Stability and instability in saddle point dynamics Part II: The subgradient method

Thomas Holding and Ioannis Lestas

***Abstract*—In part I we considered the problem of convergence to a saddle point of a concave-convex function in C^2 via gradient dynamics and an exact characterization was given to their asymptotic behaviour. In part II we consider a general class of subgradient dynamics that provide a restriction in a convex domain. We show that despite the nonlinear and non-smooth character of these dynamics their ω-limit set is comprised of solutions to only linear ODEs. In particular, we show that the latter are solutions to subgradient dynamics on affine subspaces which is a smooth class of dynamics the asymptotic properties of which have been exactly characterized in part I. Various convergence criteria are formulated using these results and several examples and applications are also discussed throughout the manuscript.**

***Index Terms*—Nonlinear systems, subgradient dynamics, saddle points, non-smooth systems, networks, optimization.**

I. Introduction

In [24] we studied the asymptotic behaviour of the gradient method when this is applied on a general concave-convex function in an unconstrained domain, and provided an exact characterization to its limiting solutions. Nevertheless, in many applications, such as primal/dual algorithms in optimization problems, it becomes necessary to constrain the system states in a prescribed convex set, e.g. positivity constraints on Lagrange multipliers or constraints on physical quantities like data flow, and prices/commodities in economics [25], [29], [43]. The subgradient method is used in such cases, which is a version of the gradient method with a projection term in the vector field additionally included, so as to ensure that the trajectories do not leave the desired set.

In discrete time, there is an extensive literature on the subgradient method, via its application in optimization problems (see e.g. [37]). However, in many applications, for example power networks [49], [14], [27], [11], [12], [28], [44], [32], [35] and classes of data network problems [29], [43], [34] continuous time models are considered. It is thus important to have a good understanding of the subgradient dynamics in a continuous time setting, which could also facilitate analysis and design by establishing links with other more abstract results in dynamical systems theory.

A main complication in the study of the subgradient method arises from the fact the this is a *non-smooth* system, i.e. a nonlinear ODE with a discontinuous vector field due to the projections involved. This complicates the analysis which is also reflected in the early work in [1]. There have since been various studies on subgradient dynamics including works that exploit tools associated with monotone operators [40], [4], [45], [20] and more recent studies that make use of tools from non-smooth analysis [6], [7], [39] (see also the introduction in part I for a more extensive discussion).

Our aim in this paper is to provide a framework of results that allows one to study the asymptotic behaviour of the subgradient method in a general setting, where the trajectories are constrained to an arbitrary convex domain, and the concave-convex function in C^2 considered does not necessarily satisfy aditional strictness properties. One of our main results is to show that despite the nonlinear and non-smooth character of the subgradient dynamics, their limiting behaviour when an equilibrium point exists, are solutions to explicit *linear* differential equations.

In particular, we show that these linear ODEs are limiting solutions of subgradient dyanmics on an *affine subspace*, which is a class of dynamics that fit within the framework studied in Part I [24]. These dynamics can therefore be exactly characterized, thus allowing to prove convergence to a saddle point for broad classes of problems.

The results in this paper are illustrated by means of examples that demonstrate also the complications in the dynamic behaviour of the subgradient method relative to the unconstrained gradient method. We also apply our results to modification schemes in network optimization, that provide convergence guarantees while maintaining a decentralized structure in the dynamics.

The methodology used for the derivations in the paper is also of independent technical interest. In particular, the notion of a face of a convex set is used to characterize the ODEs associated with the limiting behaviour of the subgradient dynamics. Furthermore, some more abstract results on corresponding semiflows have been used to address the complications associated with the non-smooth character of subgradient dynamics.

The paper is structured as follows. Section II provides preliminaries from convex analysis and dynamical systems theory that will be used within the paper. The problem formulation is given in section III and the main results are presented in section IV, where various examples that illustrate those are also discussed. Applications to modification methods in network optimization are given in section V. The proofs of the results are given in appendices A and B and an application to the problem of multipath routing is discussed in section V-B.

Thomas Holding is with the Department of Mathematics, Imperial College London, United Kingdom; t.holding@imperial.ac.uk

Ioannis Lestas is with the Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom. icl20@cam.ac.uk

This work was supported by ERC starting grant 679774.

II. Preliminaries

We use the same notation and definitions as in part I of this work [24] and we refer the reader to the preliminaries section therein. The notions below from convex analysis and analysis of dynamical systems will additionally be used throughout the paper.

A. Convex analysis

We recall first for convenience the following notions defined in part I [24] that will be frequently used in this manuscript. For a closed convex set $K \subseteq \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^n$, we denote the normal cone to K through $\mathbf{z}$ as $N_K(\mathbf{z})$. When K is an affine space $N_K(\mathbf{z})$ is independent of $\mathbf{z} \in K$ and is denoted N_K. If K is in addition non-empty, then we denote the projection of $\mathbf{z}$ onto K as $\mathbf{P}_K(\mathbf{z})$. Also for vectors $x, y \in \mathbb{R}^n$, $d(x, y)$ denotes the Euclidean metric and $|x|$ the Euclidean norm.

1) Concave-convex functions and saddle points: For a function φ that is concave-convex on $\mathbb{R}^{n+m}$ the (standard) notion of a saddle point was given in part I [24]. We now consider φ that is concave-convex in a restricted region K in which case the notion of saddle point needs to be modified.

Definition 1 (Restricted saddle point)**.** Let $K \subseteq \mathbb{R}^{n+m}$ be non-empty closed and convex. For a function $\varphi : U \to \mathbb{R}$, $K \subseteq U \subseteq \mathbb{R}^{n+m}$ that is concave-convex on K we say that $(\bar{x}, \bar{y}) \in K$ is a *K-restricted saddle point* of φ if for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ with $(x, \bar{y}), (\bar{x}, y) \in K$ we have the inequality $\varphi(x, \bar{y}) \leq \varphi(\bar{x}, \bar{y}) \leq \varphi(\bar{x}, y)$.

If in addition $\varphi \in C^1$ on an open neighbourhood of K then $\bar{\mathbf{z}} = (\bar{x}, \bar{y}) \in K$ is a K-restricted saddle point if the vector of partial derivatives $(\varphi_x(\bar{\mathbf{z}}), -\varphi_y(\bar{\mathbf{z}}))$ lies in the normal cone $N_K(\bar{\mathbf{z}})$.

If $C \subseteq K$ is closed and convex and $\bar{\mathbf{z}} \in C$ is a C-restricted saddle point, then $\bar{\mathbf{z}}$ is also a K-restricted saddle point.

A structure of K that is important in the analysis[1] of concave-convex functions is that of a Cartesian product of two convex sets ([41]), with x, y taking values in each of these two sets respectively, i.e.

$$K = K_x \times K_y,\ K_x \subset \mathbb{R}^n,\ K_y \subset \mathbb{R}^m, \quad K_x, K_y \text{ convex closed sets} \tag{1}$$

It should be noted that in this case if $\varphi \in C^1$ on an open neighbourhood of K and concave-convex on K, then $\bar{\mathbf{z}} = (\bar{x}, \bar{y}) \in K$ is a K-restricted saddle point if and only if the vector of partial derivatives $(\varphi_x(\bar{\mathbf{z}}), -\varphi_y(\bar{\mathbf{z}}))$ lies in the normal cone $N_K(\bar{\mathbf{z}})$.

In general it does not hold that if $\varphi : \mathbb{R}^{n+m} \to \mathbb{R}$ has a saddle point, and K is closed convex and non-empty, then φ has a K-restricted saddle point (an explicit example illustrating this is given later in Example 27(ii)). In this manuscript we will only consider cases where at least one K-restricted saddle point exists, leaving the problem of showing existence to the specific application.

[1] In particular, this allows concave-convex functions on K to be appropriately extended as concave-convex functions in $\mathbb{R}^{n+m}$ [41].

2) Concave programming: Concave programming (see e.g. [3]) is concerned with the study of optimization problems of the form

$$\max_{x \in C, g(x) \geq 0} U(x) \tag{2}$$

where $U : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ are concave functions and $C \subseteq \mathbb{R}^n$ is non-empty closed and convex. Under some mild assumptions, the solutions to such problems are saddle points of the Lagrangian

$$\varphi(x, y) = U(x) + y^T g(x) \tag{3}$$

where $y \in \mathbb{R}^m_+$ are the Lagrange multipliers. This is stated in the Theorem below.

Theorem 2. *Let g be concave and Slater's condition hold, i.e.*

$$\exists x' \in \operatorname{relint} C \textit{ with } g(x') > 0. \tag{4}$$

Then $\bar{x}$ is an optimum of (2) if and only if $\exists \bar{y}$ with $(\bar{x}, \bar{y})$ a $C \times \mathbb{R}^m_+$-restricted saddle point of (3).

The min-max optimization problem associated finding a $C \times \mathbb{R}^m_+$-restricted saddle point of (3) is the dual problem of (2).

