Proximal regularization for the saddle point gradient dynamics

Diego Goldsztajn and Fernando Paganini *Fellow, IEEE*,

***Abstract*—This paper concerns the solution of a convex optimization problem through the saddle point gradient dynamics. Instead of using the standard Lagrangian as is classical in this method, we consider a regularized Lagrangian obtained through a proximal minimization step. We show that, without assumptions of smoothness or strict convexity in the original problem, the regularized Lagrangian is smooth and leads to globally convergent saddle point dynamics. The method is demonstrated through an application to resource allocation in cloud computing.**

***Index Terms*—Convex optimization, proximal method, saddle point dynamics.**

I. Introduction

This paper concerns a classical method in mathematical optimization: given a convex-concave function of two vector variables, the dynamic rule that moves each variable respectively against or along the corresponding gradient. Conditions for the convergence of such dynamics to a saddle point were given in the monograph [1]. The primary motivation comes from distributed solutions to convex optimization problems, where the function in question is a Lagrangian; in this context it is called primal-dual gradient dynamics.

From its origins in economic theory, the method came more recently into the forefront in the context of resource allocation for communication networks [13], [25], [29]. Since then, a very active community has further investigated these dynamics with more advanced tools of control theory.

One line of inquiry concerns the rigorous treatment of discontinuous projections that arise when variable domains are subject to hard constraints, as happens with the dual variables of inequality constrained convex programs. Particularly challenging are LaSalle invariance results in this context. Overcoming limitations of the analysis in [13], it was shown in [6] how the results follow from the theory of projected dynamical systems [18]. More recent studies that incorporate other forms of non-smoothness are given in [9], [19].

Another subject of analysis has been the relationship between the degree of convexity or concavity of the function in question, and the achievable convergence. The global asymptotic stability results of [1] required *strict* convexity in at least one of the variables; otherwise, oscillatory behavior can be observed [13], [15]. It was shown recently in [7] that *local, strong* convexity or concavity suffice for asymptotic stability. Under global strong convexity of the optimization objective, exponential stability results can be established [19], [22].

A technique of widespread use is *regularization*, which refers to adding penalty terms to the objective function. In [1], [13] such terms were included to obtain strict convexity, without modifying the optimum. A related development are *proximal* methods [5], [21], based on minimization steps with a quadratic penalty on deviations. The recent paper [10] combines proximal methods with the saddle point dynamics, for problems involving a composite smooth/non-smooth objective and linear equality constraints. Tikhonov-type regularizations are another alternative, explored in [26].

In this paper we combine proximal methods with the saddle point dynamics, treating the very general situation of a convex optimization problem, not necessarily smooth or strictly convex, and with inequality constraints. Through a penalty on deviations in the primal variable, and partial minimization, we obtain a reduced Lagrangian which is proved to be smooth and to always generate convergent saddle point dynamics, which often lead to distributed solutions to the optimization problem. The convergence results are attained even though the reduced Lagrangian is in general not strictly convex or concave; this requires different techniques than those in the literature.

The paper is organized as follows. We end this section with a more precise background review. In Section II we introduce our problem and carry out the proximal regularization, establishing the properties of the reduced Lagrangian. The saddle point dynamics for this object are analyzed in Section III. In Section IV we apply this method to a resource allocation problem in cloud computing. Conclusions are provided in Section V and one proof is deferred to the Appendix.

A. Background and prior work

Here we provide a compact review of the most relevant literature. Consider first the equality constrained optimization

$$\min_x f(x), \quad \text{subject to} \quad g(x) = 0, \tag{1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a differentiable and convex function, and the *linear* constraints $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are independent. Assuming the corresponding Lagrangian

$$L_0(x, \nu) = f(x) + \nu^T g(x) \tag{2}$$

has a saddle point $(\hat{x}, \hat{\nu})$, the saddle point gradient dynamics

$$\dot{x} = -\frac{\partial L_0}{\partial x}, \qquad \dot{\nu} = \frac{\partial L_0}{\partial \nu},$$

are such that the Euclidean distance $\|x - \hat{x}\|^2 + \|\nu - \hat{\nu}\|^2$ is non-increasing along trajectories [1]. If $f(x)$ is also strictly convex,

D. Goldsztajn is with Eindhoven University of Technology, the Netherlands (e-mail: d.e.goldsztajn@tue.nl). Formerly at Universidad ORT Uruguay.

F. Paganini is with Universidad ORT Uruguay, Montevideo, Uruguay (e-mail: paganini@ort.edu.uy).

This work was supported in part by ANII-Uruguay under grant FCE_1_2019_1_156666.

asymptotic stability of the saddle point (unique in this case) follows by a LaSalle argument. Assuming *strong* convexity of f, exponential stability is established in [22].

An alternative from [9] is to apply the saddle point dynamics to the regularized *augmented Lagrangian*[1]

$$L_\mu(x,\nu) = f(x) + \nu^T g(x) + \frac{1}{2\mu}\|g(x)\|^2, \tag{3}$$

used in the "Method of Multipliers" [3]; exponential stability results are given for strongly convex f.

Recently, [10] considered the augmented Lagrangian when the objective f is not smooth in all its variables: namely, $f(x) = f_1(x_1) + f_2(x_2)$ where $x = (x_1, x_2)$ and only f_1 is differentiable. Let $g(x) = Ax_1 - x_2$, minimizing L_μ with respect to x_2 leads to a reduced Lagrangian in (x_1, ν) which is termed the *Proximal Augmented Lagrangian*; applying the saddle point dynamics to this object yields convergence. If f_1 is *strongly* convex exponential convergence holds [11].

The situation is more difficult when convex *inequality* constraints of the form $g(x) \leq 0$ are introduced. In this case the corresponding multipliers in (2) are constrained to the positive orthant $\{\nu \geq 0\}$. Let $[\beta]^+_\nu := \beta$ if $\nu > 0$ and $[\beta]^+_\nu := \max\{\beta, 0\}$ otherwise. The saddle point dynamics are typically modified by a positive projection of this kind:

$$\dot{\nu_j} = \left[\frac{\partial L_0}{\partial \nu_j}\right]^+_{\nu j}.$$

In these cases the saddle point dynamics have a discontinuous right-hand side, so non-trivial difficulties arise in their solution. In [6] a LaSalle invariance principle for Carathéodory solutions is applied to prove asymptotic stability, for a strictly convex objective. Alternatively, [9] uses the augmented Lagrangian with an additional penalty for inequality constraints, to obtain a strict Lyapunov function. Yet another regularized Lagrangian which avoids discontinuous projections and may achieve exponential stability is introduced in [22].

Other related work in this very active topic is to consider input-to-state stability in saddle point dynamics with injected noise [7], [24]. The connection of Robust optimization with these methods was recently considered in [12]. The approach of [10] was extended in [8] to an online tracking problem in the presence of disturbances. Another feedback-based application of the primal-dual dynamics is covered in [4].

II. Problem Formulation and regularization

Consider the general convex optimization problem

$$\underset{x}{\text{minimize}}\ f(x), \quad \text{subject to } g(x) \leq 0. \tag{4}$$

- $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ is convex, proper and closed (lower semi-continuous), see [23]. With f taking values on the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ we may encode general convex constraints. We assume that f is continuous on its effective domain (the set where it is finite), but we do not assume differentiability or strict convexity.
- $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ has convex (finite valued) components.

[1]May be interpreted as adding integral action to the control [27].

