mathbb{R}^n$ be a hypersurface for which Assumption 6 holds. The eigenvalues of the linearization matrix of the gradient descent flow (5) have strictly negative real parts at any equilibrium point except for those belonging to the consensus manifold $\mathcal{C}$.

Proof: The equilibria of the gradient descent flow (5) are critical points of the optimization problem

$$\min_{\mathbf{x}\in\mathcal{M}^N} V(\mathbf{x}) = \tfrac{1}{2}\sum_{\{i,j\}\in\mathcal{E}} a_{ij}\|\mathbf{x}_j - \mathbf{x}_i\|^2. \tag{8}$$

We will analyze (5) in an optimization framework, making use of the associated techniques and terminology. Our approach is based on the Lagrange conditions for optimality in equality constrained nonlinear programming [21].

Introduce the Lagrangian $\mathcal{L} : \mathcal{M}^N \times \mathbb{R} \to \mathbb{R}$ given by

$$\begin{aligned}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) &= V + \sum_{i\in\mathcal{V}} \lambda_i c(\mathbf{x}_i)\\ &= \tfrac{1}{2}\sum_{\{i,j\}\in\mathcal{E}} a_{ij}\|\mathbf{x}_i - \mathbf{x}_j\|^2 + \sum_{i\in\mathcal{V}} \lambda_i c(\mathbf{x}_i),\end{aligned}$$

where $\boldsymbol{\lambda} = [\lambda_i] \in \mathbb{R}^N$. The optimal solutions to (8) are critical points of $\mathcal{L}$. The critical points of $\mathcal{L}$ are exactly the eigenvalues of (5). Calculate the Euclidean gradient of $\mathcal{L}$,

$$\begin{aligned}\nabla_i\mathcal{L} &= \sum_{j\in\mathcal{N}_i} a_{ij}(\mathbf{x}_i - \mathbf{x}_j) + \lambda_i\nabla_i c(\mathbf{x}_i),\\ \tfrac{\partial}{\partial\lambda_i}\mathcal{L} &= c(\mathbf{x}_i).\end{aligned}$$

The Hessian of $\mathcal{L}$ with respect to $\mathbf{x}_i$, $\mathbf{x}_k$ is a $N(n+1)\times N(n+1)$ block matrix $\nabla^2\mathcal{L}$, with the ki block given by

$$(\nabla^2\mathcal{L})_{ki} = \begin{cases}\sum_{j\in\mathcal{N}_i} a_{ij}\mathbf{I}_{n+1} + \lambda_i\nabla_i^2 c(\mathbf{x}_i) & \text{if } k = i,\\ -a_{ik}\mathbf{I}_{n+1} & \text{if } k\in\mathcal{N}_i,\\ \mathbf{0} & \text{otherwise.}\end{cases}$$

The nullspace $\ker\nabla c_i$ of the constraint gradients is the image set of the symmetric matrix

$$\mathbf{Z}_i = \mathbf{I}_{n+1} - \mathbf{n}_i\mathbf{n}_i^\top,$$

where $\mathbf{n}_i = \mathbf{n}(\mathbf{x}_i)$ and $\mathbf{n}$ is the Gauss map. Let $\mathbf{Z}$ denote the blockdiagonal matrix with $\mathbf{Z}_i$ as the ii block. Form the matrix $\mathbf{H}(\mathbf{x}) = \mathbf{Z}\nabla^2\mathcal{L}\mathbf{Z}$ whose ki block is

$$\mathbf{Z}_k\tfrac{\partial^2\mathcal{L}}{\partial\mathbf{x}_k\partial\mathbf{x}_i}\mathbf{Z}_i = \begin{cases}\sum_{j\in\mathcal{N}_i} a_{ij}\mathbf{Z}_i + \lambda_i\mathbf{Z}_i\nabla_i^2 c\mathbf{Z}_i & \text{if } k = i,\\ -a_{ki}\mathbf{Z}_k\mathbf{Z}_i & \text{if } k\in\mathcal{N}_i,\\ \mathbf{0} & \text{otherwise,}\end{cases}$$

where we used that $\mathbf{Z}_i^2 = \mathbf{Z}_i$, which follows from $\mathbf{Z}_i$ being a projection matrix.

