Distributed and Anytime Algorithm for Network Optimization Problems with Separable Structure

Pol Mestres Jorge Cortés

Abstract— **This paper considers the problem of designing a dynamical system to solve constrained optimization problems in a distributed way and in an anytime fashion (i.e., such that the feasible set is forward invariant). For problems with separable objective function and constraints, we design an algorithm with the desired properties and establish its convergence. Simulations illustrate our results.**

I. Introduction

Distributed optimization methods are a popular tool for solving several engineering problems like parameter estimation, resource allocation in communication networks, source localization, etc. In several of these applications, feedback controllers are often implemented as the solution of such problems on a physical plant. In this context, the safe operation requirements of the system are encoded as constraints of the optimization problem. This approach is very versatile, but the implementation of such controllers faces several challenges. On the one hand, the algorithm solving the optimization problem may be terminated at any time, and hence feasibility must be maintained throughout its execution. We refer to an algorithm with this property as anytime. Moreover, it must retain its distributed and scalable character, so that each agent can implement it by communicating exclusively with its neighbors and can do so in an efficient manner independently of the size of the network. Designing algorithms that combine all these objectives together and are anytime, distributed, scalable and have provable convergence guarantees is a challenging problem.

Literature Review: In this work we take the viewpoint of optimization algorithms as continuous-time dynamical systems (cf. [1], [2]), which has recently proven to be a very powerful paradigm in applications where the optimization problem is in a feedback loop with a plant [3]. The problem of designing distributed algorithms for constrained optimization is well studied in the literature. The survey papers [4], [5] cover this topic exhaustively. Of particular interest to us are primal-dual and projected saddle-point dynamics [6], which define dynamical systems that solve constrained optimization problems. Although these works have provable convergence guarantees, the trajectories generated by the proposed dynamical systems are in general not guaranteed to be feasible throughout its execution. This anytime property is well-studied in the optimization literature and dates back to the mid-1980s (cf. [7], [8]) in the context of time-dependent planning. Our work here is related to [9]–[11], which design dynamical systems that solve nonlinear programs in continuous time with the anytime property. However, the literature on anytime algorithms generally is not concerned in making them amenable to a distributed implementation. An exception is the relaxed economic dispatch problem, which involves a convex separable objective function and globally coupling affine equality constraints. For this problem, [12] gives a distributed anytime algorithm that converges to the global optimum. The problem of designing distributed anytime algorithms for constrained optimization problems is very relevant in the control barrier function (CBF) literature [13], [14]. This is because CBF-based controllers for multiagent systems can be obtained as the solution of a network optimization problem, where the system is guaranteed to be safe only if its constraints are satisfied at all times. The works [15]–[17] tackle this problem for CBF-based quadratic programs (QPs), where a centralized QP is split into local QPs whose solution is guaranteed to preserve the safety constraints at all times. However, the solution of these local QPs might be suboptimal with respect to the centralized QP. The recent work [18] designs a distributed algorithm that is guaranteed to satisfy the constraints of the CBF-based QP at all times and converge to its state-dependent optimizer in finite time. However, their algorithm is restricted to a limited class of quadratic programs and plant dynamics and is not easily generalizable to general convex programs.

Statement of Contributions: In this paper we introduce a continuous-time dynamical system to solve convex optimization problems with separable objective function and constraints in a distributed and anytime fashion. The constraints couple the decision variables of all agents and this poses a difficulty in the design of distributed algorithms that solve such problems. We first show that the separable structure permits the intoduction of auxiliary variables to reformulate the original problem into one with local constraints while still preserving the same solution set. However, this reformulation still does not allow to fully decouple the optimization problem into one per agent because the auxiliary variables require coordination. In order to sort this hurdle, our technical approach constructs a dynamical system by combining the use of projected saddle-point dynamics, which are distributed but not anytime, and the safe gradient flow, which is anytime but not distributed. First, we establish the well-posedness of the proposed dynamical system. Second, we show that it is distributed, exhibits the anytime property and is scalable.

This work was partially supported by ARL award ARL-W911NF-22-2-0231.

P. Mestres and J. Cortés are with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, {pomestre,cortes}@ucsd.edu

Finally, we prove that all trajectories with feasible initial condition converge to a neighborhood of the optimizer, which can be made arbitrarily small by tuning a design parameter accordingly. Moreover, in the case where the feasible set is bounded, we show that all trajectories with feasible initial condition exactly converge to the optimizer.

II. Preliminaries

This section presents background on dynamical systems that solve constrained optimization problems[1].

Safe Gradient Flow: Given continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$, $h : \mathbb{R}^n \to \mathbb{R}^q$, consider the constrained nonlinear program

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \; & f(x), \\ \text{s.t.} \quad & g(x) \le 0, \\ & h(x) = 0. \end{aligned} \tag{1}$$

Let $\mathcal{F} = \{x \in \mathbb{R}^n \; : \; g(x) \le 0, \; h(x) = 0\}$ denote the constraint set and $X_{\text{KKT}} = \{x^* \in \mathbb{R}^n \; : \; \exists (u^*, v^*) \in \mathbb{R}^p \times \mathbb{R}^q$ such that (x^*, u^*, v^*) is a KKT point of (1)$\}$. We are interested in solving the optimization problem with an algorithm that respects its constraints at all times of its evolution. Given $\alpha > 0$, the *safe gradient flow*, cf. [11], is the dynamical system $\dot{x} = F_\alpha(x)$, where

$$\begin{aligned} F_\alpha(x) = \arg\min_{\xi\in\mathbb{R}^n} \; & \frac{1}{2} \left\| \xi + \nabla f(x) \right\|^2 \\ \text{s.t.} \quad & \frac{\partial g(x)}{\partial x} \xi \le -\alpha g(x), \qquad (2) \\ & \frac{\partial h(x)}{\partial x} \xi = -\alpha h(x). \end{aligned}$$

The direction prescribed by F_α can be interpreted as that closest to the gradient descent direction $-\nabla f$ while ensuring that the constraints defining $\mathcal{F}$ are not violated. The next result gathers important properties of the safe gradient flow.

Proposition 2.1: (Properties of the safe gradient flow [11, Proposition 5.1, Proposition 5.6 and Corollary 5.9]): Suppose f, g and h are continuously differentiable and their derivatives are locally Lipschitz. Then,

(i) there exists an open neighborhood U containing $\mathcal{F}$ such that (2) is well-defined;
(ii) F_α is locally Lipschitz on U;
(iii) the Lagrange multipliers of (2) are unique and locally Lipschitz as a function of x on U;
(iv) the feasible set $\mathcal{F}$ is forward invariant and asymptotically stable under (2);
(v) $F_\alpha(x^*) = 0$ if and only if $x^* \in X_{\text{KKT}}$;
(vi) if x^* is a strict local minimizer of f and an isolated equilibrium of the safe gradient flow, then x^* is asymptotically stable relative to $\mathcal{F}$.

Projected Saddle-Point Dynamics: We recall here the notion of projected saddle-point dynamics following [19]. Consider again the optimization problem (1), with continuously differentiable functions f, g and h whose derivatives are locally Lipschitz, and let $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_{\ge 0} \times \mathbb{R}^p$ be the associated Lagrangian,

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x).$$

We define the *projected saddle-point dynamics* for $\mathcal{L}$ as:

$$\dot{x} = -\nabla_x \mathcal{L}(x, \lambda, \mu), \tag{3a}$$
$$\dot{\lambda} = [\nabla_\lambda \mathcal{L}(x, \lambda, \mu)]^+_\lambda, \tag{3b}$$
$$\dot{\mu} = \nabla_\mu \mathcal{L}(x, \lambda, \mu). \tag{3c}$$

If f is strongly convex, g is convex and h is affine, the saddle point of $\mathcal{L}$ is unique and corresponds to the KKT point of (1). Moreover, [19, Theorem 5.1] ensures that the saddle point of $\mathcal{L}$ is globally asymptotically stable under the dynamics (3).

