Finite-time convergent gradient flows with applications to motion coordination ⋆

Jorge Cortés

Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, Santa Cruz, CA 95064, USA, email: `jcortes@ucsc.edu`

Abstract

This paper introduces the normalized and signed gradient dynamical systems associated with a differentiable function. Extending recent results on nonsmooth stability analysis, we characterize their asymptotic convergence properties and identify conditions that guarantee finite-time convergence. We discuss the application of the results to the design of multi-agent coordination algorithms, paying special attention to their scalability properties. Finally, we consider network consensus problems and show how the proposed nonsmooth gradient flows achieve the desired coordination task in finite time.

Key words: Gradient flows; Nonsmooth analysis; Finite-time convergence; Motion coordination; Network Consensus

1 Introduction

Problem statement. Let $f : \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, be a differentiable function. Consider the gradient flow

$$\dot{x} = -\operatorname{grad}(f)(x)\,.$$

It is well known (see e.g. [15]) that the minima of f are stable equilibria of this system, and that, if the level sets of f are bounded, then the trajectories converge asymptotically to the set of critical points of f. Gradient dynamical systems are employed in a wide range of applications, including optimization, distributed parallel computing, motion planning and control. In robotics, potential field methods are used to autonomously navigate a robot in a cluttered environment. Gradient algorithms enjoy many important features: they are naturally robust to perturbations and measurement errors, amenable to asynchronous implementations, and admit efficient numerical approximations.

In this note, we provide an answer to the following question: how could one modify the gradient vector field above so that the trajectories converge to the critical points of the function *in finite time*? - as opposed to over an infinite-time horizon. There are a number of settings where finite-time convergence is a desirable property. We study this problem with the aim of designing gradient coordination algorithms for multi-agent systems that achieve the desired task in finite time.

Our answer to the question above is the flows

$$\dot{x} = -\frac{\operatorname{grad}(f)(x)}{\|\operatorname{grad}(f)(x)\|_2}, \qquad \dot{x} = -\operatorname{sgn}(\operatorname{grad}(f)(x)),$$

where $\|\cdot\|_2$ denotes the Euclidean distance and $\operatorname{sgn}(x) = (\operatorname{sgn}(x_1), \dots, \operatorname{sgn}(x_d))$. Using tools from nonsmooth stability analysis, we show in this note that, under some assumptions on f, both systems are guaranteed to achieve the set of critical points in finite time.

Literature review. Guidelines on how to design dynamical systems for optimization purposes, with a special emphasis on gradient systems, are described in [14]. The book [3] thoroughly discusses gradient descent flows in distributed computation in settings with fixed-communication topologies. Nonsmooth analysis studies the notion and computational properties of the generalized gradient [5]. Tools for establishing stability and convergence properties of nonsmooth dynamical systems are presented in [2, 12, 21]. Finite-time discontinuous feedback stabilizers for a class of planar systems are proposed in [20]. Finite-time stability of continuous autonomous systems is rigorously studied in [4]. The reference [6] develops finite-time stabilization strategies based on time-varying feedback. Previous work on motion coordination of multi-agent systems has proposed cooperative algorithms based on gradient flows to achieve tasks such as cohesiveness [13, 18, 23], con-

⋆ This paper was not presented at any IFAC meeting. This research was supported in part by faculty research funds from the University of California, Santa Cruz, and NASA University Aligned Research Program Award TO.030.MM.D.

sensus [19], and deployment [7, 9]. The distributed algorithms in these works achieve the desired coordination task asymptotically over an infinite-time horizon.

Statement of contributions. In this paper, we introduce the normalized and signed gradient descent flows associated to a differentiable function. We characterize their convergence properties via nonsmooth stability analysis. We also identify general conditions under which these flows reach in finite time the set of critical points of the function. To do this, we extend recent results on the stability and convergence properties of general nonsmooth dynamical systems via locally Lipschitz and regular Lyapunov functions. In particular, we develop two novel results involving second-order information on the evolution of the Lyapunov function along system solutions to establish finite-time convergence. The applicability of these results is not restricted to gradient flows, and they can indeed be used in other setups with discontinuous vector fields and locally Lipschitz functions.

We discuss the applicability of the results on nonsmooth gradient flows to various design techniques for distributed multi-agent coordination. Consider a coordination algorithm defined via the gradient of an aggregate objective function that encodes a desired task. We analyze the coordination algorithms designed via the normalized and signed versions of the gradient, and characterize their scalability properties via the notion of spatially distributed map. In particular, we show how network consensus problems fit nicely into this scheme. We propose two coordination algorithms based on the Laplacian of the communication graph that are guaranteed to achieve consensus in finite time. The normalized gradient descent of the Laplacian potential is not distributed over the communication graph and achieves average-consensus, i.e., consensus at the average of the initial agents' states. The signed gradient descent of the Laplacian potential is distributed over the communication graph and achieves average-max-min-consensus, i.e., consensus at the average of the maximum and the minimum values of the initial agents' states. We also consider networks with switching connected communication topologies.

Organization. Section 2 introduces differential equations with discontinuous right-hand sides and presents various nonsmooth tools for stability analysis. In particular, we develop two novel results involving second-order information and finite-time convergence. Section 3 introduces the normalized and signed versions of the gradient descent flow of a differentiable function and characterizes their convergence properties. Conditions are given under which these flows converge in finite time. Section 4 discusses the application of the results to coordination algorithms for multi-agent systems paying special attention to distributed implementations and network consensus problems. We gather our conclusions in Section 5.

Notation. The set of strictly positive natural (respectively real) numbers is denoted by $\mathbb{N}$ (respectively $\mathbb{R}_+$). For $d \in \mathbb{N}$, let $e_1, \dots, e_d$ be the standard orthonormal basis of $\mathbb{R}^d$. For $x \in \mathbb{R}^d$, denote by $\|x\|_1$ and $\|x\|_2$ the 1-norm and the Euclidean norm of x, respectively. Denote by $v \cdot w$ the inner product of $v, w \in \mathbb{R}^d$, and by v' the transpose of $v \in \mathbb{R}^d$. For $x \in \mathbb{R}^d$, let $\operatorname{sgn}(x) = (\operatorname{sgn}(x_1), \dots, \operatorname{sgn}(x_d)) \in \mathbb{R}^d$. Let $\mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^d$. For $S \in \mathbb{R}^d$, let $\operatorname{co}(S)$ denote its convex closure. Define also $\operatorname{diag}((\mathbb{R}^d)^n) = \{(p, \dots, p) \in (\mathbb{R}^d)^n \mid p \in \mathbb{R}^d\}$ for $n \in \mathbb{N}$. Given a positive semidefinite $d \times d$-matrix A, let $H_0(A) \subset \mathbb{R}^d$ denote the eigenspace corresponding to the eigenvalue 0 (if A is positive definite, set $H_0(A) = \{0\}$). We denote by $\pi_A : \mathbb{R}^d \to H_0(A)$ the orthogonal projection onto $H_0(A)$. Let $\lambda_2(A)$ and $\lambda_d(A)$ be the smallest non-zero and greatest eigenvalue of A, respectively, i.e. $\lambda_2(A) = \min\{\lambda \mid \lambda > 0 \text{ and } \lambda \text{ eigenvalue of } A\}$ and $\lambda_d(A) = \max\{\lambda \mid \lambda \text{ eigenvalue of } A\}$. It is easy to see

$$u' A\, u \geq \lambda_2(A)\, \|u - \pi_{H_0(A)}(u)\|_2^2, \quad u \in \mathbb{R}^d. \tag{1}$$

2 Nonsmooth stability analysis

This section introduces differential equations with discontinuous right-hand sides and presents various nonsmooth tools to analyze their stability properties. We present two novel results on the second-order evolution of locally Lipschitz functions along the solutions of the system and on finite-time convergence.

2.1 *Differential equations with discontinuous right-hand sides*

For differential equations with discontinuous right-hand sides we understand the solutions in terms of differential inclusions following [12]. Let $F : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ be a set-valued map. Consider the differential inclusion

$$\dot{x} \in F(x)\,. \tag{2}$$

A solution to this equation on an interval $[t_0, t_1] \subset \mathbb{R}$ is defined as an absolutely continuous function $x : [t_0, t_1] \to \mathbb{R}^d$ such that $\dot{x}(t) \in F(x(t))$ for almost all $t \in [t_0, t_1]$. Now, consider the differential equation

$$\dot{x}(t) = X(x(t))\,, \tag{3}$$

where $X : \mathbb{R}^d \to \mathbb{R}^d$ is measurable and essentially locally bounded [12]. We understand the solution of (3) in the Filippov sense. For each $x \in \mathbb{R}^d$, consider the set

$$K[X](x) = \bigcap_{\delta > 0} \bigcap_{\mu(S) = 0} \operatorname{co}\{X(B_d(x, \delta) \setminus S)\}\,, \tag{4}$$

where μ denotes the usual Lebesgue measure in $\mathbb{R}^d$. A Filippov solution of (3) on an interval $[t_0, t_1] \subset \mathbb{R}$ is defined as a solution of the differential inclusion

$$\dot{x} \in K[X](x)\,. \tag{5}$$

Since the set-valued map $K[X] : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ is upper semicontinuous with nonempty, compact, convex values and locally bounded, the existence of Filippov solutions of (3) is guaranteed (cf. [12]). A set M is *weakly invariant* (respectively *strongly invariant*) for (3) if for each $x_0 \in M$, M contains a maximal solution (respectively all maximal solutions) of (3).