3) Faces of convex sets: Some of the main results of this manuscript refer to faces of a convex set. We refer the reader to [21, Chap. 1.8.] for further discussion of such topics.

Definition 3 (Face of a convex set)**.** Given a non-empty closed convex set K, a face F of K is a subset of K that has both the following properties:

(i) F is convex.
(ii) For any line segment $L \subseteq K$, if $(\operatorname{relint} L) \cap F \neq \emptyset$ then $L \subseteq F$.

For the readers convenience we recall some standard properties of faces:

(a) The intersection of two faces of K is a face of K.
(b) The empty set and K itself are both faces of K. If a face F is neither $\emptyset$ or K it is called a proper face.
(c) If F is a face of K and F' is a face of F, then F' is a face of K.
(d) For a face F of K, the normal cone $N_K(\mathbf{z})$ is independent of the choice of $\mathbf{z} \in \operatorname{relint}(F)$. In these cases we drop the $\mathbf{z}$ dependence and write it as N_F.
(e) K may be written as the disjoint union:

$$K = \bigcup \{\operatorname{relint} F : F \text{ is a face of } K\}. \tag{5}$$

Property (a) above leads to the following definition.

Definition 4 (Minimal face containing a set)**.** For a convex set K and a subset $A \subseteq K$ we define the *minimal face containing* A as

$$\bigcap \{F : F \text{ is a face of } K \text{ and } A \subseteq F\}$$

which is a face by property (a) above.

B. Dynamical systems

We will be using the notions of *flow, semiflow, convegence to a solution, ω-limit set of a (semi)flow, global convergence, non-expansive semiflow, Carathéodory solution* as defined in part I, section II-C.

The notions below will be additionally used in this paper.

Definition 5 (Invariant sets)**.** For a semiflow (ϕ, X, ρ) we say that a set $A \subseteq X$ is positively invariant if $\phi(\mathbb{R}_+, A) \subseteq A$. If ϕ is also a flow we say that A is negatively invariant if $\phi((-\infty, 0], A) \subseteq A$. If $\phi(t, A) = A$ for all $t \in \mathbb{R}$ then we say A is invariant.

Definition 6 (Sub-(semi)flow)**.** For a flow (resp. semiflow) (ϕ, X, ρ) and an invariant (resp. positively invariant) set $A \subseteq X$ we obtain the subflow (resp. sub-semiflow) by restricting $\phi(t, x)$ to act on $x \in A$ and denote it as (ϕ, A, ρ).

As it will be discussed in the paper, the ω-limit set of non-expansive semiflows, is comprised of semiflows of the class defined below.

Definition 7 ((Semi)Flow of isometries)**.** We say that a (semi)flow (ϕ, X, ρ) is a (semi)flow of isometries if for every $t \in \mathbb{R}$ (resp. $\mathbb{R}_+$), the function $\phi(t, \cdot) : X \to X$ is an isometry, i.e. for all $x, y \in X$ it holds that $\rho(\phi(t, x), \phi(t, y)) = \rho(x, y)$.

III. Problem formulation

The main object of study in this work is the *subgradient method* on an arbitrary concave-convex function in C^2 and an arbitrary convex domain K. We first recall the definition of the *gradient method*, which is studied in part I of this work [24].

Definition 8 (Gradient method)**.** Given φ a C^2 concave-convex function on $\mathbb{R}^{n+m}$, we define the *gradient method* as the flow on $(\mathbb{R}^{n+m}, d)$ generated by the differential equation

$$\begin{aligned} \dot{x} &= \varphi_x \\ \dot{y} &= -\varphi_y. \end{aligned} \tag{6}$$

The *subgradient method* is obtained by restricting the gradient method to a convex set K by the addition of a projection term to the differential equation (6).

Definition 9 (Subgradient method)**.** Given a non-empty closed convex set $K \subseteq \mathbb{R}^{n+m}$ and a function φ that is concave-convex on K and C^2 on an open neighbourhood of K, we define the *subgradient method on* K as a semiflow on (K, d) consisting of Carathéodory solutions of

$$\begin{aligned} \dot{\mathbf{z}} &= \mathbf{f}(\mathbf{z}) - \mathbf{P}_{N_K(\mathbf{z})}(\mathbf{f}(\mathbf{z})) \\ \mathbf{f}(\mathbf{z}) &= \begin{bmatrix} \varphi_x \\ -\varphi_y \end{bmatrix}. \end{aligned} \tag{7}$$

The equilibrium points of the subgradient method on K are K-restricted saddle points. If in addition the set K is the Cartesian product of two convex sets as in (1) then the set of equilibrium points of the subgradient method on K is equal to the set of K-restricted saddle points.

Remark 10. For (non-affine) convex sets K the subgradient method (7) is a *non-smooth* system. The vector field is discontinuous due to the convex projection term, independently of the regularity of the function φ or of the boundary of K. This is in contrast to the gradient method (6), which is a *smooth* system, as it inherits the regularity of the function φ.

We briefly summarise the contributions of this work in the bullet points below.

- We show that the subgradient dynamics, despite being nonlinear and non-smooth, have an ω-limit set that is comprised of solutions to only *linear* ODEs.
- These solutions are shown to belong to the $\omega-$limit set of the subgradient method on *affine subspaces*. This links with part I [24] of this two part work, where the limiting solutions of such systems have been exactly characterized. Based on this characterization of the limiting solutions, a convergence result for subgradient dynamics is also presented.
- Various examples that illustrate the results in the paper are presented. Applications are also provided to modification methods in network optimization that provide convergence guarantees while maintaining a decentralized structure in the dynamics. An application to the problem of multi-path routing is also discussed.

IV. Main Results

This section states the main results of the paper. The results are divided into three subsections. To facilitate the readability of section IV we outline below the main Theorems that will be presented and the way these are related.

In subsection IV-B we consider non-expansive semiflows, an abstraction we use for the subgradient dynamcis in order to develop tools for their analysis that are valid despite their non-smooth character. In particular, Proposition 14 gives an invariance principle for such semiflows, which applies without any smoothness assumption on the dynamics. We then additionally incorporate projections that constrain the trajectories within a closed convex set. Our key result, Theorem 18, says that for these semiflows the *dynamics on the ω-limit set are smooth*.

In subsection IV-C we apply these tools to the subgradient method (7). In Theorem 20 we show that the limiting solutions of the (non-smooth) subgradient method on a convex set are given by the dynamics of the (smooth) subgradient method on an *affine subspace*. This allows us to obtain Corollary 28, a criterion for global asymptotic stability of the subgradient method.

In subsection IV-D we combine Theorem 20 with the results of Part I of this work [24] (for convenience of the reader reproduced in subsection IV-A) to obtain a general convergence criterion (Theorem 31) for the subgradient method.

These results are illustrated with examples throughout. The proofs of the results are given in appendix A.

A. Subgradient method on affine subspaces

In this section we recall a result proved in part I of this work [24] on the limiting solutions of the subgradient method on affine subspaces. To state this result we recall from [24] the definition of the following matrices of partial derivatives of a concave-convex function $\varphi \in C^2$

$$\mathbf{A}(\mathbf{z}) = \begin{bmatrix} 0 & \varphi_{xy}(\mathbf{z}) \\ -\varphi_{yx}(\mathbf{z}) & 0 \end{bmatrix}$$
$$\mathbf{B}(\mathbf{z}) = \begin{bmatrix} \varphi_{xx}(\mathbf{z}) & 0 \\ 0 & -\varphi_{yy}(\mathbf{z}) \end{bmatrix}. \tag{8}$$

Consider the subgradient method (7) on an affine subspace V with normal cone N_V

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}) - \mathbf{P}_{N_V}(\mathbf{f}(\mathbf{z})) \tag{9}$$
$$\mathbf{f}(\mathbf{z}) = \begin{bmatrix} \varphi_x \\ -\varphi_y \end{bmatrix}.$$

Also let $\mathbf{\Pi} \in \mathbb{R}^{(n+m)^2}$ be the orthogonal projection matrix onto the orthogonal complement of N_V. Then the ODE (9) can be written as

$$\dot{\mathbf{z}} = \mathbf{\Pi}\mathbf{f}(\mathbf{z}) \tag{10}$$

The result is stated for $\mathbf{0}$ being an equilibrium point; the general case may be obtained by a translation of coordinates.