Let $\mathsf{T}\mathcal{M}^N$ denote the tangent bundle of $\mathcal{M}^N$,

$$\mathsf{T}\mathcal{M}^N = \{(\mathbf{x},\mathbf{v}) \,|\, \mathbf{x}\in\mathcal{M}^N, \mathbf{v}\in\mathsf{T}_{\mathbf{x}}\mathcal{M}^N\}.$$

The matrix $\mathbf{H}(\mathbf{x})$ is the Riemannian Hessian operator $\mathbf{H}(\mathbf{x}) : \mathsf{T}\mathcal{M}^N \to \mathsf{T}_{\mathbf{x}}\mathcal{M} : (\mathbf{x},\mathbf{v}) \mapsto \mathbf{H}(\mathbf{x})\mathbf{v}$ of V on $\mathcal{M}$ [22]. It also appears in the necessary second order optimality conditions for equality constrained problems, with $\mathbf{H}(\mathbf{x})$ being positive semi-definite on $\mathsf{T}\mathcal{M}^N$ if $\mathbf{x}$ is an optimal solution to (8) that satisfies some additional requirements [21].

Let $\mathbf{L}(\mathbf{x}) = -\mathbf{H}(\mathbf{x})$ be the linearization matrix of the gradient descent flow [17]. Note that $\mathbf{L}$ is symmetric wherefore its field of values

$$W(\mathbf{L}) = \{\langle\mathbf{v},\mathbf{L}\mathbf{v}\rangle \,|\, \mathbf{v}\in\mathbb{C}^{Nn}\} = \{\langle\mathbf{v},\mathbf{L}\mathbf{v}\rangle \,|\, \mathbf{v}\in\mathbb{R}^{Nn}\}$$

is real. Consider the Rayleigh quotient $R : \mathsf{T}\mathcal{M}^N \to \mathbb{R}$ given by $R(\mathbf{x},\mathbf{v}) = \langle\mathbf{v},\mathbf{L}(\mathbf{x})\mathbf{v}\rangle/\langle\mathbf{v},\mathbf{v}\rangle$. Let $\alpha(\mathbf{x})$ denote the spectral abscissa of $\mathbf{L}(\mathbf{x})$,

$$\alpha(\mathbf{x}) = \max_{\mathbf{v}\in\mathsf{T}_{\mathbf{x}}\mathcal{M}^N} R(\mathbf{x},\mathbf{v}).$$

Since $\mathbf{L}(\mathbf{x})$ is symmetric, $\alpha(\mathbf{x})$ equals the largest eigenvalue of $\mathbf{L}(\mathbf{x})$. It is bounded below as $\alpha(\mathbf{x}) \geq R(\mathbf{x},\mathbf{v})$ for all $\mathbf{v} \in \mathsf{T}_{\mathbf{x}}\mathcal{M}^N$ by the min-max theorem. It follows that $-R(\mathbf{x},\mathbf{v})$ is an upper bound on the smallest eigenvalue of $\mathbf{H}(\mathbf{x})$. If $R(\mathbf{x},\mathbf{v})$ assumes a positive value for some argument, then the $\mathbf{H}(\mathbf{x})$ cannot be positive definite and the necessary optimality conditions fails to hold.

