Limit Behavior and the Role of Augmentation in Projected Saddle Flows for Convex Optimization

Adrian Hauswirth[1], Lukas Ortmann[1], Saverio Bolognani, Florian Dörfler

*Department of Information Technology and Electrical Engineering,
ETH Zurich, 8092 Zurich, Switzerland
Email:*{hadrian,ortmannl,bsaverio,dorfler}@ethz.ch

Abstract: In this paper, we study the stability and convergence of continuous-time Lagrangian saddle flows to solutions of a convex constrained optimization problem. Convergence of these flows is well-known when the underlying saddle function is either strictly convex in the primal or strictly concave in the dual variables. In this paper, we show convergence under non-strict convexity when a simple, unilateral augmentation term is added. For this purpose, we establish a novel, non-trivial characterization of the limit set of saddle-flow trajectories that allows us to preclude limit cycles. With our presentation we try to unify several existing problem formulations as a projected dynamical system that allows projection of both the primal and dual variables, thus complementing results available in the recent literature.

Keywords: Convex optimization, dynamical systems.

1. INTRODUCTION

Saddle flows have historically been a core element in many different domains, from circuit theory (Smale, 1972; Brayton and Moser, 1964) to port-Hamiltonian systems (van der Schaft et al., 2014), and their study dates back to the seminal works by Arrow et al. (1958); Kose (1956); Venets (1985).

Currently, saddle flows are also used in the context of *feedback* (or *autonomous*) optimization, with applications in power systems, network optimizations, etc. (see Colombino et al. 2019; Dall'Anese and Andrea Simonetto 2018; Hauswirth et al. 2020; Feijer and Paganini 2010). The goal of feedback optimization is to steer a physical system to an optimal operation point while satisfying operational constraints. Some of these need to be satisfied by the system trajectory at all times, while others need to be satisfied at steady-state. Projections on the primal variables are needed to enforce trajectory constraints. On the other hand, steady-state constraints often come as inequality constraints that can be dualized. This leads to dual variables which have to be positive and therefore projections on the dual variables of the saddle flow are needed.

In this paper, we analyze the convergence and stability of projected saddle flows as they are encountered in feedback optimization. By considering projections on both the primal and the dual variables, we capture effects of physical saturation and discontinuity in the optimization strategy in a single mathematical model.

The contributions of this paper are as follows: On the technical side, we expand on existing work and characterize the zero-dissipation set of a saddle flow. This allows us to characterize the emergence of limit cycles under non-

strict convexity and show that these can be avoided with an augmentation term that does not affect the equilibria of the saddle flow. From a conceptual perspective, this shows how saddle flows can solve pure *feasibility problems*. On the educational side, we strive for a concise and accessible presentation and derive results that subsume other recent results for convex problems. In addition, we also hint at extensions of projected saddle flows on non-convex domains.

Similar to related work, our stability and convergence results are based on elementary methods such as basic convex analysis and a LaSalle invariance argument applied to a squared distance function. Beyond that, our new characterization of the zero-dissipation set requires additional facts about convexity and careful logical reasoning. The only advanced technical aspect is the use of projected dynamical systems to model the discontinuous dynamics that arise in order to enforce constraints on primal and dual variables. In contrast to other frameworks that have been exploited for this purpose, such as monotone mappings (Goebel, 2017), complementarity systems (Stegink et al., 2018), or subgradient formulations, existence and uniqueness of trajectories for projected dynamical systems is guaranteed for non-convex domains and even on abstract manifolds (Hauswirth et al., 2018). We argue that this makes them a natural choice for future research on projected saddle flows that ventures beyond convexity.

Our paper differs from related work as follows: Cherukuri et al. (2016) and Cherukuri et al. (2018) do not have primal projections. Cortés and Niederländer (2019) and (Stegink et al., 2018) do not have dual projections. Dhingra et al. (2018) works with non-convex and non-differentiable cost functions, but does not have any projections. These papers therefore do not analyze the same problem. In Tang et al. (2018), the Lagrangian is regularized with an *a priori* guess of the optimal dual variable. This makes the saddle flow strictly concave in the dual variable. This improves the convergence, but the resulting equilibrium is not the

solution of the original optimization problem. In contrast to Goebel (2017) we assume that all relevant problem components are differentiable. This allows us to characterize the limit behavior in the case where the saddle-function is not strictly concave/convex in the dual nor the primal variable. This way, we are able to extend results in Goebel (2017) and Cherukuri et al. (2015). The assumption that all relevant problem components are differentiable is motivated by the fact that feedback optimization controllers based on subgradient formulations are not easily implementable (Hauswirth et al., 2020). The saddle flow limit behavior that we identify corresponds to what has also been predicted in Holding and Lestas (2014), where different methods have been used in the proof. The proposed augmentation of the Lagrangian to ensure convergence is also different.

The rest of the paper is organized as follows. In Section 2 we fix the notation and review the basic properties of projected dynamical systems. In Section 3 we provide the formal problem description. Sections 4 and 5 discuss stability and convergence under strict convexity, and characterize the zero-dissipation set, respectively. In Section 6 we present the limit behavior under non-strict convexity and in Section 7 we give an example. Finally, we summarize our results and discuss open problems in Section 8.

2. PRELIMINARIES

2.1 Notation

We consider $\mathbb{R}^n$ with the usual inner product $\langle \cdot, \cdot \rangle$ and 2-norm $\| \cdot \|$. Hence, a linear map is given by a matrix $\mathbb{R}^{n \times m}$. We denote the *closure* of a set $\mathcal{X} \subseteq \mathbb{R}^n$ by $\operatorname{cl} \mathcal{X}$. Given a differentiable map $G : \mathbb{R}^n \to \mathbb{R}^m$ the *Jacobian of G at x* is the $n \times m$-matrix of partial derivatives denoted by $\nabla G(x)$. Namely, if $G : \mathbb{R}^n \to \mathbb{R}$ the *gradient of G at x* is $\nabla G(x)$. The partial Jacobian with respect to an argument or variable y is denoted by $\nabla_y G$. For a vector $v \in \mathbb{R}^n$ and an index set $I \subseteq \{1, \dots, n\}$ we denote by v_I the vector obtained by stacking the I-th components of v. Similarly for a matrix A, A_I denotes the matrix made up of the I-th columns of A. The *Lie derivative of G* along vector field $f : \mathbb{R}^n \to \mathbb{R}^n$ at x, denoted by $\mathcal{L}_f G(x)$ is the directional derivative of G at x in the direction of f. That is, $\mathcal{L}_f G(x) := \nabla G(x)^T f(x)$. By $\mathbb{R}^m_{\geq 0}$ we denote vectors of size m with positive entries including 0.

2.2 Projected Dynamical Systems

We quickly recall the definition and properties of projected dynamical systems on convex domains.