2.2 Stability analysis via nonsmooth Lyapunov functions

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz function. From Rademacher's Theorem [5], we know that locally Lipschitz functions are differentiable a.e. Let $\Omega_f \subset \mathbb{R}^d$ denote the set of points where f fails to be differentiable. The *generalized gradient of f at $x \in \mathbb{R}^d$* (cf. [5]) is defined by

$$\partial f(x) = \mathrm{co}\left\{ \lim_{i\to+\infty} df(x_i) \mid x_i \to x \,,\; x_i \notin S \cup \Omega_f \right\},$$

where S can be any set of zero measure. Note that if f is continuously differentiable, then $\partial f(x) = \{df(x)\}$.

Given a locally Lipschitz function f, the *set-valued Lie derivative of f with respect to X* at x (cf. [2, 7]) is

$$\begin{aligned}\widetilde{\mathcal{L}}_X f(x) = \{a \in \mathbb{R} \mid \exists v \in K[X](x) \text{ such that} \\ \zeta \cdot v = a \,,\; \forall \zeta \in \partial f(x)\}.\end{aligned}$$

For $x \in \mathbb{R}^d$, $\widetilde{\mathcal{L}}_X f(x)$ is a closed and bounded interval in $\mathbb{R}$, possibly empty. If f is continuously differentiable at x, then $\widetilde{\mathcal{L}}_X f(x) = \{df \cdot v \mid v \in K[X](x)\}$. If, in addition, X is continuous at x, then $\widetilde{\mathcal{L}}_X f(x)$ corresponds to the singleton $\{\mathcal{L}_X f(x)\}$, the usual Lie derivative of f in the direction of X at x. The next result, taken from [2], states that the set-valued Lie derivative allows us to study the evolution of a function along the Filippov solutions.

Theorem 1 *Let $x : [t_0, t_1] \to \mathbb{R}^d$ be a Filippov solution of (3). Let f be a locally Lipschitz and regular function. Then $t \mapsto f(x(t))$ is absolutely continuous, $\frac{d}{dt}\big(f(x(t))\big)$ exists a.e. and $\frac{d}{dt}\left(f(x(t))\right) \in \widetilde{\mathcal{L}}_X f(x(t))$ a.e.*

Sometimes, we can also look at second-order information for the evolution of a function along the Filippov solutions. This is what we prove in the next result.

Proposition 2 *Let $x : [t_0, t_1] \to \mathbb{R}^d$ be a Filippov solution of (3). Let f be a locally Lipschitz and regular function. Assume that $\widetilde{\mathcal{L}}_X f : \mathbb{R}^d \to 2^{\mathbb{R}}$ is single-valued, i.e., it takes the form $\widetilde{\mathcal{L}}_X f : \mathbb{R}^d \to \mathbb{R}$, and assume it is a Lipschitz and regular function. Then $\frac{d^2}{dt^2}\left(f(x(t))\right)$ exists a.e. and $\frac{d^2}{dt^2}\left(f(x(t))\right) \in \widetilde{\mathcal{L}}_X(\widetilde{\mathcal{L}}_X f)(x(t))$ a.e.*

PROOF. Applying Theorem 1 to the functions f and $\widetilde{\mathcal{L}}_X f$, respectively, we deduce that (i) the map $t \mapsto f(x(t))$ is absolutely continuous, and $\frac{d}{dt}\big(f(x(t))\big) = \widetilde{\mathcal{L}}_X f(x(t))$ a.e., and, (ii) the map $t \mapsto \widetilde{\mathcal{L}}_X f(x(t))$ is absolutely continuous, and $\frac{d}{dt}\big(\widetilde{\mathcal{L}}_X f(x(t))\big) = \widetilde{\mathcal{L}}_X(\widetilde{\mathcal{L}}_X f)(x(t))$ a.e. Since $t \mapsto \widetilde{\mathcal{L}}_X f(x(t))$ is continuous, the expression

$$\begin{aligned} f(x(t)) &= f(x(t_0)) + \int_{t_0}^{t} \frac{d}{dt}\left(f(x(s))\right) ds \\ &= f(x(t_0)) + \int_{t_0}^{t} \widetilde{\mathcal{L}}_X f(x(s)) ds. \end{aligned}$$

and the second fundamental theorem of calculus implies that $t \mapsto f(x(t))$ is actually continuously differentiable, and therefore $\frac{d}{dt}\left(f(x(t))\right) = \widetilde{\mathcal{L}}_X f(x(t))$ for all t. Now, using (ii), we conclude that $t \mapsto \frac{d}{dt}\left(f(x(t))\right)$ is differentiable a.e. and $\frac{d^2}{dt^2}\left(f(x(t))\right) \in \widetilde{\mathcal{L}}_X(\widetilde{\mathcal{L}}_X f)(x(t))$ a.e. •

The following result is a generalization of LaSalle principle for differential equations of the form (3) with nonsmooth Lyapunov functions. The formulation is taken from [2], and slightly generalizes [21].

Theorem 3 (LaSalle Invariance Principle): *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz and regular function. Let $x_0 \in S \subset \mathbb{R}^d$, with S compact and strongly invariant for (3). Assume that either $\max \widetilde{\mathcal{L}}_X f(x) \leq 0$ or $\widetilde{\mathcal{L}}_X f(x) = \emptyset$ for all $x \in S$. Let $Z_{X,f} = \{x \in \mathbb{R}^d \mid 0 \in \widetilde{\mathcal{L}}_X f(x)\}$. Then, any solution $x : [t_0, +\infty) \to \mathbb{R}^d$ of (3) starting from x_0 converges to the largest weakly invariant set M contained in $\overline{Z}_{X,f} \cap S$. Moreover, if the set M is a finite collection of points, then the limit of all solutions starting at x_0 exists and equals one of them.*

The following result is taken from [7].

Proposition 4 (Finite-time convergence with first-order information): *Under the same assumptions of Theorem 3, further assume that there exists a neighborhood U of $Z_{X,f} \cap S$ in S such that $\max \widetilde{\mathcal{L}}_X f < -\epsilon < 0$ a.e. on $U \setminus (Z_{X,f} \cap S)$. Then, any solution $x : [t_0, +\infty) \to \mathbb{R}^d$ of (3) starting at $x_0 \in S$ reaches $Z_{X,f} \cap S$ in finite time. Moreover, if $U = S$, then the convergence time is upper bounded by $f(x_0)/\epsilon$.*

Often times, first-order information is inconclusive to assess finite-time convergence. The next result uses second-order information to arrive at a satisfactory answer.

Theorem 5 (Finite-time convergence with second-order information): *Under the same assumptions of Theorem 3, further assume that (i) $x \in \mathbb{R}^d \mapsto \widetilde{\mathcal{L}}_X f(x)$ is single-valued, Lipschitz and regular; and (ii) there exists a neighborhood U of $Z_{X,f} \cap S$ in S such that $\max \widetilde{\mathcal{L}}_X(\widetilde{\mathcal{L}}_X f) > \epsilon > 0$ a.e. on $U \setminus (Z_{X,f} \cap S)$. Then, any solution $x : [t_0, +\infty) \to \mathbb{R}^d$ of (3) starting at $x_0 \in S$ reaches $Z_{X,f} \cap S$ in finite time. Moreover, if $U = S$, then the convergence time is upper bounded by $-\widetilde{\mathcal{L}}_X f(x_0)/\epsilon$.*

PROOF. Since $x \in \mathbb{R}^d \mapsto \widetilde{\mathcal{L}}_X f(x)$ is single-valued and continuous, we have $Z_{X,f} = \{x \in \mathbb{R}^d \mid \widetilde{\mathcal{L}}_X f(x) = 0\}$ and that this set is closed. Moreover, the largest weakly invariant set M contained in $Z_{X,f} \cap S$ is precisely $Z_{X,f} \cap S$

itself. Let $x : [t_0, +\infty) \to \mathbb{R}^d$ be a solution of (3) starting from $x_0 \in S \setminus Z_{X,f}$. We reason by contradiction. Assume there does not exist T such that $x(T) \in Z_{X,f}$. By the LaSalle Invariance Principle, $x(t) \to Z_{X,f}$ when $t \to +\infty$, and therefore there exists $t_* \geq t_0$ such that $x(t) \in U$ for all $t \geq t_*$. Assumption (i) guarantees that Proposition 2 can be applied. Combining this with assumption (ii), we write for $g(t) = \frac{d}{dt}\big(f(x(t))\big)$,

$$g(t) = g(t_*) + \int_{t_*}^{t} \frac{d}{ds} g(s) ds > g(t_*) + \epsilon(t - t_*)\,,\ t > t_*.$$