Under these very general conditions, we will develop a procedure for solving (4), based on saddle point gradient dynamics whose trajectories converge to an optimal solution.

Remark 1: For simplicity we have only included inequality constraints in (4); as mentioned these are the most difficult to handle. The results in this paper extend to the situation where, in addition, there are linear equality constraints with associated multipliers. We reiterate that implicit constraints may also be encoded in the objective function.

Let us introduce a regularization, adding a quadratic term to the cost and incorporating a proximal variable z:

$$\underset{x,z}{\text{minimize}}\ f(x) + \frac{1}{2}\left|\left|x - z\right|\right|^2, \quad \text{subject to } g(x) \leq 0. \tag{5}$$

Clearly, the two problems are equivalent. Specifically, they have the same infimum, and $\hat{x}$ is optimal for (4) if and only if $(\hat{x}, \hat{z})$ is optimal for (5) with $\hat{x} = \hat{z}$.

Introducing the dual variable ν, we may write the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \overline{\mathbb{R}}$ for problem (5):

$$L(x,z,\nu) = f(x) + \nu^\intercal g(x) + \frac{1}{2}\left|\left|x - z\right|\right|^2. \tag{6}$$

Note that we have not imposed yet any sign restrictions on ν.

One may consider applying the saddle point dynamics to the above Lagrangian. However, this would require dealing with the possibly non-smooth nature of f and g, as well as constraints in x implicit in f. Instead, we will attain smoothness by *reducing* the Lagrangian, as follows.

Definition 1: The function $\bar{L} : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ such that

$$\bar{L}(z,\nu) = \min_x \left\{ f(x) + \nu^\intercal g(x) + \frac{1}{2}\left|\left|x - z\right|\right|^2 \right\}.$$

is the *reduced Lagrangian* of (5).

For each fixed $(z, \nu) \in \mathbb{R}^m \times \mathbb{R}^n$, the minimized function above is strongly convex in x, so it has a unique minimizer

$$\bar{x}(z,\nu) := \underset{x}{\text{argmin}}\ L(x,z,\nu). \tag{7}$$

Assume momentarily that $\nu = \nu_0$ is fixed. Then the mapping $z \mapsto \bar{x}(z, \nu_0)$ is the *proximal* mapping [2] associated with the function $f(x) + \nu_0^T g(x)$, which is *non-expansive*:

$$\left|\left|\bar{x}(z,\nu_0) - \bar{x}(z',\nu_0)\right|\right| \leq \left|\left|z - z'\right|\right|.$$

Also, the reduced Lagrangian $\bar{L}(z, \nu_0)$ is the *Moreau-Yosida regularization* [21] of $f(x) + \nu_0^T g(x)$. In particular, it is convex and differentiable in z, with gradient

$$\frac{\partial \bar{L}}{\partial z}(z,\nu_0) = z - \bar{x}(z,\nu_0). \tag{8}$$

In what follows we will need to strengthen these properties of $\bar{x}(z,\nu)$ and $\bar{L}(z,\nu)$ to cover the case of *simultaneous* variation in (z,ν). Our first result of this kind is now stated; its proof is given in the Appendix.

Proposition 1: $\bar{x} : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is a locally Lipschitz continuous function of the variables (z,ν).

Focusing now on the reduced Lagrangian, the main result of this section is that it is smooth, over the unrestricted variables $(z,\nu) \in \mathbb{R}^n \times \mathbb{R}^m$, with locally Lipschitz gradients.

Theorem 2: $\bar{L}$ is convex in z, concave in ν, and continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m$, with locally Lipschitz gradients:

$$\frac{\partial \bar{L}}{\partial z}(z,\nu) = z - \bar{x}(z,\nu) \quad \forall\ (z,\nu) \in \mathbb{R}^m \times \mathbb{R}^n, \tag{9a}$$

$$\frac{\partial \bar{L}}{\partial \nu}(z,\nu) = g\left(\bar{x}(z,\nu)\right) \quad \forall\ (z,\nu) \in \mathbb{R}^m \times \mathbb{R}^n. \tag{9b}$$

The proof relies on Danskin's theorem. We state the following version adapted from [3].

Lemma 3: Consider a continuous $\phi : K \times \mathbb{R}^m \longrightarrow \mathbb{R}$, where $K \subset \mathbb{R}^n$ is compact. Assume $\phi(x,\nu)$ is convex in ν, and that its gradient with respect to ν exists in $K \times \mathbb{R}^m$. Define

$$\varphi(\nu) = \max_{x \in K} \phi(x,\nu).$$

Then the subgradient of φ at $\nu = \nu_0$ is the convex hull

$$\partial\varphi(\nu_0) = \mathrm{conv}\left\{ \frac{\partial \phi}{\partial \nu}(x,\nu_0) : x \in \underset{y \in K}{\mathrm{argmax}}\ \phi(y,\nu_0) \right\}. \tag{10}$$

Proof of Theorem 2: Convexity in z and the expression in (9a) follow results in [21], as already noted in equation (8). Considering now variations in ν, for a fixed $z = z_0$, we have

$$\bar{L}(z_0,\nu) = \min_x L(x,z_0,\nu); \tag{11}$$

minimum of linear functions of ν, therefore concave.

We will apply Danskin's result to establish (9b). Fix a point ν_0 and the closed ball $B := \{\nu : \|\nu - \nu_0\| \le 1\}$. Define $K = \bar{x}(z_0, B)$, compact by the continuity of Proposition 1.

Consider $\phi(x,\nu) := -L(x,z_0,\nu)$, which satisfies the hypotheses of Lemma 3 in $K \times \mathbb{R}^m$. Note that $K \subset \mathrm{dom}(f)$, effective domain of the objective, hence ϕ is continuous as required, and linear in ν, with $\frac{\partial \phi}{\partial \nu} = -g(x)$. Also, the set

$$\underset{x \in K}{\mathrm{argmax}}\ \phi(x,\nu_0) = \underset{x \in K}{\mathrm{argmin}}\ L(x,z_0,\nu_0) = \{\bar{x}(z_0,\nu_0)\},$$

a singleton. Thus, the subgradient in (10) is actually a gradient:

$$\frac{\partial \varphi}{\partial \nu}(\nu_0) = -g(\bar{x}(z_0,\nu_0)).$$

By construction, for $\nu \in B$ the minimum in (11) is achieved within K, so $\bar{L}(z_0,\nu) = -\varphi(\nu)$. This proves (9b) at $\nu = \nu_0$.

Finally, from (9), the gradients of $\bar{L}$ are locally Lipschitz: indeed, g has this property because it is convex and takes finite values, and so does $\bar{x}$ by Proposition 1. ∎

We have shown that $\bar{L}$ is convex-concave, and differentiable with locally Lipschitz gradients. It thus becomes natural to apply the saddle point gradient dynamics to this well-behaved function, in lieu of the Lagrangian (6).

Remark 2: In Section I we reviewed the "Proximal Augmented Lagrangian" method of [10], which is also based on partial minimization of a quadratically penalized Lagrangian. As explained, their problem is narrower in scope, with a specific form of the objective, and linear equality constraints; also, in their case the square of the *constraints* is the penalty function, as in the Method of Multipliers. Our regularization has the precursor [17], for a network resource allocation problem with differentiable objective, and linear inequality constraints. Our analysis generalizes this procedure.

Before proceeding to the saddle point dynamics, we establish that this method actually solves the original problem. We will use the following property of positive projections.

