

# Distributed Coordination for Nonsmooth Convex Optimization via Saddle-Point Dynamics\*

Jorge Cortés<sup>†</sup>      Simon K. Niederländer<sup>‡</sup>

January 29, 2019

## Abstract

This paper considers continuous-time coordination algorithms for networks of agents that seek to collectively solve a general class of nonsmooth convex optimization problems with an inherent distributed structure. Our algorithm design builds on the characterization of the solutions of the nonsmooth convex program as saddle points of an augmented Lagrangian. We show that the associated saddle-point dynamics are asymptotically correct but, in general, not distributed because of the presence of a global penalty parameter. This motivates the design of a discontinuous saddle-point-like algorithm that enjoys the same convergence properties and is fully amenable to distributed implementation. Our convergence proofs rely on the identification of a novel global Lyapunov function for saddle-point dynamics. This novelty also allows us to identify mild convexity and regularity conditions on the objective function that guarantee the exponential convergence rate of the proposed algorithms for convex optimization problems subject to equality constraints. Various examples illustrate our discussion.

## 1 Introduction

Distributed convex optimization problems arise in a wide range of scenarios involving multi-agent systems, including network flow optimization, control of distributed energy resources, resource allocation and scheduling, and multi-sensor fusion. In such contexts, the goals and performance metrics of the agents are encoded into suitable objective functions whose optimization may be subject to a combination of physical, communication,

---

\*This work was supported by the German National Academic Foundation, the Dr. Jürgen and Irmgard Ulderup Foundation, Award FA9550-15-1-0108, and NSF Award CNS-1329619. Preliminary versions of this manuscript appeared as [43, 44] at the IEEE Conference on Decision and Control.

<sup>†</sup>Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, United States, Email: [cortes@ucsd.edu](mailto:cortes@ucsd.edu).

<sup>‡</sup>Institute for Systems Theory and Automatic Control, University of Stuttgart, 70550 Stuttgart, Germany, Email: [simon.niederlaender@ist.uni-stuttgart.de](mailto:simon.niederlaender@ist.uni-stuttgart.de).

and operational constraints. Decentralized algorithmic approaches to solve these optimization problems yield various advantages over centralized solvers, including reduced communication and computational overhead at a single point via spatially distributed processors, robustness against malfunctions, or the ability to quickly react to changes. In this paper we are motivated by network scenarios that give rise to general nonsmooth convex optimization problems with an intrinsic distributed nature. We consider convex programs with an additively separable objective function and local coupling equality and inequality constraints. Our objective is to synthesize distributed coordination algorithms that allow each agent to find their own component of the optimal solution vector. This setup substantially differs from consensus-based distributed optimization where agents agree on the entire optimal solution vector. We also seek to provide algorithm performance guarantees by way of characterizing the convergence rate of the network state towards the optimal solution. 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 on networked systems has stimulated the synthesis of distributed strategies that have agents interacting with neighbors to coordinate their computations and solve convex optimization problems with constraints [6, 7]. A majority of works focus on consensus-based approaches, where individual agents maintain, communicate, and update an estimate of the entire solution vector of the optimization problem, implemented in discrete time, see e.g., [18, 30, 41, 42, 57] and references therein. Recent work [23, 34, 37, 54] has proposed a number of continuous-time solvers whose convergence properties can be studied using notions and tools from classical stability analysis tools. This continuous-time framework facilitates the explicit computation of the evolution of candidate Lyapunov functions and their Lie derivatives, opening the way to a systematic characterization of additional desirable algorithm properties such as speed of convergence, disturbance rejection, and robustness to uncertainty.

In contrast to consensus-based approaches, and of particular importance to our work here, are distributed strategies where each agent seeks to determine only its component of the optimal solution vector (instead of the whole one) and interchanges information with neighbors whose size is independent of the networks'. Such strategies are particularly well suited for convex optimization problems over networks that involve an aggregate objective function that does not couple the agents' decisions but local (equality or inequality) constraints that instead do. Dynamics enjoying such scalability properties include the partition-based dual decomposition algorithm for network optimization proposed in [11], the discrete-time algorithm for non-strict convex problems in [40] that requires at least one of the exact solutions of a local optimization problem at each iteration, and the inexact algorithm in [39] that only achieves convergence to an approximate solution of the optimization problem. In the context of neural networks, the work [21] proposes a generalized circuit for nonsmooth nonlinear optimization based on first-order optimality conditions with convergence guarantees. However, the proposed dynamics are not fully amenable to

distributed implementation due to the global penalty parameters involved.

A common approach to design distributed strategies relies on the saddle-point or primal-dual dynamics [1, 32, 46] corresponding to the Lagrangian of the optimization problem. In fact, recent years have seen a burgeoning activity on the use of these dynamics to solve network optimization problems arising in a number of applications, including distributed control of power networks [33, 35, 52, 55, 56], internet congestion control [14, 31], and zero-sum networked games [22, 47]. The work [20] studies primal-dual gradient dynamics for convex programs subject to inequality constraints. These dynamics are modified with a projection operator on the dual variables to preserve their nonnegativity. Although convergence in the primal variables is established, the dual variables converge to some unknown point which might not correspond to a dual solution. The work [48] introduces set-valued and discontinuous saddle-point algorithms specifically tailored for linear programs. More recently, the work [13] studies the asymptotic convergence properties of the (continuous) saddle-point dynamics associated to general saddle functions. The work [24] takes a maximal monotone mapping approach to study the robustness of pointwise asymptotic stability of saddle-point dynamics. Our present work contributes to this body of literature on distributed algorithms based on saddle-point dynamics, with the key distinctions of the generality of the problem considered and the fact that our technical analysis relies on Lyapunov, rather than LaSalle, arguments. Also, while previous work focuses on establishing asymptotic stability, the present work studies algorithm performance and, in particular, the explicit characterization of the exponential convergence rate of continuous-time algorithms for convex optimization problems subject to equality constraints. The availability of such characterizations is of critical importance to assess the speed of convergence of discrete-time implementations of the dynamics, see for instance [17, 51].

## Statement of Contributions

We consider generic nonsmooth convex optimization problems defined by an additively separable objective function and local coupling constraints. Our starting point is the characterization of the primal-dual solutions of the nonsmooth convex program as saddle points of an augmented Lagrangian which incorporates quadratic regularization and  $\ell_1$ -exact-penalty terms to eliminate the inequality constraints. This problem reformulation motivates the study of the saddle-point dynamics (gradient descent in the primal variable and gradient ascent in the dual variable) associated with the augmented Lagrangian.

Our first contribution is the identification of a novel nonsmooth Lyapunov function which allows us to establish the asymptotic correctness of the algorithm without relying on arguments based on the LaSalle Invariance Principle. With respect to the current state of the art, the availability of this function opens the way to the study of other important properties of the trajectories beyond asymptotic stability, such as the characterization of robustness or the convergence rate. In fact, our second contribution pertains the performance characterization of the proposed coordination algorithms. We restrict our study to the case when the convex optimization problem is subject to equality constraints only. For this scenario, we rely on the Lyapunov function identified in the convergence analysis to provide sufficient conditions on the objective function of the convex program that guaran-

tee the exponential convergence of the algorithm and characterize the corresponding rate. Since the proposed saddle-point algorithm relies on a priori global knowledge of a penalty parameter associated with the exact-penalty terms introduced to ensure convergence to the solutions of the optimization problem, our third contribution is an alternative, discontinuous saddle-point-like algorithm that does not require such knowledge and is fully amenable to distributed implementation over a team of agents. We show that, given any solution of the saddle-point-like algorithm, there exists a value of the penalty parameter such that the trajectory is also a solution of the saddle-point dynamics, thereby establishing that both dynamics enjoy the same convergence properties. As an additional feature, the proposed algorithm guarantees feasibility with respect to the inequality constraints for any time. Various examples illustrate our discussion.

## Organization

Section 2 introduces basic notions on nonsmooth analysis and set-valued dynamical systems. Section 3 proposes the saddle-point algorithm to solve the problem of interest, establishes its asymptotic correctness and characterizes its exponential converge rate. Section 4 proposes an alternative saddle-point-like algorithm which does not require a priori knowledge of penalty parameters and examines its convergence properties and amenability for distributed implementation. Section 5 summarizes our conclusions and ideas for future work. Finally, we gather a number of intermediate results supporting the main technical developments of the paper in an appendix.

## 2 Preliminaries

We collect here some notions on nonsmooth analysis and set-valued dynamical systems used throughout the paper for completeness. The reader familiar with these concepts can safely skip this section.

We let  $\langle \cdot, \cdot \rangle$  denote the Euclidean inner product and  $\|\cdot\|$ , respectively  $\|\cdot\|_\infty$ , denote the  $\ell_2$ - and  $\ell_\infty$ -norms in  $\mathbb{R}^n$ . The Euclidean distance from a point  $x \in \mathbb{R}^n$  to a set  $X \subset \mathbb{R}^n$  is denoted by  $\text{dist}(x, X)$ . Let  $\mathbb{1}_n = (1, \dots, 1) \in \mathbb{R}^n$ . Given  $x \in \mathbb{R}^n$ , let  $[x]^+ = (\max\{0, x_1\}, \dots, \max\{0, x_n\}) \in \mathbb{R}^n$ . Given a set  $X \subset \mathbb{R}^n$ , we denote its convex hull by  $\text{co } X$ , its interior by  $\text{int } X$ , and its boundary by  $\text{bd } X$ . The closure of  $X$  is denoted by  $\text{cl } X = \text{int } X \cup \text{bd } X$ . Let  $\mathbb{B}(x, \delta) = \{y \in \mathbb{R}^n \mid \|y - x\| < \delta\}$  and  $\overline{\mathbb{B}}(x, \delta) = \{y \in \mathbb{R}^n \mid \|y - x\| \leq \delta\}$  be the open and closed ball, centered at  $x \in \mathbb{R}^n$  of radius  $\delta > 0$ . Given  $X, Y \subset \mathbb{R}^n$ , the Minkovski sum of  $X$  and  $Y$  is defined by  $X + Y = \{x + y \mid x \in X, y \in Y\}$ .

A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is *convex* if  $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$  for all  $x, y \in \mathbb{R}^n$  and  $\theta \in [0, 1]$ . A function  $f$  is *strictly convex* if  $f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)$  for all  $x \neq y$  and  $\theta \in (0, 1)$ . A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is *positive definite with respect to*  $X \subset \mathbb{R}^n$  if  $f(x) = 0$  for all  $x \in X$  and  $f(x) > 0$  for all  $x \notin X$ . We say  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is *coercive with respect to*  $X \subset \mathbb{R}^n$  if  $f(x) \rightarrow +\infty$  when  $\text{dist}(x, X) \rightarrow +\infty$ . Given  $\gamma > 0$ , the  $\gamma$ -*sublevel set* of  $f$  is  $\text{lev}_{\leq \gamma}(f) = \{x \in \mathbb{R}^n \mid f(x) \leq \gamma\}$ . A bivariate function  $f : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$  is *convex-concave*

if it is convex in its first argument and concave in its second. A *set-valued map*  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  maps elements of  $\mathbb{R}^n$  to elements of  $2^{\mathbb{R}^n}$ . A set-valued map  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is *monotone* if  $\langle x - y, \xi_x - \xi_y \rangle \geq 0$  whenever  $\xi_x \in F(x)$  and  $\xi_y \in F(y)$ . Finally,  $F$  is *strictly monotone* if  $\langle x - y, \xi_x - \xi_y \rangle > 0$  whenever  $\xi_x \in F(x)$ ,  $\xi_y \in F(y)$  and  $x \neq y$ .