To obtain a lower bound for $\alpha(\mathbf{x})$, consider the tangent vector $\mathbf{v} = [\Pi_1\mathbf{u}\ldots\Pi_N\mathbf{u}] = [\mathbf{Z}_1\mathbf{u}\ldots\mathbf{Z}_N\mathbf{u}]$ for any $\mathbf{u} \in \mathbb{R}^n$ such that $\|\mathbf{v}\| = 1$. The intuition for this step is that all agents are located at some equilibrium $\mathbf{x}$ and that we perturb all of them in the same direction, *i.e.*, towards the consensus manifold. Because all agents move towards the same region of the consensus manifold, it is possible that cohesion is increased whereby V decreases. We calculate the effect this has on the quadratic term in the Taylor expanasion of V, *i.e.*, the term that depends on $\mathbf{H}(\mathbf{x}) = -\mathbf{L}(\mathbf{x})$. The contribution of the linear term is zero due to $\nabla V = \mathbf{0}$ at any equilibrium.

Calculate the Rayleigh quotient,

$$\begin{aligned}R(\mathbf{x},\mathbf{v}) &= \sum_{i\in\mathcal{V}}\langle\mathbf{u},(\mathbf{L}_{ii}(\mathbf{x}) + \sum_{j\in\mathcal{N}_i}\mathbf{L}_{ij})\mathbf{u}\rangle\\ &= \sum_{i\in\mathcal{V}}\langle\mathbf{u},-(\lambda_i\mathbf{Z}_i\nabla_i^2 c\mathbf{Z}_i + \sum_{j\in\mathcal{N}_i} a_{ij}(\mathbf{Z}_i - \mathbf{Z}_i\mathbf{Z}_j))\mathbf{u}\rangle.\end{aligned}$$

Denote

$$\mathbf{M}(\mathbf{x}) = -\sum_{i\in\mathcal{V}}\lambda_i\mathbf{Z}_i\nabla_i^2 c\mathbf{Z}_i + \sum_{j\in\mathcal{N}_i} a_{ij}(\mathbf{Z}_i - \mathbf{Z}_i\mathbf{Z}_j).$$

Hence $R(\mathbf{x},\mathbf{v}) = \langle \mathbf{u}, \mathbf{M}\mathbf{u}\rangle$. Let $(\mu_i, \mathbf{u}_i)$, where $\|\mathbf{u}_i\| = 1$, denote the eigenpairs of $\mathbf{M}$. Take $\mathbf{u} = \sum_{i=1}^{n+1} \mathbf{u}_i / \|\sum_{i=1}^{n+1} \mathbf{u}_i\|$ whereby $\mathbf{v} = \mathbf{Z}\mathbf{u}$ satisfies $\mathbf{v} \in \mathsf{T}_{\mathbf{x}_i}\mathcal{M}$, $\|\mathbf{v}\| = 1$ (note that $\sum_{i=1}^{n+1} \mathbf{u}_i \neq \mathbf{0}$ by linear independence). Then

$$R(\mathbf{x},\mathbf{v}) = \frac{\operatorname{tr}\mathbf{M}(\mathbf{x})}{\|\sum_{i=1}^{n}\mathbf{u}_i\|}.$$

It remains to show that $\operatorname{tr}\mathbf{M}(\mathbf{x}) \geq 0$.

For the sake of notational convenience, write

$$\mathbf{N}_i = \mathbf{n}_i\mathbf{n}_i^\top = \tfrac{\nabla_i c(\mathbf{x}_i)}{\|\nabla_i c(\mathbf{x}_i)\|}\left(\tfrac{\nabla_i c(\mathbf{x}_i)}{\|\nabla_i c(\mathbf{x}_i)\|}\right)^\top.$$

whereby $\mathbf{Z}_i = \mathbf{I}_n - \mathbf{N}_i$. Rewrite

$$\begin{aligned}\mathbf{M} = &-\sum_{i\in\mathcal{V}} \lambda_i(\mathbf{I}_n - \mathbf{N}_i)\nabla_i^2 c(\mathbf{I}_n - \mathbf{N}_i) + \\ &\sum_{j\in\mathcal{N}_i} a_{ij}(\mathbf{I}_n - \mathbf{N}_i - (\mathbf{I}_n - \mathbf{N}_i)(\mathbf{I}_n - \mathbf{N}_j)) \\ = &-\sum_{i\in\mathcal{V}} \lambda_i(\nabla_i^2 c - \mathbf{N}_i\nabla_i^2 c - \nabla_i^2 c\mathbf{N}_i + \mathbf{N}_i\nabla_i^2 c\mathbf{N}_i) + \\ &\sum_{j\in\mathcal{N}_i} a_{ij}(\mathbf{N}_j - \mathbf{N}_i\mathbf{N}_j).\end{aligned}$$