Lemma 4: For all $\nu, \mu \in \mathbb{R}^n_+$ and $\beta \in \mathbb{R}^n$ we have

$$(\nu - \mu)^\intercal \left([\beta]^+_\nu - \beta \right) \le 0.$$

Proof: The vector $[\beta]^+_\nu$ has components $[\beta_i]^+_{\nu_i}$ which only differ from β_i when $\beta_i < 0$ and $\nu_i = 0$, in this case $[\beta_i]^+_{\nu_i} = 0$. Then $(\nu - \mu)^\intercal([\beta]^+_\nu - \beta)$ is the sum of $\mu_i\beta_i$ over the indexes i such that $\beta_i < 0$ and $\nu_i = 0$, which is non-positive. ∎

Theorem 5: The following are equivalent and imply $\hat{x} = \hat{z}$.

(a) $(\hat{x}, \hat{z}, \hat{\nu})$ is a saddle point of L over $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m_+$; namely, minimum in (x,z) and maximum in ν.

(b) $(\hat{z}, \hat{\nu})$ is a saddle point of $\bar{L}$ over $\mathbb{R}^n \times \mathbb{R}^m_+$; minimum in z and maximum in ν. Also, $\bar{x}(\hat{z},\hat{\nu}) = \hat{x}$.

Proof: We prove first that (a) implies (b). Let $(\hat{x}, \hat{z}, \hat{\nu})$ be a saddle point of L, and note that in particular:

$$L(\hat{x},\hat{z},\hat{\nu}) = \min_x L(x,\hat{z},\hat{\nu}).$$

This proves that $\bar{x}(\hat{z},\hat{\nu}) = \hat{x}$, which further implies that

$$\bar{L}(\hat{z},\hat{\nu}) = L(\hat{x},\hat{z},\hat{\nu}) \le L(\bar{x}(z,\hat{\nu}),z,\hat{\nu}) = \bar{L}(z,\hat{\nu}) \quad \forall\ z \in \mathbb{R}^n.$$

Also, since the saddle is maximizing in ν we have

$$\bar{L}(\hat{z},\hat{\nu}) = L(\hat{x},\hat{z},\hat{\nu}) \ge L(\hat{x},\hat{z},\nu) \ge \bar{L}(\hat{z},\nu) \quad \forall\ \nu \in \mathbb{R}^m_+.$$

Consequently, $(\hat{z},\hat{\nu})$ is a saddle point of $\bar{L}$.

We now show that (b) implies (a). Let $\hat{x} := \bar{x}(\hat{z},\hat{\nu})$. Then

$$L(\hat{x},\hat{z},\hat{\nu}) = \bar{L}(\hat{z},\hat{\nu}) \le \bar{L}(z,\hat{\nu}) \le L(x,z,\hat{\nu}),$$

which establishes the required minimum in $(x,z) \in \mathbb{R}^n \times \mathbb{R}^n$.

For the maximum over ν, note first that since $\hat{\nu}$ maximizes the concave differentiable function $\bar{L}(\hat{z}, \cdot\,)$ over the orthant $\mathbb{R}^n_+$, then for each coordinate i we have

$$\text{either} \quad \frac{\partial \bar{L}}{\partial \nu_i}(\hat{z},\hat{\nu}) = 0, \quad \text{or} \quad \nu_i = 0,\ \frac{\partial \bar{L}}{\partial \nu_i}(\hat{z},\hat{\nu}) < 0.$$

The preceding cases are summarized by the condition

$$[g(\hat{x})]^+_{\hat{\nu}} = \left[\frac{\partial \bar{L}}{\partial \nu}(\hat{z},\hat{\nu})\right]^+_{\hat{\nu}} = 0,$$

where we have invoked the expression from Theorem 2 for the gradient $\bar{L}$, and $\bar{x}(\hat{z},\hat{\nu}) = \hat{x}$. We now see that

$$(\nu - \hat{\nu})^\intercal g(\hat{x}) = (\hat{\nu} - \nu)^\intercal \left([g(\hat{x})]^+_{\hat{\nu}} - g(\hat{x}) \right) \le 0$$

for all $\nu \in \mathbb{R}^n_+$ by Lemma 4. The maximizing part of the condition in (a) now follows from

$$\begin{aligned} L(\hat{x},\hat{z},\nu) &= L(\hat{x},\hat{z},\hat{\nu}) + (\nu - \hat{\nu})^\intercal g(\hat{x}) \\ &\le L(\hat{x},\hat{z},\hat{\nu}) \quad \forall\ \nu \in \mathbb{R}^m_+. \end{aligned}$$

Finally, condition (a) implies $\hat{x} = \hat{z}$ because the minimum of the Lagrangian (6) clearly requires this. ∎

III. Saddle point dynamics

The preceding theorem tells us that it is enough to find the saddle points of $\bar{L}$ in order to solve problem (5). Since $\bar{L}$ is differentiable by Theorem 2, this may be achieved using the following saddle point gradient dynamics:

$$\dot{z} = -\frac{\partial \bar{L}}{\partial z}(z,\nu) = \bar{x}(z,\nu) - z, \tag{12a}$$

$$\dot{\nu} = \left[\frac{\partial \bar{L}}{\partial \nu}(z,\nu)\right]^+_\nu = [g(\bar{x}(z,\nu))]^+_\nu . \tag{12b}$$

Note that these are constrained to $\mathbb{R}^n \times \mathbb{R}^m_+$ due to the positive projection in the last equation, designed to keep the multiplier $\nu \geq 0$. It is easy to see that the equilibrium points of (12) are exactly the saddle points of $\bar{L}$. We will assume the existence of at least one such saddle point.

While the computation of $\bar{x}(z,\nu)$ itself involves an optimization problem, sometimes it has closed-form solution, and often it can be solved in a distributed manner across subsets of variables, each of small dimension. This primal-dual decomposition enables the implemention of the dynamics (12) by distributed control rules, as illustrated in the application example of Section IV.

A. Background: Projected dynamics, LaSalle invariance

The precise treatment of the discontinuous projection (12b) for saddle point dynamics was carried out in [6], based on the theory of projected dynamical systems [18]. We will apply the same method here, so we begin by stating the results that follow with essentially no modification from [6]. Later on we present our new contributions.

The first result refers to solutions in the Carathéodory sense: i.e., absolutely continuous functions $\gamma(t) = (z(t),\nu(t))$ such that (12) holds at the (almost all) points of differentiability.

Proposition 6: There exists a unique solution $\gamma(t)$ to (12) for each initial condition $(z_0,\nu_0) \in \mathbb{R}^n \times \mathbb{R}^m_+$, and it is defined on $[0,\infty)$. Also, if $(z_k,\nu_k) \to (z_0,\nu_0)$ as $k \to \infty$, then the solutions $\gamma_k(t)$, with initial condition (z_k,ν_k), converge uniformly over compact sets to $\gamma(t)$.

The theory of projected dynamical systems [18] provides this kind of result when the field inside the projection is *globally* Lipschitz. For locally Lipschitz saddle point dynamics, as the ones we are considering, the corresponding extension was provided in [6, Lemma 4.3]. Briefly: since the Euclidean distance to a saddle point is non-increasing under the dynamics (see below), one can define a modified, globally Lipschitz field that is identical to the one in (12) over a compact ball, and invoke results in [18].