III. Problem Statement

Consider a network composed by agents $\{1, \dots, N\}$ whose communication topology is described by a connected undirected graph $\mathcal{G}$. An edge (i, j) represents the fact that agent i can receive information from agent j and vice versa. We refer to an algorithm run by the network as *distributed* if each agent can execute it with the information available to it and its neighbors.

For each $i \in \{1, \dots, N\}$, $k \in \{1, \dots, p\}$, $l \in \{1, \dots, q\}$ let $f_i : \mathbb{R}^n \to \mathbb{R}$ be a strongly convex and continuously differentiable function with locally Lipschitz derivatives, $g_i^k : \mathbb{R}^n \to \mathbb{R}$ be a convex and continuously differentiable function with locally Lipschitz derivatives and $h_i^l : \mathbb{R}^n \to \mathbb{R}$ be an affine function. We let $x = [x_1, \dots, x_N] \in \mathbb{R}^{nN}$. Consider

[1]Throughout the paper we denote by $\mathbb{R}$ and $\mathbb{R}_{>0}$ the set of real and nonnegative real numbers, respectively. Given $x \in \mathbb{R}^n$, $\|x\|$ denotes its Euclidean norm. For $a \in \mathbb{R}$ and $b \in \mathbb{R}_{\ge 0}$, we let

$$[a]^+_b = \begin{cases} a, & \text{if } b > 0, \\ \max\{0, a\}, & \text{if } b = 0. \end{cases}$$

For vectors $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n_{\ge 0}$, $[a]^+_b$ denotes the vector whose i-th component is $[a_i]^+_{b_i}$, for $i \in \{1, \dots, n\}$. We also write $\mathbf{0}_n = (0, \dots, 0) \in \mathbb{R}^n$. For a real-valued function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we denote by $\nabla_x F$ and $\nabla_y F$ the column vector of partial derivatives of F with respect to the first and second arguments, respectively. Given a set of functions $g^1, \dots, g^k$, we let $I_{g^1,\dots,g^k}(x) = \{1 \le i \le k \; : \; g^i(x) = 0\}$. For matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{p\times q}$, we let $A \otimes B$ denote their Kronecker product. Given a set $\mathcal{P} \subseteq \mathbb{R}^n$ and a set of variables $\xi = \{x_{i_1}, x_{i_2}, \dots, x_{i_k}\}$, we denote by $\Pi_\xi \mathcal{P} = \{(x_{i_1}, x_{i_2}, \dots, x_{i_k}) \in \mathbb{R}^k \; : \; x \in \mathcal{P}\}$ the projection of $\mathcal{P}$ onto the ξ variables. Given sets $S_1, \dots, S_k$, $\times_{i=1}^k S_i$ denotes their Cartesian product. An undirected graph is a pair $\mathcal{G} = (V, \mathcal{E})$, where $V = \{1, \dots, N\}$ is a finite set called the vertex set, $\mathcal{E} \subseteq V \times V$ is called the edge set where $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$. The set of neighbors of node i is denoted by $\mathcal{N}_i = \{j \in \mathcal{I} \; : \; (i, j) \in \mathcal{E}\}$. The adjacency matrix $A \in \mathbb{R}^{|V|\times|V|}_{\ge 0}$ of the graph $\mathcal{G}$ satisfies the property $[A]_{i,j} = [A]_{j,i} = 1$ if $(i, j) \in \mathcal{E}$ and $[A]_{i,j} = 0$ otherwise. The degree matrix D of $\mathcal{G}$ is the diagonal matrix defined by $[D]_{i,i} = |\mathcal{N}_i|$ for all $i \in \{1, \dots, |V|\}$. The Laplacian matrix of $\mathcal{G}$ is $L = D - A$. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitz vector field and consider the dynamical system $\dot{x} = F(x)$, with flow map $\Phi_t : \mathbb{R}^n \to \mathbb{R}^n$. This means that $\Phi_t(x) = x(t)$, where $x(t)$ is the unique solution of the dynamical system with $x(0) = x$. A set $\mathcal{K} \subseteq \mathbb{R}^n$ is (positively) forward invariant if $x \in \mathcal{K}$ implies that $\Phi_t(x) \in \mathcal{K}$ for all $t \ge 0$. A set $A \subseteq \mathcal{K}$ is Lyapunov stable relative to $\mathcal{K}$ if, for every open set U containing A, there exists an open set $\tilde{U}$ also containing A such that for all $x \in \tilde{U} \cap \mathcal{K}$, $\Phi_t(x) \in U \cap \mathcal{K}$ for all $t \ge 0$. A set $A \subseteq \mathcal{K}$ is asymptotically stable relative to $\mathcal{K}$ if it is Lyapunov stable relative to $\mathcal{K}$ and there is an open set U containing A such that $\Phi_t(x) \to A$ as $t \to \infty$ for all $x \in U \cap \mathcal{K}$.

the following optimization problem with separable objective function and constraints:

$$\min_{x \in \mathbb{R}^{nN}} \sum_{i=1}^{N} f_i(x_i), \tag{4}$$
$$\text{s.t.} \quad \sum_{i=1}^{N} g_i^k(x_i) \le 0, \quad k \in \{1, \dots, p\},$$
$$\sum_{i=1}^{N} h_i^l(x_i) = 0, \quad l \in \{1, \dots, q\}.$$

Since the objective function is strongly convex and the feasible set is convex, this program has a unique optimizer x^*. Note that, even though the objective function is separable, the structure of the constraints couples the decision variables of the agents. This makes challenging the design of distributed algorithmic solutions of (4).

Remark 3.1: (Separability structure): Problems of the form (4) arise in multiple applications, including communications [20], economic dispatch of power systems [12], optimal power flow [21], resource allocation [22], and safe swarm behavior using control barrier functions [15]. Also, given convex sets X_i, $i \in \{1, \dots, N\}$, a common problem considered in the distributed optimization literature [23] is

$$\min_{x \in \mathbb{R}^{n}} \sum_{i=1}^{N} f_i(x),$$
$$\text{s.t.} \quad x \in \cap_{i=1}^{N} X_i.$$

When $X_i = \{x \in \mathbb{R}^n \ : \ \bar{g}_i(x) \le 0\} \subseteq \mathbb{R}^n$ for a continuously differentiable convex function with locally Lipschitz derivatives $\bar{g}_i : \mathbb{R}^n \to \mathbb{R}^{m_i}$, for $i \in \{1, \dots, N\}$, the optimization can be reformulated as

$$\min_{x \in \mathbb{R}^{nN}} \sum_{i=1}^{N} f_i(x_i),$$
$$\text{s.t.} \quad \bar{g}_i(x_i) \le 0, \ i \in \{1, \dots, N\},$$
$$(L \otimes I_n)x = \mathbf{0}_{Nn},$$

which is of the form (4). •

Throughout the paper, we denote $f(x) = \sum_{i=1}^{N} f_i(x_i)$, $g^k(x) = \sum_{i=1}^{N} g_i^k(x_i)$ for $k \in \{1, \dots, p\}$ and $h^l(x) = \sum_{i=1}^{N} g_i^l(x_i)$ for $l \in \{1, \dots, q\}$, and write the feasible set of (4) as

$$\mathcal{F} = \{x \in \mathbb{R}^{nN} \ : g^k(x) \le 0, \ \forall k \in \{1, \dots, p\},$$
$$h^l(x) \le 0, \ \forall l \in \{1, \dots, q\}\}.$$

We also make the following assumption.

Assumption 1: (Linear independence constraint qualification for separable constraints): For all $x \in \mathbb{R}^{nN}$, the vectors $\{\nabla g^k(x)\}_{k \in I_{g^1, \dots, g^p}(x)} \cup \{\nabla h^l(x)\}_{1 \le l \le q}$ are linearly independent.

Assumption 1 is common and guarantees that the KKT conditions are necessary and sufficient for the optimality of (4).