Theorem 11. *[24, Theorem 25] Let $\mathbf{\Pi} \in \mathbb{R}^{(n+m)^2}$ be an orthogonal projection matrix, φ be C^2 and concave-convex on $\mathbb{R}^{n+m}$, and $\mathbf{0}$ be an equilibrium point of* (10)*. Then the trajectories $\mathbf{z}(t)$ of* (10) *that lie a constant distance from any equilibrium point of* (10) *are exactly the solutions to the linear ODE:*

$$\dot{\mathbf{z}}(t) = \mathbf{\Pi}\mathbf{A}(\mathbf{0})\mathbf{\Pi}\mathbf{z}(t) \tag{11}$$

that satisfy, for all $t \in \mathbb{R}$ and $r \in [0, 1]$, the condition

$$\mathbf{z}(t) \in \ker(\mathbf{\Pi}\mathbf{B}(r\mathbf{z}(t))\mathbf{\Pi}) \cap \ker(\mathbf{\Pi}(\mathbf{A}(r\mathbf{z}(t)) - \mathbf{A}(\mathbf{0}))\mathbf{\Pi}) \tag{12}$$

where $\mathbf{A}(\mathbf{z})$ and $\mathbf{B}(\mathbf{z})$ are defined by (8)*.*

Corollary 12. *Let ϕ be concave-convex on a convex set $K \subseteq \mathbb{R}^{n+m}$, and C^2 on an open neighbourhood of K. Let $\mathbf{0} \in K$, and $\mathbf{0}$ be an equilibrium point of* (10)*. Then the trajectories $\mathbf{z}(t)$ of* (10) *that lie in K for all $t \in \mathbb{R}$ and are a constant distance from any equilibrium point of* (10) *in K, satisfy* (11) *and condition* (12)*.*

Remark 13. In the remainder of this paper we show that subgradient dynamics on a general convex domain that have an equilibrium point, have an ω-limit set that is comprised of solutions of subgradient dynamics on only an affine subspace and form a flow of isometries. In particular, the ω-limit set is comprised of solutions to explicit linear ODEs, of the form described in Theorem 11, despite the subgradient dynamics being nonlinear and non-smooth.

B. Non-expansive semiflows and convex projections

If one wishes to extend the results of Part I of this work [24] to the subgradient method on a non-empty closed convex set $K \subseteq \mathbb{R}^{n+m}$, then a complication arises from the discontinuity of the vector field in (7).

The main tool used to prove the results in [24] was the non-expansive property which says that the Euclidean distance between any two solutions is non-increasing with time. One would expect that the distance between any two of the limiting solutions would be constant. A more abstract way of saying this is that the sub-flow obtained by considering the gradient method with initial conditions in the ω-limit set is a *flow of isometries*. In fact, this can be proved for any non-expansive semiflow, as stated in Proposition 14 below (proved in Appendix A-B).

Proposition 14. *Let (ϕ, X, d) be a non-expansive semiflow with $X \subseteq \mathbb{R}^{n+m}$ which has an equilibrium point $\bar{\mathbf{z}}$. Let Ω be the ω-limit set of the semiflow. Then the sub-semiflow (ϕ, Ω, d) defines a flow of isometries. Moreover, Ω is a convex set.*

Note here that (ϕ, Ω, d) is a *flow* rather than a *semiflow*. This comes from the simple observation that an isometry is always invertible, so we can define, for $t \geq 0$, $\phi(-t, \cdot) : \Omega \to \Omega$ as $\phi(t, \cdot)^{-1}$.

Remark 15. Care should be taken in interpreting the backwards flow given by Proposition 14. There could be multiple trajectories in X that meet at a point in $y \in \Omega$ at time $t = 0$, but exactly one of these trajectories will lie in Ω for all times $t \in \mathbb{R}$.

We would like to note that we are not the first to make this observation. Indeed, we deduce this result from a more general result in [10] which was published in 1970.

It should be noted that if a non-expansive semiflow has an equilibrium point then all its trajectories are bounded. This implies that each trajectory converges to its set of ω-limit points [36, Lemma 4.16]. The structure of the ω-limit set can also be used to strengthen the convergence to the ω-limit set to convergence to a solution in the ω-limit set. This is stated as Corollary 16 below (proved in Appendix A-B).

Corollary 16. *Let (ϕ, X, d) be a non-expansive semiflow with $X \subseteq \mathbb{R}^{n+m}$, which has an equilibrium point. Then each trajectory of the semiflow converges to a trajectory in its ω-limit set.*

We consider now non-expansive differential equations which are projected onto a convex set, and make the following set of assumptions.

$$\begin{gathered}(\phi, K, d) \text{ is the semiflow given by Carathéodory solutions of} \\ \dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}) - \mathbf{P}_{N_K(\mathbf{z})}(\mathbf{f}(\mathbf{z})) \text{ where,} \\ K \subseteq \mathbb{R}^{n+m}, \text{ is non-empty, closed and convex} \\ \mathbf{f} \in C^1,\ \mathbf{f} : U \to \mathbb{R}^{n+m}, U \text{ open}, K \subseteq U \subseteq \mathbb{R}^{n+m}, \\ \text{for all } \mathbf{z}, \mathbf{w} \in K \quad (\mathbf{f}(\mathbf{z}) - \mathbf{f}(\mathbf{w}))^T(\mathbf{z} - \mathbf{w}) \leq 0.\end{gathered} \tag{13}$$

Functions $-\mathbf{f}(\mathbf{z})$ such that $\mathbf{f}(\mathbf{z})$ satisfies the final inequality in (13) are referred to as monotone. A known result in the literature is the fact that the semiflow in (13) is non-expansive[2], which is stated as Lemma 17 below [20], [41]. Existence of unique solutions $\mathbf{z} : [0, \infty) \to \mathbb{R}^{n+m}$ can be deduced from

[2]In particular note that for any two trajectories $\mathbf{z}, \mathbf{z}'$ we have that $W(t) = \frac{1}{2}|\mathbf{z}(t) - \mathbf{z}'(t)|^2$, satisfies for almost all times $t \geq 0$, $\dot{W}(t) = (\mathbf{z}(t) - \mathbf{z}'(t))^T(\dot{\mathbf{z}}(t) - \dot{\mathbf{z}}'(t)) \leq 0$ where the inequality follows from (13) and the definition of the normal cone.

corresponding results for projected dynamical systems [9],[8, Theorem 3.2] (see also [41], [20], [4], [5]).

Lemma 17. *Let* (13) *hold. Then* (ϕ, K, d) *is non-expansive.*

Our main result on projected differential equations as in (13) is that, even though the projection term gives a discontinuous vector field, when we restrict our attention to the ω-limit set, the vector field is C^1. This allows us to replace *non-smooth analysis* with *smooth analysis* when studying the asymptotic behaviour of such systems.

Theorem 18. *Let* (13) *hold and assume that the semiflow* (ϕ, K, d) *has an equilibrium point. Let* Ω *be the* ω*-limit set of the semiflow. Then* (ϕ, Ω, d) *defines a flow of isometries given by solutions to the following differential equation, which has a* C^1 *vector field,*

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}) - \mathbf{P}_{N_V}(\mathbf{f}(\mathbf{z})). \tag{14}$$

Here V *is the affine span of the (unique) minimal face of* K *that contains the set of equilibrium points of the semiflow.*

The proof of Theorem 18 is provided in Appendix A-B.

Remark 19. The existence of a minimal face of K that contains the set of equilibrium points is a simple consequence of the definition of a face (see Definition 3 and the discussion that follows). The significance of minimal face flows was also noted in [48], where these have been used as a tool to deduce local stability properties for projected dynamical systems. In Theorem 18 we show that such flows can provide a characterization to the ω-limit set of dynamical systems of the form (13), which are non-expansive. Noting also that (14) is a dynamical system on an affine subspace, Theorem 11, Corollary 12 can be used to provide a characterization to the ω-limit set of subgradient dynamics as linear ODEs, as it will be discussed in the next section.

C. The subgradient method

We now apply Theorem 18 to the subgradient method. Our first result reduces the study of the convergence on general convex domains, where the subgradient method is non-smooth, to the study of convergence of the subgradient method on affine spaces, which is a smooth dynamical system studied in [24]. We also show that when an internal saddle point exists then the limiting behaviour of the subgradient method is determined by that of the corresponding unconstrained gradient method.

For a function φ that is concave-convex on a convex set $K \subseteq \mathbb{R}^{n+m}$ and C^1 on an open neighbourhood of K we denote $\mathcal{S}_K$ the set of solutions of the *gradient method* (6) (i.e. no projections included), that are in K for all times $t \in \mathbb{R}$ and are a constant distance from any equilibrium point of (6) in K.

Theorem 20. *Let function* φ *be concave-convex on on a set* $K \subseteq \mathbb{R}^{n+m}$ *as defined in* (1) *and* C^2 *on an open neighbourhood of* K*, and let* φ *have a* K*-restricted saddle point. Let* (ϕ, K, d) *denote the subgradient method* (7) *on* K *and* Ω *be its* ω*-limit set. Then* Ω *is convex, and* (ϕ, Ω, d) *defines a flow of isometries. Furthermore, the following hold:*

(i) The trajectories $\mathbf{z}(t)$ *of* (ϕ, Ω, d) *solve the ODE:*

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}) - \mathbf{P}_{N_V}(\mathbf{f}(\mathbf{z})), \tag{15}$$

where V *is the affine span of* F*, with* F *being the minimal face containing all* K*-restricted saddle points.*

(ii) If there exists a saddle point of φ *in the interior of* K*, then*

$$\Omega \subseteq \mathcal{S}_K \tag{16}$$

The proof of Theorem 20 is provided in Appendix A-C.

