Exponentially Fast Distributed Coordination for Nonsmooth Convex Optimization

Simon K. Niederländer   Frank Allgöwer   Jorge Cortés

Abstract— **This paper considers networks of agents that seek to cooperatively solve a general class of nonsmooth convex optimization problems with an inherent distributed structure. We characterize the asymptotic convergence properties of distributed continuous-time coordination algorithms whose design relies on the saddle-point dynamics associated with an augmented Lagrangian. The main technical novelty is the identification of a nonsmooth Lyapunov function which, under mild convexity and regularity assumptions on the optimization problem data, allows us to further characterize the exponential convergence rates of the proposed algorithms for optimization subject to either equality or inequality constraints.**

I. Introduction

We consider network scenarios that give rise to generic nonsmooth convex optimization problems with an inherent distributed structure. Such multi-agent optimization scenarios are motivated by various applications including network flow optimization, control of distributed energy resources, resource allocation, and multi-sensor fusion. Within such contexts, the goals and performance metrics of the individual agents are naturally encoded into suitable objective functions whose optimization may be subject to a combination of physical, communication and operational constraints. Our objective is to provide formal characterizations of the convergence and performance properties of distributed continuous-time coordination algorithms that allow each agent to find their component of the optimal solution vector. We see these characterizations as a stepping stone towards the development of strategies that are robust against disturbances and can accommodate a variety of resource constraints.

Literature Review. The interest in networked systems has stimulated the synthesis of distributed strategies where agents interact with neighbors to coordinate their computations and solve convex optimization problems [1]. A majority of works focus on consensus-based approaches, where individual agents maintain, communicate and update an estimate of the complete solution vector of the optimization problem, see e.g. [2], [3], [4] for discrete-time implementations. In contrast, recent work [5], [6], [7] has proposed continuous-time solvers whose convergence properties can be studied via classical stability analysis. Of particular importance to our work here are distributed strategies where each agent seeks to determine its own component of the solution vector and interchanges information with its neighbors, independent of the network size. A common approach to design such distributed strategies relies on the saddle-point or primal-dual dynamics [8] corresponding to the Lagrangian associated with the optimization problem. The work [9] introduces set-valued and discontinuous saddle-point dynamics specifically tailored for linear programs. The work [10] studies primal-dual dynamics for convex programs subject to inequality constraints. These dynamics are modified with a projection operator on the dual variables to preserve their nonnegativity. In the context of neural networks, the work [11] proposes a generalized circuit for nonsmooth nonlinear optimization. However, the dynamics are not fully amenable to distributed implementation in multi-agent systems due to the global penalty parameters involved. Our work builds on a class of saddle-point(-like) algorithms for nonsmooth convex optimization proposed in our earlier works [12], [13].

Statement of Contributions. We consider generic nonsmooth convex optimization problems with an inherent distributed structure. Building on saddle-point(-like) dynamics associated with an augmented Lagrangian, our first contribution is the identification of a nonsmooth Lyapunov function which allows us to establish the asymptotic correctness of the algorithms without relying on arguments based on the LaSalle invariance principle. This finding leads to our second contribution, the performance characterization of the proposed algorithms. In particular, we establish the exponential convergence rates of the algorithms for convex optimization subject to either equality or inequality constraints. The proofs are omitted for reasons of space and will appear elsewhere.

II. Preliminaries

We let $\langle \cdot, \cdot \rangle$ denote the Euclidean inner product. Let $1_n = (1, \ldots, 1) \in \mathbb{R}^n$. Given a vector $x \in \mathbb{R}^n$, let $[x]^+ = (\max\{0, x_1\}, \ldots, \max\{0, x_n\}) \in \mathbb{R}^n_{\geq 0}$. Given a set $X \subset \mathbb{R}^n$, we denote its convex hull by $\operatorname{co} X$, its interior by $\operatorname{int} X$, its boundary by $\operatorname{bd} X$, and its closure by $\operatorname{cl} X$. Let $\mathbb{B}(x, \delta) = \{y \in \mathbb{R}^n \mid \|y - x\| < \delta\}$. A *set-valued map* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ maps elements of $\mathbb{R}^n$ to elements of $2^{\mathbb{R}^n}$. The set-valued map F is *monotone* if $\langle x - y, \xi(x) - \xi(y) \rangle \geq 0$ whenever $\xi(x) \in F(x)$, $\xi(y) \in F(y)$, and *strictly monotone* if the inequality is strict when $x \neq y$. Finally, F is *strongly monotone* if there exists $\eta > 0$ such that $\langle x - y, \xi(x) - \xi(y) \rangle \geq \eta \|x - y\|^2$ whenever $\xi(x) \in F(x)$, $\xi(y) \in F(y)$.

A. Nonsmooth Analysis

We review here relevant notions from nonsmooth analysis following [14]. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *locally Lipschitz at* $x \in \mathbb{R}^n$ if there exist $\delta_x > 0$ and $L_x \geq 0$ such that

S. K. Niederländer and F. Allgöwer are with the Institute for Systems Theory and Automatic Control, University of Stuttgart, Pfaffenwaldring 9, 70550 Stuttgart, Germany. Email: {simon.niederlaender, frank.allgower}@ist.uni-stuttgart.de.