The convergence of trajectories to a saddle point relies on the LaSalle invariance principle. We will use the following version from [6], restated below in our notation.

Proposition 7: Let $K \subset \mathbb{R}^n \times \mathbb{R}^m_+$ be a compact set, invariant under (12), and assume that the following hold.

(a) For each $(z,\nu) \in K$ there exists a unique solution with initial condition (z,ν), and its omega-limit is invariant.

(b) There exists $V : \mathbb{R}^n \times \mathbb{R}^m_+ \longrightarrow \mathbb{R}_+$, a continuously differentiable function such that
$$\dot{V}(z,\nu) := \nabla V(z,\nu) \left[\nabla \bar{L}(z,\nu)\right]^\intercal \leq 0 \quad \forall\ (z,\nu) \in K;$$
note that if $\gamma(t)$ is a solution of (12), then the derivative of $V(\gamma(t))$, over time, is $\dot{V}(\gamma(t))$ almost everywhere.

Then any solution of (12), with initial condition in K, converges to the largest invariant subset of
$$\mathrm{cl}\left\{(z,\nu) \in K : \dot{V}(z,\nu) = 0\right\}.$$

Remark 3: Note that (a) holds by Proposition 6; the last statement follows as in [16, Lemma 4.1] from the continuity of solutions with respect to initial conditions.

B. Lyapunov function and negative drift

We introduce now, as is standard in this literature since [1], the Lyapunov function $V : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}_+$ given by

$$V(z,\nu) = \frac{1}{2}\left|\left|z - \hat{z}\right|\right|^2 + \frac{1}{2}\left|\left|\nu - \hat{\nu}\right|\right|^2, \tag{13}$$

where $(\hat{z},\hat{\nu})$ is a given saddle point. The next fact is standard, we present a proof for completeness and for future reference.

Proposition 8: Fix $(z,\nu) \in \mathbb{R}^m \times \mathbb{R}^n_+$ and let $p = \bar{x}(z,\nu)$.

$$\begin{aligned} \dot{V}(z,\nu) \leq{}& \bar{L}(\hat{z},\nu) - \bar{L}(\hat{z},\hat{\nu}) + \bar{L}(\hat{z},\hat{\nu}) - \bar{L}(z,\hat{\nu}) \\ &+ (\nu - \hat{\nu})^\intercal \left([g(p)]^+_\nu - g(p)\right) \leq 0. \end{aligned} \tag{14}$$

Proof: We compute:

$$\begin{aligned} \dot{V}(z,\nu) =&(\hat{z} - z)^\intercal \frac{\partial \bar{L}}{\partial z} + (\nu - \hat{\nu})^\intercal [g(p)]^+_\nu \\ =&\ (\hat{z} - z)^\intercal \frac{\partial \bar{L}}{\partial z} + (\nu - \hat{\nu})^\intercal \frac{\partial \bar{L}}{\partial \nu} \\ &+ (\nu - \hat{\nu})^\intercal \left([g(p)]^+_\nu - g(p)\right). \end{aligned}$$

Here the derivatives are evaluated at (z,ν). Using first order convexity and concavity inequalities for $\bar{L}$ we conclude that

$$\begin{aligned} \dot{V}(z,\nu) \leq&\left[\bar{L}(\hat{z},\nu) - \bar{L}(z,\nu)\right] + \left[\bar{L}(z,\nu) - \bar{L}(z,\hat{\nu})\right] \\ &+ (\nu - \hat{\nu})^\intercal \left([g(p)]^+_\nu - g(p)\right) \\ =&\ \left[\bar{L}(\hat{z},\nu) - \bar{L}(\hat{z},\hat{\nu})\right] + \left[\bar{L}(\hat{z},\hat{\nu}) - \bar{L}(z,\hat{\nu})\right] \\ &+ (\nu - \hat{\nu})^\intercal \left([g(p)]^+_\nu - g(p)\right). \end{aligned} \tag{15}$$

This proves the first inequality in (14). Moreover, the saddle point condition and Lemma 4 imply each of the three terms isolated in (15) are non-positive. ■

C. Invariant sets and convergence

The classical convergence theory for the saddle point dynamics requires *strict* convexity or concavity of the saddle function, to apply the LaSalle invariance theory; a weakening of these conditions to local ones was obtained in [7].

The main result of this section is that, under the general conditions of the original convex program (not smooth, or strictly convex), the proposed regularization yields globally convergent saddle point dynamics. The following example demonstrates that we cannot obtain the result from the standard convergence theory.

Example 1: Consider the very simple, scalar linear program $\min x$, subject to $0 \leq x \leq 1$. The standard Lagrangian is

$$L_0(x,\nu) = x - \nu_1 x + \nu_2(x-1) = x(1 - \nu_1 + \nu_2) - \nu_2,$$

where we have introduced the multipliers $\nu = (\nu_1, \nu_2)$. Applying the saddle point dynamics to this bilinear function would give an oscillatory behavior (see [13], [15]).

Adding a regularization as in (6), gives

$$L(x, z, \nu) = x(1 - \nu_1 + \nu_2) - \nu_2 + \frac{1}{2} ||x - z||^2 ;$$

minimizing over x we obtain $\bar{x}(z, \nu) = z + \nu_1 - \nu_2 - 1$, and the reduced Lagrangian

$$\bar{L}(z, \nu) = z(1 - \nu_1 + \nu_2) - \frac{1}{2}(\nu_1 - \nu_2 - 1)^2 - \nu_2.$$

This function is linear in z, so we have not gained strict convexity in the primal variable. And, although there is a negative quadratic term in ν, neither do we have strict concavity: the function is linear along the line $\nu_1 = \nu_2 + 1$.

In the absence of strict convexity or concavity, we will extend the LaSalle invariance argument by a careful consideration of the proximal regularization. In particular, recalling that the mapping $z \mapsto \bar{x}(z, \nu_0)$ is the proximal mapping associated with $L_0(\,\cdot\,, \nu_0) = f(\,\cdot\,) + \nu_0^T g(\,\cdot\,)$, we will use the fact [2] that any fixed point of $\bar{x}(\,\cdot\,, \nu_0)$ is a minimum of $L_0(\,\cdot\,, \nu_0)$.

The key step is a characterization of the set

$$E := \left\{ (z, \nu) \in \mathbb{R}^n \times \mathbb{R}^m_+ : \dot{V}(z, \nu) = 0 \right\}. \tag{16}$$

Theorem 9: If $\dot{V}(z, \nu) = 0$ and $\nu \in \mathbb{R}^m_+$, then:

$$\bar{x}(z, \nu) = z, \tag{17}$$

$$\bar{L}(z, \nu) = \bar{L}(\hat{z}, \hat{\nu}). \tag{18}$$

Proof: Let $(z, \nu) \in \mathbb{R}^n \times \mathbb{R}^m_+$ be such that $\dot{V}(z, \nu) = 0$. As before, we denote $p = \bar{x}(z, \nu)$. Recalling (14) and its proof, all three terms in (15) must be at equality:

$$\bar{L}(\hat{z}, \nu) = \bar{L}(\hat{z}, \hat{\nu}), \tag{19a}$$

$$\bar{L}(z, \hat{\nu}) = \bar{L}(\hat{z}, \hat{\nu}), \tag{19b}$$

$$(\nu - \hat{\nu})^\intercal \left[g(p)\right]^+_\nu = (\nu - \hat{\nu})^\intercal g(p). \tag{19c}$$

Claim I: $\bar{x}(z, \hat{\nu}) = z$.