## 2.1 Nonsmooth Analysis

We review here relevant basic notions from nonsmooth analysis [15] that will be most helpful in both our algorithm design and analysis. A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is *locally Lipschitz at*  $x \in \mathbb{R}^n$  if there exist  $\delta_x > 0$  and  $L_x > 0$  such that  $|f(y) - f(z)| \leq L_x \|y - z\|$  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 is locally Lipschitz (cf. [26, Theorem 3.1.1, p. 16]).

Rademacher's Theorem [15] states that locally Lipschitz functions are continuously differentiable almost everywhere (in the sense of Lebesgue measure). 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

$$\partial f(x) = \text{co} \left\{ \lim_{i \rightarrow +\infty} \nabla f(x_i) \mid x_i \rightarrow x, x_i \notin S \cup \Omega_f \right\}.$$

Note that if  $f$  is continuously differentiable at  $x \in \mathbb{R}^n$ , then  $\partial f(x)$  reduces to the singleton set  $\{\nabla f(x)\}$ . If  $f$  is convex, then  $\partial f(x)$  coincides with the *subdifferential* (in the sense of convex analysis), that is, the set of subgradients  $\xi \in \mathbb{R}^n$  satisfying

$$f(y) \geq f(x) + \langle \xi, y - x \rangle \tag{1}$$

for all  $y \in \mathbb{R}^n$  (cf. [15, Proposition 2.2.7]). With this characterization, it is not difficult to see that  $f$  is (strictly) convex if and only if  $\partial f$  is (strictly) monotone.

A set-valued map  $F$  is *upper semi-continuous* 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  $\varepsilon > 0$  and  $\delta > 0$  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 [15].

**Proposition 2.1** (Properties of the generalized gradient). *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  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 semi-continuous at  $x \in \mathbb{R}^n$ .

Let  $\mathcal{C}^{1,1}(\mathbb{R}^n, \mathbb{R})$  denote the class of functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  that are continuously differentiable and whose gradient  $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  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

$$\partial(\nabla f)(x) = \text{co} \left\{ \lim_{i \rightarrow +\infty} \nabla^2 f(x_i) \mid x_i \rightarrow x, x_i \notin \Omega_f \right\}.$$

By construction,  $\partial(\nabla f)(x)$  is a nonempty, convex and compact set of symmetric matrices which reduces to the singleton set  $\{\nabla^2 f(x)\}$  whenever  $f$  is twice continuously differentiable at  $x \in \mathbb{R}^n$  [27]. The following result is a direct extension of Lebourg's Mean-Value Theorem to vector-valued functions [50].

**Proposition 2.2** (Extended Mean-Value Theorem). *Let  $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be locally Lipschitz and let  $x, y \in \mathbb{R}^n$ . Then,*

$$\nabla f(x) - \nabla f(y) \in \text{co} \{ \partial(\nabla f)([x, y]) \} (x - y),$$

where  $\text{co} \{ \partial(\nabla f)([x, y]) \} (x - y) = \text{co} \{ H(x - y) \mid H \in \partial(\nabla f)(z) \text{ for some } z \in [x, y] \}$ , and  $[x, y] = \{x + \theta(y - x) \mid \theta \in [0, 1]\}$ .

## 2.2 Set-Valued Dynamical Systems

Throughout the manuscript, we consider set-valued and locally projected dynamical systems [16, 38] defined by differential inclusions [2]. 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

$$\begin{cases} \dot{x} \in F(x), \\ x(0) = x_0 \in X. \end{cases} \quad (\text{DI})$$

A solution of (DI) on the interval  $[0, t_+) \subset \mathbb{R}$  (if any) is an absolutely continuous mapping taking values in  $X$ , denoted by  $x \in \mathcal{AC}([0, t_+), X)$ , such that  $\dot{x}(t) \in F(x(t))$  for almost all  $t \in [0, t_+)$ . A point  $x$  is an equilibrium of (DI) if  $0 \in F(x)$ . We denote by  $\text{eq}(F)$  the set of equilibria. Given  $x_0 \in X$ , the existence of solutions of (DI) with initial condition  $x_0 \in X$  is guaranteed by the following result [2, 4, 16].

**Lemma 2.3** (Existence of local solutions). *Let the set-valued map  $F : X \rightrightarrows \mathbb{R}^n$  be locally bounded, upper semi-continuous with nonempty, convex and compact values. Then, given any  $x_0 \in X$ , there exists a solution of (DI) with initial condition  $x_0$ .*

Given a locally Lipschitz function  $V : X \rightarrow \mathbb{R}$ , the set-valued Lie derivative  $\mathcal{L}_F V : X \rightrightarrows \mathbb{R}$  of  $V$  with respect to  $F$  at  $x \in X$  is defined by

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

For each  $x \in X$ ,  $\mathcal{L}_F V(x)$  is a closed and bounded interval in  $\mathbb{R}$ , possibly empty.

Let  $G \subset \mathbb{R}^n$  be a nonempty, closed and convex set. The *tangent cone* and the *normal cone* of  $G$  at  $x \in G$  are, respectively,

$$T_G(x) = \text{cl } \mathbb{R}_{\geq 0}(G - x), \quad N_G(x) = \{ n \in \mathbb{R}^n \mid \langle n, y - x \rangle \leq 0, \forall y \in G \}.$$

Note that if  $x \in \text{int } G$ , then  $T_G(x) = \mathbb{R}^n$  and  $N_G(x) = \{0\}$ . Let  $\text{proj}_G(x) = \arg \min_{y \in G} \|x - y\|$ . 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

$$P_{T_G(x)}(F(x)) = \bigcup_{\xi \in F(x)} \lim_{\delta \searrow 0} \frac{\text{proj}_G(x + \delta \xi) - x}{\delta}. \quad (2)$$

Note that if  $x \in \text{int } G$ , then  $P_{T_G(x)}(F(x))$  reduces to the set  $F(x)$ . By definition, the orthogonal projection  $P_{T_G(x)}(F(x))$  is equivalent to the Euclidean projection of  $F(x) \subset \mathbb{R}^n$  onto the tangent cone  $T_G(x)$  at  $x \in G$ , i.e.,  $P_{T_G(x)}(F(x)) = \text{proj}_{T_G(x)}(F(x))$ , cf. [45, Remark 1.1]. Consider now the locally projected differential inclusion

$$\begin{cases} \dot{x} \in P_{T_G(x)}(F(x)), \\ x(0) = x_0 \in G. \end{cases} \quad (\text{PDI})$$

Note that, in general, the set-valued map  $x \mapsto P_{T_G(x)}(F(x))$  possesses no continuity properties and the values of  $P_{T_G(x)}(F(x))$  are not necessarily convex [2]. The following result [9, 25] states conditions under which solutions of (PDI) exist.

**Lemma 2.4** (Existence of local solutions of projected differential inclusions). *Let  $G \subset \mathbb{R}^n$  be nonempty, closed and convex, and let the set-valued map  $F : G \rightrightarrows \mathbb{R}^n$  be locally bounded, upper semi-continuous with nonempty, convex and compact values. 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 at least one solution  $x \in \mathcal{AC}([0, t_+), G)$  of (PDI) with initial condition  $x_0$ .*

### 3 Convex Optimization via Saddle-Point Dynamics

Consider the constrained minimization problem

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

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  are convex, and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$  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  $C = \{x \in \mathbb{R}^n \mid h(x) = 0_p, g(x) \leq 0_m\}$  denote the constraint set and assume that the (closed and convex) set of solutions  $S = \{x^* \in C \mid f(x^*) = \inf_C f\}$  of (P) is nonempty and bounded. Throughout the paper we assume that the constraint set  $C \subset \mathbb{R}^n$  satisfies the strong Slater assumptions [26], i.e.,

(A1)  $\text{rank}(A) = p$ , i.e., the rows of  $A \in \mathbb{R}^{p \times n}$  are linearly independent;

(A2)  $\exists x \in \mathbb{R}^n$  such that  $h(x) = 0_p$  and  $g_k(x) < 0$  for all  $k \in \{1, \dots, m\}$ .

Our main objective is to design continuous-time algorithms with performance guarantees to find the solution of the nonsmooth convex program (P). We are specifically interested in solvers that are amenable to distributed implementation by a group of agents, permitting each one of them to find their component of the solution vector. The algorithms proposed in this work build on concepts of Lagrangian duality theory and characterize the primal-dual solutions of (P) as saddle points of an augmented Lagrangian. More precisely, consider the Lagrangian function associated with (P),

$$f(x) + \langle \lambda, h(x) \rangle + \kappa \langle \mathbb{1}_m, [g(x)]^+ \rangle, \quad (3)$$

where  $\kappa > 0$  and  $\lambda \in \mathbb{R}^p$  is a Lagrange multiplier. Throughout the paper, we find it more convenient to deal with an augmented Lagrangian  $L : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$  of the form

$$L(x, \lambda) = f(x) + \frac{1}{2\mu} \|h(x)\|^2 + \langle \lambda, h(x) \rangle + \kappa \langle \mathbb{1}_m, [g(x)]^+ \rangle, \quad (4)$$

where  $\mu > 0$ . Under the regularity assumptions (A1)–(A2), for every  $x^* \in S$ , there exists  $(\lambda^*, \nu^*) \in \mathbb{R}^p \times \mathbb{R}^m$  such that  $(x^*, \lambda^*)$  is a *saddle point* of the augmented Lagrangian  $L$ , i.e.,

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*), \quad \forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^p,$$

provided that  $\kappa \geq \|\nu^*\|_\infty$ . We denote the (closed and convex) set of saddle points of  $L$  by  $\text{sp}(L)$  (note that the addition of the augmentation term  $\frac{1}{2\mu} \|h(x)\|^2$  does not change the saddle points, i.e., the original Lagrangian (3) and the augmented Lagrangian (4) have exactly the same saddle points). The following statement reveals that the converse result is also true. Its proof can be deduced from results on penalty functions in the optimization literature, see e.g. [5], and therefore we omit it.