Since $x(t_*) \notin Z_{X,f}$ by hypothesis, then $g(t_*) < 0$. Noting that $t \mapsto g(t)$ is continuous, we deduce that there exists $T < t_* - \frac{g(t_*)}{\epsilon}$ such that $g(T) = 0$, i.e., $x(T) \in Z_{X,f}$, which is a contradiction. The upper bound on the convergence time can be deduced using similar arguments.•

Remark 6 *Under the hypotheses of Theorem 5, one can see that a rescaling of the differential equation* (3) *of the form $\dot{y}(t) = s \cdot X(y(t))$, with $s \in \mathbb{R}$, results in a (finite) convergence time upper bounded by $-\widetilde{\mathcal{L}}_X f(x_0)/(s \cdot \epsilon)$.*

3 Nonsmooth gradient flows with finite-time convergence

Here, we formally introduce the normalized and signed gradient dynamical systems associated with a differentiable function. We characterize their general asymptotic convergence properties. Building on the novel results of Section 2, we identify conditions on the differentiable function under which convergence is reached in finite time. Consider the following dynamical systems on $\mathbb{R}^d$

$$\dot{x} = -\frac{\operatorname{grad}(f)(x)}{\|\operatorname{grad}(f)(x)\|_2}\,, \tag{6a}$$

$$\dot{x} = -\operatorname{sgn}(\operatorname{grad}(f)(x))\,. \tag{6b}$$

Both equations have discontinuous right-hand sides. Hence, we understand their solutions in the Filippov sense. We now describe their associated set-valued maps.

Lemma 7 *The Filippov set-valued maps associated with the discontinuous vector fields* (6a) *and* (6b) *are*

$$K\Big[\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}\Big](x) = \operatorname{co}\Big\{\lim_{i\to+\infty}\frac{\operatorname{grad}(f)(x_i)}{\|\operatorname{grad}(f)(x_i)\|_2} \mid x_i \to x,\ \operatorname{grad}(f)(x_i) \neq 0\Big\},$$

$$K\big[\operatorname{sgn}(\operatorname{grad}(f))\big](x) = \big\{v \in \mathbb{R}^d \mid v_i = \operatorname{sgn}(\operatorname{grad}_i(f)(x)) \textit{ if } \operatorname{grad}_i(f)(x) \neq 0 \textit{ and } v_i \in [-1,1] \textit{ if } \operatorname{grad}_i(f)(x) = 0,\ \textit{for } i \in \{1,\dots,d\}\big\}.$$

Note $K\big[\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}\big](x) = \frac{\operatorname{grad}(f)(x)}{\|\operatorname{grad}(f)(x)\|_2}$ *if* $\operatorname{grad}(f)(x) \neq 0$. The proof of this result follows from the definition (4) of the operator K and the particular forms of (6a) and (6b).

For a differentiable function f, let $\operatorname{Critical}(f) = \big\{x \in \mathbb{R}^d \mid \operatorname{grad}(f)(x) = 0\big\}$ denote the set of its critical points. The next result establishes the general asymptotic properties of the flows in (6).

Proposition 8 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. Let $x_0 \in S \subset \mathbb{R}^d$, with S compact and strongly invariant for* (6a) *(respectively, for* (6b)*). Then each solution of equation* (6a) *(respectively equation* (6b)*) starting from x_0 asymptotically converges to* $\operatorname{Critical}(f)$.

PROOF. For equation (6a), if $\operatorname{grad}(f)(x) \neq 0$, then

$$\begin{aligned}\widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}} f(x) &= \Big\{\frac{\operatorname{grad}(f)(x)}{\|\operatorname{grad}(f)(x)\|_2} \cdot \operatorname{grad}(f)(x)\Big\} \\ &= \big\{\|\operatorname{grad}(f)(x)\|_2\big\}.\end{aligned}$$

If, instead, $\operatorname{grad}(f)(x) = 0$, then $\widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}} f(x) = \{0\}$. Therefore, we deduce

$$\widetilde{\mathcal{L}}_{-\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}} f(x) = -\|\operatorname{grad}(f)(x)\|_2, \quad \text{for all } x \in \mathbb{R}^d.$$

Consequently, $Z_{-\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}, f} = \operatorname{Critical}(f)$ is closed, and LaSalle Invariance Principle (cf. Theorem 3) implies that each solution of (6a) starting from x_0 asymptotically converges to the largest weakly invariant set M contained in $\operatorname{Critical}(f) \cap S$, which is $\operatorname{Critical}(f) \cap S$ itself.

For equation (6b), we have $a \in \widetilde{\mathcal{L}}_{\operatorname{sgn}(\operatorname{grad}(f))} f(x)$ if and only if there exists $v \in K\big[\operatorname{sgn}(\operatorname{grad}(f))\big](x)$ such that $a = v \cdot \operatorname{grad}(f)(x)$. From Lemma 7, we deduce that $a = \operatorname{sgn}(\operatorname{grad}_1(f)(x)) \cdot \operatorname{grad}_1(f)(x) + \dots + \operatorname{sgn}(\operatorname{grad}_n(f)(x)) \cdot \operatorname{grad}_n(f)(x) = \|\operatorname{grad}(f)(x)\|_1$. Therefore, we deduce

$$\widetilde{\mathcal{L}}_{-\operatorname{sgn}(\operatorname{grad}(f))} f(x) = \{-\|\operatorname{grad}(f)(x)\|_1\}.$$

Consequently, $Z_{-\operatorname{sgn}(\operatorname{grad}(f)), f} = \operatorname{Critical}(f)$ is closed, and LaSalle Invariance Principle implies that each solution of (6b) starting from x_0 asymptotically converges to the largest weakly invariant set M contained in $\operatorname{Critical}(f) \cap S$, which is $\operatorname{Critical}(f) \cap S$ itself. •

Let us now discuss the finite-time convergence properties of the vector fields (6). Note that Proposition 4 cannot be applied. Indeed,

$$\max \widetilde{\mathcal{L}}_{-\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}} f(x) = -\|\operatorname{grad}(f)(x)\|_2\,,$$

$$\max \widetilde{\mathcal{L}}_{-\operatorname{sgn}(\operatorname{grad}(f))} f(x) = -\|\operatorname{grad}(f)(x)\|_1\,,$$

and both $\inf_{x \in U \setminus \operatorname{Critical}(f) \cap S} \|\operatorname{grad}(f)(x)\|_2 = 0$ and $\inf_{x \in U \setminus \operatorname{Critical}(f) \cap S} \|\operatorname{grad}(f)(x)\|_1 = 0$, for any neighborhood U of $\operatorname{Critical}(f) \cap S$ in S. Hence, the hypotheses of Proposition 4 are not verified by either (6a) or (6b).

Under additional conditions, one can establish stronger convergence properties of (6). We show this next.

Theorem 9 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a second-order differentiable function. Let $x_0 \in S \subset \mathbb{R}^d$, with S compact and strongly invariant for* (6a) *(respectively, for* (6b)*). Assume there exists a neighborhood V of* $\operatorname{Critical}(f) \cap S$ *in S where either one of the following conditions hold:*

(i) for all $x \in V$, the Hessian $\operatorname{Hess}(f)(x)$ *is positive definite; or*
(ii) for all $x \in V \setminus (\operatorname{Critical}(f) \cap S)$, the Hessian $\operatorname{Hess}(f)(x)$ *is positive semidefinite, the multiplicity of the eigenvalue* 0 *is constant, and* $\operatorname{grad}(f)(x)$ *is orthogonal to the eigenspace of* $\operatorname{Hess}(f)(x)$ *corresponding to the eigenvalue* 0.

Then each solution of (6a) *(respectively* (6b)*) starting from x_0 converges in finite time to a critical point of f. Furthermore, if $V = S$, then the convergence time of the solutions of* (6a) *(respectively* (6b)*) starting from x_0 is upper bounded by*

$$\frac{1}{\lambda_0}\|\operatorname{grad}(f)(x_0)\|_2 \quad \Big(\textit{respectively } \frac{1}{\lambda_0}\|\operatorname{grad}(f)(x_0)\|_1\Big),$$

where $\lambda_0 = \min_{x\in S} \lambda_2(\operatorname{Hess}(f)(x))$.