Definition 1. Given a closed convex set $\mathcal{X} \subseteq \mathbb{R}^n$, the *tangent* and *normal cone* at $x \in \mathcal{X}$ are respectively defined as

$$T_x \mathcal{X} := \operatorname{cl}\{v \in \mathbb{R}^n \mid \exists \lambda > 0 : x + \lambda v \in \mathcal{X}\} \quad (1)$$

$$N_x \mathcal{X} := \{\eta \in \mathbb{R}^n \mid \forall v \in T_x \mathcal{X} : \langle v, \eta \rangle \leq 0\} . \quad (2)$$

The cones $T_x \mathcal{X}$ and $N_x \mathcal{X}$ are both closed, convex, and they are polar to each other.

Remark 1. The tangent cone takes an explicit form for sets $\mathcal{X} := \{x \mid \zeta(x) \leq 0\}$ where $\zeta : \mathbb{R}^n \to \mathbb{R}^p$ and where $\mathcal{X}$ satisfies constraint qualifications (Rockafellar and Wets, 1998, Thm 6.14). Namely, in this case we have

$$T_x \mathcal{X} := \{v \in \mathbb{R}^n \mid \nabla \zeta_{I(x)}(x)^T v \leq 0\} ,$$

where $I(x) := \{i \mid \zeta_i(x) = 0\}$ denotes the set of active constraints.

We define a differential projection operator for a closed convex set $\mathcal{X} \subseteq \mathbb{R}^n$, $x \in \mathcal{X}$, and $v \in \mathbb{R}^n$ as

$$[v]_{\mathcal{X}}^x := \arg \min_{w \in T_x \mathcal{X}} \|v - w\|^2 ,$$

that is, $[v]_{\mathcal{X}}^x$ projects a vector v onto the tangent cone of $\mathcal{X}$ at the point x. Since $T_x \mathcal{X}$ is a closed convex set for any $x \in \mathcal{X}$, the minimum norm projection of v on $T_x \mathcal{X}$ exists and is unique, and $[v]_{\mathcal{X}}^x$ is well-defined. Furthermore, it holds that $\epsilon [v]_{\mathcal{X}}^x = [\epsilon v]_{\mathcal{X}}^x$ for all $\epsilon > 0$ since $T_x \mathcal{X}$ is a cone. Further, the following property is a consequence of Moreau's Theorem (e.g., (Rockafellar and Wets, 1998, Ex 12.22)) and has been exploited for projected dynamical systems in Cornet (1983); Aubin and Cellina (1984); Heemels et al. (2000); Hauswirth et al. (2018) and others.

Lemma 1. For a closed convex set $\mathcal{X} \subseteq \mathbb{R}^n$, $x \in \mathcal{X}$, and $v \in \mathbb{R}^n$, it holds that $v - [v]_{\mathcal{X}}^x \in N_x \mathcal{X}$.

For a continuous vector field $F : \mathcal{X} \to \mathbb{R}^n$, we define the projected dynamical system by applying the projection operator to the vector field F at every point. This leads to the initial value problem

$$\dot{x} = [F(x)]_{\mathcal{X}}^x , \quad x(0) = x_0 , \quad (3)$$

where $x_0 \in \mathcal{X}$ denotes an initial condition.

In general, $[F(x)]_{\mathcal{X}}^x$ is not continuous and standard existence results for ordinary differential equations do not apply. Instead, a (complete) *(Carathéodory) solution* to (3) is defined as an absolutely continuous function $x : [0, \infty) \to \mathcal{X}$ and $x(0) = x_0$, and for which $\dot{x}(t) = [F(x(t))]_{\mathcal{X}}^x$ holds almost everywhere, i.e., for almost all $t \in [0, \infty)$. Note that a solution to (3) has to be *viable*, i.e., remain in $\mathcal{X}$ for all $t \in [0, \infty)$ by definition.

Historically, the earliest existence results for projected dynamical systems appear to date back to Henry (1973) (for convex domains) and Cornet (1983) (non-convex, but tangentially regular sets). Projected dynamical systems also appear as special cases in the study of differential inclusions (Aubin and Cellina, 1984) and viability theory (Aubin, 1991). For more recent works that consider projected dynamical systems on manifolds and Banach spaces see Hauswirth et al. (2018); Daniele et al. (2010) and references therein. In contrast, the independent line of research in Nagurney and Zhang (1996), popular in the context of variational inequalities, uses a different approach which cannot recover the generality of the earlier works. [3]

For our purpose, we state the following existence and invariance result that can be recovered as a special case and combination of more general results in Aubin and Cellina (1984) and Hauswirth et al. (2018) that do not require $\mathcal{X}$ to be convex and F Lipschitz:

Theorem 1. Consider (3). Let $\mathcal{X}$ be closed convex, $F(x)$ locally Lipschitz, $\Psi : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable, and $\mathcal{S}_\ell := \{x \in \mathcal{X} \mid \Psi(x) \leq \ell\}$ compact for all ℓ. If $\mathcal{L}_{(3)} \Psi(x) \leq 0$ for all $x \in \mathcal{X}$, then there exists a unique complete solution of (3) for every $x_0 \in \mathcal{S}_\ell$ which converges to the largest invariant subset of $\operatorname{cl}\{x \in \mathcal{S}_\ell \mid \mathcal{L}_{(3)} \Psi(x) = 0\}$.

Note that Ψ needs to have compact sublevel sets with respect to $\mathcal{X}$, which is satisfied if $\mathcal{X}$ is itself compact.

Remark 2. Theorem 1 combines several fundamental results: First, existence of local solutions is guaranteed by

---

[3] In particular, the often-cited existence result (Nagurney and Zhang, 1996, Thm 2.5) requires (Nagurney and Zhang, 1996, Thm 2.7) which, strictly speaking, only applies to convex polyhedra.

standard viability results (e.g. Aubin 1991, Thm 3.3.4). For an invariance principle such as (Bacciotti and Cera-gioli, 1999, Thm 3) to be applicable, *continuous dependence on initial conditions* is required. This is the case because (3) is equivalent to *well-posed* differential inclusion (see Filippov 1988, Thm 3, §8 and Hauswirth et al. 2018, Thm 6.3). Finally, invariance of the compact sublevel sets of Ψ guarantees the existence of complete solutions.

3. PROBLEM FORMULATION

For the remainder of the paper, we will refer to the convex optimization problem given by

$$\underset{x}{\text{minimize}} \quad f(x) \tag{4a}$$

$$\text{subject to} \quad g(x) \leq 0 \tag{4b}$$

$$x \in \mathcal{X} . \tag{4c}$$

Here, we make the following assumption.

Assumption 1. The functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are convex and differentiable, and $\mathcal{X} \subseteq \mathbb{R}^n$ is closed and convex.