**Lemma 3.1** (Saddle Points and Solutions of the Optimization Problem). *Let  $L : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$  and let  $(x^*, \lambda^*) \in \text{sp}(L)$  with  $\kappa > \|\nu^*\|_\infty$  for some dual solution  $\nu^*$  of (P). Then,  $x^* \in S$ , i.e.,  $x^*$  is a (primal) solution of (P).*

Lemma 3.1 identifies a condition under which the penalty parameter  $\kappa$  is *exact* [5, 36]. This condition in turn depends on the dual solution set. Given this result, instead of directly solving (P), we seek to design strategies that find saddle points of  $L$ . Since the bivariate augmented Lagrangian  $L$  is, by definition, convex-concave, a natural approach to find the saddle points is via its associated saddle-point dynamics

$$\begin{cases} (\dot{x}, \dot{\lambda}) \in -(\partial_x L, -\partial_\lambda L)(x, \lambda), \\ (x(0), \lambda(0)) = (x_0, \lambda_0) \in \mathbb{R}^n \times \mathbb{R}^p. \end{cases} \quad (\text{SPD})$$

Let  $(\partial_x L, -\partial_\lambda L) : \mathbb{R}^n \times \mathbb{R}^p \rightrightarrows \mathbb{R}^n \times \mathbb{R}^p$  denote the *saddle-point operator* associated with  $L$ . Since  $L$  is convex-concave and locally Lipschitz [49, Theorem 35.1], the mapping  $(x, \lambda) \mapsto (\partial_x L, -\partial_\lambda L)(x, \lambda)$  is locally bounded, upper semi-continuous and takes nonempty, convex and compact values, cf. Proposition 2.1. The existence of local solutions  $(x, \lambda) \in \mathcal{AC}([0, t_+], \mathbb{R}^n \times \mathbb{R}^p)$  of (SPD) is guaranteed by Lemma 2.3. Moreover, we have that

$$\text{sp}(L) = (\partial_x L, -\partial_\lambda L)^{-1}(0_n, 0_p) = \text{eq}(\partial_x L, -\partial_\lambda L).$$

Consequently, our strategy to solve (P) amounts to the issue of “finding the zeros” of the saddle-point operator via (SPD). One can in fact show [49] that the saddle-point operator  $(\partial_x L, -\partial_\lambda L) : \mathbb{R}^n \times \mathbb{R}^p \rightrightarrows \mathbb{R}^n \times \mathbb{R}^p$  is maximal monotone, and thus, the existence and uniqueness of a global solution of (SPD) follows from [2, Theorem 1, p. 147].

### 3.1 Convergence Analysis

By construction of the saddle-point dynamics (SPD), it is natural to expect that its trajectories converge towards the set of saddle points of the augmented Lagrangian  $L$  as time evolves. Our proof strategy to establish this convergence result relies on Lyapunov’s direct method.



**Theorem 3.2** (Asymptotic convergence). *Let  $L : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$  with  $\mu \in (0, 1)$  and  $\kappa > \|\nu^*\|_\infty$  for some dual solution  $\nu^*$  of (P). Then, the set  $\text{sp}(L)$  is strongly globally asymptotically stable under (SPD).*

*Proof.* We start by observing that the set of saddle points  $\text{sp}(L)$  is nonempty, convex and compact given that  $\kappa > \|\nu^*\|_\infty$ , the solution set  $S \subset \mathbb{R}^n$  of (P) is nonempty and bounded, and that the strong Slater assumptions hold, cf. [26, Theorem 2.3.2]. Consider the candidate Lyapunov function  $V : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$  defined by

$$\begin{aligned} V(x, \lambda) &= f(x) - \inf_C f + \frac{1}{2\mu} \|h(x)\|^2 + \langle \lambda, h(x) \rangle + \kappa \langle \mathbb{1}_m, [g(x)]^+ \rangle \\ &\quad + \frac{1}{2} \text{dist}((x, \lambda), \text{sp}(L))^2. \end{aligned} \tag{5}$$

We start by showing that  $V(x, \lambda) > 0$  for all  $(x, \lambda) \notin \text{sp}(L)$  and  $V(x, \lambda) = 0$  if  $(x, \lambda) \in \text{sp}(L)$ . Let  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^p$  and let

$$(x^*(x, \lambda), \lambda^*(x, \lambda)) = \arg \min_{(\tilde{x}, \tilde{\lambda}) \in \text{sp}(L)} \left( \|x - \tilde{x}\|^2 + \|\lambda - \tilde{\lambda}\|^2 \right). \tag{6}$$

For notational convenience, we drop the dependency of  $(x^*, \lambda^*)$  on the argument  $(x, \lambda)$ . Note that the (unique) minimizer  $(x^*, \lambda^*)$  of (6) exists since the set  $\text{sp}(L)$  is nonempty, convex and compact.

Using the fact that  $(x^*, \lambda^*)$  is a saddle point of the original Lagrangian (3) too, and invoking Lemma 3.1, we deduce  $f(x^*) \leq f(x) + \langle \lambda^*, h(x) \rangle + \kappa \langle \mathbb{1}_m, [g(x)]^+ \rangle$ . Adding  $\langle \lambda, h(x) \rangle$  to both sides of the equation, we obtain

$$f(x) - \inf_C f + \langle \lambda, h(x) \rangle + \kappa \langle \mathbb{1}_m, [g(x)]^+ \rangle \geq \langle \lambda - \lambda^*, h(x) \rangle.$$

Using this inequality and the fact that  $\text{dist}((x, \lambda), \text{sp}(L))^2 = \|x - x^*\|^2 + \|\lambda - \lambda^*\|^2$  in the definition of  $V$ , one can write

$$\begin{aligned} V(x, \lambda) &\geq \frac{1}{2\mu} \|h(x)\|^2 + \langle \lambda - \lambda^*, h(x) \rangle + \frac{1}{2} \|x - x^*\|^2 + \frac{1}{2} \|\lambda - \lambda^*\|^2 \\ &= \frac{1}{2} \langle P(x - x^*, \lambda - \lambda^*), (x - x^*, \lambda - \lambda^*) \rangle \end{aligned}$$

where the symmetric matrix

$$P = \begin{pmatrix} \frac{1}{\mu} A^\top A + I_n & A^\top \\ A & I_p \end{pmatrix} \in \mathbb{R}^{(n+p) \times (n+p)} \tag{7}$$

is positive definite for  $\mu \in (0, 1]$  (this follows by observing that  $I_p \succ 0$  and the Schur complement [29] of  $I_p$  in  $P$ , denoted  $P/I_p = \frac{1}{\mu} A^\top A + I_n - A^\top I_p^{-1} A$ , is positive definite). Since  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^p$  is arbitrary, it follows

$$V(x, \lambda) \geq \frac{1}{2} \lambda_{\min}(P) \text{dist}((x, \lambda), \text{sp}(L))^2.$$

Thus, we have  $V(x, \lambda) > 0$  for all  $(x, \lambda) \notin \text{sp}(L)$  and  $V(x, \lambda) = 0$  if  $(x, \lambda) \in \text{sp}(L)$ . Moreover,  $V$  is coercive with respect to  $\text{sp}(L)$ , that is,  $V(x, \lambda) \rightarrow +\infty$  whenever  $\text{dist}((x, \lambda), \text{sp}(L)) \rightarrow +\infty$ .

We continue by studying the evolution of  $V$  along the solutions of the saddle-point dynamics (SPD). Let  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^p$  and let  $(x^*, \lambda^*)$  be defined by (6). Take  $\psi \in \mathcal{L}_{\text{SPD}}V(x, \lambda)$ . By definition of the set-valued Lie derivative, there exists  $(\pi_x, \pi_\lambda) \in (\partial_x L, -\partial_\lambda L)(x, \lambda)$  such that  $\psi = -\langle (\xi_x, \xi_\lambda), (\pi_x, \pi_\lambda) \rangle$  for all  $(\xi_x, \xi_\lambda) \in (\partial_x V, \partial_\lambda V)(x, \lambda)$ . Taking in particular,  $(\xi_x, \xi_\lambda) = (\pi_x, -\pi_\lambda) + (x - x^*, \lambda - \lambda^*)$ , we obtain

$$\psi = -\langle (\xi_x, \xi_\lambda), (\pi_x, \pi_\lambda) \rangle = -\|\pi_x\|^2 + \|\pi_\lambda\|^2 - \langle (\pi_x, \pi_\lambda), (x - x^*, \lambda - \lambda^*) \rangle$$

We focus on the last term. Noting that  $\pi_x = \pi_f + \frac{1}{\mu}A^\top h(x) + A^\top \lambda + \kappa \sum_{k \in \{1, \dots, m\}} \pi_{g_k}^+$ , for some  $\pi_f \in \partial f(x)$  and  $\pi_{g_k}^+ \in \partial[g_k(x)]^+$ , and  $\pi_\lambda = -h(x)$ , we have

$$\begin{aligned} -\langle (\pi_x, \pi_\lambda), (x - x^*, \lambda - \lambda^*) \rangle &= \langle \pi_f, x^* - x \rangle + \kappa \left\langle \sum_{k \in \{1, \dots, m\}} \pi_{g_k}^+, x^* - x \right\rangle \\ &\quad + \left\langle \frac{1}{\mu}A^\top h(x) + A^\top \lambda, x^* - x \right\rangle + \langle h(x), \lambda - \lambda^* \rangle. \end{aligned}$$

Using the first-order condition (1) of convexity for  $f$  and  $\langle \mathbb{1}_m, [g]^+ \rangle$  and the fact that  $x^* \in S$ , we deduce

$$\begin{aligned} -\langle (\pi_x, \pi_\lambda), (x - x^*, \lambda - \lambda^*) \rangle &\leq f(x^*) - f(x) + \kappa \langle \mathbb{1}_m, [g(x^*)]^+ \rangle - \kappa \langle \mathbb{1}_m, [g(x)]^+ \rangle \\ &\quad - \left\langle \frac{1}{\mu}A^\top h(x) + A^\top \lambda, x \right\rangle + \langle h(x), \lambda - \lambda^* \rangle \\ &= f(x^*) - f(x) - \langle \lambda^*, h(x) \rangle - \kappa \langle \mathbb{1}_m, [g(x)]^+ \rangle \\ &\quad - \left\langle \frac{1}{\mu}h(x) + \lambda, h(x) \right\rangle + \langle h(x), \lambda \rangle, \end{aligned}$$

where in the last equality we have used the fact that  $h(x) = Ax$ . Substituting this into the expression for  $\psi$  above and using again the fact that  $f(x^*) \leq f(x) + \langle \lambda^*, h(x) \rangle + \kappa \langle \mathbb{1}_m, [g(x)]^+ \rangle$ , we obtain

$$\begin{aligned} \psi &\leq - \left\| \frac{1}{\mu}A^\top h(x) + A^\top \lambda + \pi_f + \kappa \sum_{k \in \{1, \dots, m\}} \pi_{g_k}^+ \right\|^2 - \left( \frac{1}{\mu} - 1 \right) \|h(x)\|^2 \\ &\leq - \min \left\{ 1, \frac{1}{\mu} - 1 \right\} \left\| \left( \frac{1}{\mu}A^\top h(x) + A^\top \lambda + \xi_f + \kappa \sum_{k \in \{1, \dots, m\}} \xi_{g_k}^+, h(x) \right) \right\|^2. \end{aligned} \tag{8}$$