J. Cortés is with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Dr, La Jolla, California, 92093-0411, USA. Email: cortes@ucsd.edu.

$|f(y) - f(z)| \leq L_x \|y - z\|$ for all $y, z \in \mathbb{B}(x, \delta_x)$. The function f is *locally Lipschitz* if it is locally Lipschitz at x, for all $x \in \mathbb{R}^n$. A convex function $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz, cf. [15, Theorem 3.1.1, p. 16]. Let $\Omega_f \subset \mathbb{R}^n$ be the set of points at which f fails to be differentiable, and let S denote any other set of measure zero. The *generalized gradient* $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ of f at $x \in \mathbb{R}^n$ is defined by

$$\partial f(x) = \operatorname{co}\Big\{ \lim_{i \to +\infty} \nabla f(x_i) \mid x_i \to x,\ x_i \notin S \cup \Omega_f \Big\}.$$

A set-valued map $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *upper semicontinuous* if, for all $x \in \mathbb{R}^n$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $F(y) \subset F(x) + \mathbb{B}(0, \varepsilon)$ for all $y \in \mathbb{B}(x, \delta)$. We say F is *locally bounded* if, for every $x \in \mathbb{R}^n$, there exist $\delta > 0$ and $\varepsilon > 0$ such that $\|\xi\| \leq \varepsilon$ for all $\xi \in F(y)$ and all $y \in \mathbb{B}(x, \delta)$. The following result summarizes some important properties of the generalized gradient [14].

Proposition 2.1 (Properties of the generalized gradient)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz at $x \in \mathbb{R}^n$. Then,*

(i) $\partial f(x) \subset \mathbb{R}^n$ is nonempty, convex and compact, and $\|\xi\| \leq L_x$, for all $\xi \in \partial f(x)$,
(ii) $\partial f(x)$ is upper semicontinuous at $x \in \mathbb{R}^n$.

Let $\mathcal{C}^{1,1}(\mathbb{R}^n, \mathbb{R})$ denote the class of functions $f : \mathbb{R}^n \to \mathbb{R}$ that are continuously differentiable and whose gradient map $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz. The *generalized Hessian* $\partial(\nabla f) : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n}$ of f at $x \in \mathbb{R}^n$ is defined by

$$\partial(\nabla f)(x) = \operatorname{co}\Big\{ \lim_{i \to +\infty} \nabla^2 f(x_i) \mid x_i \to x,\ x_i \notin \Omega_f \Big\}.$$

By construction, $\partial(\nabla f)(x)$ is a nonempty, convex and compact set of symmetric matrices, cf. [16].

B. Set-Valued Dynamical Systems

We next consider set-valued and locally projected dynamical systems [17], [18] defined by means of differential inclusions [19]. Let $X \subset \mathbb{R}^n$ be open and let $F : X \rightrightarrows \mathbb{R}^n$ be a set-valued map. Consider the differential inclusion

$$\dot{x}(t) \in F(x(t)), \quad x(t_0) = x_0 \in X. \tag{DI}$$

A solution of (DI) on an interval $[t_0, t_1] \subset \mathbb{R}$ is an absolutely continuous map $x : [t_0, t_1] \to X$ such that $\dot{x}(t) \in F(x(t))$ for almost all (a.a.) $t \in [t_0, t_1]$. The existence of local solutions of (DI) starting from $x_0 \in X$ is guaranteed by the following result [19].

Lemma 2.2 (Existence of local solutions)**.** *Let $F : X \rightrightarrows \mathbb{R}^n$ be locally bounded, upper semicontinuous with nonempty, convex and compact values. Then, given $x_0 \in X$, there exists a local solution $x : [t_0, t_1] \to X$ of* (DI) *starting from x_0.*

Let $f : X \to \mathbb{R}$ be locally Lipschitz. The *set-valued Lie derivative* $\mathcal{L}_F f : X \rightrightarrows \mathbb{R}$ of f with respect to F at $x \in X$ is defined by

$$(\mathcal{L}_F f)(x) = \big\{ \psi \in \mathbb{R} \mid \exists \xi \in F(x) : \langle \xi, \pi \rangle = \psi, \forall \pi \in \partial f(x) \big\}.$$

Let $G \subset \mathbb{R}^n$ be a nonempty, closed and convex set. Let the *distance function* $d_G : \mathbb{R}^n \to \mathbb{R}$ be defined by $d_G(x) = \inf_{y \in G} \|x - y\|$. The *tangent cone* and the *normal cone* of G at $x \in G$ are, respectively,

$$T_G(x) = \operatorname{cl} \bigcup_{\delta > 0} \frac{1}{\delta}(G - x), \quad N_G(x) = \operatorname{cl} \bigcup_{\eta \geq 0} \eta \partial d_G(x).$$

Let $\operatorname{proj}_G(x) = \operatorname{argmin}_{y \in G} \|x - y\|$. The *orthogonal (set) projection* of a nonempty, convex and compact set $F(x) \subset \mathbb{R}^n$ at $x \in G$ with respect to $G \subset \mathbb{R}^n$ is defined by

$$\Pi_G(x, F(x)) = \bigcup_{\xi \in F(x)} \lim_{\delta \searrow 0} \frac{\operatorname{proj}_G(x + \delta \xi) - x}{\delta}. \tag{1}$$

Note that if $x \in \operatorname{int} G$, then $\Pi_G(x, F(x))$ reduces to the set $F(x)$. Consider now the projected differential inclusion

$$\dot{x}(t) \in \Pi_G(x, F(x))(t), \quad x(t_0) = x_0 \in X, \tag{PDI}$$

for a.a. $t \in [t_0, t_1]$. The following result states conditions under which local solutions of (PDI) exist [20].