In our context, $\mathcal{X}$ defines a set of constraints that have to be satisfied at all times during the evolution of the dynamical system. This is important when a physical system is actuated in closed-loop with the saddle-flow algorithm. In comparison, $g(x) \leq 0$ only has to be satisfied at steady-state.

Assumption 2. (Well-posedness). A (bounded) optimizer to (4) exists and Slater's condition holds, i.e., the feasible set $\{x \in \mathcal{X} \mid g(x) \leq 0\}$ has a non-empty relative interior.

Assumption 2 is standard to guarantee the applicability of the KKT conditions to certify global optimality (Beck, 2014):

Proposition 1. Under Assumptions 1 and 2, a feasible point $x^\star$ is a global optimizer of (4) if and only if there exists $\mu^\star \in \mathbb{R}_{\geq 0}^m$ such that

$$\nabla f(x^\star) + \nabla g(x^\star)\mu^\star \in -N_{x^\star}\mathcal{X} \tag{5}$$

and for all $i = 1, \ldots, m$

$$\mu_i^\star > 0 \quad \Rightarrow \quad g_i(x^\star) = 0 . \tag{6}$$

Throughout the paper we refer to a primal-dual pair $(x^\star, \mu^\star)$ that satisfies the conditions of Proposition 1 as a *solution* of (4), whereas an *optimizer* of (4) only refers to the primal point $x^\star$. In particular, even if (4) has a unique optimizer (e.g., due to strict convexity of f) it may have multiple solutions because $\mu^\star$ might be not unique.

Remark 3. We consider only inequality constraints $g(x) \leq 0$, since they introduce a discontinuous projection on the dual variables, which poses one of the main technical challenges. The case of equality constraints can be treated analogously with several simplifications (e.g., convex equality constraints are necessarily linear-affine and (6) is vacuous) and has been well-documented in the literature.

We now formally define the class of double-projection saddle flows that *solve* (4), i.e., whose trajectories converge to the global solutions of (4), while satisfying $x \in \mathcal{X}$ at all times.

For this we define the *state space* $\mathcal{Z} := \mathcal{X} \times \mathbb{R}_{\geq 0}^m$ and the *(primal-dual) feasible set*

$$\mathcal{F} := \{(x, \mu) \in \mathcal{Z} \mid g(x) \leq 0\}. \tag{7}$$

If $\mathcal{X}$ is known to be a closed convex set, then it is immediate that $\mathcal{Z}$ is a closed convex set with tangent cone $T_z\mathcal{Z} = T_x\mathcal{X} \times T_\mu\mathbb{R}_{\geq 0}^m$ where $z := (x, \mu)$. Hence, we can define a well-behaved projected dynamical system restricted to $\mathcal{Z}$. For this, we first dualize the inequality constraint $g(x)$ in (4b), but not the constraint in (4c) leading to a partial Lagrangian. We then augment this partial Lagrangian to get the augmented partial Lagrangian of (4) which is defined as

$$L : \quad \mathcal{Z} \to \mathbb{R}$$
$$(x, \mu) \mapsto f(x) + \mu^T g(x) + \frac{\rho}{2}\|\max\{0, g(x)\}\|^2 ,$$

where $\rho \geq 0$ is an augmentation parameter.

This Lagrangian is differentiable with derivative

$$\begin{bmatrix} \nabla_x L \\ \nabla_\mu L \end{bmatrix} = \begin{bmatrix} \nabla f(x) + \nabla g(x)(\mu + \rho\max\{0, g(x)\}) \\ g(x) \end{bmatrix} .$$

We thus study the system described by projected gradient descent on the primal variable x and a projected gradient ascent on the dual variable μ, i.e.,

$$\dot{x} = [-\nabla_x L(x, \mu)]_{\mathcal{X}}^x \tag{8a}$$

$$\dot{\mu} = [\nabla_\mu L(x, \mu)]_{\mathbb{R}_{\geq 0}^m}^\mu . \tag{8b}$$

Existence of solutions for (8) will be a by-product of the forthcoming stability analysis which uses Theorem 1. For now, note that ∇L is locally Lipschitz [4] and $\mathcal{Z}$ is closed convex, thus satisfying the requirements on $\mathcal{X}$ and F in Theorem 1.

Remark 4. Convex-concave saddle flows, such as (8), have been studied using a variety of tools. In particular, the forthcoming results in Section 4 that exploit monotonicity to establish convergence to an invariant set, can be presented using complementarity systems (Stegink et al., 2018) or maximal monotone mappings (Goebel, 2017). The latter, in particular, allows for a very clean presentation, even if the saddle function L is non-differentiable. However, when considering a non-convex problem (which is outside the scope of this paper) the monotonicity fails to hold. This not only jeopardizes stability, but also calls into question the applicability of existence results such as (Brogliato et al., 2006, Thm 1) or (Goebel, 2017, Thm 2.2).

Remark 5. The system (8) admits various variations that do not affect the main results presented in this paper. Namely, (8a) and (8b) can be subject to different time constants. Furthermore, instead of using a single augmentation parameter ρ, one can scale the augmentation term by a matrix $\Xi \succ 0$ as $\|\max\{0, g(x)\}\|_\Xi^2$ as in Stegink et al. (2018).

For simplicity of notation we henceforth use

$$F(z) := \begin{bmatrix} -\mathbb{I}_n & 0 \\ 0 & \mathbb{I}_m \end{bmatrix} \nabla L(z) . \tag{9}$$

This definition alludes to the fact that (8) describes a projected *pseudo-gradient flow* on $\mathcal{Z}$, i.e., a gradient flow in an indefinite metric (Bloch et al., 1992; van der Schaft et al., 2014; Smale, 1972). Hence, we can write (8) compactly as

$$\dot{z} = [F(z)]_{\mathcal{Z}}^z . \tag{10}$$

The following statement is well-known and can be easily derived from Proposition 1 and Lemma 1:

Proposition 2. Under Assumptions 1 and 2 the set of equilibria of (10) is equivalent to the set of global solutions of (4).

---

[4] Notice that $\max\{0, g(x)\}$ is convex and hence locally Lipschitz.

4. STABILITY OF PROJECTED SADDLE FLOWS

Next, we review the stability (but not necessarily convergence) of the dynamics (10). The results of this section use well-established tools and have been shown, in one way or another, in Cherukuri et al. (2016); Cherukuri et al. (2018); Cortés and Niederländer (2019); Stegink et al. (2018); Cherukuri et al. (2015) and exploit basic facts from convex analysis (Beck, 2014; Rockafellar and Wets, 1998). We aim for a particularly concise presentation that prepares for the main result in the next section. For completeness, the proofs for this section are included in the appendix.

As a first step, it suffices to characterize the *monotonicity* of the "unprojected" vector field $F(z)$ by exploiting the convexity of g and f. The first two points in the following proposition are well-known, the last point is new (to the best of the authors' knowledge) but easy to prove and crucial for our forthcoming results. For completeness the proof can be found in the appendix.