Remark 21. The ODE (15) is the subgradient method on the affine subspace V. A main significance of Theorem 20 is the fact that the solutions of (15) in Ω can be characterized using the results in part I [24]. In particular, it follows from Theorem 11, Corollary 12 in section IV-A that these satisfy explicit linear ODEs. This therefore shows that even though the subgradient dynamics are nonlinear and non-smooth their ω-limit set is comprised of solutions to only *linear* ODEs (stated in Corollary 34).

Remark 22. Later, in subsection IV-D we use the results in [24] on the subgradient method on affine subspaces (section IV-A) together with Theorem 20 to obtain a convergence criterion for the subgradient method. This is used subsequently to give proofs for the applications considered in section V.

Remark 23. It will be discussed in the proof of Theorem 20 that Theorem 20(ii) is a special case of Theorem 20(i) where the projection term in (15) equal to zero. In Theorem 20(ii) there is a characterization of the ω-limit set of the subgradient method, as solutions of the corresponding *gradient method*.

Remark 24. A simple consequence of (16) is the fact if the function ϕ is concave-convex on $\mathbb{R}^{n+m}$ and has a saddle point in the interior of K then the subgradient method on K is globally convergent (i.e. converges to a saddle point for any initial condition in K) if the corresponding unconstrained gradient method is globally convergent.

Remark 25. Theorem 20 follows directly from Theorem 18. Hence Theorem 20 also holds when K is an arbitrary closed convex set as in Theorem 18 (rather than just the Cartesian product of two convex sets), if the subgradient method has an equilibrium point. Note that a saddle point in the interior of K, or a K-restricted saddle point with K as in (1), is always an equilibrium point of the subgradient method on K.

We now present several examples to illustrate the application of Theorem 20 in some simple cases.

The first example corresponds to a case where the unconstrained gradient method (6) is globally convergent, but the subgradient method is not.

Example 26. *Define the concave-convex function*

$$\varphi(x_1, x_2, y) = -\frac{1}{2}|x_1|^2 + (x_1 + x_2)y \tag{17}$$

where $x_1, x_2, y \in \mathbb{R}$*. This has a single saddle point at* $(0, 0, 0)$*, and* φ *is the Lagrangian of the optimisation problem*

$$\max_{x_1+x_2=0} -\frac{1}{2}|x_1|^2 \tag{18}$$

where variable y in function φ is the Lagrange multiplier associated with the constraint. On this function the gradient method is the linear system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{y}\end{bmatrix} = \begin{bmatrix}-1 & 0 & 1\\ 0 & 0 & 1\\ -1 & -1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ y\end{bmatrix}. \tag{19}$$

It is easily verified that all the eigenvalues of this matrix lie in the left half plane, so that the gradient method is globally convergent. Now consider the family of convex sets defined by

$$K_a = \{(x_1, x_2, y) \in \mathbb{R}^3 : x_1 \geq a\} \tag{20}$$

for $a \in \mathbb{R}$. The subgradient method on K_a is given by the system

$$\begin{aligned}\dot{x}_1 &= [-x_1 + y]^+_{x_1 - a}\\ \dot{x}_2 &= y\\ \dot{y} &= -x_1 - x_2.\end{aligned} \tag{21}$$

The convergence of the subgradient method on K_a depends crucially on the value of a. There are three cases:

- *(i)* $a < 0$*: In this case the saddle point $(0, 0, 0)$ lies in the interior of K_a so that Theorem 20(ii) applies, and as the unconstrained gradient method is globally convergent, so is the subgradient method on K_a.*
- *(ii)* $a > 0$*: Here the unconstrained saddle point $(0, 0, 0)$ lies outside K_a. A simple computation shows that the point $(a, -a, 0)$ is the only K_a-restricted saddle point. Theorem 20(i) can be used here. The only proper face of K_a is the set*

$$F_a = \{(a, x_2, y) : x_2, y \in \mathbb{R}\}. \tag{22}$$

The subgradient method on F_a is the system

$$\begin{bmatrix}\dot{x}_2\\ \dot{y}\end{bmatrix} = \begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}x_2\\ y\end{bmatrix} + \begin{bmatrix}0\\ -a\end{bmatrix} \tag{23}$$

together with the equality $x_1 = a$. This matrix has imaginary eigenvalues $\pm i$, showing that the subgradient method on F_a is not globally convergent. It is easy to verify that some of these oscillatory solutions are also solutions of the subgradient method on K_a, e.g. $y(t) = a\cos(t)$, $x_2(t) = -a(1 - \sin(t))$, $x_1(t) = a$ satisfy (21)*. Therefore the subgradient method on K_a is not globally convergent when $a > 0$.*

- *(iii)* $a = 0$*: In this case the saddle point $(0, 0, 0)$ lies on the boundary of K_0. Theorem 20(i) applies, and the analysis of the subgradient method on F_0 is the same as in case (ii) above. However, when we check whether any oscillatory solutions of the subgradient method on F_0 are also solutions of the subgradient method on K_0, we find that there are no such solutions. Indeed, for a trajectory to be a solution to both the subgradient method on F_0 and the subgradient method on K_0 we must have both $x_1 = a = 0$ and $-x_1 + y \leq 0$ by* (21)*. Then* (21) *implies that $y = 0$ and then that $x_1 = 0$. So the only such solution is the saddle point. Therefore the subgradient method on K_0 is globally convergent.*

This shows that the subgradient method on K_a undergoes a bifurcation at $a = 0$.

The following example illustrates that the subgradient method can be globally convergent when the gradient method is not.

Example 27. *Define the concave-convex function*

$$\varphi(x_1, x_2, y) = -\frac{1}{2}|x_2|^2 + x_1 y. \tag{24}$$

This has a single saddle point at $(0, 0, 0)$ and corresponds to the optimisation problem

$$\max_{x_1 = 0} -\frac{1}{2}|x_2|^2 \tag{25}$$

where the constraint is relaxed via the Lagrange multiplier y. The gradient method applied to φ is the linear system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{y}\end{bmatrix} = \begin{bmatrix}0 & 0 & 1\\ 0 & -1 & 0\\ -1 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ y\end{bmatrix} \tag{26}$$

whose matrix has eigenvalues $-1, \pm i$ so the gradient method is not globally convergent. We again consider the subgradient method on the closed convex set K_a defined by (20) *for $a \in \mathbb{R}$ splitting into three cases:*

- *(i)* $a < 0$*: As in Example 26(i) the saddle point $(0, 0, 0)$ lies in the interior of K_a. As the unconstrained gradient method is not globally convergent, Theorem 20(ii) implies that the subgradient method on K_a is also not globally convergent.*
- *(ii)* $a > 0$*: The subgradient method on K_a is given by*

$$\begin{aligned}\dot{x}_1 &= [y]^+_{x_1 - a}\\ \dot{x}_2 &= -x_2\\ \dot{y} &= -x_1\end{aligned} \tag{27}$$

The saddle point $(0, 0, 0)$ lies outside K_a. For $(\bar{x}_1, \bar{x}_2, \bar{y})$ to be a K_a-restricted saddle point, (27) *implies that $\bar{x}_1 = \bar{x}_2 = 0$, but this is impossible in K_a, so there are no K_a-restricted saddle points. This can also be understood in terms of the optimisation problem* (25) *which has empty feasible set if we impose the further condition that $x_1 \geq a > 0$. This means that none of our results apply, but a direct analysis of* (27) *shows that $\dot{y} \leq -a < 0$ so that $y(t) \to -\infty$ as $t \to \infty$, and the system is not globally convergent.*

- *(iii)* $a = 0$*: Solving* (27) *for the K_0-restricted saddle points yields the continuum $\{(0, 0, y) : y \leq 0\}$. None of these lie in the interior of K_0, so Theorem 20(ii) does not apply and Theorem 20(i) is used to analyze the asymptotic behaviour. The only proper face of K_0 is F_0 defined by* (22)*. On F_0, the subgradient method is the system*

$$\begin{bmatrix}\dot{x}_2\\ \dot{y}\end{bmatrix} = \begin{bmatrix}-1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}x_2\\ y\end{bmatrix} \tag{28}$$

together with the equality $x_1 = 0$. This is globally convergent, since for all initial conditions x_2 converges to 0 and hence we have convergence to a point in the set $\{(0, 0, y) : y \in \mathbb{R}\}$, which is the set of F_0-restricted saddle points . Therefore the subgradient method on K_0 is also globally convergent.

So in this case the subgradient method on K_a starts non-convergent for $a < 0$, becomes globally convergent for $a = 0$ and finally looses all its equilibrium points when $a > 0$.