Since $(\hat{z}, \hat{\nu})$ is a saddle point of $\bar{L}$, by Theorem 5 we have $\hat{z} = \bar{x}(\hat{z}, \hat{\nu})$, a fixed point of the proximal mapping associated with $h(z) := L_0(z, \hat{\nu})$. Hence, $h(z)$ is minimized at $\hat{z}$:

$$\bar{L}(\hat{z}, \hat{\nu}) = f(\hat{z}) + \hat{\nu}^T g(\hat{z}) \leq f(x) + \hat{\nu}^T g(x) \quad \forall\, x \in \mathbb{R}^n. \tag{20}$$

Let $q := \bar{x}(z, \hat{\nu})$. By definition and (19b) we have

$$\bar{L}(\hat{z}, \hat{\nu}) = \bar{L}(z, \hat{\nu}) = f(q) + \hat{\nu}^T g(q) + \frac{1}{2} ||q - z||^2 \geq \bar{L}(\hat{z}, \hat{\nu}),$$

where the last inequality follows from (20). Therefore we must have equality and $q = z$, as claimed.

Claim II: $\bar{x}(z, \hat{\nu}) = p$.

The first order condition for convexity of $\bar{L}(\,\cdot\,, \nu)$ yields

$$\begin{aligned} \bar{L}(\hat{z}, \nu) &\geq \bar{L}(z, \nu) + (\hat{z} - z)^\intercal \frac{\partial \bar{L}}{\partial z}(z, \nu) \\ &= \bar{L}(z, \nu) + (\hat{z} - z)^\intercal (z - p) \\ &= f(p) + \nu^\intercal g(p) + \frac{1}{2} ||p - z||^2 \\ &\quad + (z - \hat{z})^\intercal (p - z). \end{aligned} \tag{21}$$

From equation (19c) we conclude that

$$\begin{aligned} 0 = \dot{V}(z, \nu) &= (z - \hat{z})^\intercal (p - z) + (\nu - \hat{\nu})^\intercal \left[g(p)\right]^+_\nu \\ &= (z - \hat{z})^\intercal (p - z) + (\nu - \hat{\nu})^\intercal g(p). \end{aligned} \tag{22}$$

We may now substitute $(z - \hat{z})^\intercal (p - z)$ by $(\hat{\nu} - \nu)^\intercal g(p)$ in equation (21), to get

$$\bar{L}(\hat{z}, \nu) \geq f(p) + \hat{\nu}^\intercal g(p) + \frac{1}{2} ||p - z||^2 .$$

Now (19a)-(19b) imply that

$$\bar{L}(\hat{z}, \nu) = \bar{L}(z, \hat{\nu}) \leq f(p) + \hat{\nu}^\intercal g(p) + \frac{1}{2} ||p - z||^2 . \tag{23}$$

Therefore there is equality above, and p is the minimizer, so $p = \bar{x}(z, \hat{\nu})$ as claimed.

Claims I and II jointly imply $p = z$ (i.e., (17) holds). In addition, returning to (23) with $p = z$, we see that

$$\begin{aligned} \bar{L}(z, \hat{\nu}) &= f(z) + \hat{\nu}^\intercal g(z) \\ &= f(z) + \nu^\intercal g(z) = \bar{L}(z, \nu), \end{aligned}$$

where we have used $(\nu - \hat{\nu})^\intercal g(z) = 0$ from (22) for $p = z$. This establishes (18) due to (19b). ∎

Proposition 10: Let $\mathcal{I}$ be an invariant subset of cl$E$, closure of the set in (16). All elements of $\mathcal{I}$ are equilibria of (12).

Proof: By Theorem 9 we know that

$$E \subset F := \left\{ (z, \nu) \in \mathbb{R}^n \times \mathbb{R}^m_+ : \text{(17) and (18) hold} \right\}.$$

The two conditions in the definition of F are continuous by Proposition 1 and Theorem 2. Thus, F is a closed set, which implies that cl$E \subset F$, and therefore $\mathcal{I} \subset F$.

Fix $(z_0, \nu_0) \in \mathcal{I}$ and let $\gamma : [0, \infty) \longrightarrow \mathcal{I}$ denote the solution to (12) with initial condition (z_0, ν_0); we will use the notation $\gamma = (\gamma_z, \gamma_\nu)$. Since $\mathcal{I} \subset F$ we have:

(i) $\gamma_z \equiv \bar{x}(\gamma)$.
(ii) $\bar{L}(\gamma) \equiv \bar{L}(\hat{z}, \hat{\nu})$.

From (i) and (12a) we see that $\dot{\gamma}_z \equiv 0$, so $\gamma_z \equiv z_0$. Also,

$$\bar{L}(\hat{z}, \hat{\nu}) \equiv \bar{L}(\gamma) \equiv f(\gamma_z) + \gamma_\nu^\intercal g(\gamma_z) \equiv f(z_0) + \gamma_\nu^\intercal g(z_0)$$

follows, invoking (i) for the second equality. Taking the derivative with respect to t on both ends we get

$$0 \equiv g(z_0)^\intercal \dot{\gamma}_v = g(z_0)^\intercal \left[g(z_0)\right]^+_{\gamma_\nu} = \left|\left| \left[g(z_0)\right]^+_{\gamma_\nu} \right|\right|^2 = ||\dot{\gamma}_\nu||^2 .$$

Then $\dot{\gamma}_\nu \equiv 0$, so (z_0, ν_0) is an equilibrium of (12). ∎

We now state our main result, on convergence of the solutions of the saddle point gradient dynamics.

Theorem 11: Each trajectory of (12) converges to an equilibrium point, a saddle point of $\bar{L}$.

Proof: Fix an initial condition $(z_0, \nu_0) \in \mathbb{R}^n \times \mathbb{R}^m_+$ and the corresponding solution $\gamma(t)$.

To apply the LaSalle invariance principle, we first identify a suitable compact invariant set, invoking Proposition 8:

$$K = \left\{ (z, \nu) \in \mathbb{R}^n \times \mathbb{R}^m_+ : V(z, \nu) \leq V(z_0, \nu_0) \right\}.$$

Note that K satisfies all the requirements of Proposition 7. Then $\gamma(t)$ converges to M, *maximal* invariant subset of

$$\text{cl} \left\{ (z, \nu) \in K : \dot{V}(z, \nu) = 0 \right\}.$$

This set is itself contained in cl$E$; so by Proposition 10, M consists exclusively of equilibrium points of (12).

It remains to prove that $\gamma(t)$ converges to a *single point* in M. Let Ω be the omega-limit set of γ, it is contained in K and invariant; therefore by maximality we must have $\Omega \subset M$, in particular all points in Ω are equilibria of (12).

Suppose Ω contained two different points $a \neq b$, say at Euclidean distance $\delta > 0$. By definition of omega-limit, there exists t_0 such that the distance $d(\gamma(t_0), a) < \delta/2$. But Proposition 8 implies this distance is non-increasing, so $\gamma(t)$ cannot accumulate in the other point b, a contradiction. Thus Ω is a singleton, it consists of a single equilibrium of (12). ■

To summarize, we have shown that each solution of (12) converges to a saddle point of $\bar{L}$, which as shown in Section II provides an optimal solution to Problem (5), and also the original Problem (4). Given the weak convexity assumptions, such solution is in general not unique; it is noteworthy that the regularized dynamics always stabilizes one such solution.