Proposition 3. Under Assumptions 1 and 2 the following holds for $F(z)$ defined in (9):

(i) For all $z, \hat{z} \in \mathcal{Z}$ the vector field $F(z)$ satisfies
$$\langle z - \hat{z}, F(z) - F(\hat{z}) \rangle \leq 0 \,.$$

(ii) If f is strictly convex, for $z, \hat{z} \in \mathcal{Z}$ with $x \neq \hat{x}$ one has
$$\langle z - \hat{z}, F(z) - F(\hat{z}) \rangle < 0 \,.$$

(iii) If $\rho > 0$, then for all $z \in \mathcal{Z}$ but $z \notin \mathcal{F}$ and all $\hat{z} \in \mathcal{F}$ we have
$$\langle z - \hat{z}, F(z) - F(\hat{z}) \rangle < 0 \,.$$

One useful property of projected dynamical systems on convex sets is the fact that monotonicity of vector fields is preserved when passing from an unprojected to a projected vector field as the following lemma shows. For completeness the proof can be found in the appendix.

Lemma 2. Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a closed convex set and $F : \mathcal{Z} \to \mathbb{R}^n$ a vector field. If for $z, \hat{z} \in \mathcal{Z}$ it holds that
$$\langle z - \hat{z}, F(z) - F(\hat{z}) \rangle \leq \alpha$$
for some $\alpha \in \mathbb{R}$, then it also holds that
$$\langle z - \hat{z}, [F(z)]_{\mathcal{Z}}^z - [F(\hat{z})]_{\mathcal{Z}}^{\hat{z}} \rangle \leq \alpha \,.$$

Lemma 2 allows to amend Proposition 3 as follows.

Corollary 1. The statements (i), (ii), and (iii) in Proposition 3 are valid if $F(z)$ and $F(\hat{z})$ are replaced with $[F(z)]_{\mathcal{Z}}^z$ and $[F(\hat{z})]_{\mathcal{Z}}^{\hat{z}}$ respectively, everywhere.

Hence, the invariance principle for projected dynamical systems (Theorem 1) can be used to conclude the following. For completeness the proof can be found in the appendix.

Theorem 2. Under Assumptions 1 and 2 there exists a unique complete solution $z : [0, \infty) \to \mathcal{Z}$ of (10) for every initial condition and z converges asymptotically to the largest invariant subset Ω of $\mathcal{M}_{z^\star}$ where
$$\mathcal{M}_{z^\star} := \mathrm{cl}\{z \in \mathcal{Z} \,|\, \langle z - z^\star, [F(z)]_{\mathcal{Z}}^z \rangle = 0\} \,, \qquad (11)$$
where $z^\star$ is a solution of (4). Moreover, all solutions $z^\star$ are stable.

If f is strictly convex and with the help of item (ii) of Proposition 3 and Corollary 1 one can show convergence to the solutions of (4). This results was established for a more general setup in Goebel (2017) for saddle-functions

that are strictly convex/concave in either the primal or dual variables. In our case, the saddle-function, which is a partial Lagrangian of (4), is never strictly concave in the dual variables. Hence, strict convexity has to stem from the primal variables.

Corollary 2. (Goebel, 2017, Thm 3.3) Under Assumptions 1 and 2 and if f is strictly convex, every trajectory of (10) converges to a primal-dual solution of (4).

Remark 6. It is worth pointing out a potential pitfall: After concluding that trajectories converge to an invariant subset Ω of $\mathcal{M}_{z^\star}$, it is tempting to study the dynamics restricted to Ω, e.g., with another LaSalle function, and conclude convergence of any trajectory to this "nested" ω-limit set. This approach is however in general not valid unless additional assumptions are satisfied (Arsie and Ebenbauer, 2010). Second, although the zero-dissipation set $\mathcal{M}_{z^\star}$ (for strictly convex f) consists only of equilibrium points, Theorem 2 does not imply convergence to a single point. To conclude pointwise convergence, one needs to exploit the definitions of stability and limit sets as done in Cherukuri et al. (2016); Goebel (2017); Stegink et al. (2018).

5. CHARACTERIZATION OF THE ZERO-DISSIPATION SET

In order to better describe the limit behavior of (10), we now introduce a novel characterization of the zero-dissipation set. The following assumption will be required below for item (vi) of Proposition 4. Whether it can be relaxed remains an open question.

Assumption 3. (Strict complementary slackness).
Under Assumption 2 there is a solution $(x^\star, \mu^\star)$ of (4) such that
$$\forall i = 1, \dots, m : \quad g_i(x^\star) = 0 \quad \Rightarrow \quad \mu_i^\star > 0 \,. \qquad (12)$$

Assumption 3 is weak, because it only needs to hold for a single solution of (4). Further, for large classes of (parametrized) optimization problems this assumption holds generically, i.e., for almost all problem instances (Spingarn and Rockafellar, 1979).

We now recall a basic, but little known, result about convex functions which is crucial for our analysis.

Lemma 3. (Beck, 2014, Ex. 7.28iii) If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable with Lipschitz gradient, then it holds $\forall x, y \in \mathbb{R}^n$ that
$$f(y) - f(x) = \nabla f(x)^T (x - y) \quad \Rightarrow \quad \nabla f(x) = \nabla f(y)$$
as well as
$$\langle x - y, \nabla f(x) - \nabla f(y) \rangle = 0 \quad \Rightarrow \quad \nabla f(x) = \nabla f(y) \,.$$

Hence, our key technical result reads as follows.

Proposition 4. (Characterization of zero-dissipation set). Let Assumptions 1 and 2 hold, let $z^\star = (x^\star, \mu^\star)$ be any solution of (4) and let $\mathcal{F}$ be defined as in (7). Then the following statements about the zero-dissipation set $\mathcal{M}_{z^\star}$ defined in (11) hold true:

(i) If $\rho > 0$ or f is strictly convex, then $\mathcal{M}_{z^\star} \subseteq \mathcal{F}$.
(ii) For all $(x, \mu) \in \mathcal{M}_{z^\star}$ it holds that $\nabla f(x) = \nabla f(x^\star)$.
(iii) For all $(x, \mu) \in \mathcal{M}_{z^\star}$ and all $i = 1, \dots, m$ we have
$$\mu_i^\star = 0 \quad \text{or} \quad g_i(x) = g_i(x^\star) + \nabla g_i(x^\star)(x - x^\star) \,.$$
(iv) For all $(x, \mu) \in \mathcal{M}_{z^\star}$ it holds that $\nabla g(x)\mu = \nabla g(x^\star)\mu$.
(v) For all $(x, \mu) \in \mathcal{M}_{z^\star}$ and all $i = 1, \dots, m$ we have
$$g_i(x^\star) < 0 \quad \Rightarrow \quad \mu_i = 0 \,.$$

(vi) If, in addition, Assumption 3 is satisfied and the zero-dissipation set $\mathcal{M}_{z^\star}$ in (11) is defined with respect to a solution $z^\star$ that satisfies (12), then it holds that for all $(x, \mu) \in \mathcal{M}_{z^\star}$ and all $i = 1, \ldots, m$ we have

$$g_i(x^\star) = 0 \quad \Rightarrow \quad \mu_i > 0 \quad \text{or} \quad g_i(x) \geq 0 \,.$$

Proof. We use the definitions of M_1, M_2, M_3 from the proof of Proposition 3 and the definitions of $\eta, \hat{\eta}$ from the proof of Lemma 2, which can be found in the appendix. Based on (11), we consider the case in which $\hat{z} = z^\star$ and $\hat{\eta} = \eta^\star$.