PROOF. Our strategy is to show that the hypotheses of Theorem 5 are verified by both vector fields. From Proposition 8, we know that each solution of (6a) (respectively (6b)) starting from x_0 converges $\operatorname{Critical}(f)$. Let us take an open set $U \subset S$ such that $\operatorname{Critical}(f)\cap S \subset U \subset \overline{U} \subset V$. Since S is compact, then $\overline{U}$ is also compact. By continuity, under either assumption (i) or assumption (ii), the function $\lambda_2(\operatorname{Hess}(f)) : \overline{U} \to \mathbb{R}$, $x \mapsto \lambda_2(\operatorname{Hess}(f)(x))$, reaches its minimum on $\overline{U}$, i.e, there exists $\lambda_0 > 0$ such that $\lambda_2(\operatorname{Hess}(f)(x)) \geq \lambda_0$ for all $x \in \overline{U}$. Moreover, from (1), we have for all $u \in \mathbb{R}^d$,

$$\begin{aligned} &u' \operatorname{Hess}(f)(x)\, u \geq \\ &\quad\geq \lambda_2(\operatorname{Hess}(f)(x))\, \|u - \pi_{H_0(\operatorname{Hess}(f)(x))}(u)\|_2^2. \end{aligned} \tag{7}$$

For (6a), recall from the proof of Proposition 8, that the function $x \in \mathbb{R}^d \mapsto \widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}} f(x) = \|\operatorname{grad}(f)(x)\|_2$ is single-valued, locally Lipschitz and regular, and hypothesis (i) in Theorem 5 is satisfied. Additionally, $Z_{-\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}, f} = \operatorname{Critical}(f)$. Let us take $x \notin \operatorname{Critical}(f)$, and let us compute $\widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}}(\widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}} f)(x)$. Noting

$$\partial(\|\operatorname{grad}(f)\|_2)(x) = \left\{ \operatorname{Hess}(f)(x)\frac{\operatorname{grad}(f)(x)}{\|\operatorname{grad}(f)(x)\|_2} \right\},$$

we deduce

$$\begin{aligned} &\widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}}(\widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}} f)(x) = \\ &\quad \frac{\operatorname{grad}(f)(x)'}{\|\operatorname{grad}(f)(x)\|_2}\operatorname{Hess}(f)(x)\frac{\operatorname{grad}(f)(x)}{\|\operatorname{grad}(f)(x)\|_2}. \end{aligned} \tag{8}$$

Let $x \in \overline{U} \setminus (\operatorname{Critical}(f) \cap S)$. Now, note that under any of the two assumptions in the statement of the theorem, we have $\pi_{H_0(\operatorname{Hess} f(x))}\big(\frac{\operatorname{grad}(f)(x)}{\|\operatorname{grad}(f)(x)\|_2}\big) = 0$. Therefore, using (7) in equation (8), we conclude

$$\begin{aligned} &\widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}}(\widetilde{\mathcal{L}}_{\frac{\operatorname{grad}(f)}{\|\operatorname{grad}(f)\|_2}} f)(x) \geq \\ &\qquad \geq \lambda_2(\operatorname{Hess}(f)(x))\left\|\frac{\operatorname{grad}(f)(x)}{\|\operatorname{grad}(f)(x)\|_2}\right\|_2^2 = \\ &\qquad = \lambda_2(\operatorname{Hess}(f)(x)) \geq \lambda_0 > 0, \end{aligned}$$

for $x \in \overline{U} \setminus (\operatorname{Critical}(f) \cap S)$. Hence, hypothesis (ii) in Theorem 5 is also verified, and we deduce that the set $\operatorname{Critical}(f)$ is reached in finite time, which in particular implies that the limit of any solution of equation (6a) starting from $x_0 \in S$ exists and is reached in finite time.

For (6b), recall from the proof of Proposition 8, that the function $x \in \mathbb{R}^d \mapsto \widetilde{\mathcal{L}}_{\operatorname{sgn}(\operatorname{grad}(f))} f(x) = \|\operatorname{grad}(f)(x)\|_1$ is single-valued, locally Lipschitz and regular, and hypothesis (i) in Theorem 5 is satisfied. Additionally, $Z_{-\operatorname{sgn}(\operatorname{grad}(f)),f} = \operatorname{Critical}(f)$. Let us take $x \notin \operatorname{Critical}(f)$, and let us compute $\widetilde{\mathcal{L}}_{\operatorname{sgn}(\operatorname{grad}(f))}(\widetilde{\mathcal{L}}_{\operatorname{sgn}(\operatorname{grad}(f))} f)(x)$. By definition, $a \in \widetilde{\mathcal{L}}_{\operatorname{sgn}(\operatorname{grad}(f))}(\widetilde{\mathcal{L}}_{\operatorname{sgn}(\operatorname{grad}(f))} f)(x)$ if and only if there exists $v \in K\big[\operatorname{sgn}(\operatorname{grad}(f))\big](x)$ such that $a = v \cdot \zeta$, for any $\zeta \in \partial(\|\operatorname{grad}(f)\|_1)(x)$. Note that

$$\begin{aligned} &\partial(\|\operatorname{grad}(f)\|_1)(x) = \\ &\big\{\zeta \in \mathbb{R}^d \mid \zeta = \operatorname{Hess}(f)(x)\,\eta, \text{ for some } \eta \in \mathbb{R}^d \text{ with} \\ &\quad \eta_i = \operatorname{sgn}(\operatorname{grad}_i(f)(x)) \text{ if } \operatorname{grad}_i(f)(x) \neq 0 \text{ and} \\ &\quad \eta_i \in [-1, 1] \text{ if } \operatorname{grad}_i(f)(x) = 0, \text{ for } i \in \{1, \dots, d\}\big\}. \end{aligned}$$

In particular, $\operatorname{Hess}(f)(x)\, v \in \partial(\|\operatorname{grad}(f)\|_1)(x)$. Then

$$a = v' \operatorname{Hess}(f)(x)\, v.$$

Let us now decompose v as $v = \pi_{H_0(x)}(v) + (v - \pi_{H_0(x)}(v))$, where $\pi_{H_0(x)}(v) \in H_0(x)$ and $v - \pi_{H_0(x)}(v) \in H_0(x)^\perp$. Because $v \in K\big[\operatorname{sgn}(\operatorname{grad}(f))\big](x)$, we deduce $v \cdot \operatorname{grad}(f)(x) = \|\operatorname{grad}(f)(x)\|_1$. Let $x \in \overline{U} \setminus (\operatorname{Critical}(f) \cap S)$. Under either assumption (i) or (ii),

$$\begin{aligned} \|\operatorname{grad}(f)(x)\|_1 &= v \cdot \operatorname{grad}(f)(x) \\ &= (v - \pi_{H_0(x)}(v)) \cdot \operatorname{grad}(f)(x) \\ &\leq \|v - \pi_{H_0(x)}(v)\|_2 \|\operatorname{grad}(f)(x)\|_2. \end{aligned}$$

Using $\|u\|_1 \geq \|u\|_2$ for any $u \in \mathbb{R}^d$, we deduce from this equation that $\|v - \pi_{H_0(x)}(v)\|_2 \geq 1$. Therefore, using (7)

$$\begin{aligned} a = v' \operatorname{Hess}(f)(x)\, v &\geq \lambda_2(\operatorname{Hess}(f)(x))\, \|v - \pi_{H_0(x)}(v)\|_2^2 \\ &\geq \lambda_2(\operatorname{Hess}(f)(x)) \geq \lambda_0 > 0, \end{aligned}$$

for $x \in \overline{U} \setminus (\operatorname{Critical}(f) \cap S)$. Consequently, we get $\max \widetilde{\mathcal{L}}_{\operatorname{sgn}(\operatorname{grad}(f))}(\widetilde{\mathcal{L}}_{\operatorname{sgn}(\operatorname{grad}(f))} f) > \lambda_0 > 0$ on

$\overline{U} \setminus (\text{Critical}(f) \cap S)$. Hence, hypothesis (ii) in Theorem 5 is also verified, and we deduce that the set Critical(f) is reached in finite time, which in particular implies that the limit of any solution of equation (6b) starting from $x_0 \in S$ exists and is reached in finite time.

The upper bounds on the convergence time of the solutions of both flows also follow from Theorem 5. •

Corollary 10 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a second-order differentiable function. Let $x_0 \in S \subset \mathbb{R}^d$, with S compact and strongly invariant for* (6a) *(respectively, for* (6b)*). Assume that for each $x \in \text{Critical}(f) \cap S$, the Hessian* Hess$(f)(x)$ *is positive definite. Then each solution of equation* (6a) *(respectively equation* (6b)*) starting from x_0 converges in finite time to a minimum of f.*

4 Applications to motion coordination

Here we discuss the application of the results on the proposed nonsmooth gradient dynamical systems to the design of multi-agent coordination algorithms. We start by presenting the notion of proximity graphs and of spatially distributed map. These concepts will allow us to characterize the scalability properties of coordination algorithms designed via the normalized and signed gradients of suitable objective functions. We end the section illustrating our results in network consensus problems.