Lemma 2.3 (Existence of local solutions of projected differential inclusions)**.** *Let $G \subset \mathbb{R}^n$ and let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfy the hypothesis of Lemma 2.2. If there exists $c > 0$ such that, for every $x \in G$,*

$$\sup_{\xi \in F(x)} \|\xi\| \leq c(1 + \|x\|),$$

then, for any $x_0 \in G$, there exists a local solution $x : [t_0, t_1] \to G$ of (PDI) *starting from x_0.*

III. Problem Statement

Consider the constrained minimization problem

$$\min\{f(x) \mid h(x) = 0_p,\ g(x) \leq 0_m\}, \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are convex and locally Lipschitz, and $h : \mathbb{R}^n \to \mathbb{R}^p$ is affine, i.e., $h(x) = Ax - b$, with $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$, where $p \leq n$. Let the equality and inequality constraint sets be defined by $H = \{x \in \mathbb{R}^n \mid h(x) = 0_p\}$ and $G = \{x \in \mathbb{R}^n \mid g(x) \leq 0_m\}$, respectively. We assume that the (closed and convex) set of optimal solutions $S = \{x^* \in H \cap G \mid f(x^*) \leq f(x),\ \forall x \in H \cap G\}$ is nonempty. Throughout the paper we assume that (P) satisfies the strong Slater condition [15], i.e.,

(H1) $\operatorname{rank}(A) = p$, i.e., the rows of matrix $A \in \mathbb{R}^{p \times n}$ are linearly independent,
(H2) $\exists x \in \mathbb{R}^n$ such that $Ax = b$ and $g_k(x) < 0$ for all $k \in \{1, \ldots, m\}$.

Let $\kappa > 0$ and let the augmented Lagrangian $L_\kappa : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ associated with (P) be defined by

$$L_\kappa(x, \lambda) = f(x) + \frac{1}{2}\|h(x)\|^2 + \langle \lambda, h(x) \rangle + \kappa \langle 1_m, [g(x)]^+ \rangle,$$

where $\lambda \in \mathbb{R}^p$ is a Lagrange multiplier. Under the regularity assumptions (H1)–(H2), for every $x^* \in S$ there exist $(\lambda^*, \mu^*) \in \mathbb{R}^p \times \mathbb{R}^m_{\geq 0}$ with $(\lambda^*, \mu^*) \neq (0_p, 0_m)$ such that (x^*, λ^*) is a saddle point of L_κ, i.e.,

$$L_\kappa(x^*, \lambda) \leq L_\kappa(x^*, \lambda^*) \leq L_\kappa(x, \lambda^*),$$

for all $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^p$, given that $\kappa \geq \|\mu^*\|_\infty$. We denote the (closed and convex) set of saddle points of L_κ

by $\mathrm{sp}(L_\kappa)$. The following result from [12] reveals an intimate relationship between saddle points of L_κ and optimal solutions of (P).

Lemma 3.1 (Saddle-point Theorem)**.** *Let $L_\kappa : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ and let $(x^*, \lambda^*) \in \mathrm{sp}(L_\kappa)$ with $\kappa > \|\mu^*\|_\infty$ for some dual solution μ^* of* (P). *Then, x^* is an optimal solution of* (P).

Lemma 3.1 identifies a condition under which the penalty parameter κ is *exact* [21]. Given this result, instead of solving (P), we seek to design strategies that find saddle points of L_κ. Since the bivariate augmented Lagrangian L_κ is, by definition, convex-concave, a natural approach to find the saddle points is via its associated saddle-point dynamics.

IV. Continuous-time Distributed Optimization

We recall the optimization algorithms developed in our previous work [13]. The main technical novelty is the identification of a nonsmooth Lyapunov function which allows us to establish the asymptotic correctness of the algorithms through a Lyapunov, rather than a LaSalle, argument. As we show later, this function plays a key role in characterizing the exponential convergence rates of the algorithms.

A. Saddle-Point Dynamics

Consider the saddle-point dynamics associated with the augmented Lagrangian L_κ defined over $\mathbb{R}^n \times \mathbb{R}^p$,

$$\begin{cases} \dot{x}(t) \in -\partial_x L_\kappa(x,\lambda)(t), & x(t_0) = x_0 \in \mathbb{R}^n, \\ \dot{\lambda}(t) \in +\partial_\lambda L_\kappa(x,\lambda)(t), & \lambda(t_0) = \lambda_0 \in \mathbb{R}^p, \end{cases} \tag{SPD}$$

for a.a. $t \in [t_0, t_1]$. Note that the existence of local solutions $(x,\lambda) : [t_0,t_1] \to \mathbb{R}^n \times \mathbb{R}^p$ of (SPD) starting from $(x_0,\lambda_0) \in \mathbb{R}^n \times \mathbb{R}^p$ is guaranteed by Proposition 2.1 and Lemma 2.2. Our next result characterizes the asymptotic convergence of the solutions of (SPD) to the set of saddle points of L_κ.