Since  $\mu \in (0, 1)$ , and using the fact that  $\text{sp}(L) = (\partial_x L, -\partial_\lambda L)^{-1}(0_n, 0_p)$ , we deduce that the right-hand-side of (8) equals to zero if and only if  $(x, \lambda) \in \text{sp}(L)$ . Since  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^p$  and  $\psi \in \mathcal{L}_{\text{SPD}}V(x, \lambda)$  are arbitrary, it follows  $\mathcal{L}_{\text{SPD}}V(x, \lambda) \subset (-\infty, 0)$  for all  $(x, \lambda) \notin \text{sp}(L)$ . Hence, the set of saddle points  $\text{sp}(L)$  is strongly globally asymptotically stable under (SPD), concluding the proof.  $\square$

**Remark 3.3** (Alternative convergence proof via the LaSalle Invariance Principle). The strong global asymptotic stability of  $\text{sp}(L)$  under (SPD) can also be established using the alternative *non-strict* Lyapunov function

$$V(x, \lambda) = \frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|\lambda - \lambda^*\|^2, \quad (9)$$

where  $(x^*, \lambda^*) \in \text{sp}(L)$  is arbitrary. In fact, from the proof of Theorem 3.2, one can deduce that the Lie derivative of  $V$  along (SPD) is only negative semidefinite, implying stability, albeit not asymptotic stability. To conclude the latter, one can invoke the LaSalle Invariance Principle for differential inclusions [3] to identify the limit points of the trajectories as the set of saddle points. In fact, this is the approach traditionally taken in the literature characterizing the convergence properties of saddle-point dynamics, see e.g., [13, 20, 24, 28, 54] and references therein. This approach has the disadvantage that  $V$ , not being a strict Lyapunov function, cannot be used to characterize properties of the solutions of (SPD) beyond asymptotic convergence. By contrast, the Lyapunov function (5) identified in the proof of Theorem 3.2 is *strict*. Albeit this observation might seem like a minor point, it is actually quite significant, as no such function has been identified in the study of primal-dual dynamics, going back to the seminal works [1, 32]. The availability of a strict Lyapunov function opens the way to the study of other properties of the solutions such as the characterization of the rate of convergence (the subject of our next section), the robustness against disturbances via the notion of input-to-state stability, or the design of opportunistic state-triggered implementations that naturally result in aperiodic discrete-time algorithms. For general saddle-point dynamics, not necessarily arising from convex optimization problems, the work [12] shows that the function (9) is a Lyapunov function under the stringent condition that the Lagrangian giving rise to the dynamics is strongly convex in  $x$  and strongly concave in  $\lambda$  (which does not hold for Lagrangians arising from optimization problems, which are linear in the dual variable). •

It is worth noticing that the convergence argument with the non-strict Lyapunov function (9) of Remark 3.3 ensures the strong global asymptotic stability of  $\text{sp}(L)$  under (SPD) for any value of the parameter  $\kappa > 0$ . However, only when this parameter is exact, cf. Lemma 3.1, it is guaranteed that saddle points correspond to (primal) solutions of (P).

Point-wise convergence of the solutions of (SPD) in the set  $\text{sp}(L)$  follows from the stability of the each individual saddle point and the asymptotic stability of the set  $\text{sp}(L)$  established in Theorem 3.2, as stated in the following result. The proof is analogous to the case of ordinary differential equations, cf. [8, Corollary 5.2]), and hence we omit it for reasons of space.

**Corollary 3.4** (Point-wise asymptotic convergence). *Any solution  $(x, \lambda) \in \mathcal{AC}([0, +\infty), \mathbb{R}^n \times \mathbb{R}^p)$  of (SPD) starting from  $\mathbb{R}^n \times \mathbb{R}^p$  converges asymptotically to a point in the set  $\text{sp}(L)$ .*

## 3.2 Performance Characterization for Equality-Constrained Problems

In this section, we characterize the exponential convergence rate of solutions of the saddle-point dynamics (SPD) for the case when the convex optimization problem (P) is subject to

equality constraints only. In order to do so, we pose additional convexity and regularity assumptions on the objective function of (P). We have gathered in the Appendix various intermediate results to ease the exposition of the following result.

**Theorem 3.5** (Exponential convergence). *Let  $L : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$  with  $\mu \in (0, 1)$ . Suppose that  $f \in \mathcal{C}^{1,1}(\mathbb{R}^n, \mathbb{R})$  and  $\partial(\nabla f) \succ 0$ . Then, the (singleton) set  $\text{sp}(L)$  is exponentially stable under (SPD).*

*Proof.* Under the assumptions of the result, note that the dynamics (SPD) take the form of a differential equation,

$$\begin{cases} (\dot{x}, \dot{\lambda}) = -(\nabla_x L, -\nabla_\lambda L)(x, \lambda), \\ (x(0), \lambda(0)) = (x_0, \lambda_0) \in \mathbb{R}^n \times \mathbb{R}^p. \end{cases}$$

Let  $L : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$  with  $\mu \in (0, 1)$ . Since  $\partial(\nabla f) \succ 0$ , it follows that  $\nabla f$  is strictly monotone [27, Example 2.2]. Therefore, for any fixed  $\lambda$ , the mapping  $x \mapsto \nabla_x L(x, \lambda)$  is strictly monotone as well. Thus, by assumption (A1), the set of saddle points of  $L$  is a singleton, i.e.,  $\text{sp}(L) = \{x^*\} \times \{\lambda^*\}$ , where  $\lambda^* = -(AA^\top)^{-1}A\nabla f(x^*)$ . In this case, the Lyapunov function  $V$  as defined in (5) reads

$$V(x, \lambda) = f(x) - f(x^*) + \frac{1}{2\mu} \|h(x)\|^2 + \langle \lambda, h(x) \rangle + \frac{1}{2} \text{dist}((x, \lambda), \text{sp}(L))^2.$$

and we readily obtain from the proof of Theorem 3.2 that

$$V(x, \lambda) \geq \frac{1}{2} \lambda_{\min}(P) \text{dist}((x, \lambda), \text{sp}(L))^2,$$

for all  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^p$ , where  $P \in \mathbb{R}^{(n+p) \times (n+p)}$  is defined as in (7).

Our next objective is to upper bound the evolution of  $V$  along the solutions of (SPD) in terms of the distance to  $\text{sp}(L)$ . Let  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^p$  and consider the Lie derivative of  $V$  with respect to (SPD) at  $(x, \lambda)$ , i.e.,

$$\mathcal{L}_{\text{SPD}} V(x, \lambda) = -\langle x - x^*, \nabla f(x) - \nabla f(x^*) \rangle - \|\nabla_x L(x, \lambda)\|^2 - \left(\frac{1}{\mu} - 1\right) \|\nabla_\lambda L(x, \lambda)\|^2,$$

where we have used the fact that  $0 = \nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \frac{1}{\mu} A^\top h(x^*) + A^\top \lambda^*$  and  $0 = \nabla_\lambda L(x^*, \lambda^*) = h(x^*)$ . Now, since  $f \in \mathcal{C}^{1,1}(\mathbb{R}^n, \mathbb{R})$ , the gradient  $\nabla f$  is locally Lipschitz and thus, the extended Mean-Value Theorem, cf. Proposition 2.2, yields

$$\nabla f(x) - \nabla f(x^*) \in \text{co} \{ \partial(\nabla f([x, x^*])) \} (x - x^*),$$

where  $[x, x^*] = \{x + \theta(x^* - x) \mid \theta \in [0, 1]\}$ . By Lemma A.1, there exists a symmetric and positive definite matrix  $H(x) \in \text{co} \{ \partial(\nabla f([x, x^*])) \}$  such that  $\nabla f(x) - \nabla f(x^*) = H(x)(x - x^*)$ . Hence, after some computations, one can obtain

$$\mathcal{L}_{\text{SPD}} V(x, \lambda) = -\langle Q(x)(x - x^*, \lambda - \lambda^*), (x - x^*, \lambda - \lambda^*) \rangle,$$

where the symmetric matrix  $Q(x) \in \mathbb{R}^{(n+p) \times (n+p)}$  is given by

$$Q = \begin{pmatrix} H + \left(H + \frac{1}{\mu} A^\top A\right)^\top \left(H + \frac{1}{\mu} A^\top A\right) + \left(\frac{1}{\mu} - 1\right) A^\top A & H^\top A^\top + \frac{1}{\mu} A^\top A A^\top \\ AH + \frac{1}{\mu} A A^\top A & A A^\top \end{pmatrix}.$$

By assumption (A1), we have that  $AA^\top \succ 0$ . The Schur complement of  $AA^\top$  in  $Q(x)$ , denoted by  $Q(x)/AA^\top$ , reads

$$Q(x)/AA^\top = H(x) + H(x)^\top (I_n - A^\top (AA^\top)^{-1} A) H(x) + \left(\frac{1}{\mu} - 1\right) A^\top A.$$

Since  $I_n - A^\top (AA^\top)^{-1} A$  is a projection matrix, i.e., symmetric and idempotent, it follows that  $I_n - A^\top (AA^\top)^{-1} A \succeq 0$ . Moreover, since  $H(x) \succ 0$  (cf. Lemma A.1) and  $\mu \in (0, 1)$ , we conclude that  $Q(x)/AA^\top$  is positive definite, and so is  $Q(x)$ . Thus, we have  $\mathcal{L}_{\text{SPD}} V(x, \lambda) \leq -\lambda_{\min}(Q(x)) \text{dist}((x, \lambda), \text{sp}(L))^2$ . More generally, since any  $\gamma$ -sublevel set  $\text{lev}_{\leq \gamma} V$  is compact and positively invariant under (SPD), and  $\text{co}\{\partial(\nabla f([x, x^*]))\} \succ 0$  for all  $x, x^* \in \mathbb{R}^n$  (cf. Lemma A.1), we conclude

$$\mathcal{L}_{\text{SPD}} V(x, \lambda) \leq -\eta \text{dist}((x, \lambda), \text{sp}(L))^2,$$

for all  $(x, \lambda) \in \text{lev}_{\leq \gamma} V$ , where

$$\eta = \min_{(x, \lambda) \in \text{lev}_{\leq \gamma} V} \min_{H(x) \in \text{co}\{\partial(\nabla f([x, x^*]))\}} \lambda_{\min}(Q(x)) > 0.$$

Note that both minima are attained since  $\text{lev}_{\leq \gamma} V$  and  $\text{co}\{\partial(\nabla f([x, x^*]))\}$  are compact sets, cf. Lemma A.2.