4.1 Proximity graphs and spatially-distributed maps

We introduce some concepts regarding proximity graphs for point sets in $\mathbb{R}^d$. We assume the reader is familiar with the standard notions of graph theory as defined in [11, Chapter 1]. Given a vector space $\mathbb{V}$, let $\mathbb{F}(\mathbb{V})$ be the collection of finite subsets of $\mathbb{V}$. Accordingly, $\mathbb{F}(\mathbb{R}^d)$ is the collection of finite point sets in $\mathbb{R}^d$; elements of $\mathbb{F}(\mathbb{R}^d)$ are of the form $\{p_1, \dots, p_m\} \subset \mathbb{R}^d$, where $p_1, \dots, p_m$ are distinct points in $\mathbb{R}^d$. Let $\mathbb{G}(\mathbb{R}^d)$ be the set of undirected graphs whose vertex set is an element of $\mathbb{F}(\mathbb{R}^d)$. Finally, let $i_{\mathbb{F}} : (\mathbb{R}^d)^n \to \mathbb{F}(\mathbb{R}^d)$ be the natural immersion, i.e., $i_{\mathbb{F}}(P)$ is the point set that contains only the distinct points in $P = (p_1, \dots, p_n) \in (\mathbb{R}^d)^n$. The cardinality of $i_{\mathbb{F}}(p_1, \dots, p_n)$ is in general less than or equal to n.

A *proximity graph function* $\mathcal{G} : (\mathbb{R}^d)^n \to \mathbb{G}(\mathbb{R}^d)$ associates to a tuple $P \in (\mathbb{R}^d)^n$ an undirected graph with vertex set $i_{\mathbb{F}}(\mathcal{P})$ and edge set $\mathcal{E}_{\mathcal{G}}(P)$, where $\mathcal{E}_{\mathcal{G}} : (\mathbb{R}^d)^n \to \mathbb{F}(\mathbb{R}^d \times \mathbb{R}^d)$. The edge set of a proximity graph depends on the location of its vertices. Examples include the complete graph, the r-disk graph, the Euclidean Minimum Spanning Three, the Delaunay graph, etc. see [9, 10, 16]. To each proximity graph $\mathcal{G}$, one associates the *set of neighbors map* $\mathcal{N}_{\mathcal{G}} : (\mathbb{R}^d)^n \to (\mathbb{F}(\mathbb{R}^d))^n$, defined by

$$\mathcal{N}_{\mathcal{G},i}(P) = \{p_j \in i_{\mathbb{F}}(P) \mid j \neq i \text{ and } (p_i, p_j) \in \mathcal{E}_{\mathcal{G}}(P)\}.$$

Any standard directed graph G with vertex set $\{1, \dots, n\}$ and edge set $E \subset \{1, \dots, n\} \times \{1, \dots, n\}$ can be seen as a proximity graph where, for each $P \in (\mathbb{R}^d)^n$, $(p_i, p_j) \in \mathcal{E}_{\mathcal{G}}(P)$ if and only if $(i, j) \in E$. In this case, $\mathcal{N}_{\mathcal{G},i}(P) = \mathcal{N}_{G,i} = \{j \in \{1, \dots, n\} \mid (i, j) \in E\}$.

Given a set Y and a proximity graph function $\mathcal{G}$, a map $T : (\mathbb{R}^d)^n \to Y^n$ is *spatially distributed over* $\mathcal{G}$ if there exist a map $\widetilde{T} : \mathbb{R}^d \times \mathbb{F}(\mathbb{R}^d) \to Y$, with the property that, for all $(p_1, \dots, p_n) \in (\mathbb{R}^d)^n$ and for all $j \in \{1, \dots, n\}$,

$$T_j(p_1, \dots, p_n) = \widetilde{T}(p_j, \mathcal{N}_{\mathcal{G},j}(p_1, \dots, p_n)),$$

where T_j is the jth-component of T. In other words, the jth component of a spatially distributed map at $(p_1, \dots, p_n)$ can be computed with the knowledge of the vertex p_j and the neighboring vertices in $\mathcal{G}(p_1, \dots, p_n)$.

4.2 Gradient-based coordination algorithms

There are a number of gradient-based coordination algorithms proposed in the literature to optimize aggregate objective functions encoding various network tasks. We roughly distinguish between two classes: (i) algorithms directly derived from the gradient of a network-wide aggregate objective function (e.g., the deployment algorithms in [9], the consensus algorithm in [19] and the cohesiveness algorithms in [13, 18, 23]), and algorithms where each agent follows the gradient of a meaningful local objective function whose optimization helps the network achieve the desired global task (e.g. the rendezvous strategies in [1, 8, 17], and the interaction laws in [7]).

The general idea is as follows: consider a network composed of n agents with sensing, computing, communication, and motion control capabilities. The state of the ith agent, denoted $p_i \in \mathbb{R}^d$, might correspond, depending on the specific problem, to the location of the agent in space, or to other physical quantities. The state p_i evolves according to a first-order continuous dynamics

$$\dot{p}_i(t) = u_i. \tag{9}$$

The control u_i takes values in a bounded subset of $\mathbb{R}^d$. Additionally, the network communication topology is described by a proximity graph $\mathcal{G}$. Specifically, the ith agent is capable of transmitting information to the jth agent if and only if $p_j \in \mathcal{N}_{\mathcal{G},p_i}(\mathcal{P})$. Typical proximity graphs are the r-disk graph (where two agents are neighbors if they lie within a distance $r \in \mathbb{R}_+$ from each other) or the visibility graph (where two agents are neighbors if they are visible to each other), see [9].

The last ingredient is an aggregate objective function $\mathcal{H} : (\mathbb{R}^d)^n \to \mathbb{R}$ encoding the desired coordination task, i.e., such that its critical points correspond to the network configurations where the coordination task is achieved.

Algorithms designed from the aggregate objective function. Assume the function $\mathcal{H}$ has the additional property that its gradient grad$(\mathcal{H}) : (\mathbb{R}^d)^n \to (\mathbb{R}^d)^n$ is spatially distributed over the proximity graph $\mathcal{G}$. One then can set up the gradient coordination algorithm

$$\dot{p}_i(t) = -\frac{\partial \mathcal{H}}{\partial p_i}(p_1(t), \dots, p_n(t)), \quad i \in \{1, \dots, n\}, \tag{10}$$

which is spatially distributed over $\mathcal{G}$.

Following (6), consider the normalized and signed versions of the gradient coordination algorithm (10)

$$\dot{p}_i(t) = -\frac{\frac{\partial \mathcal{H}}{\partial p_i}(p_1(t), \dots, p_n(t))}{\| \frac{\partial \mathcal{H}}{\partial P}(p_1(t), \dots, p_n(t)) \|_2}, \tag{11a}$$

$$\dot{p}_i(t) = -\operatorname{sgn}\Big(\frac{\partial \mathcal{H}}{\partial p_i}(p_1(t), \dots, p_n(t))\Big). \tag{11b}$$

Although both vector fields enjoy similar convergence properties (cf. Proposition 8), there is a fundamental difference between them, as the following results states.

Proposition 11 *Let $\mathcal{G}$ be a proximity graph other than the complete graph. Let $\mathcal{H} : (\mathbb{R}^d)^n \to \mathbb{R}$ be such that $\operatorname{grad}(\mathcal{H}) : (\mathbb{R}^d)^n \to (\mathbb{R}^d)^n$ is spatially distributed over $\mathcal{G}$. Then the coordination algorithm in* (11a) *is not spatially distributed over $\mathcal{G}$, and the coordination algorithm in* (11b) *is spatially distributed over $\mathcal{G}$.*

PROOF. In the algorithm (11a), each agent must compute the norm of $\operatorname{grad}(\mathcal{H})$, and therefore needs to know in general all agents' states. Since $\mathcal{G}$ is not complete, this computation is not spatially distributed over it. On the other hand, the algorithm (11b) is spatially distributed over $\mathcal{G}$ because $\operatorname{grad}(\mathcal{H})$ is spatially distributed over $\mathcal{G}$.•

Algorithms designed from local objective functions. To derive a control law for each specific agent, one assumes that its neighboring agents (with respect to a given proximity graph) remains fixed. One then identifies a local objective function, which is somehow related with the global aggregate objective function $\mathcal{H}$, and devises a control law to optimize it. The optimization of these local objective functions must help the network achieve the desired global task. The specific control strategy for each agent might be heuristically derived or arise naturally from the gradient information of the local objective function. In the latter case, the finite-time convergence properties of the normalized and signed gradient flows (6a) and (6b) can be invaluable in characterizing the asymptotic convergence of the algorithm. In both cases, the resulting algorithms are spatially distributed with respect to the selected proximity graph.