Our goal is to design an algorithm, in the form of a locally Lipschitz dynamical system, such that

(i) is *distributed*, i.e., each agent can execute it with locally available information;
(ii) is *anytime*, i.e., the feasible set $\mathcal{F}$ is forward invariant;
(iii) *solves* (4), i.e., all trajectories starting in $\mathcal{F}$ converge to its optimizer.

Even though algorithmic solutions exist in the literature that enjoy some of these properties (e.g., the projected saddle-point dynamics enjoys (i) and (iii) for certain classes of optimization problems), the design of an algorithm that enjoys all three is challenging.

IV. Design of Algorithmic Solution

Here we propose an algorithmic solution to the constrained program (4) to meet the requirements stated in Section III. Our exposition proceeds by first reformulating the optimization problem and then building on the projected saddle-point dynamics and the safe gradient flow to synthesize a coordination algorithm with the desired properties.

A. Reformulation using constraint mismatch variables

In this section we provide an equivalent formulation of (4) that addresses the coupling among the agents' decision variables arising from the structure of the constraints. The basic idea to "decouple" them is to introduce, following [24], *constraint-mismatch variables* which help agents keep track of local constraints while collectively satisfying the original constraints. Formally, to the state of each agent, we add one variable per constraint: y_i^k for agent i and the kth inequality constraint and z_j^l for agent i and the lth equality constraint. For convenience, we use the notation $x = [x_1, \dots, x_N]$, $y_i = [y_i^1, \dots, y_i^p]$, $z_i = [z_i^1, \dots, z_i^p]$, $y = [y_1, \dots, y_N]$, $z = [z_1, \dots, z_N]$. Consider then the following problem

$$\min_{x \in \mathbb{R}^{nN}, y \in \mathbb{R}^{Np}, z \in \mathbb{R}^{Nq}} \sum_{i=1}^{N} f_i(x_i), \tag{5}$$
$$\text{s.t.} \quad g_i^k(x_i) + \sum_{j \in \mathcal{N}_i} (y_i^k - y_j^k) \le 0,$$
$$h_i^l(x_i) + \sum_{j \in \mathcal{N}_i} (z_i^k - z_j^k) = 0,$$
$$i \in \{1, \dots, N\}, \ k \in \{1, \dots, p\}, \ l \in \{1, \dots, q\}.$$

Note that, in this formulation, constraints are now locally expressible, meaning that agent $i \in \{1, \dots, N\}$ can evaluate the ones corresponding to g_i^k and h_i^l with information from its neighbors. Let $\mu_i = [\mu_i^1, \dots, \mu_i^p]$, $\lambda_i = [\lambda_i^1, \dots, \lambda_i^p]$, $\lambda = [\lambda_1, \dots, \lambda_N]$ and $\mu = [\mu_1, \dots, \mu_N]$ be the Lagrange multipliers for the constraints in (5). We next show that the optimizer in x of (5) is x^*, the optimizer of (4).

Proposition 4.1: (Equivalence between the two formulations): Let $\mathcal{F}_r^*$ be the solution set of (5). Then, $x^* = \Pi_x(\mathcal{F}_r^*)$.

Proof: Note that (4) is equivalent to

$$\min_{x \in \mathbb{R}^{nN}, s \in \mathbb{R}^{p}} \sum_{i=1}^{N} f_i(x_i),$$
$$\text{s.t.} \quad \sum_{i=1}^{N} g_i^k(x_i) + s^k = 0, \ s^k \ge 0,$$
$$\sum_{i=1}^{N} h_i^l(x_i) = 0, \tag{6}$$
$$k \in \{1, \dots, p\}, \ l \in \{1, \dots, q\}.$$

and (5) is equivalent to

$$\begin{aligned}&\min_{x\in\mathbb{R}^{Nn},y\in\mathbb{R}^{Np},s\in\mathbb{R}^{Np},z\in\mathbb{R}^{Nq}}\sum_{i=1}^{N}f_i(x_i),\\ &\text{s.t.}\quad g_i^k(x_i)+\sum_{j\in\mathcal{N}_i}(y_i^k-y_j^k)+s_i^k=0,\ s_i^k\geq 0,\\ &\qquad h_i^l(x_i)+\sum_{j\in\mathcal{N}_i}(z_i^l-z_j^l)=0,\\ &\quad i\in\{1,\dots,N\},\ k\in\{1,\dots,p\},\ l\in\{1,\dots,q\}.\end{aligned}\tag{7}$$

Now the proof follows a similar reasoning as the proof from [24, Proposition 4.2]. We only need to show that the feasible sets of (6) and (7) are the same, because their objective functions coincide. First, if $(\hat{x},\hat{y},\hat{s},\hat{z})\in\mathbb{R}^{Nn}\times\mathbb{R}^{Np}\times\mathbb{R}^{Np}\times\mathbb{R}^{Nq}$ is a feasible point for (7), then by adding up all constraints for $i\in\{1,\dots,N\}$ and letting $\bar{s}^k=\sum_{i=1}^N\hat{s}_i^k$, $\bar{s}=[\bar{s}^1,\dots,\bar{s}^p]$ it follows that $(\hat{x},\bar{s})$ is a feasible point for (6). Now, let $(\tilde{x},\tilde{s})$ be a feasible point for (6). Let $v=[g_1^1(\tilde{x}_1),\dots,g_i^k(\tilde{x}_N),\dots;g_N^p(\tilde{x}_N)]\in\mathbb{R}^{Np}$ and $\breve{s}=[\frac{\tilde{s}^1}{N},\dots,\frac{\tilde{s}^1}{N},\dots,\frac{\tilde{s}^p}{N},\dots,\frac{\tilde{s}^p}{N}]\in\mathbb{R}^{Np}$. Note that $\mathbf{1}_{Np}^T(v+\breve{s})=0$. This implies that $v+\breve{s}$ belongs to the range space of the Laplacian L of the communication graph and hence there exists $\tilde{y}$ such that $-L\tilde{y}=v+\breve{s}$. By a similar argument, by letting $w=[h_1^1(\tilde{x}_1),\dots,h_i^l(\tilde{x}_N),\dots;g_N^p(\tilde{x}_N)]$ there exists $\tilde{z}$ such that $-L\tilde{z}=w$. Now it follows that $(\tilde{x},\tilde{y},\breve{s},\tilde{z})$ is feasible for (6), hence proving that the feasible sets of (6) and (7) are the same. ■

Proposition 4.1 implies that (5) has a unique optimizer in the variables x. However, since the objective function in (5) is not strongly convex in y and z, the optimizer in the variables y and z might not be unique. Hence, for the results that follow, we take $\epsilon>0$ and define $f_i^\epsilon(x_i,y_i,z_i)=f_i(x_i)+\frac{\epsilon}{2}\|y_i\|^2+\frac{\epsilon}{2}\|z_i\|^2$, $f^\epsilon(x,y,z)=\sum_{i=1}^N f_i^\epsilon(x_i,y_i,z_i)$. Consider the following regularized version of (5),

$$\begin{aligned}&\min_{x\in\mathbb{R}^{Nn},y\in\mathbb{R}^{Np},z\in\mathbb{R}^{Nq}}\sum_{i=1}^{N}f_i^\epsilon(x_i,y_i,z_i),\\ &\text{s.t.}\quad g_i^k(x_i)+\sum_{j\in\mathcal{N}_i}(y_i^k-y_j^k)\leq 0,\\ &\qquad h_i^l(x_i)+\sum_{j\in\mathcal{N}_i}(z_i^k-z_j^k)=0,\\ &\quad i\in\{1,\dots,N\},\ k\in\{1,\dots,p\},\ l\in\{1,\dots,q\}.\end{aligned}\tag{8}$$

Let $(x^{*,\epsilon},y^{*,\epsilon},z^{*,\epsilon})\in\mathbb{R}^{Nn}\times\mathbb{R}^{Np}\times\mathbb{R}^{Nq}$ be the optimizer of (8), which is unique because the objective function is strongly convex and the feasible set is convex. Next we establish a sensitivity result for the regularized optimization problem (8).