Theorem 4.1 (Asymptotic convergence)**.** *Let $\kappa > 0$ and let $(\tilde{x}, \tilde{\lambda}) \in \mathrm{sp}(L_\kappa)$. Define the map $V_\kappa : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ by*

$$\begin{aligned} V_\kappa(x,\lambda) = &\min_{(x^*,\lambda^*) \in \mathrm{sp}(L_\kappa)} \Big(\frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|\lambda - \lambda^*\|^2 \Big) \\ &+ L_\kappa(x,\lambda) - L_\kappa(\tilde{x},\tilde{\lambda}). \end{aligned}$$

If ∂f is strictly monotone, then $(\mathcal{L}_{SPD} V_\kappa)(x,\lambda) \subset (-\infty, 0)$ for all $(x,\lambda) \in (\mathbb{R}^n \times \mathbb{R}^p) \setminus \mathrm{sp}(L_\kappa)$.

The theorem states that the set of saddle points of L_κ is strongly globally asymptotically stable under (SPD), given that f is strictly convex. Point-wise convergence of the solutions of (SPD) in the set $\mathrm{sp}(L_\kappa)$ follows from the stability of each individual saddle point of L_κ and the asymptotic stability of the set $\mathrm{sp}(L_\kappa)$, established in Theorem 4.1.

Corollary 4.2 (Point-wise asymptotic convergence)**.** *Any solution $(x,\lambda) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p$ of* (SPD) *starting from $(x_0,\lambda_0) \in \mathbb{R}^n \times \mathbb{R}^p$ converges asymptotically to a point in the set $\mathrm{sp}(L_\kappa)$.*

Remark 4.3 (Convergence for linear programs). The strong global asymptotic stability of $\mathrm{sp}(L_\kappa)$ under (SPD) can also be established using the alternative Lyapunov function $V : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ defined by

$$V(x,\lambda) = \frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|\lambda - \lambda^*\|^2,$$

where (x^*, λ^*) is an arbitrary saddle point of L_κ. In fact, one can deduce that the Lie derivative of V is negative semidefinite, implying stability, albeit not asymptotic stability. To conclude the latter, one can resort [12] to the LaSalle invariance principle for differential inclusions [22]. This approach has (i) the advantage that it only requires monotonicity of ∂f, instead of strict monotonicity, and is therefore applicable to linear programs, and (ii) the disadvantage that, V not being a strict Lyapunov function, it cannot be used to characterize the convergence rate of the algorithm. •

We note that the saddle-point dynamics (SPD) need not converge to an optimal solution of (P), unless the penalty parameter κ is exact, cf. Lemma 3.1. In fact, the lower bound on κ in Lemma 3.1 is characterized by some dual solution μ^* of (P), which is unknown a priori. Our forthcoming discussion proposes dynamics that do not rely on the penalty parameter κ, yet enjoy the same convergence properties as the saddle-point dynamics (SPD).

B. Saddle-Point-Like Dynamics

Recall that $G \subset \mathbb{R}^n$ denotes the (closed and convex) inequality constraint set associated with (P). Let the set-valued flow $F : G \times \mathbb{R}^p \rightrightarrows \mathbb{R}^n$ be defined by

$$F(x,\lambda) = -\nabla\Big(\frac{1}{2}\|h(x)\|^2\Big) - \nabla_x \langle \lambda, h(x) \rangle - \partial f(x).$$

The definition of F is motivated by the fact that, for $(x,\lambda) \in \mathrm{int}\, G \times \mathbb{R}^p$, we have $-\partial_x L_\kappa(x,\lambda) = \{F(x,\lambda)\}$. Consider now the saddle-point-like dynamics defined over $G \times \mathbb{R}^p$,

$$\begin{cases} \dot{x}(t) \in \Pi_G(x, F(x,\lambda))(t), & x(t_0) = x_0 \in G, \\ \dot{\lambda}(t) \in +\partial_\lambda L_\kappa(x,\lambda)(t), & \lambda(t_0) = \lambda_0 \in \mathbb{R}^p, \end{cases} \tag{SPLD}$$

for a.a. $t \in [t_0, t_1]$, where Π_G is defined in (1). The existence of local solutions $(x,\lambda) : [t_0,t_1] \to G \times \mathbb{R}^p$ of (SPLD) is guaranteed by Proposition 2.1 and Lemma 2.3.