Solve $\nabla_i\mathcal{L} = \mathbf{0}$ for

$$\lambda_i = \tfrac{1}{\|\nabla_i c\|^2}\sum_{j\in\mathcal{N}_i} a_{ij}\langle \nabla_i c(\mathbf{x}_i), \mathbf{x}_j - \mathbf{x}_i\rangle.$$

Note that λ_i is well-defined since $\mathcal{M}$ is nonsingular by assumption, which implies $\nabla_i c \neq \mathbf{0}$. Note that $\operatorname{tr}\mathbf{N}_i = 1$. Let $\Delta_i c = \operatorname{tr}\nabla_i^2 c(\mathbf{x}_i)$ denote the Laplace-Beltrami operator acting on c. Calculate

$$\begin{aligned}\operatorname{tr}\mathbf{M} = &-\sum_{i\in\mathcal{V}} \lambda_i(\Delta_i c - \langle \mathbf{n}_i, \nabla_i^2 c\mathbf{n}_i\rangle) + \\ &\sum_{j\in\mathcal{N}_i} a_{ij}(1 - \langle \mathbf{n}_i, \mathbf{n}_j\rangle^2) \\ = &\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i} a_{ij}[-1 + \langle \mathbf{n}_i, \mathbf{n}_j\rangle^2 + \\ &\tfrac{1}{\|\nabla_i c\|^2}\langle \mathbf{x}_i - \mathbf{x}_j, \nabla_i c\rangle(\Delta_i c - \langle \mathbf{n}_i, \nabla_i^2 c\mathbf{n}_i\rangle)]\end{aligned}$$

The sum in the expression for $\operatorname{tr}\mathbf{M}$ is positive if every term is positive, *i.e.*, if

$$\langle \mathbf{n}_i, \mathbf{n}_j\rangle^2 + \tfrac{\langle \mathbf{x}_i - \mathbf{x}_j, \nabla_i c\rangle(\Delta_i c - \langle \mathbf{n}_i, \nabla_i^2 c\mathbf{n}_i\rangle)}{\|\nabla_i c\|^2} \geq 1$$

with equality only when $\mathbf{x}_i = \mathbf{x}_j$. This relation holds by Assumption 6 on the geometry of $\mathcal{M}$. ■

V. Convexity

A. Convex sets

Assumption 6 allows for a geometric interpretation. Recall that by the Jordan-Brouwer separation theorem, a compact hypersurface $\mathcal{M}$ separates $\mathbb{R}^n$ into two connected sets, one interior set which is bounded, $\mathcal{K}$, and one exterior set which is unbounded $(\mathcal{K}\cup\mathcal{M})^{\mathsf{c}}$. The inequality in Assumption 6 implies that $\mathcal{K}$ is convex, *i.e.*, that $\mathcal{M} = \partial\mathcal{K}$ is the boundary of a convex set. To show this, we first need a lemma.

Lemma 11: Let $\mathcal{M} \subset \mathbb{R}^n$ be a nonsingular hypersurface given by $\mathcal{M} = \{\mathbf{y} \in \mathbb{R}^n \,|\, c(\mathbf{y}) = 0\}$, where c is $\mathcal{C}^1$. Take any $\mathbf{z} \in \mathbb{R}^n$. The vector $\mathbf{v}$ of shortest length $\|\mathbf{v}\|$ such that $\mathbf{y} + \mathbf{v} = \mathbf{z}$ for some $\mathbf{y} \in \mathcal{M}$ is parallel to the normal of $\mathcal{M}$ given by $\nabla c(\mathbf{y})$.