Although the minimal face F in Theorem 20(i) is given as the intersection of all faces that contain K-restricted saddle points, it can be useful to obtain convergence criteria that do not depend upon knowledge of all K-restricted saddle points. We note that if the subgradient method is globally convergent on any affine span of a face of K, then global convergence is implied.

Corollary 28. *Let function φ be concave-convex on a set $K \subseteq \mathbb{R}^{n+m}$ as defined in* (1)*, and C^2 on an open neighbourhood of K. Let φ have a K-restricted saddle point. Assume that, for any face F of K that contains a K-restricted saddle point, the subgradient method on $\mathrm{aff}(F)$ is globally convergent. Then the subgradient method on K is globally convergent.*

Example 29. *To illustrate this result, let us consider the case of positivity constraints, where (x, y) are restricted to $K = \mathbb{R}^n_+ \times \mathbb{R}^m_+$. Here the faces of K are given by sets of the form*

$$\{(x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^m_+ : x_i = 0, y_j = 0 \text{ for } i \notin I, j \notin J\}$$

where $I \subseteq \{1, \ldots, n\}$ and $J \subseteq \{1, \ldots, m\}$ are sets of indices. The affine span of such a face is then given by

$$\{(x, y) \in \mathbb{R}^{n+m} : x_i = 0, y_j = 0 \text{ for } i \notin I, j \notin J\}. \tag{29}$$

Thus, by Corollary 28, checking convergence of the subgradient method in this case may be done by checking convergence of the gradient method with any arbitrary set of coordinates fixed as zero[3]*.*

In some cases the faces of the constraint set K have an interpretation in terms of the specific problem.

Example 30. *Consider the optimisation problem*

$$\max_{g_j(x) \geq 0, j \in \{1,\ldots,m\}} U(x) \tag{30}$$

where $U, g_j : \mathbb{R}^n \to \mathbb{R}$ are concave functions in C^2. This is associated with the Lagrangian

$$\varphi(x, y) = U(x) + \sum_{j \in \{1,\ldots,m\}} y_j g_j(x) \tag{31}$$

where $y \in \mathbb{R}^m$ is a vector of Lagrange multipliers[4]*. To ensure that the Lagrange multipliers are non-negative we define the constraint set $K = \mathbb{R}^n \times \mathbb{R}^m_+$. As in Example 29 the affine spans of the faces of K are given by* (29) *for $I = \{1, \ldots, m\}$ and J any subset of $\{1, \ldots, m\}$. The subgradient method applied on such a face corresponds to the gradient method on the modified Lagrangian*

$$\varphi'(x, y) = U(x) + \sum_{j \in J} y_j g_j(x) \tag{32}$$

which is associated with the modified optimisation problem

$$\max_{g_j(x) = 0, j \in J} U(x) \tag{33}$$

where, compared to (30)*, the inequality constraints are replaced by equality constraints, and some subset of the constraints are removed.*

If φ is concave-convex on $\mathbb{R}^{n+m}$ then Corollary 28 applies. We obtain that the subgradient method on K applied to φ is globally convergent, if, for any $J \subseteq \{1, \ldots, m\}$, the gradient method applied to the Lagrangian φ' corresponding to the modified optimisation problem (33) *is globally convergent.*

[3] This result was presented previously by the authors in [22].

[4] For simplicity of presentation we shall assume throughout the example that there is no duality gap in the problems considered.

D. A general convergence criterion

By combining Theorem 20 with the results on the limiting solutions of the (smooth) subgradient method on affine subspaces given in [24] (recalled in section IV-A) we obtain the following convergence criterion for the subgradient method on arbitrary convex sets and arbitrary concave-convex functions in C^2. This states that the subgradient method is globally convergent, if it has no trajectory satisfying an explicit linear ODE.

The theorem is stated under the assumption that $\mathbf{0} \in K$ is a K-restricted saddle point. The general case is obtained by a translation of coordinates.

Theorem 31. *Let function φ be concave-convex on a set $K \subseteq \mathbb{R}^{n+m}$ as defined in* (1)*, and C^2 on an open neighbourhood of K, and let $\mathbf{0} \in K$ be a K-restricted saddle point of φ. Let F be the minimal face of K that contains all K-restricted saddle points and let V be the affine span of F. Let $\mathbf{\Pi}$ be the orthogonal projection matrix onto the orthogonal complement of N_V. Let also $\mathbf{A}(.)$ and $\mathbf{B}(.)$ be the matrices defined in* (8)*.*

Then if the subgradient method (7) *on K applied to φ has no non-constant trajectory $\mathbf{z}(t)$ that satisfies both the following*

(i) the linear ODE

$$\dot{\mathbf{z}}(t) = \mathbf{\Pi}\mathbf{A}(\mathbf{0})\mathbf{\Pi}\mathbf{z}(t) \tag{34}$$

(ii) for all $r \in [0, 1]$ and $t \in \mathbb{R}$,

$$\mathbf{z}(t) \in \ker(\mathbf{\Pi}\mathbf{B}(r\mathbf{z}(t))\mathbf{\Pi}) \cap \ker(\mathbf{\Pi}(\mathbf{A}(r\mathbf{z}(t)) - \mathbf{A}(\mathbf{0}))\mathbf{\Pi}), \tag{35}$$

then the subgradient method is globally convergent.

The proof of Theorem 31 is provided in Appendix A-D.

Remark 32. Although the condition (35) appears difficult to verify, it is only necessary to show that the condition *does not* hold (by non-trivial trajectories) in order to prove global convergence. This turns out to be easy in many cases, for example in the proofs of the convergence of the modification methods discussed in section V (Theorem 41). In particular, these are examples where global convergence is desired to a saddle point without knowing the saddle points a priori and without function φ satisfying strictness properties that guarantee convergence. The derivations exploit the structure of the matrices $\mathbf{A}, \mathbf{B}, \mathbf{\Pi}$ to prove that (34) and (35) are satisfied only by saddle points.

Remark 33. It should be noted that (34) and (35) are satisfied by all trajectories $\mathbf{z}(t)$ in the ω-limit set of the subgradient method. This follows from Theorem 20 and Theorem 11, Corollary 12 and is stated in the corollary below (proved in Appendix A-D).

Corollary 34. *Consider the subgradient method* (7) *and let* $\mathbf{0}$ *be a* K*-restricted saddle point. Then any trajectory* $\mathbf{z}(t)$ *in the* ω*-limit set satisfies* (34) *and* (35)*, i.e. it is a solution of a linear ODE.*

V. Applications

In this section we apply the results of section IV to obtain global convergence in a number cases. In particular, we look at examples of modification methods, relevant in network optimization, where the concave-convex function is modified to provide guarantees of convergence. The application of one such modification method to the problem of multi-path routing is also discussed.

The proofs for this section are provided in appendix B.

A. Modification methods for convergence

We will consider methods for modifying φ so that the (sub)gradient method converges to a saddle point. The methods that will be discussed are relevant in network optimisation (see e.g. [1], [16]), as they preserve the localised structure of the dynamics. It should be noted that these modifications do not necessarily render the function strictly[5] concave-convex and hence convergence proofs are more involved. We show below that the results in section IV provide a systematic and unified way of proving convergence by making use of Theorem 31, while also allowing to consider these methods in a generalized setting of a general convex domains for the variables x, y respectively, in the concave-convex function $\varphi(x, y)$.

1) Auxiliary variables method: Given function φ concave-convex on a convex set K as in (1), and C^2 on an open neighbourhood U of K, we define the modified function $\varphi' : \mathbb{R}^{n'} \times U \to \mathbb{R}$ as

$$\begin{aligned} \varphi'(x', x, y) &= \varphi(x, y) + \psi(Mx - x') \\ \psi : \mathbb{R}^{n'} \to \mathbb{R}, \psi \in C^2, &\text{ is strictly concave} \\ \text{with } \psi(0) = 0, \psi(u) &\le 0, \end{aligned} \tag{36}$$

where x' is a vector of n' auxiliary variables, and $M \in \mathbb{R}^{n' \times n}$ is a constant matrix that satisfies $\ker(M) \cap \ker(\varphi_{xx}(\bar{\mathbf{z}})) = \{0\}$ for a K-restricted saddle point $\bar{\mathbf{z}}$ of φ.

We define the augmented convex domain as $K' = \mathbb{R}^{n'} \times K$. Note that the additional auxiliary variables are not restricted and are allowed to take values in the whole of $\mathbb{R}^{n'}$. Also note that the $n \times n$ identity matrix always satisfies the assumptions upon M above.

Remark 35. An important feature of this modification (and also the ones that will be considered below) is the fact that there is a correspondence between K-restricted saddle points of φ and K'-restricted saddle points of φ', with the values of x, y at the saddle points remaining unchanged. In particular, if $(\bar{x}, \bar{y})$ is a K-restricted saddle point of φ, then $(M\bar{x}, \bar{x}, \bar{y})$ is a K'-restricted saddle point of φ'. In the reverse direction, if $(\bar{x}', \bar{x}, \bar{y})$ is a K'-restricted saddle point of φ' then $M\bar{x} = \bar{x}'$ and $(\bar{x}, \bar{y})$ is a K-restricted saddle point of φ.