We now investigate the geometric interpretation of Π_G for locally projected dynamical systems [18]. Let the set of unit outward normals to G at $x \in \mathrm{bd}\, G$ be defined by

$$N_G^\sharp(x) = N_G(x) \cap \mathrm{bd}\, \overline{\mathbb{B}}(0,1).$$

Note that if $(x,\lambda) \in \mathrm{int}\, G \times \mathbb{R}^p$, then, by definition of Π_G, it follows $\Pi_G(x, F(x,\lambda)) = F(x,\lambda)$. However, if $(x,\lambda) \in \mathrm{bd}\, G \times \mathbb{R}^p$, then

$$\Pi_G(x, F(x,\lambda)) = \bigcup_{\xi \in F(x,\lambda)} \xi - \max\big\{0, \langle \xi, n^*(x,\xi) \rangle\big\} n^*(x,\xi),$$

where

$$n^*(x,\xi) \in \underset{n \in N_G^\sharp(x)}{\mathrm{argmax}} \langle \xi, n \rangle. \tag{2}$$

Note that if $\{\xi\} \cap T_G(x) \neq \emptyset$ for some $(x,\lambda) \in \mathrm{bd}\, G \times \mathbb{R}^p$ and $\xi \in F(x,\lambda)$, then $\sup_{n \in N_G^\sharp(x)} \langle \xi, n \rangle \leq 0$, and by definition of Π_G, no projection needs to be performed. The following result establishes the existence and uniqueness of the maximizer $n^*(x, \xi)$ of (2) whenever $\{\xi\} \cap T_G(x) = \emptyset$, i.e., whenever the projection needs to be computed.

Lemma 4.4 (Existence and uniqueness)**.** *Let $(x, \lambda) \in \operatorname{bd} G \times \mathbb{R}^p$. If there exists $\xi \in F(x, \lambda)$ such that $\sup_{n \in N_G^\sharp(x)} \langle \xi, n \rangle > 0$, then the maximizer $n^*(x, \xi)$ of* (2) *exists and is unique.*

We note that the computational complexity to solve (2) not only depends on the problem dimensions $n, p, m \in \mathbb{N}$, but also on the convexity and regularity assumptions on the optimization problem data, i.e., f, h and g. Our strategy to show that the saddle-point-like dynamics (SPLD) enjoy the same convergence properties as the saddle-point dynamics (SPD) is to establish that, in fact, its solutions are also solutions of (SPD), given that the penalty parameter κ is sufficiently large. The following result makes this precise.

Proposition 4.5 (Relationship of solutions)**.** *Let $(x, \lambda) : [t_0, +\infty) \to G \times \mathbb{R}^p$ be any solution of* (SPLD) *starting from $(x_0, \lambda_0) \in G \times \mathbb{R}^p$. Then, there exists $\kappa > 0$ such that the solution is also a solution of* (SPD).

Proposition 4.5 states that the set of solutions of (SPD) is, in general, richer than the set of solutions of (SPLD), given that κ exceeds a certain threshold. We note that the saddle-point-like dynamics (SPLD) do not incorporate any knowledge of the penalty parameter κ (as the saddle-point dynamics (SPD)) and are therefore amenable to distributed implementation in multi-agent systems.

C. Distributed Implementation

Consider a network of $n \in \mathbb{N}$ agents whose communication topology is represented by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, i.e., $\mathcal{V} = \{1, \ldots, n\} \subset \mathbb{N}$ and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is symmetric. The objective of the agents is to cooperatively solve the constrained minimization problem (P). We assume that the aggregate objective function f is additively separable, i.e., $f(x) = \sum_{i=1}^n f_i(x_i)$, where f_i and $x_i \in \mathbb{R}$ denote the local objective function and state associated with agent $i \in \{1, \ldots, n\}$, respectively. Additionally, we assume that the constraints of (P) are compatible with the network topology described by $\mathcal{G}$. Formally, we say the inequality constraints $g_k(x) \le 0$, $k \in \{1, \ldots, m\}$, are compatible with $\mathcal{G}$ if g_k can be expressed as a function of some components of the network state $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, which induce a complete subgraph of $\mathcal{G}$. A similar definition can be stated for the equality constraints $h_\ell(x) = 0$, $\ell \in \{1, \ldots, p\}$.

We now show that the saddle-point-like dynamics (SPLD) are well-suited for distributed implementation. If $(x, \lambda) \in \operatorname{int} G \times \mathbb{R}^p$, then each agent $i \in \{1, \ldots, n\}$ implements

$$\dot{x}_i + \sum_{\{\ell : a_{\ell i} \neq 0\}} a_{\ell i} \Bigg(\sum_{\{j : a_{\ell j} \neq 0\}} a_{\ell j} x_j - b_\ell + \lambda_\ell \Bigg) \in -\partial f_i(x_i),$$

and some dual dynamics

$$\dot{\lambda}_\ell = \sum_{\{i : a_{\ell i} \neq 0\}} a_{\ell i} x_i - b_\ell.$$

If $(x, \lambda) \in \operatorname{bd} G \times \mathbb{R}^p$, then each agent i implements

$$\dot{x}_i \in \bigcup_{\xi_i \in F_i(x, \lambda)} \xi_i - \max \Bigg\{ 0, \sum_{\{j : n_j^* \neq 0\}} \xi_j n_j^*(x, \xi) \Bigg\} n_i^*(x, \xi),$$

and some dual dynamics. We say that the saddle-point-like dynamics (SPLD) are distributed over $\mathcal{G}$ if the following conditions are satisfied:

- (C1) The network constraints h, g are compatible with $\mathcal{G}$.
- (C2) Agent i knows its state x_i and objective function f_i.
- (C3) Agent i knows its neighbors' states x_j, their objective functions f_j, and
  - (i) the non-zero elements of every row of A and every b_ℓ for which $a_{\ell i} \neq 0$, and
  - (ii) the active inequality constraints g_k in which it is involved.