Lemma 4.2: (Sensitivity of regularized problem): Given $\delta>0$, there exists $\bar{\epsilon}>0$ so that if $\epsilon<\bar{\epsilon}$, then $\|x^{*,\epsilon}-x^*\|<\delta$.

Proof: Let (x^*,y^*,z^*) be an optimizer of (5) with $m=f(x^*,y^*,z^*)$. Since x^* is unique, there exists $\beta>0$ such that $f(x)=f^0(x,y,z)\geq m+\beta$, for all $x\in\mathcal{F}$, $y\in\mathbb{R}^{Np}$, $z\in\mathbb{R}^{Nq}$ whenever $\|x-x^*\|=\delta$. Hence, for any $\epsilon>0$ it follows that

$$f^\epsilon(x,y,z)\geq f^0(x,y,z)\geq m+\beta,$$

for all $x\in\mathcal{F}$, $y\in\mathbb{R}^{Np}$, $z\in\mathbb{R}^{Nq}$ whenever $\|x-x^*\|=\delta$. On the other hand, since f is continuous with respect to ϵ, for any $\delta>0$ we can find $\bar{\epsilon}$ such that

$$f^\epsilon(x^*,y^*,z^*)\leq m+\frac{\beta}{2}\quad\forall\epsilon<\bar{\epsilon}.$$

Hence, by taking $\epsilon<\bar{\epsilon}$ we can ensure that the set

$$\{(x,y,z)\in\mathcal{F}\times\mathbb{R}^{Np}\times\mathbb{R}^{Nq}\ :\ \|x-x^*\|\leq\delta\}$$

contains the local minimizer of f^ϵ for any $\epsilon<\bar{\epsilon}$. Thus, $\|x^{*,\epsilon}-x^*\|\leq\delta$ for all $\epsilon<\bar{\epsilon}$, as stated. ■

Given Lemma 4.2, in what follows we focus on solving (8) and assume that ϵ is taken sufficiently small to guarantee a desired maximum distance between $x^{*,\epsilon}$ and x^*.

B. Cascade of saddle-point dynamics and safe gradient flow

In this section, we build on the reformulation presented above to design our proposed algorithmic solution. Note that, if we had knowledge of the optimizers $y^{*,\epsilon},z^{*,\epsilon}$ of Problem (8), we could break the optimization into N, one per agent $i\in\{1,\dots,N\}$, decoupled optimization problems as follows,

$$\begin{aligned}&\min_{x_i\in\mathbb{R}^n}f_i^\epsilon(x_i),\\ &\text{s.t.}\quad g_i^k(x_i)+\sum_{j\in\mathcal{N}_i}((y_i^k)^{*,\epsilon}-(y_j^k)^{*,\epsilon})\leq 0,\\ &\qquad h_i^l(x_i)+\sum_{j\in\mathcal{N}_i}((z_i^k)^{*,\epsilon}-(z_j^k)^{*,\epsilon})=0,\\ &\quad k\in\{1,\dots,p\},\ l\in\{1,\dots,q\}.\end{aligned}\tag{9}$$

In turn, each of these problems could be solved in an anytime fashion by having each agent execute the corresponding safe gradient flow, cf. Proposition 2.1. However, since $y^{*,\epsilon}$ and $z^{*,\epsilon}$ are not readily available, agents need to interact with their neighbors to compute them. Since this would require an iterative algorithm, this means agents will face evolving y and z in the corresponding formulation of (9), which raises the additional challenge of ensuring the anytime nature of the safe gradient flow is preserved. We tackle these challenges next.

To generate the update law for y and z, we propose to use the projected saddle-point dynamics of (8). By [19, Theorem 5.1], these are guaranteed to converge to its optimizers. Simultaneously, we implement the safe gradient flow of (9) with the current values of y and z, i.e., (with the notation $y_{\mathcal{N}_i}=y_i\cup\{y_j\}_{j\in\mathcal{N}_i}$, $z_{\mathcal{N}_i}=z_i\cup\{z_j\}_{j\in\mathcal{N}_i}$):

$$\begin{aligned}&S_\alpha^i(x_i,y_{\mathcal{N}_i},z_{\mathcal{N}_i})=\\ &\arg\min_{\xi_i\in\mathbb{R}^n}\frac{1}{2}\|\xi_i+\nabla f_i(x_i)\|^2,\\ &\text{s.t.}\ \nabla g_i^k(x_i)\xi_i\leq-\alpha(g_i^k(x_i)+\sum_{j\in\mathcal{N}_i}(y_i^k-y_j^k)),\\ &\qquad\nabla h_i^l(x_i)\xi_i=-\alpha(h_i^l(x_i)+\sum_{j\in\mathcal{N}_i}(z_i^l-z_j^l)),\\ &\qquad\qquad k\in\{1,\dots,p\},\ l\in\{1,\dots,q\},\end{aligned}\tag{10}$$

for all $i\in\{1,\dots,N\}$. We denote $S_\alpha(x,y,z)=[S_\alpha^1(x_1,y_{\mathcal{N}_1},z_{\mathcal{N}_1}),\dots,S_\alpha^N(x_N,y_{\mathcal{N}_N},z_{\mathcal{N}_N})]$. To add more flexibility to our design, we add a timescale separation parameter $\tau > 0$ that allows the projected saddle-point dynamics to be run at a faster rate relative to the safe gradient flow. This leads to the cascaded dynamical system:

$$\tau \dot{v}_i = -\nabla f_i(v_i) - \sum_{k=1}^{p} \lambda_i^k \nabla g_i^k(v_i), \tag{11a}$$

$$\tau \dot{y}_i^k = -\epsilon y_i^k - \sum_{j \in \mathcal{N}_i} (\lambda_i^k - \lambda_j^k), \tag{11b}$$

$$\tau \dot{z}_i^l = -\epsilon z_i^l - \sum_{j \in \mathcal{N}_i} (\mu_i^l - \mu_j^l), \tag{11c}$$

$$\tau \dot{\lambda}_i^k = [g_i^k(v_i) + \sum_{j \in \mathcal{N}_i} (y_i^k - y_j^k)]^+_{\lambda_i^k}, \tag{11d}$$

$$\tau \dot{\mu}_i^l = h_i^l(v_i) + \sum_{j \in \mathcal{N}_i} (z_i^k - z_j^k), \tag{11e}$$

$$\dot{x}_i = S_\alpha^i(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i}), \tag{11f}$$

for $i \in \{1, \ldots, N\}$, $k \in \{1, \ldots, p\}$ and $l \in \{1, \ldots, q\}$, where v_i are *virtual* variables that play the role of x_i in the projected saddle-point dynamics. Since (11) results from the cascaded interconnection of saddle-point dynamics and the safe gradient flow, we refer to it as SP-SGF.

Remark 4.3: (Scalability and distributed character of SP-SGF): In algorithm (11), each agent has a state variable of dimension $2n + 2p + 2q$. To compute the evolution of these state variables, each agent only requires information provided by its neighbors in $\mathcal{G}$. Therefore, the algorithm is distributed. In addition, since the memory needed by each agent to run (11) remains constant as the network size N increases, the algorithm is also scalable. •

Remark 4.4: (Algorithm implementation): Note that the execution of SP-SGF requires solving the optimization problem (10), for each $i \in \{1, \ldots, N\}$, which is a quadratic program and hence can be solved efficiently. In fact, if the number of constraints is low, closed-form expressions for its solution [25, Theorem 1] are available. •

In what follows, we assume that for all $i \in \{1, \ldots, N\}$, the set of initial conditions for v_i, y_i, z_i, λ_i and μ_i in (11) lie in compact sets $\mathcal{V}_i$, $\mathcal{Y}_i$, $\mathcal{Z}_i$, Λ_i and M_i respectively. This means that the initial conditions v, y, z, λ, μ lie in compact sets $\mathcal{V} := \times_{i=1}^N \mathcal{V}_i$, $\mathcal{Y} := \times_{i=1}^N \mathcal{Y}_i$, $\mathcal{Z} := \times_{i=1}^N \mathcal{Z}_i$, $\Lambda = \times_{i=1}^N \Lambda_i$, $M = \times_{i=1}^N M_i$. Since the projected saddle-point dynamics (11a)-(11e) are convergent by [19, Theorem 5.1], there exist compact sets $\bar{\mathcal{V}}$, $\bar{\mathcal{Y}}$, $\bar{\mathcal{Z}}$, $\bar{\Lambda}$, $\bar{M}$ such that the trajectories of v, y, z, λ, μ under (11) stay in $\bar{\mathcal{V}}$, $\bar{\mathcal{Y}}$, $\bar{\mathcal{Z}}$, $\bar{\Lambda}$ and $\bar{M}$ respectively for all positive times. In what follows, we make the following assumption regarding the feasibility of (10).