4.3 Network consensus problems

Here we focus on consensus problems. Let $G = (\{1, \dots, n\}, E)$ be an undirected graph with n vertices. The graph Laplacian matrix L_G associated with G (see, for instance, [11]) is defined as $L_G = \Delta_G - A_G$, where Δ_G is the degree matrix and A_G is the adjacency matrix of the graph. The Laplacian matrix has the following relevant properties: it is symmetric, positive semidefinite and has $\lambda = 0$ as an eigenvalue with eigenvector $\mathbf{1}$. More importantly, the graph G is connected if and only if $\operatorname{rank}(L_G) = n - 1$, i.e., if the eigenvalue 0 has multiplicity one. This is the reason why the eigenvalue $\lambda_2(L_G) = \min\{\lambda \mid \lambda > 0 \text{ and } \lambda \text{ eigenvalue of } L_G\}$ is termed the *algebraic connectivity* of the graph G.

In this setup, the agents' states p_i, $i \in \{1, \dots, n\}$, evolve in $\mathbb{R}$, $p_i \in \mathbb{R}$. The variable p_i does not necessarily refer to physical variables such as spatial coordinates or velocities. Two agents p_i and p_j are said to *agree* if and only if $p_i = p_j$. A meaningful function that quantifies the group disagreement in a network is the *Laplacian potential* $\Phi_G : \mathbb{R}^n \to \overline{\mathbb{R}}_+$ associated with G (see [19]),

$$\Phi_G(p_1, \dots, p_n) = \frac{1}{2} P' L_G P = \frac{1}{2} \sum_{(i,j) \in E} (p_j - p_i)^2,$$

with $P = (p_1, \dots, p_n)' \in \mathbb{R}^n$. It is clear that $\Phi_G(p_1, \dots, p_n) = 0$ if and only if all neighboring nodes in the graph G agree. If G is connected, then all nodes agree and a consensus is reached. Therefore, we want the network to reach the critical points of Φ_G. Assume G is connected. The Laplacian potential is smooth, and its gradient is $\operatorname{grad}(\Phi_G)(P) = L_G P$, which is clearly spatially distributed over the proximity graph induced by G. The gradient coordination algorithm (10) reads in this case

$$\dot{p}_i(t) = -\frac{\partial \Phi_G}{\partial p_i} = \sum_{j \in \mathcal{N}_{G,i}} (p_j(t) - p_i(t)), \tag{12}$$

for $i \in \{1, \dots, n\}$, and asymptotically converges to the critical points of Φ_G, i.e., asymptotically achieves consensus. Actually, since the system is linear, the convergence is exponential with rate equal to the algebraic connectivity of the graph. Additionally, the fact that $\mathbf{1} \cdot (L_G P) = 0$ implies that $\sum_{i=1}^n p_i$ is constant along the solutions. Therefore, each solution of (12) is convergent to a point of the form $(p_*, \dots, p_*)$, with $p_* = \frac{1}{n} \sum_{i=1}^n p_i(0)$ (this is called *average-consensus*).

Now, consider the discontinuous differential equations corresponding to (6), for $i \in \{1, \dots, n\}$,

$$\dot{p}_i(t) = \frac{\sum_{j \in \mathcal{N}_{G,i}} (p_j(t) - p_i(t))}{\| L_G P(t) \|_2}, \tag{13a}$$

$$\dot{p}_i(t) = \operatorname{sgn}\big(\sum_{j \in \mathcal{N}_{G,i}} (p_j(t) - p_i(t)) \big). \tag{13b}$$

Before analyzing the convergence properties of these flows, we identify a conserved quantity for each of them.

Proposition 12 *Define $g_1 : \mathbb{R}^n \to \mathbb{R}$, $g_2 : \mathbb{R}^n \to \mathbb{R}$ by*

$$g_1(p_1, \dots, p_n) = \sum_{i=1}^n p_i,$$
$$g_2(p_1, \dots, p_n) = \max_{i \in \{1, \dots, n\}} \{p_i\} + \min_{i \in \{1, \dots, n\}} \{p_i\}.$$

Then g_1 is constant along the solutions of (13a) *and g_2 is constant along the solutions of* (13b).

PROOF. The function g_1 is differentiable, with $\operatorname{grad}(g_1)(P) = \mathbf{1}$. For any $P = (p_1, \dots, p_n) \in \mathbb{R}^n$, $\widetilde{\mathcal{L}}_{-\frac{\operatorname{grad}(\Phi_G)}{\|\operatorname{grad}(\Phi_G)\|_2}} g_1(P) = \{0\}$. Therefore, from Theorem 1, we conclude that g_1 is constant along the solutions of (13a). On the other hand, from [5, Proposition 2.3.12], one can deduce that g_2 is locally Lipschitz and regular, with $\partial g_2(P) = \mathrm{co}\{e_j \in \mathbb{R}^n \mid j \in \{1,\dots,n\}$ with $p_j = \min_{i\in\{1,\dots,n\}}\{p_i\}\} + \mathrm{co}\{e_k \in \mathbb{R}^n \mid k \in \{1,\dots,n\}$ with $p_k = \max_{i\in\{1,\dots,n\}}\{p_i\}\}$. Let $a \in \widetilde{\mathcal{L}}_{-\operatorname{sgn}(\operatorname{grad}(\Phi_G))} g_2(P)$. By definition, there exists $v \in K\big[-\operatorname{sgn}(\operatorname{grad}(\Phi_G))\big](P)$ with

$$a = v\cdot\zeta, \quad \text{for all } \zeta \in \partial g_2(P). \tag{14}$$

If $P \notin \operatorname{diag}(\mathbb{R}^n)$, then $\partial g_2(P) = \mathbb{R}^d$, and, for (14) to hold, necessarily $v = (0,\dots,0)$. Therefore, $a = 0$. If $P \notin \operatorname{diag}(\mathbb{R}^n)$, there exist $j,k \in \{1,\dots,n\}$ with $p_j = \min_{i\in\{1,\dots,n\}}\{p_i\}$, $p_k = \max_{i\in\{1,\dots,n\}}\{p_i\}$ such that

$$\sum_{i\in\mathcal{N}_{G,j}} (p_i - p_j) > 0, \qquad \sum_{i\in\mathcal{N}_{G,k}} (p_i - p_k) < 0,$$

and hence, from Lemma 7, $v_j = 1$ and $v_k = -1$. Therefore, we deduce $a = v\cdot(e_j + e_k) = 1 - 1 = 0$. Note that $\widetilde{\mathcal{L}}_{-\operatorname{sgn}(\operatorname{grad}(\Phi_G))} g_2(P) \neq \emptyset$ because $\operatorname{sgn}(\operatorname{grad}(\Phi_G))\cdot\zeta = 0$ for all $\zeta \in \partial g_2(P)$, and hence $0 \in \widetilde{\mathcal{L}}_{-\operatorname{sgn}(\operatorname{grad}(\Phi_G))} g_2(P)$. Finally, we conclude $\widetilde{\mathcal{L}}_{-\operatorname{sgn}(\operatorname{grad}(\Phi_G))} g_2(P) = \{0\}$, and therefore g_2 is constant along the solutions of (13b). •

The following theorem completely characterizes the asymptotic convergence properties of the flows in (13).

Theorem 13 *Let $G = (\{1,\dots,n\},E)$ be a connected undirected graph. Then, the flows in* (13) *achieve consensus in finite time. More specifically, for $P_0 = ((p_1)_0,\dots,(p_n)_0) \in \mathbb{R}^n$,*

(i) the solutions of (13a) *starting from P_0 converge in finite time to $(p_*,\dots,p_*)$, with $p_* = \frac{1}{n}\sum_{i=1}^n (p_i)_0$ (average-consensus). The convergence time is upper bounded by $\|L_G P_0\|_2/\lambda_2(L_G)$;*

(ii) the solutions of (13b) *starting from P_0 converge in finite time to $(p_*,\dots,p_*)$, with $p_* = \frac{1}{2}\big(\max_{i\in\{1,\dots,n\}}\{(p_i)_0\} + \min_{i\in\{1,\dots,n\}}\{(p_i)_0\}\big)$ (average-max-min-consensus). The convergence time is upper bounded by $\|L_G P_0\|_1/\lambda_2(L_G)$.*