In contrast to consensus-based distributed algorithms where each agent maintains, communicates and updates an estimate of the complete solution vector, the saddle-point-like dynamics (SPLD) only require each agent to store and communicate its own component of the solution vector.

V. Performance Characterization

In this section, we characterize the performance properties of the saddle-point(-like) dynamics. The nonsmooth Lyapunov function proposed in Section IV-A allows us to go beyond the qualitative statement of asymptotic convergence and instead precisely characterize the exponential convergence rates of the algorithms. We separate our study in two cases depending on whether the optimization is subject to equality or inequality constraints.

A. Distributed Optimization under Inequality Constraints

Consider the unconstrained minimization problem

$$\min\{f(x) + \kappa\langle 1_m, [g(x)]^+ \rangle \mid x \in \mathbb{R}^n\}, \tag{PI}$$

where $\kappa > 0$. Let $F_\kappa : \mathbb{R}^n \to \mathbb{R}$ be defined by $F_\kappa(x) = f(x) + \kappa\langle 1_m, [g(x)]^+ \rangle$. The dynamics associated with (PI) reduce to the gradient dynamics defined over $\mathbb{R}^n$,

$$\dot{x}(t) \in -\partial F_\kappa(x)(t), \quad x(t_0) = x_0 \in \mathbb{R}^n, \tag{GD}$$

for a.a. $t \in [t_0, t_1]$. The existence of local solutions $x : [t_0, t_1] \to \mathbb{R}^n$ of (GD) is guaranteed by Proposition 2.1 and Lemma 2.2. Our analysis relies on the following assumption:

- (H3) The set of equilibria $\operatorname{eq}(\partial F_\kappa) = \{x^* \in \mathbb{R}^n \mid 0 \in \partial F_\kappa(x^*)\}$ of (GD) is nonempty, convex and compact.

Our next result characterizes the performance properties of the solutions of (GD).

Theorem 5.1 (Performance characterization)**.** *Let $\kappa > 0$ and let $F_\kappa : \mathbb{R}^n \to \mathbb{R}$. The following statements hold:*

- *(i) If ∂f is monotone, then $\operatorname{eq}(\partial F_\kappa)$ is strongly stable under* (GD).
- *(ii) If ∂f is strictly monotone, then $\operatorname{eq}(\partial F_\kappa)$ is strongly asymptotically stable under* (GD).
- *(iii) If ∂f is strongly monotone, then $\operatorname{eq}(\partial F_\kappa)$ is strongly exponentially stable under* (GD).

The result states that the performance properties of the solutions of (GD) merely depend on the convexity of the objective function. In particular, if f is strongly convex, then the convergence rate of (GD) is exponential, determined by the strong convexity modulus $\eta > 0$. Point-wise convergence of the solutions of (GD) in the set $\mathrm{eq}(\partial F_\kappa)$ follows from Corollary 4.2. The following result is an immediate consequence of Proposition 4.5 and Theorem 5.1(iii).

Corollary 5.2 (Locally projected gradient dynamics)**.** *Let $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and consider the dynamics*

$$\dot{x}(t) \in \Pi_G(x, -\partial f(x))(t), \quad x(t_0) = x_0 \in G, \tag{PGD}$$

for a.a. $t \in [t_0, +\infty)$. If ∂f is strongly monotone, then any solution $x : [t_0, +\infty) \to G$ starting from $x_0 \in G$ converges exponentially fast to the singleton set $\mathrm{eq}(\Pi_G) = \{x^* \in G \mid 0 \in \Pi_G(x^*, -\partial f(x^*))\}$.

The following example illustrates that the locally projected gradient dynamics (PGD) perform well within the exponential performance bound obtained in Theorem 5.1(iii).

Example 1 (Projected gradient dynamics for inequality constrained optimization). Consider a network of $n = 10$ agents that seek to cooperatively solve the minimization problem

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}} && \sum_{i \in \{1,\dots,n\}} x_i^2/2 + |x_i| \\ &\text{subject to} && \|(x_1 - 2, \dots, x_5 - 2)\|_\infty \leq 1, \\ &&& \|(x_6 + 2, \dots, x_n + 2)\|_\infty \leq 1, \end{aligned} \tag{EX1}$$

where $x_i \in \mathbb{R}$ denotes the state associated with agent $i \in \{1, \dots, n\}$. The generalized gradient of $x_i \mapsto f_i(x_i) = x_i^2/2 + |x_i|$ at $x_i \in \mathbb{R}$ is

$$\partial f_i(x_i) = \begin{cases} \{x_i + 1\}, & \text{if } x_i > 0, \\ [-1, 1], & \text{if } x_i = 0, \\ \{x_i - 1\}, & \text{if } x_i < 0. \end{cases}$$

Figure 1 illustrates the execution of (PGD). •

B. Distributed Optimization under Equality Constraints

Consider the constrained minimization problem

$$\min \Big\{ f(x) + \frac{1}{2}\|h(x)\|^2 \;\Big|\; h(x) = 0_p \Big\}. \tag{PE}$$

The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ associated with (PE) takes the form

$$L(x, \lambda) = f(x) + \frac{1}{2}\|h(x)\|^2 + \langle \lambda, h(x) \rangle,$$

Within this section we rely on the following assumptions:

(H4) The function f belongs to class $\mathcal{C}^{1,1}(\mathbb{R}^n, \mathbb{R})$.