Assumption 2: (Feasibility of S_α): For all $i \in \{1, \ldots, N\}$, (10) is feasible for all $(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i}) \in \Pi_{x_i}\mathcal{F} \times \Pi_{y_{\mathcal{N}_i}}\bar{\mathcal{Y}} \times \Pi_{z_{\mathcal{N}_i}}\bar{\mathcal{Z}}$.

The following result gives a sufficient condition under which Assumption 2 holds.

Lemma 4.5: (Sufficient condition for feasibility of S_α): Suppose that the vectors $\{\nabla g_i^k(x_i)\}_{k=1}^p \cup \{\nabla h_i^l(x_i)\}_{l=1}^q$ are linearly independent for all $x_i \in \Pi_{x_i}\mathcal{F}$. Then, for all $i \in \{1, \ldots, N\}$, (10) is feasible for all $(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i}) \in \Pi_{x_i}\mathcal{F} \times \Pi_{y_{\mathcal{N}_i}}\bar{\mathcal{Y}} \times \Pi_{z_{\mathcal{N}_i}}\bar{\mathcal{Z}}$.

Proof: By considering the inequality constraints in (10) as equality constraints, and since $p + q \leq n$ necessarily, (10) consists of a linear system of equations with at least as many unknowns as equations. If the number of equations is strictly less than the number of unknowns (i.e., $p + q < n$), (10) is feasible. If the number of equations is equal to the number of unknowns, (i.e., $p + q = n$), (10) is feasible because $\{\nabla g_i^k(x_i)\}_{k=1}^p \cup \{\nabla h_i^l(x_i)\}_{l=1}^q$ are linearly independent for all $x_i \in \Pi_{x_i}\mathcal{F}$. ■

The next result establishes some feasibility and regularity properties of S_α. Its proof follows an argument analogous to the proof of [11, Proposition 5.3].

Proposition 4.6: (Well-posedness and regularity of SP-SGF): Under Assumption 2, the following statements hold:

- There exists an open neighborhood U containing $\bar{\mathcal{V}} \times \bar{\mathcal{Y}} \times \bar{\mathcal{Z}} \times \bar{\Lambda} \times \bar{M} \times \mathcal{F}$ such that (11) is well-defined on U;
- The dynamical system (11) is locally Lipschitz on U;
- The Lagrange multipliers of (10) are unique and locally Lipschitz as a function of x, y and z on U.

V. Invariance and Convergence Analysis

Having established the distributed character of the algorithm (11), here we show the forward invariance of the feasible set and the asymptotic convergence to the optimizer.

We start by introducing some useful notation. For $i \in \{1, \ldots, N\}$, $k \in \{1, \ldots, p\}$ and $l \in \{1, \ldots, q\}$, we let $\psi_{y_i^k}(t; p_0), \psi_{z_j^l}(t; p_0), \psi_{x_i}(t; p_0)$ be the solution of (11b), (11c), (11f) respectively for initial conditions $p_0 = (v_0, y_0, z_0, \lambda_0, \mu_0, x_0) \in \mathcal{P} := \mathcal{V} \times \mathcal{Y} \times \mathcal{Z} \times \Lambda \times M \times \mathcal{F}$. We also let $\psi_y(t; p_0) = [\psi_{y_1^1}(t; p_0), \ldots, \psi_{y_1^p}(t; p_0), \ldots, \psi_{y_N^1}(t; p_0), \ldots, \psi_{y_N^p}(t; p_0)]$, and define $\psi_z(t; p_0)$ and $\psi_x(t; p_0)$ analogously. The next result establishes the *anytime* nature of SP-SGF.

Lemma 5.1: (Anytime property): Suppose that $x_0 \in \mathcal{F}$ and Assumption 2 holds. Then, the trajectories of (11) satisfy $\psi_x(t; p_0) \in \mathcal{F}$ for all $t \geq 0$.

Proof: Since Assumption 2 holds, the dynamics (11) are well-defined on a neighborhood U, cf. Proposition 4.6. If, at some $\bar{t}$, $\sum_{i=1}^N g_i^k(\psi_{x_i}(\bar{t}; p_0)) = 0$ for $k \in \{1, \ldots, p\}$, then because of the constraints in (10),

$$\begin{aligned}
&\frac{d}{dt} \sum_{i=1}^N g_i^k(\psi_{x_i}(t; p_0))|_{t=\bar{t}} \\
&= \sum_{i=1}^N \nabla g_i^k(\psi_{x_i}(\bar{t}; p_0)) S_\alpha^i(\psi_{x_i}(\bar{t}; p_0), \psi_{y_{\mathcal{N}_i}}(\bar{t}; p_0), \psi_{z_{\mathcal{N}_i}}(\bar{t}; p_0)) \\
&\leq -\alpha \sum_{i=1}^N \Big(g_i^k(\psi_{x_i}(\bar{t}; p_0)) + \sum_{j \in \mathcal{N}_i} (\psi_{y_i^k}(\bar{t}; p_0) - \psi_{y_j^k}(\bar{t}; p_0)) \Big) = 0.
\end{aligned}$$

Hence by Brezis' Theorem [26], it follows that $\sum_{i=1}^N g_i^k(\psi_{x_i}(t; p_0)) \leq 0$ for all $t \geq 0$, $p_0 \in \mathcal{P}$ and $k \in \{1, \ldots, p\}$. By a similar argument, $\frac{d}{dt} \sum_{i=1}^N h_i^l(\psi_{x_i}(t; p_0))|_{t=\bar{t}} = 0$. Hence, it follows that $\sum_{i=1}^N h_i^l(\psi_{x_i}(t; p_0)) = 0$ for all $t \geq 0$, $p_0 \in \mathcal{P}$ and $l \in \{1, \ldots, q\}$. ■

Next, we turn to the study of the convergence properties of (11). The next result establishes a connection between the equilibrium points of S_α and the optimizers of (8).

Proposition 5.2: (Relationship between equilibria and optimizers): Let $x \in \mathcal{F}$. Then, $S_\alpha(x, y^{*,\epsilon}, z^{*,\epsilon}) = 0$ if and only if $x = x^{*,\epsilon}$.

Proof: Note that $S^i_\alpha(x_i, y^{*,\epsilon}_{\mathcal{N}_i}, z^{*,\epsilon}_{\mathcal{N}_i})$ is the safe gradient flow associated to the optimization problem (9), which by Proposition 4.1 has $x^{*,\epsilon}_i$ as the unique optimizer. The result then follows from Proposition 2.1(v). ■

Next we show that the trajectories of x in (11) converge to the optimizer of (8).