For zero dissipation we need $M_1 = M_2 = M_3 = 0$ and

$$\langle z - \hat{z}, \eta \rangle = \langle z - \hat{z}, \hat{\eta} \rangle = 0 \,. \tag{13}$$

First, if $\rho > 0$, then $M_2 = 0$ is only the case when $g(x) \leq 0$ and hence $z \in \mathcal{F}$. Otherwise, if f is strictly convex the primal optimizer of (4) is unique and $M_1 < 0$ for all $x \neq \hat{x}$ and in particular for $x \notin \mathcal{F}$. This establishes (i).

Second, for $M_1 = \langle x - \hat{x}, \nabla f(x) - \nabla f(\hat{x}) \rangle = 0$ we can use Lemma 3 to conclude (ii).

Third, for $M_3 = 0$ to hold we require that

$$\mu^T \left(-g(\hat{x}) + g(x) + \nabla g(x)^T(\hat{x} - x) \right) = 0 \tag{14}$$

$$\hat{\mu}^T \left(-g(x) + g(\hat{x}) + \nabla g(\hat{x})^T(x - \hat{x}) \right) = 0 \,. \tag{15}$$

From (15) we directly infer (iii).

From (14), we establish, using Lemma 3, that

$$\mu_i = 0 \quad \text{or} \quad \nabla g_i(x) = \nabla g_i(\hat{x}) \,,$$

which implies (iv).

Next, for (v), recall that $\mathcal{Z} := \mathcal{X} \times \mathbb{R}^m_{\geq 0}$. Therefore, the projection on $\mathcal{Z}$ can be decomposed into a projection on $\mathcal{X}$ and m projections on $\mathbb{R}_{\geq 0}$. Therefore, $\langle z - \hat{z}, \eta \rangle = 0$ implies that $\langle \mu_i - \hat{\mu}_i, \eta_{\mu i} \rangle = 0$ for all $i = 1, \ldots, m$ where $\eta_{\mu i}$ denotes the i-th dual components of $\eta, \hat{\eta}$, respectively. The analogous statement holds for $\hat{\eta}_{\mu i}$.

Combining these statements, we have, for all $i = 1, \ldots, m$,

$$\eta_{\mu i} \neq 0 \quad \text{or} \quad \eta_{\hat{\mu} i} \neq 0 \quad \Rightarrow \quad \mu_i = \hat{\mu}_i \,. \tag{16}$$

Note that for $\mathbb{R}_{\geq 0}$ we have $N_x \mathbb{R}_{\geq 0} = \mathbb{R}$ for all $x > 0$ and $N_0 \mathbb{R}_{\geq 0} = \mathbb{R}_{\geq 0}$. Thus, using Lemma 1, we have that

$$\hat{\eta}_{\mu i} = \begin{cases} 0 & \text{if } g_i(\hat{x}) \geq 0 \text{ or } \hat{\mu}_i > 0 \\ g_i(\hat{x}) & \text{otherwise} \,. \end{cases}$$

The statement (v) follows immediately because, since $\hat{x}$ is a solution of (4), $g_i(\hat{x}) < 0$ implies $\hat{\mu}_i = 0$ by complementary slackness (6) and therefore $\eta_{\mu i} \neq 0$, and with (16) it follows that $\mu_i = \hat{\mu}_i$.

Finally, for (vi) we work with the contraposition of (16) and the strict complementarity slackness assumption on $(\hat{x}, \hat{\mu})$. Namely, if $g_i(\hat{x}) = 0$ we have that $\hat{\eta}_{\mu i} = 0$ and $\hat{\mu}_i > 0$ by (12). If $\mu_i = 0$, then (16) implies that $\eta_{\mu i} \neq 0$ which in turn implies that $g_i(x) \geq 0$. ∎

Proposition 4 is instrumental to study the occurrence of limit cycles for non-strictly convex cost functions.

6. LIMIT BEHAVIOR UNDER NON-STRICT CONVEXITY

When f is not strictly convex it is well-known that convergence to equilibrium points is not in general guaranteed and limit cycles can occur. In the following, we do not consider projections of the primal variables and hence make the assumption that $\mathcal{X} = \mathbb{R}^n$. Furthermore, we make use of Assumption 3.

With the help of Proposition 4 we can show that on the zero dissipation set $\mathcal{M}_{z^\star}$ the dynamics (10) follow simple linear Hamiltonian dynamics. Since we are only interested in invariant subsets of $\mathcal{M}_{z^\star}$ we can then establish a contradiction which leads us to conclude that, in the presence of augmentation of the Lagrangian, the only invariant subsets of $\mathcal{M}_{z^\star}$ are equilibrium points.

Proposition 5. Let Assumptions 1, 2 and 3 hold, let $\mathcal{X} = \mathbb{R}^n$, and let the zero-dissipation set $\mathcal{M}_{z^\star}$ in (11) be defined with respect to a solution $z^\star$ that satisfies (12). Then, for all $z \in \mathcal{M}_{z^\star}$, the dynamical system (10) reduces to

$$\begin{bmatrix} \dot{x} \\ \dot{\mu}_{I^\star} \end{bmatrix} = \begin{bmatrix} 0 & -A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} x \\ \mu_{I^\star} \end{bmatrix} - \begin{bmatrix} c \\ d \end{bmatrix} \,, \tag{17}$$

where $c = \nabla f(x^\star)$, $d = \nabla g_{I^\star}(x^\star)^T x^\star$, and $A = \nabla g_{I^\star}(x^\star)$, with $I^\star := \{i \mid g_i(x^\star) = 0\}$ and $\mu_i = 0$ for all $i \notin I^\star$.