(H5) The set of saddle points of L is nonempty, convex and compact.

Under the assumption (H4), the saddle-point dynamics associated with the Lagrangian L defined over $\mathbb{R}^n \times \mathbb{R}^p$ read

$$\begin{cases} \dot{x}(t) + \nabla_x L(x, \lambda)(t) = 0_n, & x(t_0) = x_0 \in \mathbb{R}^n, \\ \dot{\lambda}(t) - \nabla_\lambda L(x, \lambda)(t) = 0_p, & \lambda(t_0) = \lambda_0 \in \mathbb{R}^p, \end{cases} \tag{SP}$$

for all $t \in [t_0, t_1]$. The existence and uniqueness of local solutions $(x, \lambda) : [t_0, t_1] \to \mathbb{R}^n \times \mathbb{R}^p$ of (SP) is guaranteed by assumption (H4) and the Picard-Lindelöf theorem, cf. [23]. Our next result characterizes the performance of (SP).

Theorem 5.3 (Performance characterization)**.** *Let $L : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$. The following statements holds:*

(i) If ∇f is monotone, then $\mathrm{sp}(L)$ is stable under (SP).

(ii) If ∇f is strictly monotone, then $\mathrm{sp}(L)$ is asymptotically stable under (SP).

(iii) If $\partial(\nabla f) \succ 0$, then $\mathrm{sp}(L)$ is exponentially stable under (SP).

Theorem 5.3(iii) states that if the generalized Hessian of f is positive definite, then the convergence rate of the solutions of (SP) is exponential. However, the performance bound not only depends on the initial condition $(x_0, \lambda_0) \in \mathbb{R}^n \times \mathbb{R}^p$, but also on the convexity and regularity assumptions on the objective function. The following example illustrates the result obtained in Theorem 5.3(iii).

Example 2 (Saddle-point dynamics for equality constrained optimization). Consider a group of $n = 10$ agents whose objective is to cooperatively solve the minimization problem

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}} && \sum_{i \in \{1,\dots,n\}} x_i^2/2 \\ &\text{subject to} && \mathrm{Circ}_n(0, 1, 1/2)x = 1_n, \end{aligned} \tag{EX2}$$

where Circ_n denotes the tridiagonal circulant matrix of dimension $n \times n$. The network topology is encoded in the sparsity structure of Circ_n. The generalized Hessian of $x_i \mapsto f_i(x_i) = x_i^2/2$ at $x_i \in \mathbb{R}$ reduces to the singleton set

$$\partial(\nabla f_i)(x_i) = \{\nabla^2 f_i(x_i)\}.$$

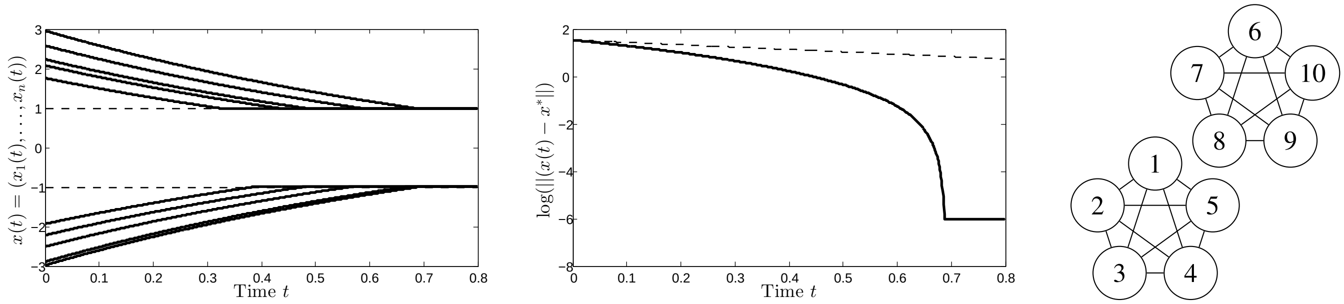
Figure 2 illustrates the execution of (SP). •

VI. Conclusions

We have studied the asymptotic convergence properties of distributed continuous-time coordination algorithms for networks of agents that seek to collectively solve a class of nonsmooth convex optimization problems. We have identified a nonsmooth Lyapunov function which, under mild convexity and regularity assumptions on the optimization problem data, allows us to go beyond the qualitative statement of asymptotic convergence. In particular, we have explicitly characterized the exponential convergence rates of the proposed algorithms for optimization subject to either equality or inequality constraints. Future work will characterize the convergence rates of the algorithms for problems with both equality and inequality constraints, study the robustness properties of the proposed algorithms against disturbances and link failures, design opportunistic state-triggered implementations, and explore the extension to optimization scenarios defined over infinite-dimensional state spaces.