[5]It should be noted that the relaxed strictness conditions in [45] are also not necessarily satisfied.

Remark 36. The significance of this method will become more clear in the multipath routing problem discussed in Appendix V-B. In particular, this method allows convergence to be guaranteed in network optimization problems without introducing additional information transfer among nodes. Special cases of this method have also been used in [13], [23] in applications in economic and power networks.

2) Penalty function method: For this and the next method we will assume that the concave-convex functions φ is a Lagrangian originating from a concave optimization problem (see subsection II-A2). We will assume that the Lagrangian φ satisfies

$$\begin{aligned} &\varphi(x, y) = U(x) + y^T g(x) \\ &C^2 \ni U : \mathbb{R}^n \to \mathbb{R} \text{ is concave} \\ &C^2 \ni g : \mathbb{R}^n \to \mathbb{R}^m \text{ is concave.} \end{aligned} \tag{37}$$

We consider a so called penalty method (see e.g. [18]). This method adds a penalising term to the Lagrangian based directly on the constraint functions. The new Lagrangian φ' is defined by

$$\begin{aligned} \varphi'(x, y) &= \varphi(x, y) + \psi(g(x)) \\ C^2 \ni \psi : \mathbb{R}^m \to \mathbb{R} &\text{ is strictly concave with } \psi_u > 0 \\ \psi(u) = 0 &\iff u \ge 0. \end{aligned} \tag{38}$$

It is easy to see that the saddle points of φ and φ' are the same.

Remark 37. This modification method is also often applied to network optimization problems, i.e. problems where $U(x)$ is of the form $U(x) = \sum_i U_i(x)$ and each of the $U_i(x)$ is a function of only a few of the components of x. Similarly each component, $g_i(x)$, of the constraints $g(x)$ depends on only a few of the components of x. The subgradient method for such problems applied to (37) has a decentralized structure. When applied to the modified version (38) the dynamics will still have a decentralized structure, but will often also involve additional information exchange between neighboring nodes, e.g. when $g(x)$ is linear, due to the nonlinearity of the function $\psi(.)$.

Remark 38. This method has been considered previously (see [16] and the references therein[6]), either without constraints, or with positivity constraints, i.e. $K = \mathbb{R}^n_+ \times \mathbb{R}^m_+$. Theorem 41 below applies to all non-empty closed sets $K \subseteq \mathbb{R}^{n+m}$ which are a product set of convex sets as in (1).

3) Constraint modification method: We next recall a method proposed in [1]. Here we instead modify the constraints to enforce strict concavity. The Lagrangian (37) is modified to become:

$$\begin{aligned} &\varphi'(x, y) = U(x) + y^T \psi(g(x)) \\ &C^2 \ni U : \mathbb{R}^n \to \mathbb{R} \text{ is concave} \\ &C^2 \ni g : \mathbb{R}^n \to \mathbb{R}^m \text{ is concave} \\ &C^2 \ni \psi = [\psi^1, \dots, \psi^m]^T : \mathbb{R}^m \to \mathbb{R}^m \\ &\psi^j(0) = 0, \psi^j_u \ge 0 \text{ and } \psi^j_{uu} < 0 \text{ for } j = 1, \dots m. \end{aligned} \tag{39}$$

[6]Note that a related modification method in discrete time is the ADMM method [2], [19].

It is clear that the value of x at the saddle points of the modified and original Lagrangian will be the same. In analogy with Remark 37, this method also preserves the decentralized structure of the subgradient method for network optimization problems, but may require additional information transfer.

Remark 39. Previous works [1],[6] have proved convergence of this method with positivity constraints, i.e. $K = \mathbb{R}^n_+ \times \mathbb{R}^m_+$. Theorem 41 below applies to any constraint set K which is a product set of convex sets as in (1).

Remark 40. It should be noted that even though the three methods described in this section all provide global convergence guarantees, they lead to different information structures in the underlying dynamics when applied to network optimization problems. In particular, the auxiliary variable method leads to fully decentralized implementations, whereas the other two can require additional information transfer among nodes. This will be illustrated in subsection V-B where the multipath routing problem will be discussed.

4) Convergence results: We now give a global convergence result for each of the methods described above on general convex domains.

Theorem 41 (Convergence of modification methods)**.** *Let $K \subseteq \mathbb{R}^{n+m}$ be a non-empty closed set that is the Cartesian product of two convex sets as in* (1)*, and assume that φ, φ' satisfy one of the following:*

1) Auxiliary variable method: *Let $\varphi \in C^2$ be concave-convex and φ', K' be defined by* (36) *and the text directly below it.*
2) Penalty function method: *Let φ have the form* (37)*, and φ' be defined by* (38)*.*
3) Constraint modification method: *Let φ have the form* (37)*, and φ' be given by* (39)*.*

Also assume that φ has a K-restricted saddle point. Then the subgradient method (7) *applied to φ' on K' in* 1) *and K in* 2), 3) *is globally convergent.*

Remark 42. Each of the convergence results in Theorem 41 is proved using Theorem 31. It should also be noted that the modification methods do not necessarily produce a strictly concave-convex function φ', or lead to a φ' that strictly decreases (increases) for all deviations from the set of saddle points due to a change in x, x' (change in y). Global convergence to a saddle point is still though guaranteed by ensuring that no trajectory, other than saddle points, satisfy conditions (34), (35) in Theorem 31. A specific such example will be studied in the next section.

B. Multi-path congestion control

Combined control of routing and flow is a problem that has received considerable attention within the communications literature due to the significant advantages it can provide relative to congestion control algorithms that use single paths [30]. Nevertheless its implementation is not directly obvious as the availability of multiple routes can render the network prone to route flapping instabilities [47], [26], [46].

A classical approach to analyse such algorithms is to formulate them as solving a corresponding network optimization problem [29], [43] with primal/dual update rules leading to a decentralized implementation. This optimization problem is, however, not strictly concave and modifications that make it strictly concave can lead to a deviation from the optimal solution [46], [17].

Here we consider a multi-path routing problem with a fixed number of routes per source/destination pair, as in [29], [46], [31], [33]. For such schemes we investigate algorithms that allow the corresponding network optimization problem to be solved without requiring any relaxation in its solution or any additional information exchange.

1) Problem formulation: We consider a network that consists of sources $s_1, \dots, s_m$, routes $r_1, \dots, r_n$, and links $l_1, \dots, l_l$. Each source s_i is associated with a unique destination for a message which is to be routed each source also has a fixed set of routes associated with it.. Every route r_j has a unique source s_i, and we write $r_j \sim s_i$ to mean that s_i is the source associated with route r_j. Routes r_j each use a number of links, and we write $r_j \sim l_k$ to mean that the link l_k is used by the route r_j. The desired running capacity of the link l_k is denoted C_k, and $0 \leq C \in \mathbb{R}^l$ is the vector of these capacities. We let A be the connectivity matrix, so that $A_{kj} = 1$ if $l_k \sim r_j$ and 0 otherwise. In the same way we set $H_{ij} = 1$ if $s_i \sim r_j$ and 0 otherwise. x_j denotes the current usage of the route r_j. We associate to each source s_i a strictly concave, increasing utility function U_i.

We consider the problem of maximising total utility

$$\max_{x \geq 0, Ax \leq C} \sum_{s_i} U_i \left(\sum_{r_j \sim s_i} x_j \right). \tag{40}$$

Here the first sum is over all sources s_i, and the second over routes r_j with $r_j \sim s_i$ (we shall use such notation throughout this section). This optimisation problem is associated with the Lagrangian

$$\varphi(x, y) = \sum_{s_i} U_i \left(\sum_{r_j \sim s_i} x_j \right) + y^T (C - Ax). \tag{41}$$

where $y \in \mathbb{R}^l_+$ are the Lagrange multipliers. Note that even though $U_i(.)$ is strictly concave, this is not strictly concave in (40) with respect to the decision variables x_i, hence the Lagrangian $\varphi(x, y)$ is not strictly concave-convex.

A common approach in the context of congestion control is to consider primal-dual dynamics originating from this Lagrangian so as to deduce decentralized algorithms for solving the network optimisation problem (40) [29],[43]. This gives rise to the subgradient dynamics

$$\begin{aligned} \dot{x}_j &= \left[U_i' \left(\sum_{s_i \sim r_k} x_k \right) - \sum_{l_k \sim r_j} y_k \right]^+_{x_j} \\ \dot{y}_k &= \left[\sum_{l_k \sim r_j} x_j - C_k \right]^+_{y_k} \end{aligned} \tag{42}$$

where $s_i \sim x_j$ in the equation for $\dot{x}_j$ and U_i' is the derivative of the utility function U_i. Note that the equilibrium points of (42) are saddle points of the Lagrangian (under the positivity constraints on x and y) and hence also solutions of the optimization problem (40) (Slater's condition is assumed to hold).