We now proceed to quadratically upper bound the function  $V$ . By convexity of  $f$  and  $h$ , we obtain

$$\begin{aligned} V(x, \lambda) &\leq \langle x - x^*, \nabla f(x) - \nabla f(x^*) \rangle + \frac{1}{2\mu} \|h(x)\|^2 + \langle \lambda - \lambda^*, h(x) \rangle \\ &\quad + \frac{1}{2} \text{dist}((x, \lambda), \text{sp}(L))^2 \\ &= \frac{1}{2} \langle R(x)(x - x^*, \lambda - \lambda^*), (x - x^*, \lambda - \lambda^*) \rangle, \end{aligned}$$

where the symmetric and positive definite matrix  $R(x) \in \mathbb{R}^{(n+p) \times (n+p)}$  is given by

$$R(x) = \begin{pmatrix} 2H(x) + \frac{1}{\mu} A^\top A + I_n & A^\top \\ A & I_p \end{pmatrix}.$$

Similar arguments as above yield

$$V(x, \lambda) \leq \frac{1}{2} \vartheta \text{dist}((x, \lambda), \text{sp}(L))^2$$

for all  $(x, \lambda) \in \text{lev}_{\leq \gamma} V$ , where

$$\vartheta = \max_{(x, \lambda) \in \text{lev}_{\leq \gamma} V} \max_{H(x) \in \text{co}\{\partial(\nabla f([x, x^*]))\}} \lambda_{\max}(R(x)) > 0.$$

Note again that both maxima are attained since  $\text{lev}_{\leq \gamma} V$  and  $\text{co}\{\partial(\nabla f([x, x^*]))\}$  are compact sets, cf. Lemma A.2.

Since  $\frac{d}{dt}V(x(t), \lambda(t)) = \mathcal{L}_{\text{SPD}}V(x(t), \lambda(t))$  for all  $t \in [0, +\infty)$ , we have

$$\frac{d}{dt}V(x(t), \lambda(t)) \leq -\eta \text{dist}((x(t), \lambda(t)), \text{sp}(L))^2 \leq -\frac{2\eta}{\vartheta}V(x(t), \lambda(t)),$$

for all  $t \in [0, +\infty)$ . Integration yields

$$V(x(t), \lambda(t)) \leq V(x_0, \lambda_0) \exp\left(-\frac{2\eta}{\vartheta}t\right),$$

and therefore,

$$\text{dist}((x, \lambda), \text{sp}(L)) \leq \sqrt{\frac{\vartheta}{\lambda_{\min}(P)}} \text{dist}((x_0, \lambda_0), \text{sp}(L)) \exp\left(-\frac{\eta}{\vartheta}t\right),$$

for all  $t \in [0, +\infty)$ . Therefore, the singleton set  $\text{sp}(L)$  is exponentially stable and the convergence rate of solutions of (SPD) is upper bounded by  $\eta/\vartheta$ .  $\square$

The exponential convergence rate in Theorem 3.5 depends not only 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. However, if  $f \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$  is quadratic, then the convergence rate is determined by  $\lambda_{\min}(Q)/\lambda_{\max}(R)$ , independently of  $x \in \mathbb{R}^n$ . We also note that the exponential convergence guarantee does not rule out the possibility of complex dynamical behavior of the algorithm trajectories, possibly involving chaotic transients, depending on the complexity of the optimization problems, and in particular, the network constraints, see e.g. [19].

**Remark 3.6** (Connection with discrete-time algorithms). Compared to discrete-time solvers such as the augmented Lagrangian method (originally known as the method of multipliers), the exponential convergence rate for the continuous-time dynamics here corresponds to a linear convergence rate for its first-order Euler discretization. It is worth mentioning that, exclusively for the case of equality-constrained problems, we obtain the exponential rate under slightly weaker conditions than the ones usually stated in the literature for the augmented Lagrangian method, namely Lipschitz continuity of  $\nabla f$  and strong convexity of  $f$ .  $\bullet$

**Remark 3.7** (Performance analysis under inequality and equality constraints). Our performance analysis here only considers convex optimization scenarios subject to equality constraints. A natural question is whether this analysis can be extended to the general case including both inequality and equality constraints. Clearly, the performance bound

derived in Theorem 3.5 holds true whenever the inequality constraints are inactive. However, one faces various technical challenges in what concerns the analysis of the nonsmooth Lyapunov function (5) whenever the  $\ell_1$ -exact-penalty terms kick in. For example, given a large value of the penalty parameter  $\kappa$ , it is challenging to quadratically upper bound the function  $V$ . The performance characterization in the presence of inequality constraints remains an open question. •

## 4 Convex Optimization via Saddle-Point-Like Dynamics

As we noted in Section 3, the condition identified in Lemma 3.1 for the penalty parameter  $\kappa$  to be exact relies on knowledge of the dual solution set. In turn, exactness is required to ensure that the saddle-point dynamics (SPD) converges to a solution of (P). Motivated by these observations and building on our results above, in this section we propose discontinuous saddle-point-like dynamics that do not rely on a priori knowledge of the penalty parameter  $\kappa$ . We establish that these new dynamics are also correct, i.e., converge to a solution of the optimization problem and discuss conditions under which the dynamics can be implemented in a distributed fashion.

Let  $G = \{x \in \mathbb{R}^n \mid g(x) \leq 0_m\}$  denote the inequality constraint set associated with the convex program (P). Define the set-valued flow  $F : G \times \mathbb{R}^p \rightrightarrows \mathbb{R}^n$ ,

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

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

$$\begin{cases} (\dot{x}, \dot{\lambda}) \in (P_{T_G}(F), \partial_\lambda L)(x, \lambda), \\ (x(0), \lambda(0)) = (x_0, \lambda_0) \in G \times \mathbb{R}^p, \end{cases} \quad (\text{SPLD})$$

where the projection operator  $P_{T_G}$  is defined in (2). Since the mapping  $(x, \lambda) \mapsto F(x, \lambda)$  is locally bounded, upper semi-continuous and takes nonempty, convex and compact values, Lemma 2.4 guarantees the existence of local solutions  $(x, \lambda) \in \mathcal{AC}([0, t_+], G \times \mathbb{R}^p)$  of (SPLD).

### 4.1 Convergence Analysis

Our strategy to show that the saddle-point-like dynamics (SPLD) also converge to the set of saddle points is to establish that, in fact, its solutions are also solutions of the saddle-point dynamics (SPD) if the penalty parameter  $\kappa$  is sufficiently large. We establish this formally in the following result.

**Proposition 4.1** (Relationship of solutions). *Let  $(x, \lambda) : [0, +\infty) \rightarrow G \times \mathbb{R}^p$  be a 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).*

*Proof.* Since (SPD) and (SPLD) are identical on  $\text{int } G \times \mathbb{R}^p$ , it suffices to focus our attention on showing  $P_{T_G(x)}(F(x, \lambda)) \subset -\partial_x L(x, \lambda)$  for points  $(x, \lambda) \in \text{bd } G \times \mathbb{R}^p$ . Let the set of unit outward normals to  $G$  at  $x \in \text{bd } G$  be defined by  $N_G^\sharp(x) = N_G(x) \cap \text{bd } \overline{\mathbb{B}}(0, 1)$ . Take  $\xi \in F(x, \lambda)$  and suppose  $\{\xi\} \cap T_G(x) \neq \emptyset$ . In this case, it follows  $\sup_{n \in N_G^\sharp(x)} \langle \xi, n \rangle \leq 0$ , and by definition of  $P_{T_G}$ , we have  $P_{T_G}(\{\xi\}) = \xi$ . Clearly,

$$P_{T_G}(\{\xi\}) \in \xi - \kappa \sum_{k \in K^\circ(x)} \text{co}\{\{0\}, \partial g_k(x)\},$$

for any  $\kappa > 0$ , where  $K^\circ(x) = \{k \in \{1, \dots, m\} \mid g_k(x) = 0\}$ . Suppose now  $\{\xi\} \cap T_G(x) = \emptyset$ , i.e.,  $\sup_{n \in N_G^\sharp(x)} \langle \xi, n \rangle > 0$ . By Lemma A.3, there exists a unique  $n^*(x, \xi) \in N_G^\sharp(x)$  maximizing (12) such that  $P_{T_G}(\{\xi\}) = \xi - \max\{0, \langle \xi, n^*(x, \xi) \rangle\} n^*(x, \xi)$ . Now, by re-scaling  $n^*(x, \xi)$  by some constant  $\sigma^*(x, n^*) > 0$  minimizing

$$\min \left\{ \sigma \mid \frac{1}{\sigma} n^*(x, \xi) \in \sum_{k \in K^\circ(x)} \text{co}\{\{0\}, \partial g_k(x)\} \right\},$$

the choice  $\kappa \geq \sigma^*(x, n^*) \max\{0, \langle \xi, n^*(x, \xi) \rangle\}$  guarantees that

$$P_{T_G}(\{\xi\}) \in \xi - \kappa \sum_{k \in K^\circ(x)} \text{co}\{\{0\}, \partial g_k(x)\}.$$

Now, since  $\xi \in F(x, \lambda)$  is arbitrary, if  $\kappa \geq \max_{\xi \in F(x, \lambda)} \sigma^*(x, n^*) \max\{0, \langle \xi, n^*(x, \xi) \rangle\}$ , where  $F(x, \lambda)$  is nonempty, convex and compact, we conclude that  $P_{T_G(x)}(F(x, \lambda)) \subset -\partial_x L(x, \lambda)$ . More generally, let  $V$  be defined as in Remark 3.3 for some primal-dual solution  $(x^*, \lambda^*)$  of (P). Then, any choice

$$\kappa \geq \max_{(x, \lambda) \in \text{lev}_{\leq \gamma} V \cap (G \times \mathbb{R}^p)} \max_{\xi \in F(x, \lambda)} \sigma^*(x, n^*) \max\{0, \langle \xi, n^*(x, \xi) \rangle\},$$

guarantees that any solution of (SPLD) starting in  $\text{lev}_{\leq \gamma} V \cap (G \times \mathbb{R}^p)$  is also a solution of (SPD), concluding the proof.  $\square$

The arbitrariness of the choice of  $\gamma$  in Proposition 4.1 ensures that, given any solution of (SPLD), there exists  $\kappa$  such that the solution is also a solution of (SPD). Note that, in general, the set of solutions of (SPD) is richer than the set of solutions of (SPLD). Figure 1 illustrates the effect of an increasing penalty parameter  $\kappa$  on (SPD). The combination of Theorem 3.2 and Proposition 4.1 leads immediately to the following result.