Theorem 5.3: (Convergence to optimizer): Suppose Assumption 2 holds. For any $\delta > 0$ and compact set Ω containing $\{x \in \mathcal{F} \ : \ \|x - x^{*,\epsilon}\| \le \delta\}$, there exists $\tau_{\delta,\Omega} > 0$ and $T_{\delta,\Omega}$ such that if $\tau < \tau_{\delta,\Omega}$, then under the dynamics (11):

$$\|\psi_x(t; p_0) - x^{*,\epsilon}\| < \delta,$$

for all $t \ge T_{\delta,\Omega}$ and $p_0 = (v_0, y_0, z_0, \lambda_0, \mu_0, x_0) \in \mathcal{P}$. Moreover, if $\mathcal{F}$ is bounded, then for any $\tau > 0$,

$$\lim_{t\to\infty} \|\psi_x(t; p_0) - x^{*,\epsilon}\| = 0.$$

for all $p_0 \in \mathcal{P}$.

Proof: Since the dynamics in (11) are not differentiable, the standard version of Tikhonov's theorem for singular perturbations [27, Theorem 11.2] is not applicable. Instead we use [28, Corollary 3.4], which gives a Tikhonov-type singular perturbation statement for differential inclusions. In the case of non-smooth ODEs for which the fast dynamics do not depend on the slow variable, like (11), we need to check the following assumptions. First, that the dynamics (11) are Lipschitz. Note that local Lipschitzness of (11) follows from Proposition 4.6, the Lipschitzness of the gradients of f and g and the Lipschitzness of the max operator. Moreover, since $\bar{\mathcal{V}}$, $\bar{\mathcal{Y}}$, $\bar{\mathcal{Z}}$, $\bar{\Lambda}$, $\bar{M}$ and Ω are compact, we can redefine the dynamics (11) outside of $\bar{\mathcal{V}}\times\bar{\mathcal{Y}}\times\bar{\mathcal{Z}}\times\bar{\Lambda}\times\bar{M}\times\Omega$ so that they are globally Lipschitz while still keeping the same dynamics for initial conditions in $\mathcal{V}\times\mathcal{Y}\times\mathcal{Z}\times\Lambda\times M\times\Omega$. Second, existence and uniqueness of the equilibrium of the fast dynamics. This follows from the fact that (8) has a strongly convex objective function and convex constraints, which implies that it has a unique KKT point. Third, Lipschitzness and asymptotic stability of the reduced-order model

$$\dot{\bar{x}} = S_\alpha(\bar{x}, y^{*,\epsilon}, z^{*,\epsilon}) \tag{12}$$

Lipschitzness follows from Proposition 4.6 and asymptotic stability follows from [11, Theorem 5.6] and the fact that Ω is compact. Fourth, asymptotic stability of the fast dynamics. This follows from [19, Theorem 5.1]. Finally, note that $x^{*,\epsilon}$ is the only equilibrium point of (12) and hence the result follows from [28, Corollary 3.4].

Now suppose that $\mathcal{F}$ is compact. Pick an arbitrary $\theta > 0$. Since f is continuous and $x^{*,\epsilon}$ is the unique minimizer of (8), there exist constants $a_\theta > 0$, $b_\theta > 0$ such that the sets

$$A_\theta = \{x \in \mathcal{F} \ : \ f(x) - f(x^{*,\epsilon}) \le a_\theta\}$$
$$B_\theta = \{x \in \mathcal{F} \ : \ \|x - x^{*,\epsilon}\| \le b_\theta\}$$
$$C_\theta = \{x \in \mathcal{F} \ : \ \|x - x^{*,\epsilon}\| \le \theta\}$$

satisfy $B_\theta \subseteq A_\theta \subseteq C_\theta$. Next, we show that there exists $T_\theta > 0$ such that $\psi_x(t; p_0) \in C_\theta$ for $t \ge T_\theta$ and all $p_0 \in \mathcal{P}$ (i.e., C_θ is asymptotically stable relative to $\mathcal{F}$). Since θ is arbitrary, this completes the proof. Let $(\{\phi^k_i(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i})\}_{k=1}^p, \{\chi^l_i(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i})\}_{l=1}^q)$ be the Lagrange multipliers associated to the optimization problem defining $S^i_\alpha(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i})$, which are unique and locally Lipschitz by Proposition 4.6. Then, by following an argument analogous to the one in the proof of [11, Lemma 5.8]:

$$\begin{aligned} \frac{d}{dt}(f(x) - f(x^{*,\epsilon})) &\le -\|S_\alpha(x, y, z)\|^2 \\ &+ \sum_{i=1}^N \sum_{k=1}^p \phi^k_i(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i})\alpha(g^k_i(x_i) + \sum_{j\in\mathcal{N}_i}(y^k_i - y^k_j)) \\ &+ \sum_{i=1}^N \sum_{l=1}^q \chi^l_i(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i})\alpha(h^l_i(x_i) + \sum_{j\in\mathcal{N}_i}(z^k_i - z^k_j)). \end{aligned} \tag{13}$$

Since (11a)-(11e) are the projected saddle-point dynamics of (8) and the objective function of (8) is strongly convex, by [19, Theorem 5.1], the variables v, y, z, λ, μ converge to the KKT point of (8) for all $\tau > 0$. Moreover, since $S_\alpha(x, y^{*,\epsilon}, z^{*,\epsilon}) = 0$ if and only if $x = x^{*,\epsilon}$ by Proposition 5.2, S_α is continuous by Proposition 4.6 and $\mathcal{P}$ is compact, for any fixed $\tau > 0$, there exist $\sigma_{\theta,\tau}$ and $T_{1,\theta,\tau}$ such that for all $t \ge T_{1,\theta,\tau}$ and $p_0 \in \mathcal{P}$, $\|\psi_x(t; p_0) - x^{*,\epsilon}\| > b_\theta$ implies $\|S_\alpha(\psi_x(t; p_0), \psi_y(t; p_0), \psi_z(t; p_0))\| > \sigma_{\theta,\tau}$.

Now, define

$$\begin{aligned} \hat{g}^k_i(t, p_0) &= g^k_i(\psi_{x_i}(t; p_0)) + \sum_{j\in\mathcal{N}_i}(\psi_{y^k_i}(t; p_0) - \psi_{y^k_j}(t; p_0)), \\ \hat{h}^l_i(t, p_0) &= h^l_i(\psi_{x_i}(t; p_0)) + \sum_{j\in\mathcal{N}_i}(\psi_{z^l_i}(t; p_0) - \psi_{z^l_j}(t; p_0)), \\ \hat{\phi}^k_i(t, p_0) &= \phi^k_i(\psi_{x_i}(t; p_0), \psi_{y_{\mathcal{N}_i}}(t; p_0), \psi_{z_{\mathcal{N}_i}}(t; p_0)), \\ \hat{\chi}^l_i(t, p_0) &= \chi^l_i(\psi_{x_i}(t; p_0), \psi_{y_{\mathcal{N}_i}}(t; p_0), \psi_{z_{\mathcal{N}_i}}(t; p_0)), \end{aligned}$$

and let us show that there exists a time $T_{2,\theta,\tau} > 0$ such that

$$\alpha \sum_{i=1}^N \sum_{k=1}^p \hat{\phi}^k_i(t, p_0)\hat{g}^k_i(t, p_0) + \alpha \sum_{i=1}^N \sum_{l=1}^q \hat{\chi}^l_i(t, p_0)\hat{h}^l_i(t, p_0) < \frac{\sigma_{\theta,\tau}}{2}, \tag{14}$$

for all $t \ge T_{2,\theta}$ and $p_0 \in \mathcal{P}$. First define

$$c_\phi := \max_{\substack{(x,y,z)\in\mathcal{F}\times\bar{\mathcal{Y}}\times\bar{\mathcal{Z}} \\ i\in\{1,\dots,N\} \\ k\in\{1,\dots,p\}}} |\phi^k_i(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i})|,$$

$$c_\chi := \max_{\substack{(x,y,z)\in\mathcal{F}\times\bar{\mathcal{Y}}\times\bar{\mathcal{Z}} \\ i\in\{1,\dots,N\} \\ l\in\{1,\dots,q\}}} |\chi^l_i(x_i, y_{\mathcal{N}_i}, z_{\mathcal{N}_i})|.$$