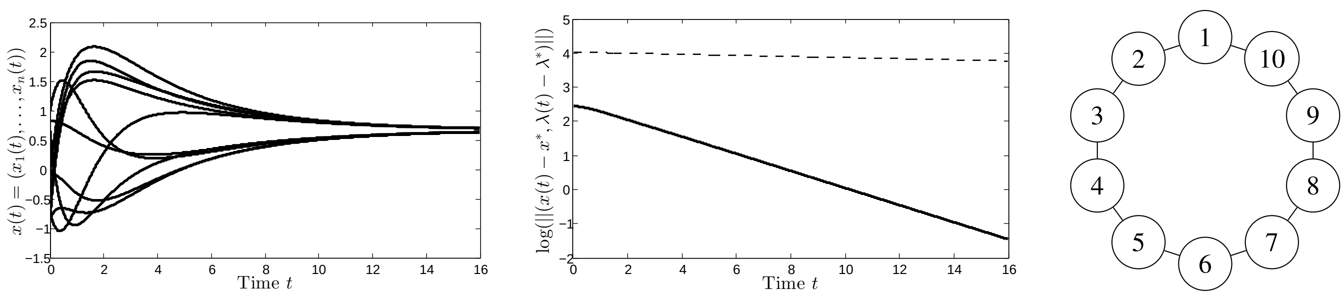
Acknowledgments

Research supported by the German Research Foundation (DFG) within the Cluster of Excellence in Simulation Technology (EXC 310/2) at the University of Stuttgart, Germany, and NSF Award CMMI-1300272.



(a) Network state evolution (b) Performance bound (c) Network topology

Fig. 1. (a) Solutions of the locally projected gradient dynamics (PGD) solving the nonsmooth convex program (EX1). The local projection prevents the solutions of (PGD) from violating the inequality constraints (depicted by the dashed lines) of (EX1) at any time instance. The initial conditions are randomly chosen within the feasible set. (b) Since the aggregate objective function $x \mapsto f(x) = 1/2\|x\|^2 + \|x\|_1$ of (EX1) is strongly convex, the hypothesis of Corollary 5.2 is satisfied and thus, the solutions of (PGD) converge to the singleton set $\mathrm{eq}(\Pi_G)$ within the exponential performance bound (depicted by the dashed line). (c) The network topology under which the the locally projected gradient dynamics (PGD) are amenable to distributed implementation.



(a) Network state evolution (b) Performance bound (c) Network topology

Fig. 2. (a) The network state evolution of the saddle-point dynamics (SP) solving the convex program (EX2). The initial conditions are randomly chosen within the interval $[-1, 1]$. (b) Since $\partial(\nabla f)(x) \succ 0$ for all $x \in \mathbb{R}^n$ (in fact, the aggregate objective function $x \mapsto f(x) = 1/2\|x\|^2$ of (EX2) is strongly convex), the hypothesis of Theorem 5.3(iii) is satisfied and therefore, the solutions of (SP) converge to the singleton set $\mathrm{sp}(L)$ within the exponential performance bound (depicted by the dashed line). (c) The network topology under which the saddle-point dynamics (SP) are distributed implementable.

References

[1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[2] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[3] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[4] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Control and Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.

[5] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.

[6] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *IEEE Conf. on Decision and Control*, Orlando, Florida, 2011, pp. 3800–3805.

[7] F. Zanella, D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-Raphson consensus for distributed convex optimization," in *IEEE Conf. on Decision and Control*, Orlando, Florida, Dec. 2011, pp. 5917–5922.

[8] K. Arrow, L. Hurwitz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*. Stanford, California: Stanford University Press, 1958.

[9] D. Richert and J. Cortés, "Robust distributed linear programming," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2567–2582, 2015.

[10] D. Feijer and F. Paganini, "Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, pp. 1974–1981, 2010.

[11] M. Forti, P. Nistri, and M. Quincampoix, "Generalized neural network for nonsmooth nonlinear programming problems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 51, no. 9, pp. 1741–1754, 2004.

[12] S. K. Niederländer and J. Cortés, "Distributed coordination for separable convex optimization with coupling constraints," in *IEEE Conf. on Decision and Control*, Osaka, Japan, 2015, pp. 694–699.

[13] S. K. Niederländer and J. Cortés, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *SIAM Journal on Control and Optimization*, 2016, submitted.

[14] F. H. Clarke, *Optimization and Nonsmooth Analysis*, ser. Canadian Mathematical Society Series of Monographs and Advanced Texts. Wiley, 1983.

[15] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, ser. Grundlehren Text Editions. New York: Springer, 1993.

[16] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V. H. Nguyen, "Generalized Hessian matrix and second-order optimality conditions for problems with $\mathcal{C}^{1,1}$ data," *Applied Mathematics and Optimization*, vol. 11, pp. 43–56, 1984.

[17] J. Cortés, "Discontinuous dynamical systems - a tutorial on solutions, nonsmooth analysis, and stability," *IEEE Control Systems Magazine*, vol. 28, no. 3, pp. 36–73, 2008.

[18] A. Nagurney and D. Zhang, *Projected Dynamical Systems and Variational Inequalities with Applications*, ser. International Series in Operations Research and Management Science. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996, vol. 2.

[19] J. P. Aubin and A. Cellina, *Differential Inclusions*, ser. Grundlehren der mathematischen Wissenschaften. New York: Springer, 1984, vol. 264.

[20] C. Henry, "An existence theorem for a class of differential equations with multivalued right-hand side," *Journal of Mathematical Analysis and Applications*, vol. 41, pp. 179–186, 1973.

[21] D. P. Bertsekas, "Necessary and sufficient conditions for a penalty method to be exact," *Mathematical Programming*, vol. 9, no. 1, pp. 87–99, 1975.

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

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