Remark 43. The dynamics (42) are the subgradient method (7) applied to the Lagrangian (41) on the positive orthant $\mathbb{R}_+^{n+l}$.

The dynamics (42) are also localised in the sense that the update rules for x_j depend only on the current usage, x_k, of routes with the same source and of the congestion signals associated with links on these routes. In the same way the update rules for congestion signals y_k depend only on the usage of routes using the associated link.

2) Instability: The dynamics (42) inherit the stability properties of the subgradient method discussed in section IV. In particular the distance of $(x(t), y(t))$ from any saddle point $(\bar{x}, \bar{y})$ is non-increasing. However, the lack of strict concavity of the Lagrangian (41) leads to a lack of global convergence of the dynamics (42) in some situations as we shall describe below.

We assume for simplicity that there is a strictly positive saddle point $\bar{\mathbf{z}} > 0$. In this situation Theorem 20(ii) applies, and the convergence properties are the same as those of the unconstrained gradient method. The structure of the problem suggests an application of [24, Theorem 21]. Here a simple computation yields that $\mathcal{S}_{\text{linear}}$ is equal to $\bar{\mathcal{S}}$ (we use the notation of [24]) unless the following *algebraic* condition on the network topology holds:

$$\exists u \in \ker(H) \setminus \{0\}, \lambda > 0 \text{ such that } A^T A u = \lambda u. \quad (43)$$

[24, Theorem 21] tells us that global convergence holds if (43) does not hold, but in fact more is true.

Proposition 44. *Let* $\bar{\mathbf{z}} = (\bar{x}, \bar{y}) > 0$ *be a saddle point of* φ *defined by* (41) *and* $U_i \in C^2$ *be be strictly concave and strictly increasing. Then the dynamics* (42) *are globally convergent if and only if* (43) *does not hold.*

The algebraic criterion (43) on the network topology is satisfied by many networks, for example the network in Figure 1.

We also remark that under the condition (43), the system is sensitive to noise in the sense that the unconstrained dynamics satisfy the conditions of [24, Theorem 22].

3) Modified dynamics: Here we present a modification of the dynamics (42), that, while still fully localised, gives guaranteed convergence to an optimal solution of (40).

We use the auxiliary variables method described in subsection V-A1 and define the modified optimisation problem

$$\max_{\substack{x \geq 0, x' \in \mathbb{R}^n \\ Ax \leq C}} \sum_{s_i} U_i \left(\sum_{r_j \sim s_i} x_j \right) - \frac{1}{2} \sum_{r_k} \kappa_k |x_k' - x_k|^2 \quad (44)$$

where $x' \in \mathbb{R}^n$ is an additional vector to be optimised over, and $\kappa_k > 0$ are arbitrary constants. It is important to note that this has the same optimal x points as (40). This gives rise to a modified Lagrangian

$$\begin{aligned} \varphi'(x', x, y) = \sum_{s_i} U_i \left(\sum_{r_j \sim s_i} x_j \right) + y^T (C - Ax) \\ - \frac{1}{2} \sum_{r_k} \kappa_k |x_k' - x_k|^2. \end{aligned} \quad (45)$$

The new dynamics are given by the following subgradient method.

$$\begin{aligned} \dot{x}_j &= \left[U_i' \left(\sum_{s_i \sim r_k} x_k \right) - \sum_{l_k \sim r_j} y_k + \kappa_j (x_j' - x_j) \right]_{x_j}^+ \\ \dot{x}_j' &= \kappa_j (x_j - x_j') \\ \dot{y}_k &= \left[\sum_{l_k \sim r_j} x_j - C_k \right]_{y_k}^+ . \end{aligned} \quad (46)$$

Remark 45. The dynamics (46) are the subgradient method (7) applied to the modified Lagrangian (45) on $\mathbb{R}_+^n \times \mathbb{R}^n \times \mathbb{R}_+^l$. The Lagrangian (45) corresponds to (36) with $\psi(z) = -|z|^2/2$ and M the $n \times n$ identity matrix.

It is apparent (as discussed in subsection V-A1) that the equilibrium points of the modified dynamics (46) and the original dynamics (42) are in correspondence. We remark that the new dynamics are analogous to the addition of a low pass filter to the unmodified dynamics (42).

These dynamics are still localised. Each route r_k is now associated with its usage, x_k, and a new variable x_k'. To update x_k the only additional information required over the unmodified scheme is the value of x_k', and to update x_k' one only needs x_k. Thus the new variables x_k' are local to the updaters of x_k.

It should be noted that if instead the other two modification methods described in subsection V-A were used, then the modified gradient dynamics would require additional information transfer among nodes for their implementation. In particular, due to the nonlinearity of the function ψ in (38), (39), the ODE for the x_i updates would have required also the flows x_j from neighboring nodes, which is practically undesirable as such information is not available in existing implementations of congestion control algorithms.

Convergence of the modified dynamics (46) to an optimum of the original problem now follows immediately from Theorem 41.1).

Proposition 46. *Let* $U_i \in C^2$ *be strictly concave and strictly increasing. Then solutions of* (46) *converge as* $t \to \infty$ *to maxima of the original problem* (40).

Remark 47. The use of derivative action to damp oscillatory behaviour has been studied previously in the context of node based multi-path routing in [38] by incorporating derivative action in a price signal that gets communicated (i.e. a form of prediction is needed) and a local stability result was derived. This has also been used in gradient dynamics in game theory in [42]. A control scheme similar to (46) for multi-path routing was proposed in [33] and studied in discrete and continuous time. In [33] the scheme differs from (46) in that the x_j variables are updated instantaneously. In our context this would be

$$x(t) = \underset{x \geq 0, Ax \leq C}{\operatorname{argmax}} \ \varphi'(x'(t), x, y(t)). \tag{47}$$

4) Numerical examples: In this subsection we present numerical simulations to illustrate the results described above. We consider the network in Figure 1 with two sources (nodes 1, 2) and two destinations (nodes 3, 4). The capacities are all set to 1, and the utility functions are chosen as $\log(1+x)$ and $1-e^{-x}$ for the sources at 1 and 2 respectively. The parameters κ_j were all set to 1. This network satisfies the condition (43) and this is apparent in the oscillating modes of the unmodified dynamics (42), shown in Figure 2, that do not decay. However, when we apply the modified dynamics (46) to this network, we obtain the rapid convergence to the equilibrium shown in Figure 3. Simulations demonstrate improved performance of the modified dynamics also in cases the unmodified dynamics lead to decaying oscillations, by providing improved damping to those (omitted due to page constraints).

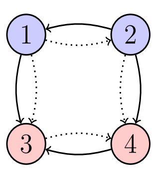

Fig. 1. First example network. Sources at 1 and 2 transmit to the destinations 4 and 3 respectively. Each has a choice of two routes. Routes associated with the source at 1 are dotted lines, while those associated with the source at 2 are solid lines.

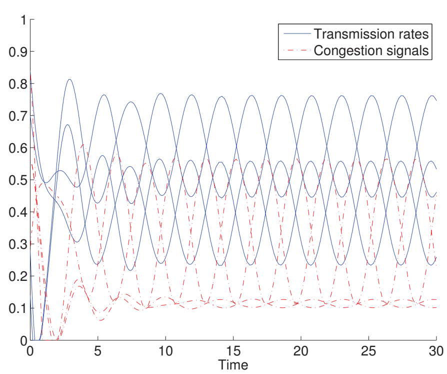

Fig. 2. The unmodified dynamics (42) running on the network given in Figure 1 with all link capacities set to 1 and the utility functions are $\log(1+x)$ and $1 - e^{-x}$ for the sources at 1 and 2 respectively. In this network the condition (43) holds, and there is oscillatory behaviour which does not decay.

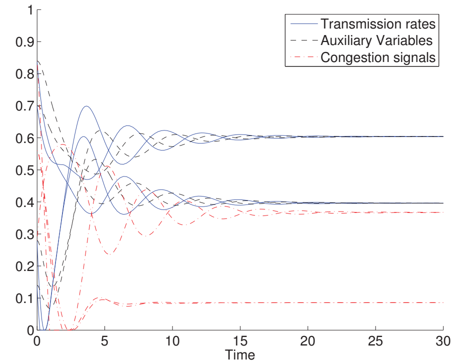

Fig. 3. The modified dynamics (46) running on the network given in Figure 1 with all link capacities set to 1, $\kappa_j = 1$ for all j. The utility functions are $\log(1+x)$ and $1-e^{-x}$ for the sources at 1 and 2 respectively. In this network the condition (43) holds, but the modification of the dynamics causes rapid convergence to equilibrium.