**Corollary 4.2** (Asymptotic convergence). *Any solution  $(x, \lambda) \in \mathcal{AC}([0, +\infty), G \times \mathbb{R}^p)$  of (SPLD) starting from a point in  $G \times \mathbb{R}^p$  converges asymptotically to a point in the set  $S \times \mathbb{R}^p$ , where  $S \subset \mathbb{R}^n$  is the set of solutions of (P).*



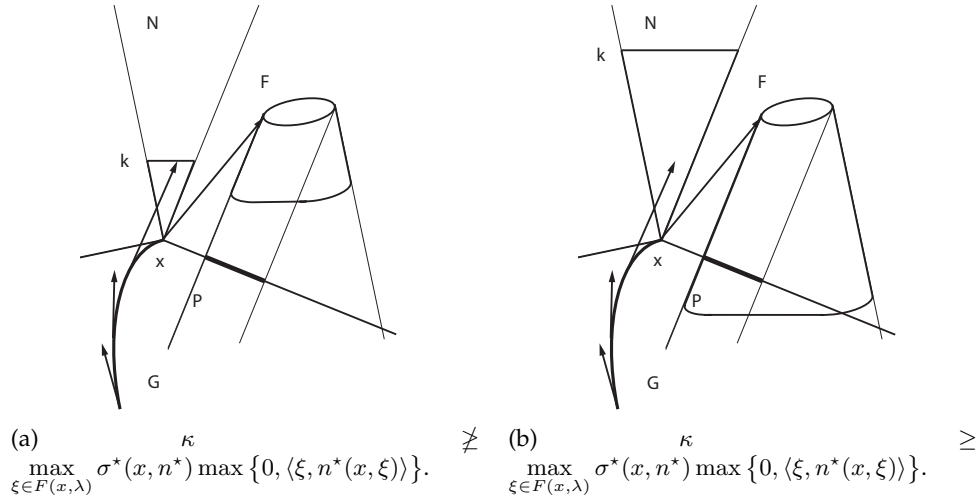


Figure 1: Illustration of the effect of an increasing penalty parameter  $\kappa > 0$  on (SPD). For a fixed  $(x, \lambda) \in \text{bd} G \times \mathbb{R}^p$ , the set  $-\partial_x L(x, \lambda) = F(x, \lambda) - \kappa \sum_{k=1}^m \text{co}\{\{0\}, \partial g_k(x)\}$  enlarges as the penalty parameter  $\kappa > 0$  increases (a), until the exact-penalty value is reached/exceeded (b) where the inclusion  $P_{T_G(x)}(F(x, \lambda)) \subset -\partial_x L(x, \lambda)$  holds.

## 4.2 Distributed Implementation

In this section, we discuss conditions under which the proposed dynamics can be implemented in a distributed way over a multi-agent system. First, we note that the implementation of (SPD) requires the use of an exact penalty parameter  $\kappa$ , whose determination in turn requires a dedicated distributed algorithm to ensure that agents collectively compute a value that satisfies the condition of Lemma 3.1. By contrast, the fact that the saddle-point-like dynamics (SPLD) do not incorporate any knowledge of the penalty parameter  $\kappa$  makes them more easily amenable to distributed implementation. We next describe in detail the requirements on the problem data that ensure that this is the case.

Consider a network of  $n \in \mathbb{N}$  agents whose communication topology is represented by an undirected and connected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = \{1, \dots, n\}$  is the vertex set and  $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$  is the (symmetric) edge set. The objective of the agents is to cooperatively solve the constraint 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  $i \in \{1, \dots, n\}$ , 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  $g_k(x) \leq 0$ ,  $k \in \{1, \dots, m\}$ , are compatible with  $\mathcal{G}$  if  $g_k$  can be expressed as a function of some components of the network state  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ , which induce a complete subgraph of  $\mathcal{G}$ . A similar definition can be stated for the equality constraints  $h_\ell(x) = 0$ ,  $\ell \in \{1, \dots, p\}$ . We assume that individual agents either have access to or enjoy enough computational capabilities to compute closed-form expressions for the (generalized) gradients of functions they are involved with.

In this network scenario, if  $(x, \lambda) \in \text{int } G \times \mathbb{R}^p$ , then each agent  $i \in \{1, \dots, n\}$  implements its primal dynamics (SPLD), where  $P_{T_G(x)}(F(x, \lambda)) = F(x, \lambda)$ , i.e.,

$$\dot{x}_i + \sum_{\{\ell: a_{\ell i} \neq 0\}} a_{\ell i} \left( \frac{1}{\mu} \left( \sum_{\{j: a_{\ell j} \neq 0\}} a_{\ell j} x_j - b_\ell \right) + \lambda_\ell \right) \in -\partial f_i(x_i),$$

and some dual dynamics (SPLD), i.e.,

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

where  $\ell \in \{1, \dots, p\}$ , corresponding to the Lagrange multipliers for the constraints that the agent is involved in (alternatively)

Hence, in order for agent  $i$  to be able to implement its corresponding primal dynamics, it also needs access to certain dual components  $\lambda_\ell$  for which  $a_{\ell i} \neq 0$ . If  $(x, \lambda) \in \text{bd } G \times \mathbb{R}^p$ , then each agent  $i \in \{1, \dots, n\}$  implements the locally projected dynamics (SPLD), i.e.,

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

and the dual dynamics (SPLD) described above. Hence, if the states of some agents are, at some time instance, involved in the active inequality constraints, the respective agents need to solve (12). We say that the saddle-point-like algorithm (SPLD) is distributed over  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  when the following conditions are satisfied:

- (C1) The network constraints  $h$  and  $g$  are compatible with the graph  $\mathcal{G}$ ;
- (C2) Agent  $i$  knows its state  $x_i \in \mathbb{R}$  and its objective function  $f_i$ ;
- (C3) Agent  $i$  has access to its neighbors' states  $x_j \in \mathbb{R}$ , their objective functions  $f_j$ , and
  - (i) the non-zero elements of every row of  $A \in \mathbb{R}^{p \times n}$ , and every  $b_\ell \in \mathbb{R}$  for which  $a_{\ell i} \neq 0$ , and
  - (ii) the active inequality constraints  $g_k$  in which agent  $i$  and its neighbors are involved.

(Note that, under these assumptions, it is also possible to have, for each Lagrange multiplier, only one agent implement the corresponding dual dynamics and then share the computed value with its neighboring agents – which are the ones involved in the corresponding equality constraint).

Note that the saddle-point-like algorithm (SPLD) can solve optimization scenarios where the agents' states belong to an arbitrary Euclidean space. In contrast to consensus-based distributed algorithms where each agent maintains, communicates and updates an estimate of the complete solution vector of the optimization problem, the saddle-point-like algorithm (SPLD) only requires each agent to store and communicate its own component of the solution vector. Thus, the algorithm scales well with respect to the number of agents in the network. The following examples illustrate an application of the above results to nonsmooth convex optimization scenarios over a network of agents.

**Example 4.3** (Saddle-point-like dynamics for nonsmooth convex optimization). Consider a network of  $n = 50$  agents that seek to cooperatively solve the nonsmooth convex optimization problem

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && \sum_{i \in \{1, \dots, n\}} x_i^4/4 + |x_i| \\ & \text{subject to} && \text{Circ}_n(0, 1, 1/2)x = \mathbb{1}_n/5, \\ & && x_i \leq 1/2, \quad i \in \{1, \dots, n\}, \end{aligned} \tag{10}$$

where  $x_i \in \mathbb{R}$  denotes the state associated with agent  $i \in \{1, \dots, n\}$ , and  $\text{Circ}_n(0, 1, 1/2)$  is the tridiagonal circulant matrix [10] encoding the network topology in its sparsity structure. Although this specific example is academic, examples belonging to the same class of optimization problems arise in a variety of networked scenarios, see e.g. [53]. The generalized gradient of  $f_i$  at  $x_i \in \mathbb{R}$  is

$$\partial f_i(x_i) = \begin{cases} \{x_i^3 + 1\}, & \text{if } x_i > 0, \\ [-1, 1], & \text{if } x_i = 0, \\ \{x_i^3 - 1\}, & \text{if } x_i < 0. \end{cases}$$

Figure 2 illustrates the asymptotic convergence of solutions of the saddle-point-like dynamics (SPLD) to the set  $\text{sp}(L) = \{x^*\} \times M$ , where

$$M = \{\lambda^* \in \mathbb{R}^p \mid \lambda^* \in -(\text{Circ}_n(0, 1, 1/2) \text{Circ}_n(0, 1, 1/2)^\top)^{-1} \text{Circ}_n(0, 1, 1/2) \partial f(x^*)\}.$$

Following up on our observations in Remark 3.7, it is interesting to observe that, even though the inequality constraints become active during the evolution, cf. Figure 2(a), the convergence is exponential, cf. Figure 2(c). •

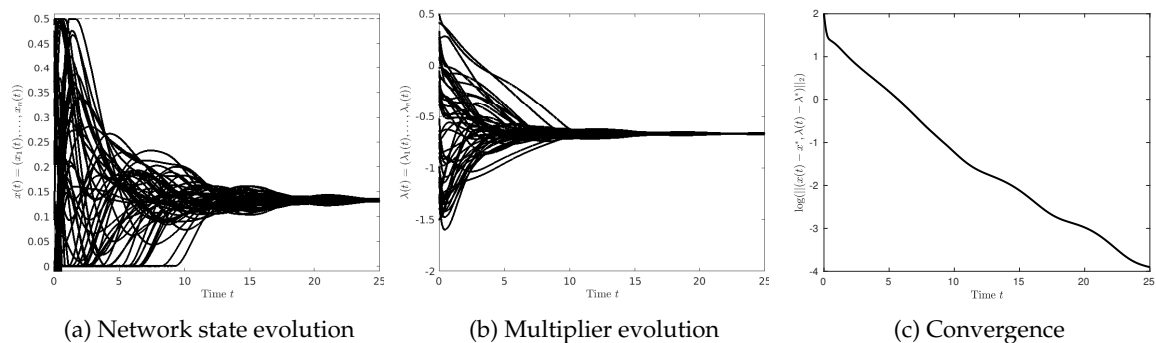


Figure 2: (a) The network state evolution of algorithm (SPLD) solving the nonsmooth convex program (10). The projection operator prevents the solutions of (SPLD) from violating the inequality constraints (depicted with a dashed line) of (10). (b) The Lagrange multiplier evolution associated with the equality constraints of (10). The initial conditions are randomly chosen from the interval  $[-3/2, 1/2]$ . (c) Since the aggregate objective function of (10) is strictly convex, the solutions of (SPLD) converge asymptotically to the set  $\text{sp}(L)$ .