Proof. First, note that (v) in Proposition 4 implies that $\mu_i = 0$ for all $i \notin I^\star$ on all of $\mathcal{M}_{z^\star}$. Second, using (i), (ii), (iv), and (v) in Proposition 4 we know that for all $(x, \mu) \in \mathcal{M}_{z^\star}$ we have

$$-\nabla_x L(x, \mu) = -(\nabla f(x) + \nabla g(x)\mu)$$
$$= -(\nabla f(x^\star) + \nabla g(x^\star)\mu)$$
$$= -(\nabla f(x^\star) + \nabla g_{I^\star}(x^\star)\mu_{I^\star}) \,.$$

Furthermore, we know that $i \in I^\star$ implies $\mu_i^\star > 0$ according to (12). Hence, we apply (iii) in Proposition 4 to conclude that for all $i \in I^\star$ we have

$$g_i(x) = g_i(x^\star) + \nabla g_i(x^\star)^T(x - x^\star) = \nabla g_i(x^\star)^T(x - x^\star) \,,$$

that is, g_i is linear on $\mathcal{M}_{z^\star}$ and equal to its linearization at $x^\star$.

Finally, with (vi) of Proposition 4 we know that $\mu_i > 0$ or $g_i(x) \geq 0$ holds on all of $\mathcal{M}_{z^\star}$ and therefore the projection of the dual variables in $I^\star$ is never active, i.e.,

$$\dot{\mu}_i = [g_i(x)]_+^{\mu_i} = g_i(x) \quad \text{for } i \in I^\star$$

and hence, on $\mathcal{M}_{z^\star}$ the system (10) reduces to (17). ∎

Note that Proposition 5 states that trajectories, while in $\mathcal{M}_{z^\star}$, satisfy (17), but that does not imply that $\mathcal{M}_{z^\star}$ is invariant. In this sense, Proposition 5 does not prove the existence of limit cycles. In fact, as we show next, in the presence of an augmentation term the only invariant subsets of $\mathcal{M}_{z^\star}$ are equilibrium points.

Theorem 3. Under Assumptions 1, 2 and 3, if $\mathcal{X} = \mathbb{R}^n$ and $\rho > 0$, then the largest invariant set $\Omega \subseteq \mathcal{M}_{z^\star}$ in Theorem 2 is equivalent to the set of equilibria of (10). Furthermore, all trajectories of (10) converge asymptotically to the set of solutions of (4).

Proof. The linear system (17) has analytic solutions

$$x(t) := x^\star + V \begin{bmatrix} \text{diag}(\beta) \sin(\sigma t + \phi) \\ \gamma \end{bmatrix}$$

$$\mu_I(t) := \mu^\star - U \left[\text{diag}(\beta) \cos(\sigma t + \phi) \right] \,,$$

where A has the singular value decomposition $A = U\Sigma V^T$ with $\Sigma = [\text{diag}(\sigma) \ 0]$, cos and sin apply componentwise, and $\beta, \phi \in \mathbb{R}^{|I^\star|}$, $\gamma \in \mathbb{R}^{n-|I^\star|}$ depend on initial conditions.

For any initial condition and any $\beta \neq 0$ there exists t such that for at least one component $i \in I^\star$ we have that $\dot{\mu}_i(t) > 0$. Recall that, on $\mathcal{M}_{z^\star}$,

$$\dot{\mu}_i(t) = Ax - d = \nabla g_i(x^\star)(x - x^\star) = g_i(x) \,,$$

and therefore, unless $\beta = 0$, there exists at least one oscillating mode that drives the trajectory outside of the feasible set. However, by (i) in Proposition 4 we have that $\mathcal{M}_{z^\star} \subseteq \mathcal{F}$ and therefore such an orbit is not invariant with respect to $\mathcal{M}_{z^\star}$. Hence, the only invariant subsets of $\mathcal{M}_{z^\star}$ are equilibrium points. Conversely, any equilibrium point is invariant and contained in $\mathcal{M}_{z^\star}$. Since Ω is the *largest* invariant subset of $\mathcal{M}_{z^\star}$ it has to contain all equilibria. ■

7. ILLUSTRATIVE EXAMPLE

To illustrate the occurrence of limit cycles and that we do not experience such behavior with an augmented Lagrangian we analyze the stylized problem

$$\max_{x \leq 0} \quad x$$

with the unique solution $(x^\star, \mu^\star) = (0, 1)$. We dualize the constraint and augment the Lagrangian. Hence, the projected saddle flow (10) for this problem is defined on $\mathcal{Z} = \mathbb{R} \times \mathbb{R}_{\geq 0}$ and given by

$$\dot{z} = \begin{bmatrix} \dot{x} \\ \dot{\mu} \end{bmatrix} = \begin{bmatrix} 1 - \mu - \rho \max\{0, x\} \\ [x]_{\mathbb{R}_{\geq 0}}^{\mu} \end{bmatrix}, \qquad (18)$$

where $\rho \geq 0$ is the augmentation parameter. A LaSalle function is given by

$$V(x, \mu) = \frac{1}{2}\|x - x^\star\|^2 + \frac{1}{2}\|\mu - \mu^\star\|^2 \geq 0,$$

and the Lie derivative of $V(x, \mu)$ is

$$\mathcal{L}_{(18)}V(x, \mu) = (x - [x]_{\mathbb{R}_{\geq 0}}^{\mu})(1 - \mu) - \rho x \max\{0, x\}$$

$$= \begin{cases} x - \max\{0, x\} - \rho x \max\{0, x\} & \mu = 0 \\ -\rho x \max\{0, x\} & \mu > 0. \end{cases}$$

For $\rho = 0$ (no augmentation) the zero-dissipation set is $\mathcal{M}_{z^\star} = \{z \in \mathcal{Z}\} \setminus \{x < 0, \mu = 0\}$, and therefore there is only dissipation on the negative x-axis. The red line in Fig. 1 shows that in this case all trajectories converge to the circle around the solution $(0, 1)$ with radius 1. On this set the trajectories form periodic orbits around the solution. With augmentation the zero-dissipation set is $\mathcal{M}_{z^\star} = \{z \in \mathcal{Z} \mid x \leq 0, \mu > 0\}$. Due to the dissipation on the negative x-axis and in the first quadrant, all trajectories converge to the largest invariant subset Ω, which contains only the solution. Compare the blue line in Fig. 1.

8. CONCLUSION

In this paper, we proposed a unified formulation of saddle dynamics in the presence of projection of both the primal and the dual variables.

Such a unified approach, based on the formalism of projected dynamical systems, allows us to derive stability and convergence results that subsume more specific results that have been presented in the literature.

Also, we characterize the zero-dissipation set of a saddle flow and the emergence of limit cycles under non-strict convexity. We show that oscillations can be avoided with a simple augmentation term that does not affect the equilibrium of the saddle flow.