Proof: The Lagrange conditions for optimality in the nonlinear optimization problem

$$\min_{y\in\mathbb{R}^n} \tfrac{1}{2}\|\mathbf{z} - \mathbf{y}\|^2 \text{ subject to } c(\mathbf{y}) = 0$$

are necessary since $\mathcal{M}$ is nonsingular (*i.e.*, all points on $\mathcal{M}$ are regular). Form the Lagrangian function $\mathcal{L}(\mathbf{y},\lambda) = \|\mathbf{z} - \mathbf{y}\|^2 + \lambda c(\mathbf{y})$. The Lagrange conditions state that

$$\mathbf{z} - \mathbf{y} + \lambda\nabla c(\mathbf{y}) = \mathbf{0}$$

from which it follows that $\mathbf{v} = \mathbf{z} - \mathbf{y} = -\lambda\nabla c(\mathbf{y})$. ■

Theorem 12: Suppose Assumption 6 holds and that $\mathcal{M}$ is a closed manifold, then $\mathcal{M}$ is the boundary of a convex set.

Proof: Note that in order for Assumption 6 to hold, since $\langle \mathbf{n}(\mathbf{y}), \mathbf{n}(\mathbf{z})\rangle^2 \leq 1$, it is necessary that $\langle \mathbf{y} - \mathbf{z}, \nabla c(\mathbf{y})\rangle$ and $\Delta c(\mathbf{y}) - \langle \mathbf{n}(\mathbf{y}), \nabla^2 c(\mathbf{y})\mathbf{n}(\mathbf{y})\rangle$ have the same sign. The latter expression only depends on $\mathbf{y}$ wherefore the sign of $\langle \mathbf{y} - \mathbf{z}, \nabla c(\mathbf{y})\rangle$ cannot vary with $\mathbf{z}$, *i.e.*, either

$$\langle \mathbf{y} - \mathbf{z}, \nabla c(\mathbf{y})\rangle \geq 0 \tag{9}$$

or

$$\langle \mathbf{y} - \mathbf{z}, \nabla c(\mathbf{y})\rangle \leq 0 \tag{10}$$

holds for all $\mathbf{z} \in \mathcal{M}$ at any $\mathbf{y} \in \mathcal{M}$.

Recall that we have chosen the sign of c such that for all $\mathbf{y} \in \mathcal{M}$, $\nabla c(\mathbf{y})$ points towards the exterior of the two sets separated by $\mathcal{M}$. Let $\mathcal{K}$ denote the interior set. Following the negative normal $-\nabla c(\mathbf{y})$ on a line segment from $\mathbf{y}$ through the interior set $\mathcal{K}$, we find another point $\mathbf{z} \in \mathcal{M}$ (otherwise the interior set would be unbounded). Note that $\mathbf{y} - \mathbf{z}$ is aligned with the normal at $\mathbf{y}$. Hence $\langle \mathbf{y} - \mathbf{z}, \nabla c(\mathbf{y})\rangle \geq 0$ wherefore we can exclude the case of (10).

By (9), for each $\mathbf{y} \in \mathcal{M}$, there is an affine hyperplane through $\mathbf{y}$ with normal $\nabla c(\mathbf{y})$. This hyperplane divides $\mathbb{R}^{n+1}$ into a closed set containing $\mathcal{M}$ and an open set which is disjoint from $\mathcal{M}$. Let $\mathcal{H}(\mathbf{y})$ denote the closed half-space which contains $\mathcal{M}$, *i.e.*,

$$\mathcal{H}(\mathbf{y}) = \{\mathbf{w} \in \mathbb{R}^n \,|\, \langle \mathbf{y} - \mathbf{w}, \nabla c(\mathbf{y})\rangle \geq 0\}.$$

Form

$$\mathcal{S} = \cap_{\mathbf{y}\in\mathcal{M}}\mathcal{H}(\mathbf{y}).$$

Since $\mathcal{S}$ is an intersection of convex sets, it is convex. We will show that $\mathcal{K} = \mathcal{S}$.