IV. Application example

In our recent papers [14], [20] we considered a resource allocation problem in the area of cloud computing. It concerns the joint optimization of: capacity scaling in a set of computer clusters, load balancing between them, and task scheduling within each cluster. The following formulation is taken from the preceding references.

Problem 1: $\underset{s}{\text{minimize}}\ \sum_{j=1}^{N} c_j(s_j)$, subject to

$$\sum_{j=1}^{N} \lambda_{ij} = \lambda_i \quad \forall\ i = 1, \ldots, M; \tag{24a}$$

$$\sum_{i=1}^{M} \frac{r_{ij}}{\mu_{ij}} = s_j \quad \forall\ j = 1, \ldots, N; \tag{24b}$$

$$0 \leq \lambda_{ij} \leq r_{ij} \quad \forall\ i = 1, \ldots, M;\ j = 1, \ldots, N. \tag{24c}$$

Here the index i represents job *type*; λ_i is the (given) arrival rate of such jobs into the system. The index j represents the computer cluster, and μ_{ij} is the service rate it can provide, per server, to tasks of type i. The decision variables are:

- The matrix $\Lambda = (\lambda_{ij})$ of rates at which tasks of type i are dispatched to cluster j; it is subject to the constraints (24a), and its selection is the *load balancing* problem.
- The number of server instances s_j which are active in cluster j; its selection is the *right sizing* problem. An increasing, convex cost $c_j(\,\cdot\,)$ is applied, which may describe for instance the energy expenditure of maintaining this active capacity. Natural models could be a linear cost c_j, or a piecewise linear cost reflecting the existence of cheaper and more expensive server instances.
- The matrix $R = (r_{ij})$ of service rates that cluster j allocates to jobs of type i. This corresponds to an assignment of r_{ij}/μ_{ij} server instances; equation (24b) totals instances per cluster. The selection of R subject to this constraint is the *scheduling* problem.

Since the scale of this resource allocation problem is often large, rather than a centralized solution we seek a suitable decomposition. In [20] we used Lagrange duality to justify the decomposition between Join-the-Shortest-Queue load balancing and MaxWeight scheduling, a proposal from [28] in a queueing theory context; its combination with capacity right-sizing is also studied. To avoid the fast switching ("chattering") of solutions based on JSQ/MW, we investigated in [14] the use of proximal regularizations; results were stated for a smooth and strictly convex cost function.

The theory in the present paper allows us to fully justify the procedure in [14], as well as extend it to a more general cost function. We begin by reformulating the problem as follows:

$$\underset{\Lambda, R}{\text{minimize}} \sum_{j=1}^{N} c_j \left(\sum_{i=1}^{M} \frac{r_{ij}}{\mu_{ij}} \right) + \chi_{\Delta}(\Lambda) + \chi_{\mathbb{R}_+^{M \times N}}(R),$$
$$\text{subject to } \lambda_{ij} \leq r_{ij}\ \forall\ i, j.$$

Above, we have substituted (24b) into the cost function, and made the constraints (24a) and $\Lambda \geq 0$ implicit through the convex indicator function $\chi_{\Delta}(\,\cdot\,)$ corresponding to the set

$$\Delta := \left\{ \Lambda \in \mathbb{R}_+^{M \times N} : \sum_{j=1}^{N} \lambda_{ij} = \lambda_i,\ \forall\ i = 1, \ldots, M \right\}.$$

Similarly, we have the implicit constraint $R \geq 0$ through $\chi_{\mathbb{R}_+^{M \times N}}(\,\cdot\,)$. This optimization problem fits into the class of Problem (4): the objective is a closed, proper convex function of $x = (\Lambda, R)$, and the inequality constraints are in this case linear. Adding the regularization variable $z = (\alpha, \beta)$ and the multiplier ν we write the Lagrangian from (6):

$$L(x, z, \nu) = \sum_j c_j \left(\sum_i \frac{r_{ij}}{\mu_{ij}} \right) + \chi_{\Delta}(\Lambda) + \chi_{\mathbb{R}_+^{M \times N}}(R)$$
$$+ \sum_{i,j} \nu_{ij}(\lambda_{ij} - r_{ij}) + \frac{1}{2} \sum_{i,j} \left[(\lambda_{ij} - \alpha_{ij})^2 + (r_{ij} - \beta_{ij})^2 \right].$$

After completing squares, we arrive at the following expression for the reduced Lagrangian:

$$\bar{L}(\alpha, \beta, \nu) = \sum_{i,j} \left[(\alpha_{ij} - \beta_{ij})\nu_{ij} - \nu_{ij}^2 \right]$$
$$+ \min_{\Lambda \in \Delta} \left\{ \frac{1}{2} \sum_{i,j} (\alpha_{ij} - \nu_{ij} - \lambda_{ij})^2 \right\}$$
$$+ \min_{R \geq 0} \left\{ \sum_j c_j \left(\sum_i \frac{r_{ij}}{\mu_{ij}} \right) + \frac{1}{2} \sum_{i,j} (\beta_{ij} + \nu_{ij} - r_{ij})^2 \right\}.$$

The minimization in Λ amounts to a projection on the set Δ:

$$\bar{\Lambda}(\alpha, \nu) = \pi_{\Delta}(\alpha - \nu) = \underset{\Lambda \in \Delta}{\text{argmin}}\ \|\Lambda - (\alpha - \nu)\|. \tag{25}$$

Note that constraints decouple over the rows of Λ: for each job type i we have a projection on a simplex, which can be computed through finitely many operations (see [14]).

In turn, the minimization in R decouples over j. At each cluster we solve for the column $\bar{R}_j(\beta, \nu) \geq 0$ that minimizes

$$c_j \left(\sum_i \frac{r_{ij}}{\mu_{ij}} \right) + \frac{1}{2} \sum_i (\beta_{ij} + \nu_{ij} - r_{ij})^2. \tag{26}$$

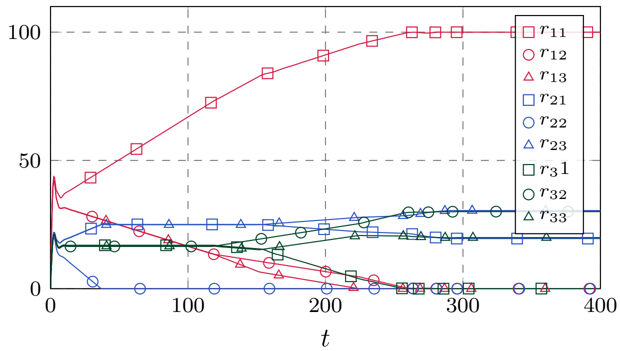

Fig. 1. Evolution of the scheduling $R = (r_{ij})$ over time for the system considered in Section IV-A; $r_{ij}(t)$ converges to an equilibrium for each i, j.

If, for instance, the cost is piecewise linear, the above is a quadratic program.