We expect to extend the analysis of these dynamics to non-convex domains, profiting from the well-posedness of projected dynamical systems on these domains. This extension has the potential to support the application of these methods for the analysis and design of feedback optimization schemes for a wide class of systems.
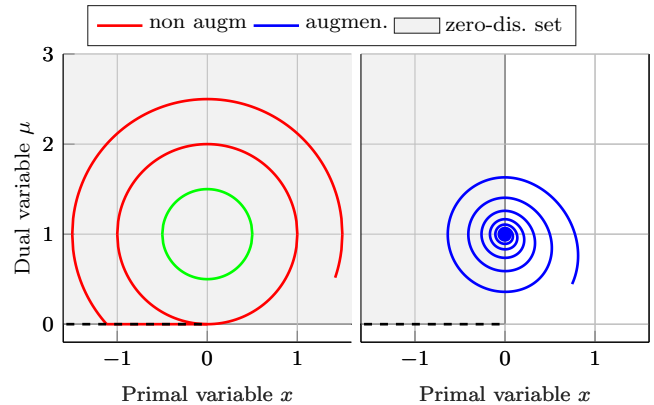

Fig. 1. Left plot without augmentation ($\rho = 0$): Trajectory (red) of (18) converging to the boundary of the largest invariant subset $\Omega = \{(x, u) \in \mathbb{R} \times \mathbb{R}_{\geq 0} \mid \|(x, \mu - 1)\|_2 < 1\}$. Exemplary trajectory (green) of (18) inside of Ω. The zero-dissipative set $\mathcal{M}_{z^\star}$ is everything besides the negative x-axis. Right plot with augmentation ($\rho > 0$): Trajectory (blue) of (18) converging to the largest invariant subset $\Omega = \{(x^\star, \mu^\star)\}$. The zero-dissipative set $\mathcal{M}_{z^\star}$ is the second quadrant without the negative x-axis.

REFERENCES

Arrow, K.J., Hurwicz, L., and Uzawa, H. (1958). *Studies in Linear and Nonlinear Programming.* Stanford University Press, Stanford.

Arsie, A. and Ebenbauer, C. (2010). Locating omega-limit sets using height functions. *Journal of Differential Equations*, 248(10), 2458–2469.

Aubin, J.P. and Cellina, A. (1984). *Differential Inclusions: Set-Valued Maps and Viability Theory.* Grundlehren Der Mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg.

Aubin, J.P. (1991). *Viability Theory.* Systems & Control: Foundations & Applications. Springer Science & Business Media, Boston.

Bacciotti, A. and Ceragioli, F. (1999). Stability and Stabilization of Discontinuous Systems and Nonsmooth Lyapunov Functions. *ESAIM: COCV*, 4, 361–376.

Beck, A. (2014). *Introduction to Nonlinear Optimization.* MOS-SIAM Series on Optimization. SIAM, Philadelphia, PA, 1st ed.

Bloch, A.M., Brockett, R.W., and Ratiu, T.S. (1992). On the geometry of saddle point algorithms. In *Proc. 31st IEEE Conference on Decision and Control*, 1482–1487.

Brayton, R. and Moser, J. (1964). A theory of nonlinear networks. I. *Quarterly of Appl. Math*, 22(1), 1–33.

Brogliato, B., Daniilidis, A., Lemaréchal, C., and Acary, V. (2006). On the equivalence between complementarity systems, projected systems and differential inclusions. *Systems & Control Letters*, 55(1), 45–51.

Cherukuri, A., Mallada, E., and Cortés, J. (2015). Convergence of Caratheodory solutions for primal-dual dynamics in constrained concave optimization. In *SIAM Conf. on Control and Its Applications 2015*, 290–296.

Cherukuri, A., Mallada, E., Low, S., and Cortés, J. (2018). The role of convexity in saddle-point dynamics: Lyapunov function and robustness. *IEEE Trans. Autom. Control*, 63(8), 2449–2464.

Cherukuri, A., Mallada, E., and Cortés, J. (2016). Asymptotic convergence of constrained primal–dual dynamics. *Systems & Control Letters*, 87, 10 – 15.

Colombino, M., Dall'Anese, E., and Bernstein, A. (2019). Online optimization as a feedback controller: Stability and tracking. *IEEE Trans. Control Netw. Syst.*, 422–432.

Cornet, B. (1983). Existence of slow solutions for a class of differential inclusions. *Journal of Mathematical Analysis and Applications*, 96(1), 130–147.

Cortés, J. and Niederländer, S.K. (2019). Distributed coordination for nonsmooth convex optimization via saddle-point dynamics. *Journal of Nonlinear Science*, 29(4), 1247–1272.

Dall'Anese, E. and Andrea Simonetto (2018). Optimal power flow pursuit. *IEEE Trans. Smart Grid*, 9(2), 942–952.

Daniele, P., Giuffré, S., Maugeri, A., and Pia, S. (2010). A Panoramic View on Projected Dynamical Systems. In P.M. Pardalos, T.M. Rassias, and A.A. Khan (eds.), *Nonlinear Analysis and Variational Problems*, number 35 in Springer Optimization and Its Applications, 235–258. Springer New York.

Dhingra, N.K., Khong, S.Z., and Jovanovic, M.R. (2018). The proximal augmented lagrangian method for nonsmooth composite optimization. *IEEE Trans. Autom. Control*, 2861–2868.

Feijer, D. and Paganini, F. (2010). Stability of primal–dual gradient dynamics and applications to network optimization. *Automatica*, 46(12), 1974–1981.

Filippov, A.F. (1988). *Differential Equations with Discontinuous Righthand Sides: Control Systems*. Mathematics and Its Applications (Soviet Series). Springer Netherlands.

Goebel, R. (2017). Stability and robustness for saddle-point dynamics through monotone mappings. *Systems & Control Letters*, 108, 16 – 22.

Hauswirth, A., Bolognani, S., and Dörfler, F. (2018). Projected Dynamical Systems on Irregular, Non-Euclidean Domains for Nonlinear Optimization. *ArXiv180904831 Math*.

Hauswirth, A., Bolognani, S., Hug, G., and Dörfler, F. (2020). Timescale separation in autonomous optimization. *IEEE Trans. Autom. Control*. To appear.

Heemels, W.P.M.H., Schumacher, J.M., and Weiland, S. (2000). Projected dynamical systems in a complementarity formalism. *Operations Research Letters*, 27(2), 83–91.

Henry, C. (1973). An existence theorem for a class of differential equations with multivalued right-hand side. *Journal of Mathematical Analysis and Applications*, 41(1), 179–186.

Holding, T. and Lestas, I. (2014). On the convergence to saddle points of concave-convex functions, the gradient method and emergence of oscillations. In *Proc. 53rd IEEE Conference on Decision and Control*, 1143–1148.

Kose, T. (1956). Solutions of saddle value problems by differential equations. *Econometrica*, 59–70.

Nagurney, A. and Zhang, D. (1996). *Projected Dynamical Systems and Variational Inequalities with Applications*. Springer, 1 ed.