**Example 4.4** (Saddle-point dynamics for equality constrained optimization). Consider a network of  $n = 50$  agents whose objective is to cooperatively solve the convex optimization problem

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && \sum_{i \in \{1, \dots, n\}} f_i(x_i) \\ & \text{subject to} && Ax = \mathbb{1}_n, \end{aligned} \tag{11}$$

where  $f_i \in \mathcal{C}^{1,1}(\mathbb{R}, \mathbb{R})$  defined by

$$f_i(x_i) = \begin{cases} x_i^2, & \text{if } x_i \geq 0, \\ x_i^2/2, & \text{if } x_i < 0, \end{cases}$$

is the objective function associated with agent  $i \in \{1, \dots, n\}$ . We consider two cases for the matrix  $A$  (whose sparsity structure encodes the network topology): in the first case,  $\text{Trid}_n(1/2, 1, -1/10)$  is the tridiagonal Toeplitz matrix [10] of dimension  $n \times n$ ; in the second case, we modify the tridiagonal Toeplitz matrix, setting to  $1/10$  its entries at positions  $(11, 19)$ ,  $(21, 28)$ ,  $(31, 39)$ ,  $(2, 40)$  and  $(5, 49)$ . This addition has the effect of significantly increasing the connectivity of the network topology (effectively reducing the graph diameter from 49 in the first case to 23 in the second). The gradient of  $f_i$  at  $x_i \in \mathbb{R}$  is  $\nabla f_i(x_i) = \max\{x_i, 2x_i\}$ , and the generalized Hessian of  $f_i$  at  $x_i \in \mathbb{R}$  is

$$\partial(\nabla f_i)(x_i) = \begin{cases} \{2\}, & \text{if } x_i > 0, \\ [1, 2], & \text{if } x_i = 0, \\ \{1\}, & \text{if } x_i < 0. \end{cases}$$

Figures 3 and 4 illustrate the convergence and performance of solutions of the saddle-point dynamics (SPD) to the singleton set  $\text{sp}(L)$  for the first and second selection of constraint matrix, respectively. •

## 5 Conclusions

We have investigated the design of continuous-time solvers for a class of nonsmooth convex optimization problems. Our starting point was an equivalent reformulation of this problem in terms of finding the saddle points of an augmented Lagrangian function. This reformulation has naturally led us to study the associated saddle-point dynamics, for which we established convergence to the set of solutions of the nonsmooth convex program. The novelty of our analysis relies on the identification of a global Lyapunov function for the saddle-point dynamics. Based on these results, we have introduced a discontinuous saddle-point-like algorithm that enjoys the same convergence properties and is fully amenable to distributed implementation over a group of agents that seeks to collectively solve the optimization problem. With respect to consensus-based approaches, the novelty of our design is that it allows each individual agent to asymptotically find its component of the solution by interacting with its neighbors, without the need to maintain, communicate, or update a global estimate of the complete solution vector. We also established the

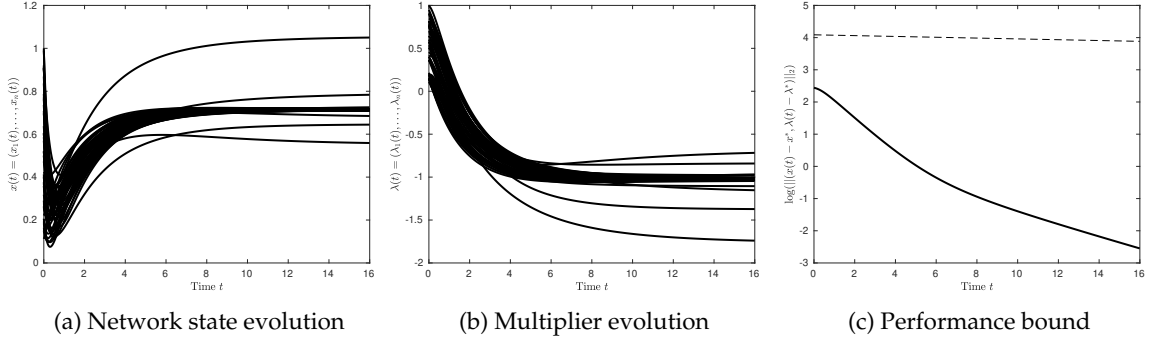


Figure 3: (a) The network state evolution of algorithm (SPD) solving the convex program (11) with  $A = \text{Trid}_n(1/2, 1, -1/10)$ . (b) The Lagrange multiplier evolution associated with the equality constraints of (11). The initial conditions are randomly chosen in the interval  $(0, 1]$ . (c) Since  $\partial(\nabla f) \succ 0$  (in fact, the aggregate objective function of (11) is strongly convex), the hypothesis of Theorem 3.5 is satisfied and thus, the solutions of (SPD) converge to the singleton set  $\text{sp}(L)$  within the exponential performance bound (depicted by the dashed line).

performance properties of the proposed coordination algorithms for convex optimization scenarios subject to equality constraints. In particular, we explicitly characterized the exponential convergence rate under mild convexity and regularity conditions on the objective functions. Future work will characterize the rate of convergence for nonsmooth convex optimization problems subject to both inequality and equality constraints, study the robustness properties of the proposed algorithms against disturbances and link failures, identify suitable (possibly aperiodic) stepsizes that guarantee convergence of the discretization of our dynamics, design opportunistic state-triggered implementations to efficiently use the capabilities of the network agents, and explore the extension of our analysis and algorithm design to optimization problems defined over infinite-dimensional state spaces.

## Appendix

We gather in this appendix various intermediate results used in the derivation of the main results of the paper.

### Auxiliary Results for Performance Characterization

The following two results characterize properties of the generalized Hessian of  $\mathcal{C}^{1,1}$  functions. In each case, let  $f \in \mathcal{C}^{1,1}(\mathbb{R}^n, \mathbb{R})$  and consider the set-valued map  $\partial(\nabla f) : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n}$  as defined in Section 2.1. For  $x, y \in \mathbb{R}^n$ , we let  $[x, y] = \{x + \theta(y - x) \mid \theta \in [0, 1]\}$  and study the set

$$\partial(\nabla f([x, y])) = \bigcup_{z \in [x, y]} \partial(\nabla f)(z).$$

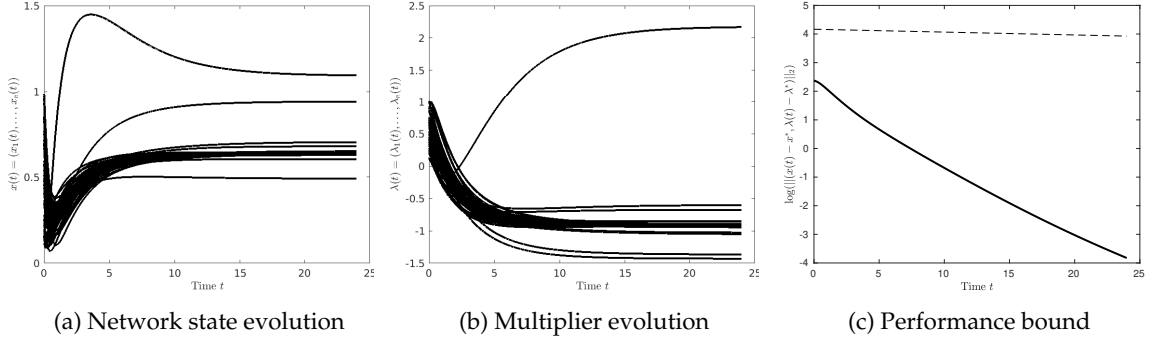


Figure 4: (a) The network state evolution of algorithm (SPD) solving the convex program (11), where  $A$  is obtained by modifying various entries of  $\text{Trid}_n(1/2, 1, -1/10)$ . (b) The Lagrange multiplier evolution associated with the equality constraints of (11). (c) The solutions of (SPD) converge to the singleton set  $\text{sp}(L)$  within the exponential performance bound (depicted by the dashed line).

**Lemma A.1** (Positive definiteness). *Let  $f \in \mathcal{C}^{1,1}(\mathbb{R}^n, \mathbb{R})$  and suppose  $\partial(\nabla f) \succ 0$ . Then,  $\text{co} \{ \partial(\nabla f([x, y])) \} \succ 0$  for all  $x, y \in \mathbb{R}^n$ .*

*Proof.* Since  $\partial(\nabla f)(x) \succ 0$  for all  $x \in \mathbb{R}^n$ , it follows  $\bigcup_{z \in [x, y]} \partial(\nabla f)(z) \succ 0$ . Moreover, since the cone of symmetric positive definite matrices is convex itself, it contains all convex combinations of elements in  $\bigcup_{z \in [x, y]} \partial(\nabla f)(z)$ , i.e., in particular  $\text{co} \{ \partial(\nabla f([x, y])) \} \succ 0$ , concluding the proof.  $\square$

**Lemma A.2** (Compactness). *Let  $f \in \mathcal{C}^{1,1}(\mathbb{R}^n, \mathbb{R})$ . Then, the set  $\partial(\nabla f([x, y]))$  and its convex closure are both compact.*

*Proof.* Boundedness of the set  $\partial(\nabla f([x, y]))$  follows from [50, Proposition 5.15]. To show that  $\partial(\nabla f([x, y]))$  is closed, take  $\nu \in \text{cl} \partial(\nabla f([x, y]))$ . By definition, there exists  $\{\nu_n\} \subset \partial(\nabla f([x, y]))$  such that  $\nu_n \rightarrow \nu$ . Since  $\nu_n$  belongs to  $\partial(\nabla f([x, y]))$ , let us denote  $\nu_n \in \partial(\nabla f)(z_n)$ , i.e.,  $\nu_n$  is a vector based at  $z_n$ . Similarly, let  $z$  be the point at which the vector  $\nu$  is based. Following the above arguments, we have  $z_n \rightarrow z$ . Since  $[x, y]$  is compact and  $z_n \in [x, y]$ , we deduce  $z \in [x, y]$ . Assume, by contradiction, that  $\nu \notin \partial(\nabla f)(z)$ . Then, since  $\partial(\nabla f)(z)$  is closed, there exists  $\varepsilon > 0$  such that  $\{\nu\} \cap \partial(\nabla f)(z) + \mathbb{B}(0, \varepsilon) = \emptyset$ . Using upper semi-continuity, there exists  $N \in \mathbb{N}$  such that if  $n \geq N$ , then  $\partial(\nabla f)(z_n) \subset \partial(\nabla f)(z) + \mathbb{B}(0, \varepsilon)$ . This fact is in contradiction with  $\nu_n \rightarrow \nu$ . Therefore, it follows  $\nu \in \partial(\nabla f)(z) \subset \partial(\nabla f([x, y]))$ , and we conclude  $\partial(\nabla f([x, y]))$  is closed. Thus,  $\partial(\nabla f([x, y]))$  is compact, and so is  $\text{co} \{ \partial(\nabla f([x, y])) \}$ , concluding the proof.  $\square$

## Auxiliary Results for Convergence Analysis of Saddle-Point-Like Dynamics

Here, we investigate the explicit computation of the projection operator  $P_{T_G}$ , which plays a key role in the dynamics (SPLD). Recall that  $T_G(x)$  and  $N_G(x)$  denote the tangent and

normal cone of  $G \subset \mathbb{R}^n$  at  $x \in G$ , respectively. The following geometric interpretation of  $P_{T_G}$  is well-known in the literature of locally projected dynamical systems [38, 45]:

- (i) if  $(x, \lambda) \in \text{int } G \times \mathbb{R}^p$ , then  $P_{T_G(x)}(F(x, \lambda)) = F(x, \lambda)$ ;
- (ii) if  $(x, \lambda) \in \text{bd } G \times \mathbb{R}^p$ , then

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

where

$$n^*(x, \xi) \in \arg \max_{n \in N_G^\sharp(x)} \langle \xi, n \rangle. \quad (12)$$

Note that if  $\{\xi\} \cap T_G(x) \neq \emptyset$  for some  $(x, \lambda) \in \text{bd } G \times \mathbb{R}^p$  and  $\xi \in F(x, \lambda)$ , then  $\sup_{n \in N_G^\sharp(x)} \langle \xi, n \rangle \leq 0$ , and by definition of  $P_{T_G}$ , no projection needs to be performed. The following result establishes the existence and uniqueness of the maximizer  $n^*(x, \xi)$  of (12) whenever  $\{\xi\} \cap T_G(x) = \emptyset$ .