PROOF. Let $\Phi_G^{-1}(\leq \Phi_G(P_0)) = \{(p_1,\dots,p_n) \in \mathbb{R}^d \mid \Phi_G(p_1,\dots,p_n) \leq \Phi_G(P_0)\}$. Clearly, this set is strongly invariant for both flows. Since L_G is positive semidefinite, $\Phi_G(p_1,\dots,p_n) \geq \lambda_2(L_G)\,\|P - \pi_{H_0(A)}(P)\|_2^2$. Then, $\|P - \pi_{H_0(A)}(P)\|_2^2 \leq \Phi_G(P_0)/\lambda_2(L_G)$ for $P \in \Phi_G^{-1}(\leq \Phi_G(P_0))$. Consider also the closed set

$$\begin{aligned} W_1(P_0) &= \big\{P \in \mathbb{R}^n \mid P\cdot\mathbf{1} = P_0\cdot\mathbf{1}\big\},\\ W(P_0) &= \big\{P \in \mathbb{R}^n \mid \min_{i\{1,\dots,n\}}\{(p_i)_0\} \leq \\ &\qquad \frac{1}{n}P\cdot\mathbf{1} \leq \max_{i\in\{1,\dots,n\}}\{(p_i)_0\}\big\}. \end{aligned}$$

One can see that $W(P_0)$ is strongly invariant for (13a) and for (13b). Now, define the set $S = W(P_0)\cap\Phi_G^{-1}(\leq \Phi_G(P_0))$. From the previous discussion, we deduce that S is strongly invariant for (13a) and (13b). Clearly, S is closed. Furthermore, using $P = \pi_{H_0(L_G)}(P) + P - \pi_{H_0(L)}(P)$, and noting $\pi_{H_0(L_G)}(P) = \frac{P\cdot\mathbf{1}}{n}\mathbf{1}$, we deduce

$$\begin{aligned} \|P\|_2 &= \|\pi_{H_0(L)}(P)\|_2 + \|P - \pi_{H_0(L)}(P)\|_2 \\ &\leq \max\big\{|\min_{i\{1,\dots,n\}}\{(p_i)_0\}|, |\max_{i\{1,\dots,n\}}\{(p_i)_0\}|\big\} + \frac{\Phi_G(P_0)}{\lambda_2(L)} \end{aligned}$$

for $P \in S$. Therefore, S is bounded, and hence compact. Now, the Hessian $\operatorname{Hess}(\Phi_G)(P) = L_G$ is positive semidefinite at any $P \in \mathbb{R}^n$, with the eigenvalue 0 having a constant multiplicity 1. Additionally, for $P \notin \operatorname{Critical}(\Phi_G)$, $\operatorname{grad}(\Phi_G)(P) = L_G P \neq 0$ is orthogonal to $\operatorname{span}\{\mathbf{1}\}$, the eigenspace of L_G corresponding to the eigenvalue 0. Therefore Proposition 12 and Theorem 9 with $V = S$ imply (i) and (ii) in the statement. •

Fig. 1 illustrates the evolution of the differential equations (12), (13a) and (13b). As stated in Theorem 13, the agents' states evolving under (13a) achieve consensus in finite time at $\frac{1}{n}\sum_{i=1}^n (p_i)_0$, and the agents' states evolving under (13b) achieve consensus in finite time at $\frac{1}{2}\big(\max_{i\in\{1,\dots,n\}}\{(p_i)_0\} + \min_{i\in\{1,\dots,n\}}\{(p_i)_0\}\big)$.

Networks with switching communication topologies. For networks with switching connected topologies, one can derive a similar result to Theorem 13. Consider, following [19], the next setup. Let Γ_n be the finite set of connected undirected graphs with vertices $\{1,\dots,n\}$,

$$\Gamma_n = \{G = (\{1,\dots,n\},E) \mid G \text{ connected and undirected}\}.$$

Let $I_{\Gamma_n} \subset \mathbb{N}$ be an index set associated with the elements of Γ_n. A *switching signal* σ is a map $\sigma : \overline{\mathbb{R}}_+ \to I_{\Gamma_n}$. For each time $t \in \overline{\mathbb{R}}_+$, the switching signal σ establishes the communication graph $G_{\sigma(t)} \in \Gamma_n$ employed by the network agents. Now, consider a network subject to the switching communication topology defined by σ and executing one of the coordination algorithms introduced above. In other words, consider the switching system

$$\dot p_i(t) = -\frac{\partial \Phi_{G_{\sigma(t)}}}{\partial p_i} = \sum_{j\in\mathcal{N}_{G_{\sigma(t)},i}} (p_j(t) - p_i(t)), \tag{15}$$

for $i \in \{1,\dots,n\}$, and the switching systems

$$\dot p_i(t) = \frac{\sum_{j\in\mathcal{N}_{G_{\sigma(t)},i}} (p_j(t) - p_i(t))}{\|L_{G_{\sigma(t)}} P(t)\|_2}, \tag{16a}$$

$$\dot p_i(t) = \operatorname{sgn}\Big(\sum_{j\in\mathcal{N}_{G_{\sigma(t)},i}} (p_j(t) - p_i(t))\Big). \tag{16b}$$

The switching system (15) asymptotically achieves average-consensus for an arbitrary switching signal σ.

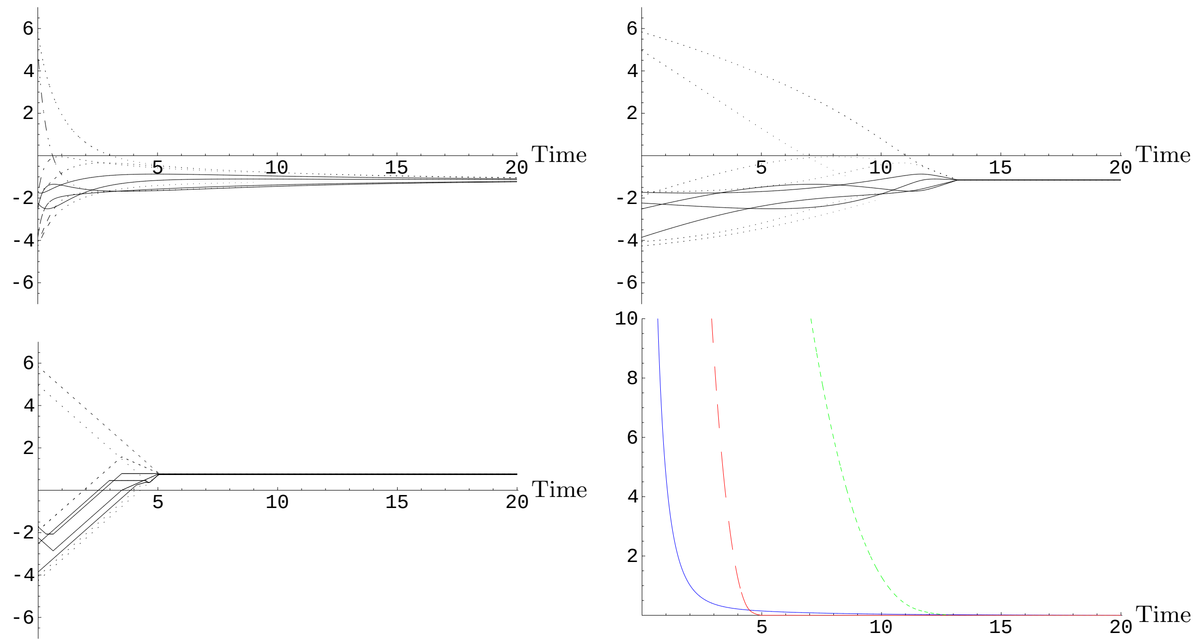

Fig. 1. Evolution of (12) (top left), (13a) (top right) and (13b) (bottom left) for 10 agents starting from a randomly generated initial configuration with $p_i \in [-7, 7]$, $i \in \{1, \dots, 10\}$. The evolution of the Laplacian potential (bottom right) for each flow is plotted in solid, dashed and dotted lines, respectively. The graph $G = (\{1, \dots, 10\}, E)$ has edge set $E = \{(1, 4), (1, 10), (2, 10), (3, 6), (3, 9), (4, 8), (5, 6), (5, 9), (7, 10), (8, 9)\}$. The algebraic connectivity of G is $\lambda_2(L_G) = 0.12$.

Let $G_* \in \Gamma_n$ be such that

$$\frac{\lambda_2(L_{G_*})}{\lambda_n(L_{G_*})} = \min_{G \in \Gamma_n} \Big\{ \frac{\lambda_2(L_G)}{\lambda_n(L_G)} \Big\}.$$

For the systems in (16), one can deduce the next result.