The regularization terms yield a locally Lipschitz mapping $\bar{x}(z,\nu) = (\bar{\Lambda}(\alpha,\nu), \bar{R}(\beta,\nu))$, as follows from Proposition 1. The saddle point gradient dynamics of (12) become:

$$\dot{\alpha}_{ij} = \bar{\lambda}_{ij}(\alpha,\nu) - \alpha_{ij}, \tag{27}$$

$$\dot{\beta}_{ij} = \bar{r}_{ij}(\beta,\nu) - \beta_{ij}, \tag{28}$$

$$\dot{\nu}_{ij} = \left[\bar{\lambda}_{ij}(\alpha,\nu) - \bar{r}_{ij}(\beta,\nu)\right]^{+}_{\nu_{ij}}. \tag{29}$$

We remark that (29) constitutes a fluid model of a *queue*: ν_{ij} is the queue of type i jobs at cluster j. With this identification:

- (26) and (28) provide the control of scheduling, decentralized over cluster j, as a function of local queues.
- (25) and (27) provide a recipe for load balancing, decoupled over job type i. Its implementation requires feeding back the queue information to the load balancer.

These are the same information requirements used by the Max Weight scheduler and Join-the-Shortest-Queue load balancer, studied in [20], [28]. However, in contrast to that case in which Λ and R switch arbiratrily fast around an equilibrium with tied queues, here we obtain a smooth evolution in time, guided by the regularization variables α and β.

A. Numerical example

Let us consider a system with $M = N = 3$,

$$\lambda = \begin{bmatrix} 100 \\ 50 \\ 50 \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix}.$$

Suppose that the costs c_j are piecewise linear: specifically, $c_j(s) = s$ for $s \leq 50$ and $c_j(s) = 2(s-50)+50$ for $s > 50$, for all $j = 1, \ldots, M$. Following the procedure explained above we compute $\Lambda(t), R(t), \alpha(t), \beta(t)$ and $\nu(t)$ numerically. We used Matlab's `ode45` function to discretize the dynamics.

Due to space limitations we only show $R(t)$ in Fig. 1. We see that $r_{ij}(t)$ converges to an equilibrium for each i and j, where the cluster sizes are $s_1^* \cong 57$, $s_2^* \cong 30$ and $s_3^* \cong 30$.

V. Conclusion

This paper proposes a solution method for a general convex optimization problem: the objective is closed, proper, convex (not necessarily smooth), taking values in $\overline{\mathbb{R}}$ and possibly enconding implicit constraints. The explicit constraints are convex and real-valued; both equality and inequality constraints are allowed. We show that a regularized Lagrangian obtained after a proximal minimization step, is smooth and leads to a globally convergent saddle point dynamics; its equilibrium points are optima of the original problem.

We have used the standard discontinuous projections to enforce sign restrictions on multipliers; an open question for future work is whether smoother alternatives as those in [22] could extend to general convex constraints. Another question would be to identify minimal conditions under which the regularized Lagrangian yields exponential stability.

Appendix
Proof of Proposition 1

Proof: Fix $(z_0, \nu_0) \in \mathbb{R}^n \times \mathbb{R}^m$ and let $\bar{x}_0 := \bar{x}(z_0, \nu_0)$; recall it is the unique minimizer of

$$L(x, z_0, \nu_0) = f(x) + \frac{1}{2}\left|\left|x - z_0\right|\right|^2 + \nu_0^{\intercal} g(x)$$

over $x \in \mathbb{R}^n$. Since this function is strongly convex in x, then, for all $x \in \mathbb{R}^n$, we have

$$L(x, z_0, \nu_0) \geq \underbrace{L\left(\bar{x}_0, z_0, \nu_0\right)}_{\bar{L}(z_0,\nu_0)} + \frac{1}{2}\left|\left|x - \bar{x}_0\right|\right|^2. \tag{30}$$

Claim I: $\bar{x}(z,\nu)$ is continuous at (z_0, ν_0). Namely, given some $\varepsilon > 0$, we will find $\delta > 0$ such that:

$$||z_1 - z_0|| < \delta, \ ||\nu_1 - \nu_0|| < \delta \implies ||\bar{x}(z_1, \nu_1) - \bar{x}_0|| < \varepsilon.$$

To prove this, let γ be an upper bound of the continuous function $||g(x)||$ in the ball $B(\bar{x}_0, \varepsilon)$, and write

$$\begin{aligned} L\left(\bar{x}_0, z_0, \nu_1\right) &= L\left(\bar{x}_0, z_0, \nu_0\right) + (\nu_1 - \nu_0)^{\intercal} g\left(\bar{x}_0\right) \\ &\leq \bar{L}(z_0, \nu_0) + \gamma\left|\left|\nu_1 - \nu_0\right|\right|. \end{aligned} \tag{31}$$

Also, for all x such that $||x - \bar{x}_0|| < \varepsilon$ we have:

$$\begin{aligned} L(x, z_0, \nu_1) &= L(x, z_0, \nu_0) + (\nu_1 - \nu_0)^{\intercal} g(x) \\ &\geq \bar{L}(z_0, \nu_0) + \frac{1}{2}\left|\left|x - \bar{x}_0\right|\right|^2 - \gamma\left|\left|\nu_1 - \nu_0\right|\right| \\ &\geq L\left(\bar{x}_0, z_0, \nu_1\right) + \frac{1}{2}\left|\left|x - \bar{x}_0\right|\right|^2 - 2\gamma\left|\left|\nu_1 - \nu_0\right|\right|. \end{aligned}$$

The last step uses (31). Now choose $\delta \in (0, \varepsilon/2)$ such that $2\gamma\delta < \varepsilon^2/8$. Let ν_1 satisfy $||\nu_1 - \nu_0|| < \delta$. We conclude that

$$L(x, z_0, \nu_1) > L(\bar{x}_0, z_0, \nu_1)$$

for any x satisfying $||x - \bar{x}_0|| = \varepsilon/2$, and the same inequality must hold for $||x - \bar{x}_0|| \geq \varepsilon/2$. Indeed, since the sublevel set $\{x : L(x, z_0, \nu_1) \leq L(\bar{x}_0, z_0, \nu_1)\}$ is convex and contains $\bar{x}_0$, it cannot contain points outside $B(\bar{x}_0, \varepsilon/2)$ without intersecting the boundary of this ball. We conclude that the minimizer $\bar{x}(z_0, \nu_1)$ of $x \mapsto L(x, z_0, \nu_1)$ lies inside $B(\bar{x}_0, \varepsilon/2)$.

The proof of the claim now follows from the non-expansiveness of the proximal mapping $z \mapsto \bar{x}(z, \nu_1)$: for all (z_1, ν_1) such that $||z_1 - z_0|| < \delta$, $||\nu_1 - \nu_0|| < \delta$ we have

$$\begin{aligned} ||\bar{x}(z_1, \nu_1) - \bar{x}_0|| &\leq ||\bar{x}(z_1, \nu_1) - \bar{x}(z_0, \nu_1)|| \\ &\quad + ||\bar{x}(z_0, \nu_1) - \bar{x}_0|| < ||z_0 - z_1|| + \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

Claim II: There exist $\rho > 0$ and $K \geq 1$ such that

$$||z_1 - z_0|| < \rho,\ ||\nu_1 - \nu_0|| < \rho,\ ||\nu_2 - \nu_0|| < \rho \qquad (32)$$
$$\implies ||\bar{x}(z_1, \nu_2) - \bar{x}(z_1, \nu_1)|| \leq K\, ||\nu_2 - \nu_1||\,.$$

We define these constants as follows. First, since g has finite convex components, it is locally Lipschitz [23]; let $M > 0$ be the Lipschitz constant in $B(\bar{x}_0, 2\varepsilon)$ for some $\varepsilon > 0$. Introduce: $\delta > 0$ from Claim I for this ε,

$$K > \max\{2M, 1\} \quad \text{and} \quad \rho < \min\left\{\frac{\varepsilon}{2K}, \delta\right\}.$$