Since $\mathcal{M} \subset \mathcal{H}(\mathbf{y})$ for all $\mathbf{y} \in \mathcal{M}$, it follows that $\mathcal{M} \subset \mathcal{S}$. Hence $\mathcal{K} \subset \mathcal{S}$. To show $\mathcal{S} \subset \mathcal{K}$, suppose by way of contradiction that there is a $\mathbf{s} \in \mathcal{S}\backslash\mathcal{K}$. There is a point $\mathbf{y} \in \mathcal{M}$ which minimizes the Euclidean distance to $\mathbf{s}$. By Lemma 11, this point satisfies $\mathbf{s} = \mathbf{y} + \lambda\nabla c(\mathbf{y})$ for some $\lambda \in \mathbb{R}$. Because $\mathbf{s} \notin \mathcal{K}$ and $\nabla c(\mathbf{y})$ points away from $\mathcal{K}$ at $\mathbf{y}$, it must be the case that $\lambda > 0$. Then

$$\langle \mathbf{y} - \mathbf{s}, \nabla c(\mathbf{y})\rangle = -\lambda\|\nabla c(\mathbf{y})\|^2 < 0.$$

This implies that $\mathbf{s} \notin \mathcal{H}(\mathbf{y})$ and hence $\mathbf{s} \notin \mathcal{S}$, which contradicts the assumption that $\mathbf{s} \in \mathcal{S} \backslash \mathcal{K}$. ■

B. Strongly convex functions

Conversely, we could assume that c is a convex function on all of $\mathbb{R}^n$. However, c being convex does not imply that Assumption 13 holds. A counter example is given by $c : \mathbb{R}^2 \to \mathbb{R} : \mathbf{x} \mapsto \|\mathbf{x}\|^2 - r^2$, which yields the Kuramoto model on $\mathcal{S}^1$. Consider the class of strongly convex functions. A strongly convex function f with parameter m satisfies

$$f(\mathbf{z}) \geq f(\mathbf{y}) + \langle \mathbf{z} - \mathbf{y}, \nabla f(\mathbf{y}) \rangle + \tfrac{m}{2} \|\mathbf{z} - \mathbf{y}\|^2$$

at all points $\mathbf{y}$, $\mathbf{z}$ in its domain. That c is strongly convex on $\mathcal{M}$ implies

$$\langle \mathbf{y} - \mathbf{z}, \nabla c(\mathbf{y}) \rangle \geq \tfrac{m}{2} \|\mathbf{z} - \mathbf{y}\|^2.$$

Equivalently, any continuous function f on a compact domain is strongly convex if $m\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq M\mathbf{I}$.

Assumption 13: Let c be a strongly convex function,

$$m\mathbf{I} \preceq \nabla^2 c(\mathbf{y}) \preceq M\mathbf{I}.$$

Moreover, suppose that c satisfies

$$\tfrac{m((n+1)m-M)}{(LK)^2} \geq 2,$$

where $n = \dim \mathcal{M}$, L is a global Lipschitz constant of the Gauss map $\mathbf{n} : \mathcal{M} \to \mathcal{S}^{n-1}$, *i.e.*,

$$\|\mathbf{n}(\mathbf{y}) - \mathbf{n}(\mathbf{z})\| \leq L\|\mathbf{y} - \mathbf{z}\|,$$

for all $\mathbf{y}, \mathbf{z} \in \mathcal{M}$, and $K = \max_{\mathbf{y} \in \mathcal{M}} \|\nabla c(\mathbf{y})\|$.

Proposition 14: Assumption 13 implies Assumption 6.