Note that such c_ϕ, c_χ exist because $\mathcal{F}$, $\bar{\mathcal{Y}}$ and $\bar{\mathcal{Z}}$ are compact. Now note that

$$\begin{aligned} \frac{d}{dt}(\hat{g}^k_i(t, p_0)) &\le -\alpha\hat{g}^k_i(t, p_0) + \sum_{j\in\mathcal{N}_i}(\dot{\psi}_{y^k_i}(t; p_0) - \dot{\psi}_{y^k_j}(t)), \\ \frac{d}{dt}(\hat{h}^l_i(t, p_0)) &\le -\alpha\hat{h}^l_i(t, p_0) + \sum_{j\in\mathcal{N}_i}(\dot{\psi}_{z^l_i}(t; p_0) - \dot{\psi}_{z^l_j}(t)). \end{aligned}$$

Since the variables y_i^k, z_i^k are convergent by [19, Theorem 5.1],

$$\lim_{t\to\infty} \dot{\psi}_{y_i^k}(t;p_0) = 0, \quad \forall i \in \{1,\dots,N\},\ k \in \{1,\dots,p\},$$
$$\lim_{t\to\infty} \dot{\psi}_{z_i^l}(t;p_0) = 0, \quad \forall i \in \{1,\dots,N\},\ l \in \{1,\dots,q\}.$$

for all $p_0 \in \mathcal{P}$. Hence, there exists a time $\hat{T}_{2,\theta,\tau} > 0$ such that

$$\sum_{j\in\mathcal{N}_i} (\dot{\psi}_{y_i^k}(t;p_0) - \dot{\psi}_{y_j^k}(t)) \le \frac{\sigma_{\theta,\tau}}{8\alpha N p c_\phi},$$
$$\sum_{j\in\mathcal{N}_i} (\dot{\psi}_{z_i^l}(t;p_0) - \dot{\psi}_{z_j^l}(t)) \le \frac{\sigma_{\theta,\tau}}{8\alpha N p c_\chi}$$

for all $i \in \{1,\dots,N\}$, $k \in \{1,\dots,p\}$, $l \in \{1,\dots,q\}$ and $t \ge \hat{T}_{2,\theta,\tau}$, $p_0 \in \mathcal{P}$. By the Comparison Lemma [27, Lemma 3.4], it holds that

$$\hat{g}_i^k(t,p_0) \le \hat{g}_i^k(\hat{T}_{2,\theta,\tau},p_0)e^{-\alpha(t-\hat{T}_{2,\theta,\tau})} + \frac{\sigma_{\theta,\tau}}{8Npc_\phi},$$
$$\hat{h}_i^l(t,p_0) \le \hat{h}_i^l(\hat{T}_{2,\theta,\tau},p_0)e^{-\alpha(t-\hat{T}_{2,\theta,\tau})} + \frac{\sigma_{\theta,\tau}}{8Npc_\chi}.$$

Since $\mathcal{F}$ is compact, $\psi_x(t;p_0) \in \mathcal{F}$ by Lemma 5.1, $\psi_y(t;p_0) \in \bar{\mathcal{Y}}$ and $\psi_z(t;p_0) \in \bar{\mathcal{Z}}$ for all $t \ge 0$, this implies that there exists a time $T_{2,\theta,\tau} > 0$ such that (14) holds for all $t \ge T_{2,\theta,\tau}$ and $p_0 \in \mathcal{P}$. Now, let $T_{\theta,\tau} = \max\{T_{1,\theta,\tau}, T_{2,\theta,\tau}\}$. Then, it holds that for all $t \ge T_{\theta,\tau}$, $\frac{d}{dt}(f(\psi_x(t;p_0)) - f(x^{*,\epsilon})) < 0$ if $\|\psi_x(t;p_0) - x^{*,\epsilon}\| > b_\theta$. Since $B_\theta \subseteq A_\theta$, this implies that A_θ is asymptotically stable relative to $\mathcal{F}$. Since $A_\theta \subseteq C_\theta$, it follows that C_θ is asymptotically stable relative to $\mathcal{F}$, hence completing the proof. Note that this argument is valid for all fixed $\tau > 0$. ■

By Theorem 5.3, the trajectories of the x variable in SP-SGF converge arbitrarily close to the optimizer $x^{*,\epsilon}$ provided that the timescale parameter τ is small enough. Moreover, if the feasible set $\mathcal{F}$ is bounded, asymptotic convergence holds for any timescale. The combination of the scalable and distributed character, cf. Remark 4.3, the anytime nature, cf. Lemma 5.1, and the convergence properties, cf. Theorem 5.3 means that SP-SGF provides the algorithmic solution with the properties stated in Section III.

Example 5.4: (Resource allocation): We illustrate the behavior of SP-SGF in a resource allocation example. Consider 13 agents whose communication graph is an undirected line graph. Solving distributed optimization problems with this particular topology is challenging due to its low connectivity. Each agent's state variable is $x_i = [x_{i,1}, x_{i,2}] \in \mathbb{R}^2$, where $x_{i,1}$ (resp. $x_{i,2}$) corresponds to the amount of resource 1 (resp. 2) allocated by agent i. Resource 1 is subject to an equality constraint and resource 2 is subject to an inequality constraint. Hence, the agents solve the optimization problem,

$$\min_{\{x_i\}_{i=1}^{13}} \sum_{i=1}^{13} \frac{1}{2}\|x_i\|^2, \tag{15}$$
$$\text{s.t. } h(\{x_i\}_{i=1}^{13}) = 5 - \sum_{i=1}^{13} p_i x_{i,1} = 0,$$
$$g(\{x_i\}_{i=1}^{13}) = -3 + \sum_{i=1}^{13} e^{-x_{i,2}} \le 0.$$

with $p_1 = 1$, $p_2 = 3$, $p_3 = 2$, $p_4 = 1$, $p_5 = 1$, $p_6 = 1$, $p_7 = 2$, $p_8 = 4$, $p_9 = 1$, $p_{10} = 1$, $p_{11} = 0.5$, $p_{12} = 2$, $p_{13} = 1$. Note that the condition in Lemma 4.5 holds and hence Assumption 2 holds. This implies by Proposition 5.3 that (11) is well-defined for (15). We use $\epsilon = 0.0001$ and $\alpha = 1$. Figure 1 illustrates the convergence of the x variables under SP-SGF.

Since the feasible set of (15) is unbounded, Proposition 5.3 states that convergence arbitrarily close to the optimizer can be achieved by taking τ sufficiently small. Figure 2 illustrates the convergence of the quantities $\sum_{i=1}^{13} x_{i,1}^2$ and $\sum_{i=1}^{13} x_{i,2}^2$ for different values of τ and shows that this quantity converges exactly to its optimal value for a wide range of values of τ, suggesting that the statement in Proposition 5.3 might be too conservative.