To establish the claim we begin with the identities

$$L(x, z_1, \nu_2) = L(x, z_1, \nu_1) + (\nu_2 - \nu_1)^T g(x);$$
$$L(\bar{x}_1, z_1, \nu_2) = \bar{L}(z_1, \nu_1) + (\nu_2 - \nu_1)^T g(\bar{x}_1),$$

with the notation $\bar{x}_1 := \bar{x}(z_1, \nu_1)$. Note that if (z_1, ν_1) are as in (32) with $\rho < \delta$, then $||\bar{x}_1 - \bar{x}_0|| < \varepsilon$ from Claim I. Subtracting the two equations and using the bound analogous to (30) we obtain, for any $x \in B(\bar{x}_0, 2\varepsilon)$, the bound:

$$\begin{aligned} L(x, z_1, &\nu_2) - L(\bar{x}_1, z_1, \nu_2) \\ &\geq \frac{1}{2}||x - \bar{x}_1||^2 + (\nu_2 - \nu_1)^\intercal [g(x) - g(\bar{x}_1)] \\ &\geq \frac{1}{2}||x - \bar{x}_1||^2 + \|\nu_2 - \nu_1\| M \|x - \bar{x}_1\|. \end{aligned}$$

Any point in $B(\bar{x}_1, \varepsilon)$ is in $B(\bar{x}_0, 2\varepsilon)$, in particular, this is the case for any x satisfying

$$||x - \bar{x}_1|| = K\, ||\nu_2 - \nu_1|| < 2K\rho < \varepsilon.$$

Applying the preceding bound to such a point yields

$$\begin{aligned} L(x, z_1, \nu_2) &- L(\bar{x}_1, z_1, \nu_2) \\ &\geq \left[\frac{K^2}{2} - MK\right] ||\nu_2 - \nu_1||^2 > 0, \end{aligned}$$

since $K > 2M$; we are assuming $\nu_1 \neq \nu_2$, otherwise the claim is trivial. This implies that the minimizer of $L(x, z_1, \nu_2)$ cannot lie in the boundary of $B(\bar{x}_1, K\|\nu_1 - \nu_2\|)$; reasoning as in Claim I it cannot lie outside the ball either, so we have $||\bar{x}(z_1, \nu_2) - \bar{x}_1|| \leq K\, ||\nu_2 - \nu_1||$, which proves Claim II.

To complete the proof of the proposition we take ρ and K as in Claim II. For $i = 1, 2$ let (z_1, ν_1) and (z_2, ν_2) be such that $||z_i - z_0|| < \rho$ and $||\nu_i - \nu_0|| < \rho$. We have:

$$\begin{aligned} ||\bar{x}(z_2, \nu_2) - \bar{x}(z_1, \nu_1)|| &\leq ||\bar{x}(z_2, \nu_2) - \bar{x}(z_1, \nu_2)|| \\ &\quad + ||\bar{x}(z_1, \nu_2) - \bar{x}(z_1, \nu_1)|| \\ &\leq ||z_2 - z_1|| + K\, ||\nu_2 - \nu_1|| \\ &\leq K\sqrt{2}\, ||(z_2, \nu_2) - (z_1, \nu_1)||\,, \end{aligned}$$

which invokes the non-expansiveness of the proximal mapping, the Euclidean norm and the fact that $K \geq 1$. ■

References

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

[2] H. H. Bauschke, P. L. Combettes *et al.*, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011, vol. 408.

[3] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[4] C.-Y. Chang, M. Colombino, J. Cortés, and E. Dall Anese, "Saddle-flow dynamics for distributed feedback-based optimization," *IEEE Control Systems Letters*, vol. 3, no. 4, pp. 948–953, 2019.

[5] G. Chen and M. Teboulle, "A proximal-based decomposition method for convex minimization problems," *Mathematical Programming*, vol. 64, no. 1-3, pp. 81–101, 1994.

[6] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *Systems & Control Letters*, vol. 87, pp. 10–15, 2016.

[7] A. Cherukuri, E. Mallada, S. 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, 2017.

[8] M. Colombino, E. Dall Anese, and A. Bernstein, "Online optimization as a feedback controller: Stability and tracking," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 1, pp. 422–432, 2019.

[9] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *Journal of Nonlinear Science*, vol. 29, no. 4, pp. 1247–1272, 2019.

[10] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented lagrangian method for nonsmooth composite optimization," *IEEE Trans. on Automatic Control*, vol. 64, no. 7, pp. 2861–2868, 2018.

[11] D. Ding and M. R. Jovanović, "Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented lagrangian: A lyapunov-based approach," *arXiv preprint arXiv:1910.00783*, 2019.

[12] K. Ebrahimi, N. Elia, and U. Vaidya, "A continuous time dynamical system approach for solving robust optimization," in *2019 18th European Control Conference (ECC)*. IEEE, 2019, pp. 1479–1485.

[13] D. Feijer and F. Paganini, "Stability of primal–dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.

[14] D. Goldsztajn, F. Paganini, and A. Ferragut, "Proximal optimization for resource allocation in distributed computing systems with data locality," in *57th Annual Allerton Conference on Communication, Control, and Computing*, 9 2019, pp. 773–780.

[15] T. Holding and I. Lestas, "On the convergence to saddle points of concave-convex functions, the gradient method and emergence of oscillations," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 1143–1148.

[16] H. K. Khalil, "Nonlinear systems," *Upper Saddle River*, 2002.

[17] X. Lin and N. B. Shroff, "An optimization-based approach for qos routing in high-bandwidth networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, no. 6, pp. 1348–1361, 2006.

[18] A. Nagurney and D. Zhang, *Projected dynamical systems and variational inequalities with applications*. Springer Science & Business Media, 2012, vol. 2.

[19] S. K. Niederländer, F. Allgöwer, and J. Cortés, "Exponentially fast distributed coordination for nonsmooth convex optimization," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 1036–1041.

[20] F. Paganini, D. Goldsztajn, and A. Ferragut, "An optimization approach to load balancing, scheduling and right sizing of cloud computing systems with data locality," in *IEEE Conf. on Decision & Control*, 2019.

[21] N. Parikh, S. Boyd *et al.*, "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[22] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 43–48, 2018.

[23] R. T. Rockafellar, *Convex analysis*. Princeton university press, 1970, vol. 28.

[24] J. W. Simpson-Porco, "Input/output analysis of primal-dual gradient algorithms," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2016, pp. 219–224.

[25] R. Srikant and L. Ying, *Communication networks: an optimization, control, and stochastic networks perspective*. Cambridge University Press, 2013.

[26] M. Vaquero and J. Cortés, "Distributed augmentation-regularization for robust online convex optimization," *IFAC-PapersOnLine*, vol. 51, no. 23, pp. 230–235, 2018.

[27] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *50th IEEE conference on decision and control*. IEEE, 2011, pp. 3800–3805.

[28] W. Wang, K. Zhu, L. Ying, J. Tan, and L. Zhang, "Maptask scheduling in mapreduce with data locality: Throughput and heavy-traffic optimality," *IEEE/ACM Trans. Networking*, vol. 24(1), pp. 190–203, 2016.

[29] J. T. Wen and M. Arcak, "A unifying passivity framework for network flow control," *IEEE Trans. on Automatic Control*, vol. 49, no. 2, pp. 162–174, 2004.