Proof: Consider the last term in the inequality of Assumption 6. Strong convexity of c implies that

$$\tfrac{\langle \mathbf{y}-\mathbf{z}, \nabla c(\mathbf{y}) \rangle (\Delta c(\mathbf{y}) - \langle \mathbf{n}(\mathbf{y}), \nabla^2 c(\mathbf{y}) \mathbf{n}(\mathbf{y}) \rangle)}{\|\nabla c(\mathbf{y})\|^2} \geq \tfrac{m(nm-M)\|\mathbf{z}-\mathbf{y}\|^2}{2\|\nabla c(\mathbf{y})\|^2}.$$

Since $\mathcal{M}$ is nonsingular by assumption, *i.e.*, $\nabla c(\mathbf{y}) \neq \mathbf{0}$ for all $\mathbf{y} \in \mathcal{M}$, the Gauss map

$$\mathbf{n} : \mathbf{y} \mapsto \tfrac{\nabla c(\mathbf{y})}{\|\nabla c(\mathbf{y})\|}$$

is locally Lipschitz on $\mathcal{M}$. Since $\mathcal{M}$ is a closed manifold there is a global Lipschitz constant L of $\mathbf{n}$ over all points on $\mathcal{M}$. It follows that

$$\tfrac{m((n+1)m-M)\|\mathbf{z}-\mathbf{y}\|^2}{2\|\nabla c(\mathbf{y})\|^2} \geq \tfrac{m((n+1)m-M)\|\mathbf{n}(\mathbf{z})-\mathbf{n}(\mathbf{y})\|^2}{2L^2K^2}$$

where we also utilized the definition of K.

Let ϑ denote the angle between $\mathbf{n}(\mathbf{y})$ and $\mathbf{n}(\mathbf{z})$. For Assumption (6) we find that

$$\begin{aligned}
\langle \mathbf{n}(\mathbf{y}), \mathbf{n}(\mathbf{z}) \rangle^2 + \tfrac{\langle \mathbf{y}-\mathbf{z}, \nabla c(\mathbf{y}) \rangle (\Delta c(\mathbf{y}) - \langle \mathbf{n}(\mathbf{y}), \nabla^2 c(\mathbf{y}) \mathbf{n}(\mathbf{y}) \rangle)}{\|\nabla c(\mathbf{y})\|^2} &\geq \\
\cos^2 \vartheta + \tfrac{m((n+1)m-M)\|\mathbf{z}-\mathbf{y}\|^2}{2K^2} &\geq \\
\cos^2 \vartheta + \tfrac{m((n+1)m-M)\|\mathbf{n}(\mathbf{z})-\mathbf{n}(\mathbf{y})\|^2}{2L^2K^2} &= \\
\cos^2 \vartheta + \tfrac{m((n+1)m-M)(1-\cos\vartheta)}{(LK)^2} &= \\
\cos^2 \vartheta + \alpha(1 - \cos\vartheta) &\geq 1,
\end{aligned}$$

where $\alpha = \frac{m((n+1)m-M)}{(LK)^2}$, if α is sufficiently large.

Denote $g(\vartheta, \alpha) = \cos^2 \vartheta + \alpha(1 - \cos\vartheta)$. We minimize this expression with respect to θ to find the range of α for which $g(\vartheta, \alpha) \geq 1$ for all $\vartheta \in [0, \pi]$. Hence

$$\tfrac{\partial g(\vartheta, \alpha)}{\partial \vartheta} = -2 \sin\vartheta \cos\vartheta + \alpha \sin\theta = 0.$$

Either $\sin\vartheta = 0$ or $\cos\vartheta = \frac{\alpha}{2}$ for $\alpha \in [0, 2]$. In the first case $\cos\theta \in \{-1, 1\}$, which results in either $1 + 2\alpha \geq 1$ or $1 \geq 1$. The condition on α is $\alpha \geq 0$. In the second case

$$g(\vartheta, \alpha) = \tfrac{\alpha^2}{4} + \alpha(1 - \tfrac{\alpha}{2}) = \alpha - \tfrac{\alpha^2}{4} \geq 1,$$

which yields $\alpha \geq 2$. Hence we require $\frac{m((n+1)m-M)}{(LK)^2} \geq 2$. ■

References

[1] W. Song, J. Markdahl, S. Zhang, X. Hu, and Y. Hong. Intrinsic reduced attitude formation with ring inter-agent graph. *Automatica*, 85:193–201, 2017.