Rockafellar, R.T. and Wets, R.J.B. (1998). *Variational Analysis*. Springer, Berlin Heidelberg, Germany, 3rd printing ed.

Smale, S. (1972). On the mathematical foundations of electrical circuit theory. *Journal of Differential Geometry*, 7(1-2), 193–210.

Spingarn, J.E. and Rockafellar, R.T. (1979). The generic nature of optimality conditions in nonlinear programming. *Math. of Operations Research*, 4(4), 425–430.

Stegink, T.W., Van Damme, T., and De Persis, C. (2018). Convergence of projected primal-dual dynamics with applications in data centers. *IFAC-PapersOnLine*, 51(23), 88–93.

Tang, Y., Dall'Anese, E., Bernstein, A., and Low, S. (2018). Running Primal-Dual Gradient Method for Time-Varying Nonconvex Problems. *ArXiv181200613 Math*.

van der Schaft, A., Jeltsema, D., et al. (2014). Port-hamiltonian systems theory: An introductory overview. *Foundations and Trends® in Systems and Control*, 1(2-3), 173–378.

Venets, V. (1985). Continuous algorithms for solution of convex optimization problems and finding saddle points of convex-concave functions with the use of projection operations. *Optimization*, 16(4), 519–533.

Appendix A. PROOF OF PROPOSITION 3

The proof mainly exploits the definition of convexity and follows ideas from Cherukuri et al. (2016); Stegink et al. (2018) and others. Using the definition of ∇L we can decompose

$$\langle z - \hat{z}, F(z) - F(\hat{z}))\rangle = -M_1 - M_2 - M_3 \,,$$

where we define

$$M_1 := \langle x - \hat{x}, \nabla f(x) - \nabla f(\hat{x})\rangle$$
$$M_2 := \rho \langle x - \hat{x}, \nabla g(x) \max\{0, g(x)\}\rangle$$
$$\quad - \rho \langle x - \hat{x}, \nabla g(\hat{x}) \max\{0, g(\hat{x})\}\rangle$$
$$M_3 := \langle x - \hat{x}, \nabla g(x)\mu - \nabla g(\hat{x})\hat{\mu}\rangle$$
$$\quad - \langle \mu - \hat{\mu}, g(x) - g(\hat{x})\rangle \,.$$

Hence, for the first term we have by the definition of convexity of f that $M_1 \geq 0$, and if f is strictly convex we have $M_1 > 0$ for all $x \neq \hat{x}$.

For M_2 we notice that if $\hat{z} \in \mathcal{F}$ we have that $\max\{0, g(\hat{x})\} = 0$. Furthermore, we define $I(x) := \{i \,|\, g_i(x) = 0\}$ and notice that if $z \in \mathcal{X}$, but $z \notin \mathcal{F}$, then for all $i \in I(x)$, where $I(x) \neq \emptyset$, we have $g_i(x) > 0$. Therefore, using also convexity of g, it holds that

$$M_2 = \rho \langle x - \hat{x}, \nabla g(x) \max\{0, g(x)\}\rangle$$
$$= \rho \sum_{i \in I(x)} (x_i - \hat{x}_i)^T \nabla g_i(x) g_i(x)$$
$$\geq \rho \sum_{i \in I(x)} \underbrace{(g_i(x) - g_i(\hat{x}))}_{>0 \quad \leq 0} \underbrace{g_i(x)}_{>0} > 0 \,.$$

Finally, for M_3 we have

$$M_3 = \mu^T \nabla g(x)^T (x - \hat{x}) - \hat{\mu}^T \nabla g(\hat{x})^T (x - \hat{x})$$
$$\quad - \mu^T (g(x) - g(\hat{x})) + \hat{\mu}^T (g(x) - g(\hat{x}))$$
$$= \mu^T \underbrace{\left(g(\hat{x}) - g(x) - \nabla g(x)^T (\hat{x} - x)\right)}_{\geq 0 \text{ (by convexity)}}$$
$$\quad + \hat{\mu}^T \underbrace{\left(g(x) - g(\hat{x}) - \nabla g(\hat{x})^T (x - \hat{x})\right)}_{\geq 0 \text{ (by convexity)}},$$

and since $\mu, \hat{\mu} \geq 0$ it follows that $M_3 \geq 0$. Hence, (i), (ii), and (iii) follow immediately.

Appendix B. PROOF OF LEMMA 2

Lemma 1 states that there are normal vectors $\eta \in N_z \mathcal{Z}$ and $\hat{\eta} \in N_{\hat{z}} \mathcal{Z}$ such that $[F(z)]_{\mathcal{Z}}^z = F(z) - \eta$ and $[F(\hat{z})]_{\mathcal{Z}}^{\hat{z}} = F(\hat{z}) - \hat{\eta}$. Further, by definition of the normal cone to a convex set we have $\langle \eta, \hat{z} - z\rangle \leq 0$ and $\langle \hat{\eta}, z - \hat{z}\rangle \leq 0$ $\forall z, \hat{z} \in \mathcal{Z}$. Hence,

$$\langle z - \hat{z}, [F(z)]_{\mathcal{Z}}^z - [F(\hat{z})]_{\mathcal{Z}}^{\hat{z}}\rangle$$
$$= \langle z - \hat{z}, F(z) - \eta - F(\hat{z}) + \hat{\eta}\rangle$$
$$= \underbrace{\langle z - \hat{z}, F(z) - F(\hat{z})\rangle}_{\leq \alpha} + \underbrace{\langle \hat{z} - z, \eta\rangle}_{\leq 0} + \underbrace{\langle z - \hat{z}, \hat{\eta}\rangle}_{\leq 0}$$

which immediately proves the lemma.

Appendix C. PROOF OF THEOREM 2

By Assumption 2 there exists an optimizer $x^\star$ to (4) with dual solution $\mu^\star$ satisfying the conditions of Proposition 1 and $(x^\star, \mu^\star)$ is an equilibrium of (10) by Proposition 2. Using Proposition 1 we know that for every (local) solution $z : [0, T] \to \mathcal{Z}$ of (10) we have for almost all $t \in [0, T)$ that

$$\frac{d}{dt}\|z(t) - z^\star\|^2 = 2\left\langle z - z^\star, [F(z)]_{\mathcal{Z}}^z - [F(z^\star)]_{\mathcal{Z}}^{z^\star}\right\rangle$$
$$= 2\langle z - z^\star, [F(z)]_{\mathcal{Z}}^z\rangle \leq 0 \,.$$

Since $\|z - z^\star\|^2$ has compact level sets as a function of z on $\mathcal{Z}$ it follows from Theorem 1 that all trajectories are complete and converge to the largest invariant subset Ω of $\mathcal{M}_{z^\star}$. The non-positivity of the Lie derivative of $\|z(t) - z^\star\|^2$ guarantees stability via a standard Lyapunov argument.