**Lemma A.3** (Computation of the projection operator). *Let  $(x, \lambda) \in \text{bd } G \times \mathbb{R}^p$ . 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 (12) exists and is unique.*

*Proof.* Let  $(x, \lambda) \in \text{bd } G \times \mathbb{R}^p$  and suppose there exists  $\xi \in F(x, \lambda)$  such that  $\sup_{n \in N_G^\sharp(x)} \langle \xi, n \rangle > 0$ . By definition, the normal cone  $N_G(x)$  of  $G$  at  $x \in \text{bd } G$  is closed and convex. Existence of  $n^*(x, \xi)$  follows from compactness of the set  $N_G^\sharp(x)$ . Now, let  $\tilde{n}^*(x, \xi)$  and  $\hat{n}^*(x, \xi)$  be two distinct maximizer of (12) such that  $\langle \xi, \tilde{n}^*(x, \xi) \rangle > 0$  and  $\langle \xi, \hat{n}^*(x, \xi) \rangle > 0$ . Convexity implies  $(\tilde{n}^*(x, \xi) + \hat{n}^*(x, \xi)) / \|\tilde{n}^*(x, \xi) + \hat{n}^*(x, \xi)\| \in N_G^\sharp(x)$ . Therefore, it follows

$$\frac{\langle \xi, \tilde{n}^*(x, \xi) + \hat{n}^*(x, \xi) \rangle}{\|\tilde{n}^*(x, \xi) + \hat{n}^*(x, \xi)\|} = \frac{2\langle \xi, \tilde{n}^*(x, \xi) \rangle}{\|\tilde{n}^*(x, \xi) + \hat{n}^*(x, \xi)\|} > \langle \xi, \tilde{n}^*(x, \xi) \rangle,$$

which contradicts the fact that  $\tilde{n}^*(x, \xi)$  maximizes (12).  $\square$

We note that the computational complexity of solving (12) depends not only on the problem dimensions  $n, p, m > 0$ , but also on the convexity and regularity assumptions of the problem data, i.e., on  $f, h$  and  $g$ .

## References

- [1] K. Arrow, L Hurwitz, and H. Uzawa. *Studies in Linear and Non-Linear Programming*. Stanford University Press, Stanford, California, 1958.
- [2] J. P. Aubin and A. Cellina. *Differential Inclusions*, volume 264 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York, 1984.

- [3] 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.
- [4] A. Bacciotti and L. Rosier. *Liapunov Functions and Stability in Control Theory*. Communications and Control Engineering. Springer, New York, 2nd edition, 2005.
- [5] D. P. Bertsekas. Necessary and sufficient conditions for a penalty method to be exact. *Mathematical Programming*, 9(1):87–99, 1975.
- [6] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 2nd edition, 1999.
- [7] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [8] S. P. Bhat and D. S. Bernstein. Nontangency-based Lyapunov tests for convergence and stability in systems having a continuum of equilibria. *SIAM Journal on Control and Optimization*, 42(5):1745–1775, 2003.
- [9] B. Brogliato, A. Daniilidis, C. Lemaréchal, and V. Acary. On the equivalence between complementarity systems, projected systems, and differential inclusions. *Systems & Control Letters*, 55(1):45–51, 2006.
- [10] F. Bullo, J. Cortés, and S. Martínez. *Distributed Control of Robotic Networks*. Applied Mathematics Series. Princeton University Press, 2009. Electronically available at <http://coordinationbook.info>.
- [11] R. Carli and G. Notarstefano. Distributed partition-based optimization via dual decomposition. In *IEEE Conf. on Decision and Control*, pages 2979–2984, Firenze, Italy, 2013.
- [12] J. Chen and V. K. N. Lau. Convergence analysis of saddle point problems in time varying wireless systems - control theoretical approach. *IEEE Transactions on Signal Processing*, 60(1):443–452, 2012.
- [13] A. Cherukuri, B. Gharesifard, and J. Cortés. Saddle-point dynamics: conditions for asymptotic stability of saddle points. *SIAM Journal on Control and Optimization*, 55(1):486–511, 2017.
- [14] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255–312, 2007.
- [15] F. H. Clarke. *Optimization and Nonsmooth Analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. Wiley, 1983.
- [16] J. Cortés. Discontinuous dynamical systems - a tutorial on solutions, nonsmooth analysis, and stability. *IEEE Control Systems*, 28(3):36–73, 2008.



- [17] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović. The proximal augmented Lagrangian method for nonsmooth composite optimization. *IEEE Transactions on Automatic Control*, page To appear, 2018.
- [18] J. C. Duchi, A. Agarwal, and M. J. Wainwright. Dual averaging for distributed optimization: convergence analysis and network scaling. *IEEE Transactions on Automatic Control*, 57(3):592–606, 2012.
- [19] M. Ercsey-Ravasz and Z. Toroczkai. The chaos within Sudoku. *Scientific Reports*, 2:725, 2012.
- [20] D. Feijer and F. Paganini. Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica*, 46:1974–1981, 2010.
- [21] 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*, 51(9):1741–1754, 2004.
- [22] B. Gharesifard and J. Cortés. Distributed convergence to Nash equilibria in two-network zero-sum games. *Automatica*, 49(6):1683–1692, 2013.
- [23] B. Gharesifard and J. Cortés. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3):781–786, 2014.
- [24] R. Goebel. Stability and robustness for saddle-point dynamics through monotone mappings. *Systems & Control Letters*, 108:16–22, 2017.
- [25] C. Henry. An existence theorem for a class of differential equations with multivalued right-hand side. *Journal of Mathematical Analysis and Applications*, 41:179–186, 1973.
- [26] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Grundlehren Text Editions. Springer, New York, 1993.
- [27] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V. H. Nguyen. Generalized Hessian matrix and second-order optimality conditions for problems with  $C^{1,1}$  data. *Applied Mathematics and Optimization*, 11:43–56, 1984.
- [28] T. Holding and I. Lestas. On the convergence of saddle points of concave-convex functions, the gradient method and emergence of oscillations. In *IEEE Conf. on Decision and Control*, pages 1143–1148, Los Angeles, CA, 2014.
- [29] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [30] B. Johansson, M. Rabi, and M. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Control and Optimization*, 20(3):1157–1170, 2009.
- [31] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control in communication networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, 1998.

- [32] T. Kose. Solutions of saddle value problems by differential equations. *Econometrica*, 24(1):59–70, 1956.
- [33] N. Li, C. Zhao, and L. Chen. Connecting automatic generation control and economic dispatch from an optimization view. *IEEE Transactions on Control of Network Systems*, 3(3):254–264, 2016.
- [34] J. Lu and C. Y. Tang. Zero-gradient-sum algorithms for distributed convex optimization: the continuous-time case. *IEEE Transactions on Automatic Control*, 57(9):2348–2354, 2012.
- [35] E. Mallada, C. Zhao, and S. H. Low. Optimal load-side control for frequency regulation in smart grids. *IEEE Transactions on Automatic Control*, 62(12):6294–6309, 2017.
- [36] O. L. Mangasarian. Sufficiency of exact penalty minimization. *SIAM Journal on Control and Optimization*, 23(1):30–37, 1985.
- [37] D. Mateos-Núñez and J. Cortés. Noise-to-state exponentially stable distributed convex optimization on weight-balanced digraphs. *SIAM Journal on Control and Optimization*, 54(1):266–290, 2016.
- [38] A. Nagurney and D. Zhang. *Projected Dynamical Systems and Variational Inequalities with Applications*, volume 2 of *International Series in Operations Research and Management Science*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.
- [39] I. Necoara and V. Nedelcu. Rate analysis of inexact dual first order methods. *IEEE Transactions on Automatic Control*, 59(5):1232–1243, 2014.
- [40] I. Necoara and J. Suykens. Application of a smoothing technique to decomposition in convex optimization. *IEEE Transactions on Automatic Control*, 53(11):2674–2679, 2008.
- [41] A. Nedic and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.
- [42] A. Nedic, A. Ozdaglar, and P. A. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, 2010.
- [43] S. K. Niederländer, F. Allgöwer, and J. Cortés. Exponentially fast distributed coordination for nonsmooth convex optimization. In *IEEE Conf. on Decision and Control*, pages 1036–1041, Las Vegas, NV, 2016.
- [44] S. K. Niederländer and J. Cortés. Distributed coordination for separable convex optimization with coupling constraints. In *IEEE Conf. on Decision and Control*, pages 694–699, Osaka, Japan, 2015.
- [45] M. Pappalardo and M. Passacantando. Stability for equilibrium problems: From variational inequalities to dynamical systems. *Journal of Optimization Theory & Applications*, 113(3):567–582, 2002.

- [46] B. T. Polyak. Iterative methods using Lagrange multipliers for solving extremal problems with constraints of the equation type. *USSR Computational Mathematics and Mathematical Physics*, 10(5):1098–1106, 1970.
- [47] L. J. Ratliff, S. A. Burden, and S. S. Sastry. On the characterization of local Nash equilibria in continuous games. *IEEE Transactions on Automatic Control*, 61(8):2301–2307, 2016.
- [48] D. Richert and J. Cortés. Robust distributed linear programming. *IEEE Transactions on Automatic Control*, 60(10):2567–2582, 2015.
- [49] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [50] R. T. Rockafellar and R. J. B. Wets. *Variational Analysis*, volume 317 of *Comprehensive Studies in Mathematics*. Springer, New York, 1998.
- [51] W. Shi, Q. Ling, G. Wu, and W. Yin. EXTRA: an exact first-order algorithm for decentralized consensus optimization. *SIAM Journal on Optimization*, 25(2):944–966, 2015.
- [52] T. Stegink, C. D. Persis, and A. J. van der Schaft. A unifying energy-based approach to stability of power grids with market dynamics. *IEEE Transactions on Automatic Control*, 62(6):2612–2622, 2017.
- [53] P. Wan and M. D. Lemmon. Event-triggered distributed optimization in sensor networks. In *Symposium on Information Processing of Sensor Networks*, pages 49–60, San Francisco, CA, 2009.
- [54] J. Wang and N. Elia. A control perspective for centralized and distributed convex optimization. In *IEEE Conf. on Decision and Control*, pages 3800–3805, Orlando, Florida, 2011.
- [55] X. Zhang and A. Papachristodoulou. A real-time control framework for smart power networks: Design methodology and stability. *Automatica*, 58:43–50, 2015.
- [56] C. Zhao, U. Topcu, N. Li, and S. H. Low. Design and stability of load-side primary frequency control in power systems. *IEEE Transactions on Automatic Control*, 59(5):1177–1189, 2014.
- [57] M. Zhu and S. Martínez. On distributed convex optimization under inequality and equality constraints. *IEEE Transactions on Automatic Control*, 57(1):151–164, 2012.