Figure 3 compares the evolution of the constraints of (15) under SP-SGF against two other algorithms: the projected saddle point dynamics (3) (abbreviated SP), which is not distributed, and the projected saddle-point dynamics (abbreviated SP-CM) for its reformulation with constraint mismatch variables as in (5), which is distributed. SP-SGF satisfies the constraints at all times whereas SP and SP-CM do not. We note that, in this case, SP-SGF requires running a dynamical system with 104 scalar variables (8 for each agent), SP-CM requires running a dynamical system of 78 scalar variables (6 for each agent) and SP requires running a dynamical system with 28 scalar variables. •

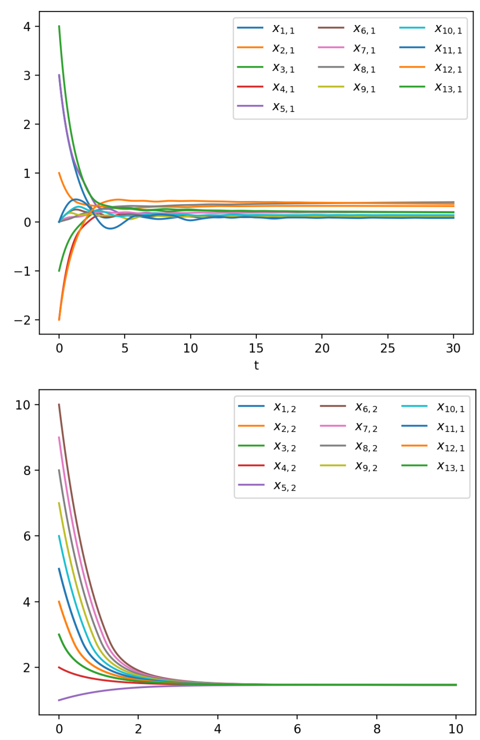

Fig. 1: Evolution of the variables $x_{1,i}$ (top) and $x_{2,i}$ (bottom) for $i \in \{1,\dots,13\}$ under SP-SGF for (15) with initial conditions $x_1 = [3,5]$, $x_2 = [1,4]$, $x_3 = [-1,3]$, $x_4 = [-2,2]$, $x_5 = [3,1]$, $x_6 = [0,10]$, $x_7 = [0,9]$, $x_8 = [0,8]$, $x_9 = [0,7]$, $x_{10} = [0,6]$, $x_{11} = [0,5]$, $x_{12} = [-2,4]$, $x_{13} = [4,3]$, $v_{i,1} = v_{i,2} = z_i = y_i = \lambda_i = \mu_i = 0$ for all $i \in \{1,\dots,13\}$ and $\tau = 1$.

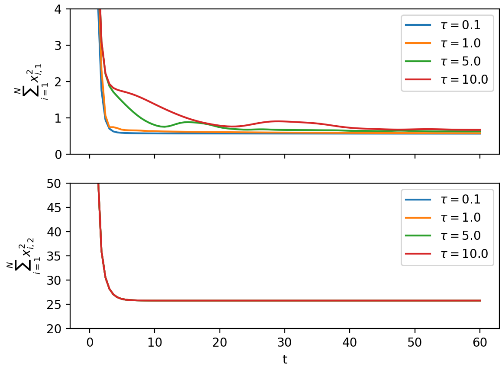

Fig. 2: This plot shows the evolution of $\sum_{i=1}^{13} x_{i,1}^2$ and $\sum_{i=1}^{13} x_{i,2}^2$ under SP-SGF with initial conditions as in Figure 1 for different values of τ.

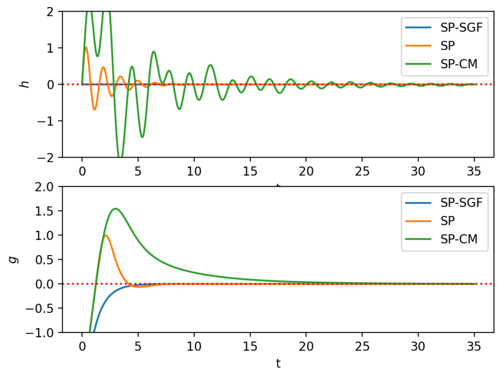

Fig. 3: This plot shows the evolution of the constraints of (15) for SP-SGF with the same initial conditions as in Figure 1, SP with the same primal initial conditions as in Figure 1 and $\lambda = \mu = 0$ and SP-CM with the same initial conditions as in Figure 1 for x_i, z_i, y_i, λ_i and μ_i for $i \in \{1, \dots, 13\}$.

VI. Conclusions

We have introduced a continuous-time dynamical system that solves network optimization problems with separable objective function and constraints in a distributed and anytime fashion. We have achieved this by combining the projected saddle-point dynamics and the safe gradient flow in a cascaded system. We have argued the scalable nature of the algorithm execution from the point of view of individual agents and established practical convergence to the optimizer when the feasible is unbounded, and exact convergence when it is bounded. Future work will consider other network optimization problems, refine the convergence guarantees presented here and possibly design new distributed, anytime algorithms, and investigate discretization schemes for the continuous-time dynamics. We also plan to apply our coordination algorithms in the implementation of optimization-based controllers arising from safety certificates for multi-agent systems.

References

[1] R. W. Brockett, "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems," *Linear Algebra and its Applications*, vol. 146, pp. 79–91, 1991.

[2] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*. Springer, 1994.

[3] A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Optimization algorithms as robust feedback controllers," *preprint*, 2021.

[4] A. Nedić and J. Liu, "Distributed optimization for control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 77–103, 2018.

[5] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, "A survey of distributed optimization," *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

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

[7] T. Dean, *Intractability and time-dependent planning*. Morgan Kaufmann, 1987.

[8] T. Dean and M. Boddy, "An analysis of time-dependent planning," *Proceedings of the Seventh AAAI National Conference on Artificial Intelligence*, pp. 49–54, 1988.

[9] Y. S. Xia and J. Wang, "On the stability of globally projected dynamical systems," *Journal of Optimization Theory & Applications*, vol. 106, no. 1, pp. 129–150, 2000.

[10] M. Muehlebach and M. I. Jordan, "On constraints in first-order optimization: A view from non-smooth dynamical systems," *Journal of Machine Learning Research*, vol. 23, pp. 1–47, 2022.

[11] A. Allibhoy and J. Cortés, "Control barrier function-based design of gradient flows for constrained nonlinear programming," *IEEE Transactions on Automatic Control*, 2022, submitted.

[12] A. Cherukuri and J. Cortés, "Distributed generator coordination for initialization and anytime optimization in economic dispatch," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 3, pp. 226–237, 2015.

[13] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: theory and applications," in *European Control Conference*, Naples, Italy, 2019, pp. 3420–3431.

[14] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2017.

[15] U. Borrmann, L. Wang, A. D. Ames, and M. Egerstedt, "Control barrier certificates for safe swarm behavior," *IFAC-PapersOnLine*, vol. 48, no. 27, pp. 68–73, 2015.

[16] L. Wang, A. Ames, and M. Egerstedt, "Safety barrier certificates for collisions-free multirobot systems," *IEEE Transactions on Robotics*, vol. 33, no. 3, pp. 661–674, 2017.

[17] M. Jankovic and M. Santillo, "Collision avoidance and liveness of multi-agent systems with cbf-based controllers," in *IEEE Conf. on Decision and Control*, Austin, TX, USA, 2021, pp. 6822–6828.

[18] X. Tan and D. V. Dimarogonas, "Distributed implementation of control barrier functions for multi-agent systems," *IEEE Control Systems Letters*, vol. 6, pp. 1879–1884, 2022.

[19] A. Cherukuri, E. Mallada, S. H. Low, and J. Cortés, "The role of convexity in saddle-point dynamics: Lyapunov function and robustness," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2449–2464, 2018.

[20] 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*, vol. 49, no. 3, pp. 237–252, 1998.

[21] T. Erseghe, "Distributed optimal power flow using ADMM," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2370–2380, 2014.

[22] L. Xiao and S. Boyd, "Optimal scaling of a gradient method for distributed resource allocation," *Journal of Optimization Theory & Applications*, vol. 129, no. 3, pp. 469–488, 2006.

[23] 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.

[24] A. Cherukuri and J. Cortés, "Distributed algorithms for convex network optimization under non-sparse equality constraints," in *Allerton Conf. on Communications, Control and Computing*, Monticello, IL, Sept. 2016, pp. 452–459.

[25] X. Tan and D. V. Dimarogonas, "On the undesired equilibria induced by control barrier function based quadratic programs," *arXiv preprint arXiv:2104.14895*, 2021.

[26] H. Brezis, "On a characterization of flow-invariant sets," *Communications on Pure and Applied Mathematics*, vol. 23, pp. 261–263, 1970.

[27] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.

[28] F. Watbled, "On singular perturbations for differential inclusions on the infinite interval," *Journal of Mathematical Analysis and Applications*, vol. 310, no. 2, pp. 362–378, 2005.