[2] S. Al-Abri, W. Wu, and F. Zhang. A gradient-free 3-dimensional source seeking strategy with robustness analysis. *IEEE Transactions on Automatic Control*, 2018.

[3] A. Aydogdu, S.T. McQuade, and N.P. Duteil. Opinion dynamics on a general compact Riemannian manifold. *Networks & Heterogeneous Media*, 12(3):489–523, 2017.

[4] A. Crnkić and V. Jaćimović. Swarms on the 3-sphere with adaptive synapses: Hebbian and anti-Hebbian learning rule. *Systems & Control Letters*, 122:32–38, 2018.

[5] S.-Y. Ha and D. Kim. A second-order particle swarm model on a sphere and emergent dynamics. *SIAM Journal on Applied Dynamical Systems*, 18(1):80–116, 2019.

[6] A. Sarlette and R. Sepulchre. Consensus optimization on manifolds. *SIAM Journal on Control and Optimization*, 48(1):56–76, 2009.

[7] J. Markdahl, J. Thunberg, and J. Gonçalves. Almost global consensus on the n-sphere. *IEEE Transactions on Automatic Control*, 63(6):1664–1675, 2018.

[8] J. Markdahl. Synchronization on Riemannian manifolds: Multiply connected implies multistable. *arXiv preprint arXiv:1906.07452*, 2019.

[9] Johan Markdahl, Johan Thunberg, and Jorge Goncalves. High-dimensional kuramoto models on stiefel manifolds synchronize complex networks almost globally. *Automatica*, 113:108736, 2020.

[10] J. Zhu. Synchronization of Kuramoto model in a high-dimensional linear space. *Physics Letters A*, 377(41):2939–2943, 2013.

[11] J. Zhu. High-dimensional Kuramoto model limited on smooth curved surfaces. *Physics Letters A*, 378(18-19):1269–1280, 2014.

[12] J. Zhang, Y. Wang, and J. Zhu. Synchronisation of Lohe model on smooth curved surfaces. *The Journal of Engineering*, 2019(22):8343–8347, 2019.

[13] E.L. Lima. The Jordan-Brouwer separation theorem for smooth hypersurfaces. *The American Mathematical Monthly*, 95(1):39–42, 1988.

[14] H.K. Khalil. *Nonlinear systems*. Prentice Hall, 2002.

[15] R. Olfati-Saber. Swarms on the sphere: A programmable swarm with synchronous behaviors like oscillator networks. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 5060–5066, 2006.

[16] M.A. Lohe. Quantum synchronization over quantum networks. *Journal of Physics A: Mathematical and Theoretical*, 43(46):465301, 2010.

[17] U. Helmke and J.B. Moore. *Optimization and Dynamical Systems*. Springer, 2012.

[18] K. Kurdyka, T. Mostowski, and A. Parusinski. Proof of the gradient conjecture of R. Thom. *Annals of Mathematics*, 152(3):763–792, 2000.

[19] R.A. Freeman. A global attractor consisting of exponentially unstable equilibria. In *Proceedings of the 31st American Control Conference*, pages 4855–4860, 2013.

[20] C. Lageman. *Convergence of Gradient-Like Dynamical Systems and Optimization Algorithms*. PhD thesis, University of Würzburg, 2007.

[21] J. Nocedal and S.J. Wright. Numerical optimization. *Springer*, 1999.

[22] P. Birtea and D. Comănescu. Hessian operators on constraint manifolds. *Journal of Nonlinear Science*, 25(6):1285–1305, 2015.