Corollary 14 *Let $\sigma : \overline{\mathbb{R}}_+ \to I_{\Gamma_n}$ be a switching signal. Then, the flow (16a) achieves average-consensus in finite time upper bounded by $\frac{\lambda_n(L_{G_*})}{\lambda_2(L_{G_*})} \big\| P(0) - \frac{1}{n} \sum_{i=1}^n (p_i)_0 \, \mathbf{1} \big\|_2$, and the flow (16b) achieves average-max-min-consensus in finite time equal to $\frac{1}{2} \big(\max_{i \in \{1,\dots,n\}} \{(p_i)_0\} - \min_{i \in \{1,\dots,n\}} \{(p_i)_0\} \big)$.*

PROOF. For the flow (16a), consider the candidate Lyapunov function $V_1 : \mathbb{R}^n \to \mathbb{R}$

$$V_1(P) = \frac{1}{2} \big\| P - \frac{1}{n} \sum_{i=1}^n p_i \, \mathbf{1} \big\|_2^2 .$$

The first-order evolution of this function along the network trajectories is determined by $\widetilde{\mathcal{L}}_{-\frac{\operatorname{grad}(\Phi_G)}{\|\operatorname{grad}(\Phi_G)\|_2}} V_1(P) = -\frac{P' L_G P}{\|L_G P\|_2}$, for each $G \in \Gamma_n$, which is single-valued, Lipschitz and regular. Additionally, for any $G \in \Gamma_n$,

$$\widetilde{\mathcal{L}}_{-\frac{\operatorname{grad}(\Phi_G)}{\|\operatorname{grad}(\Phi_G)\|_2}} V_1(P) \leq -\frac{\lambda_2(L_G)}{\lambda_n(L_G)} \big\| P - \frac{1}{n} \sum_{i=1}^n p_i \, \mathbf{1} \big\|_2 \leq 0.$$

The application of the LaSalle Invariance Principle ensures that the flow (16a) achieves average-consensus. From the previous inequality and the definition of V_1,

$$\big\| P(t) - \frac{1}{n} \sum_{i=1}^n p_i \, \mathbf{1} \big\|_2 \leq \big\| P(0) - \frac{1}{n} \sum_{i=1}^n p_i \, \mathbf{1} \big\|_2 - \frac{\lambda_2(L_{G_*})}{\lambda_n(L_{G_*})} t,$$

which implies the result.

For the flow (16b), consider the candidate Lyapunov function $V_2 : \mathbb{R}^n \to \mathbb{R}$

$$\begin{aligned} V_2(P) &= \big\| P - \frac{1}{2} \big(\max_{i \in \{1,\dots,n\}} \{p_i\} + \min_{i \in \{1,\dots,n\}} \{p_i\} \big) \, \mathbf{1} \big\|_\infty \\ &= \frac{1}{2} \big(\max_{i \in \{1,\dots,n\}} \{p_i\} - \min_{i \in \{1,\dots,n\}} \{p_i\} \big). \end{aligned}$$

This function is locally Lipschitz and regular. Let $a \in \widetilde{\mathcal{L}}_{-\operatorname{sgn}(\operatorname{grad}(\Phi_G))} V_2(P)$. Then, there exists $v \in K\big[-\operatorname{sgn}(\operatorname{grad}(\Phi_G)) \big](P)$ such that $a = v \cdot \zeta$, for all $\zeta \in \partial V_2(P)$. Take P with $V_2(P) \neq 0$. Let $j, k \in \{1, \dots, n\}$ such that $p_j = \min_{i \in \{1,\dots,n\}} \{p_i\}$, $p_k = \max_{i \in \{1,\dots,n\}} \{p_i\}$. Then $\frac{1}{2}(e_k - e_j) \in \partial V_2(P)$, $v_j = 1$ and $v_k = -1$. Therefore $a = -1$. The result follows from Proposition 4. •

Remark 15 *Note that in the proof of Corollary 14, we have explicitly computed the convergence time of the* *flow* (13b) *to achieve average-max-min-consensus to be*

$$\frac{1}{2}\big(\max_{i\in\{1,\dots,n\}}\{(p_i)_0\} - \min_{i\in\{1,\dots,n\}}\{(p_i)_0\}\big).$$

If the network agents had the capability to decide exactly when convergence has been achieved (for instance, by running in parallel another consensus algorithm), this information and the consensus value would enable them to compute $\max_{i\in\{1,\dots,n\}}\{(p_i)_0\}$ *and* $\min_{i\in\{1,\dots,n\}}\{(p_i)_0\}$.•

5 Conclusions

We have introduced the normalized and signed versions of the gradient descent flow of a differentiable function. We have characterized the general asymptotic convergence properties of these nonsmooth gradient flows, and identified suitable conditions on the differentiable function that guarantee that convergence to the critical points is achieved in finite time. We have obtained these results building on two novel nonsmooth analysis results on finite-time convergence and second-order information about the evolution of the Lyapunov function along the solutions of the system. The applicability of these results is not restricted to gradient flows, and they can indeed be used in other setups involving discontinuous vector fields, locally Lipschitz functions and finite-time convergence. We have discussed the application of the results to gradient coordination algorithms for multi-agent systems, and, in particular, to network consensus problems.

Future work will be devoted to explore (i) the use of the upper bounds on the (finite) convergence time of the proposed nonsmooth flows in assessing the (time) complexity of a variety of distributed coordination algorithms for robotic networks; (ii) the application of the results to distributed sensor fusion algorithms based on consensus (e.g. [22, 24]), coordination problems such as formation control, deployment and rendezvous, and other problems where gradient systems play an important role; (iii) the identification of other nonsmooth spatially distributed algorithms based on gradient information with similar convergence properties.

References

[1] H. Ando, Y. Oasa, I. Suzuki, and M. Yamashita. Distributed memoryless point convergence algorithm for mobile robots with limited visibility. *IEEE Transactions on Robotics and Automation*, 15(5):818–828, 1999.

[2] A. Bacciotti and F. Ceragioli. Stability and stabilization of discontinuous systems and nonsmooth Lyapunov functions. *ESAIM. Control, Optimisation & Calculus of Variations*, 4:361–376, 1999.

[3] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, MA, 1997. ISBN 1886529019.

[4] S. P. Bhat and D. S. Bernstein. Finite-time stability of continuous autonomous systems. *SIAM Journal on Control and Optimization*, 38(3):751–766, 2000.

[5] F. H. Clarke. *Optimization and Nonsmooth Analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, 1983. ISBN 047187504X.

[6] J. M. Coron. On the stabilization in finite time of locally controllable systems by means of continuous time-varying feedback laws. *SIAM Journal on Control and Optimization*, 33:804–833, 1995.

[7] J. Cortés and F. Bullo. Coordination and geometric optimization via distributed dynamical systems. *SIAM Journal on Control and Optimization*, 2005. To appear.

[8] J. Cortés, S. Martínez, and F. Bullo. Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions. *IEEE Transactions on Automatic Control*, July 2004. To appear.

[9] J. Cortés, S. Martínez, and F. Bullo. Spatially-distributed coverage optimization and control with limited-range interactions. *ESAIM. Control, Optimisation & Calculus of Variations*, 2005. To appear.

[10] M. de Berg, M. van Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Springer Verlag, New York, 1997. ISBN 354061270X.

[11] R. Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer Verlag, New York, second edition, 2000.

[12] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides*, volume 18 of *Mathematics and Its Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.

[13] V. Gazi and K. M. Passino. Stability analysis of swarms. *IEEE Transactions on Automatic Control*, 48(4):692–697, 2003.

[14] U. Helmke and J.B. Moore. *Optimization and Dynamical Systems*. Springer Verlag, New York, 1994. ISBN 0387198571.

[15] W. M. Hirsch and S. Smale. *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press, New York, NY, 1974.

[16] J. W. Jaromczyk and G. T. Toussaint. Relative neighborhood graphs and their relatives. *Proceedings of the IEEE*, 80(9):1502–1517, 1992.

[17] J. Lin, A. S. Morse, and B. D. O. Anderson. The multi-agent rendezvous problem. In *IEEE Conf. on Decision and Control*, pages 1508–1513, Maui, Hawaii, December 2003.

[18] P. Ögren, E. Fiorelli, and N. E. Leonard. Cooperative control of mobile sensor networks: adaptive gradient climbing in a distributed environment. *IEEE Transactions on Automatic Control*, 49(8):1292–1302, 2004.

[19] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49 (9):1520–1533, 2004.

[20] E. P. Ryan. Finite-time stabilization of uncertain nonlinear planar systems. *Dynam. Control*, 1:83–94, 1991.

[21] D. Shevitz and B. Paden. Lyapunov stability theory of nonsmooth systems. *IEEE Transactions on Automatic Control*, 39(9):1910–1914, 1994.

[22] D. P. Spanos, R. Olfati-Saber, and R. M. Murray. Distributed sensor fusion using dynamic consensus. In *IFAC World Congress*, Prague, CZ, July 2005. Electronic proceedings.

[23] H. Tanner, A. Jadbabaie, and G. J. Pappas. Flocking in fixed and switching networks. *IEEE Transactions on Automatic Control*, 2005. Submitted.

[24] L. Xiao, S. Boyd, and S. Lall. A scheme for asynchronous distributed sensor fusion based on average consensus. In *Fourth International Conference on Information Processing in Sensor Networks (IPSN'05)*, Los Angeles, CA, 2005. To appear.