VI. Conclusion

In this paper we considered the problem of convergence to a saddle point of a concave convex function via subgradient dynamics that provide a restriction in an arbitrary convex domain. We showed that despite the nonlinear and non-smooth character of these dynamics, when these have an equilibrium point their ω-limit set is comprised of trajectories that are solutions to only linear ODEs. In particular, we showed that these ODEs are subgradient dynamics on affine subspaces which is a class of dynamics the asymptotic properties of which have been exactly characterized in part I. Various convergence criteria have been deduced from these results that can guarantee convergence to a saddle point. Several examples have also been discussed throughout the manuscript to illustrate the results in the paper.

References

[1] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in linear and non-linear programming*. Standford University Press, 1958.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[4] H Brézis. *Operateurs Maximaux Monotones Et Semi-Groupes De Contractions Dans Les Espaces De Hilbert*. North-Holland Mathematics Studies 5. North Holland Publishing, 1973.

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

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

[7] A. Cherukuri, E. Mallada, S. Low, and J. Corteés. The role of convexity on saddle-point dynamics: Lyapunov function and robustness. *arXiv preprint arXiv:1608.08586*, 2016.

[8] M-G. Cojocaru. Projected dynamical systems on Hilbert spaces. *PhD Thesis. Queen's University, Canada*, 2002.

[9] M-G. Cojocaru and L. Jonker. Existence of solutions to projected differential equations in Hilbert spaces. *Proceedings of the American Mathematical Society*, 132(1):183–193, 2004.

[10] G. Della Riccia. Equicontinuous semi-flows (one-parameter semi-groups) on locally compact or complete metric spaces. *Mathematical systems theory*, 4(1):29–34, 1970.

[11] E. Devane, A. Kasis, M. Antoniou, and I. Lestas. Primary frequency regulation with load-side participation Part II: Beyond passivity approaches. *IEEE Transactions on Power Systems*, 32(5):3519–3528, 2017.

[12] E. Devane, A. Kasis, C. Spanias, M. Antoniou, and I. Lestas. Distributed frequency control and demand-side management. *Smarter Energy: From Smart Metering to the Smart Grid*, pages 245–268, 2016.

[13] E. Devane and I. Lestas. Stability and convergence of distributed algorithms for the OPF problem. In *52nd IEEE Conference on Decision and Control*, December 2013.

[14] F. Dörfler, J. Simpson-Porco, and F. Bullo. Breaking the hierarchy: Distributed control and economic optimality in microgrids. *IEEE Transactions on Control of Network Systems*, 3(3):241–253, 2016.
[15] R. Ellis. *Lectures on topological dynamics*. Mathematics lecture note series. W. A. Benjamin, 1969.
[16] D. Feijer and F. Paganini. Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica*, 46(12):1974–1981, 2010.
[17] W.J. Feng, L. Wang, and Q.G. Wang. A family of multi-path congestion control algorithms with global stability and delay robustness. *Automatica*, 50(12):3112 – 3122, 2014.
[18] R.M. Freund. Penalty and barrier methods for constrained optimization. *Lecture notes for nonlinear programming. MIT.*, 2004.
[19] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1):17–40, 1976.
[20] R. Goebel. Stability and robustness for saddle-point dynamics through monotone mappings. *Systems & Control Letters*, 108:16–22, 2017.
[21] P.M. Gruber and J.M. Wills. *Handbook of Convex Geometry*. Convex Geometry. North-Holland, 1993.
[22] T. Holding and I. Lestas. Stability and instability in primal-dual algorithms for multi-path routing. In *54th IEEE Conference on Decision and Control*, December 2015.
[23] T. Holding and I. Lestas. On the emergence of oscillations in distributed resource allocation. *Automatica*, 85:22–33, 2017.
[24] T. Holding and I. Lestas. Stability and instability in saddle point dynamics - Part I. *IEEE Transactions on Automatic Control*, 2020.
[25] L. Hurwicz. The design of mechanisms for resource allocation. *The American Economic Review*, 63(2):1–30, May 1973.
[26] K. Kar, S. Sarkar, and L. Tassiulas. Optimization based rate control for multipath sessions. Technical report, Univ. of Maryland, Inst. Systems Research, 2001. http://hdl.handle.net/1903/6225.
[27] A. Kasis, E. Devane, C. Spanias, and I. Lestas. Primary frequency regulation with load-side participation Part I: Stability and Optimality. *IEEE Transactions on Power Systems*, 32(5):3505–3518, 2017.
[28] A. Kasis, N. Monshizadeh, E. Devane, and I. Lestas. Stability and optimality of distributed secondary frequency control schemes in power networks. *IEEE Transactions on Smart Grid*, 10(2):1747–1761, 2019.
[29] F. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, March 1998.
[30] G. Lee and J. Choi. A survey of multipath routing for traffic engineering.
[31] I. Lestas and G. Vinnicombe. Combined control of routing and flow: a multipath routing approach. In *43rd IEEE Conference on Decision and Control*, December 2004.
[32] N. Li, C. Zhao, and L. Chen. Connecting automatic generation control and economic dispatch from an optimization view. *IEEE Transactions on Control of Network Systems*, 3(3):254–264, 2016.
[33] X. Lin and N.B. Shroff. Utility maximization for communication networks with multipath routing. *IEEE Transactions on Automatic Control*, 51(5):766–781, 2006.
[34] S. H. Low and D. E. Lapsley. Optimization flow controlI: basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, 7(6):861–874, 1999.
[35] E. Mallada, C. Zhao, and S. Low. Optimal load-side control for frequency regulation in smart grids. *IEEE Transactions on Automatic Control*, 62(3):6294–6309, 2017.
[36] J. D. Meiss. *Differential dynamical systems*, volume 14. Siam, 2007.
[37] A. Nedić and A. Ozdaglar. Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142(1):205–228, 2009.
[38] F. Paganini and E. Mallada. A unified approach to congestion control and node-based multipath routing. *IEEE/ACM Transactions on Networking*, 17(5):1413–1426, October 2009.
[39] D. Richert and J. Cortés. Robust distributed linear programming. *IEEE Transactions on Automatic Control*, 60(10):2567–2582, 2015.
[40] R. T. Rockafellar. Saddle-points and convex analysis. *Differential games and related topics*, 109, 1971.
[41] R. T. Rockafellar. *Convex analysis*. Princeton University Press, 2nd edition, 1972.
[42] J.S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to nash equilibria. *IEEE Transactions on Automatic Control*, 50(3):312–327, March 2005.
[43] R. Srikant. *The mathematics of Internet congestion control*. Birkhauser, 2004.
[44] T. Stegink, C. De Persis, and A. van der Schaft. A unifying energy-based approach to stability of power grids with market dynamics. *IEEE Transactions on Automatic Control*, 62(6):2612–2622, 2017.
[45] V.I. Venets. Continuous algorithms for solution of convex optimization problems and finding saddle points of contex-coneave functions with the use of projection operations. *Optimization*, 16(4):519–533, 1985.
[46] T. Voice. Stability of multi-path dual congestion control algorithms. *IEEE/ACM Transactions on Networking*, 15(6):1231–1239, 2007.
[47] Z. Wang and J. Crowcroft. Analysis of shortest-path routing algorithms in a dynamic network environment. *ACM SIGCOMM Computer Communication Review*, 22(2):63–71, 1992.
[48] D Zhang and A Nagurney. On the stability of projected dynamical systems. *Journal of Optimization Theory and Applications*, 85(1):97–124, 1995.
[49] C. Zhao, U. Topcu, N. Li, and S.H. Low. Design and stability of load-side primary frequency control in power systems. *IEEE Transactions on Automatic Control*, 59(5):1177–1189, 2014.

Appendix A
Proofs of the main results

In this appendix we prove the main results of the paper, which are stated in section IV in the main text.

A. Outline of the proofs

We first give a brief outline of the derivations of the results to improve their readability.

1) Non-expansive semiflows and convex projections: In section A-B of this appendix we prove the results described in subsection IV-B in the main text.

We revisit some of the literature on topological dynamical systems [10], quoting a more general result Theorem 51, from which Proposition 14 is deduced. These results allow us to prove the main result of the subsection, Theorem 18, using the fact that the convex projection term cannot break the isometry property of the flow on the ω-limit set.

2) Subgradient method: In sections A-C, A-D in this appendix we prove the results in subsections IV-C, IV-D, respectively, in the main text using the results in subsection IV-B.

B. Non-expansive semiflows and convex projections

In this section we provide the proofs of Proposition 14 and Theorem 18.

B.1. Convergence to a flow of isometries

We begin by revisiting the literature on topological dynamical systems, in which a type of incremental stability is studied, and show how this leads to an invariance principle for non-expansive semiflows.

Definition 48 (Equicontinuous semiflow)**.** We say that a flow (resp. semiflow) (ϕ, X, ρ) is *equicontinuous* if for any $x(0) \in X$ and $\varepsilon > 0$ there is a $\delta = \delta(x(0), \varepsilon)$ such that if $\rho(x'(0), x(0)) < \delta$ then

$$\rho(x(t), x'(t)) \leq \varepsilon \text{ for all } t \in \mathbb{R} \text{ (resp. } \mathbb{R}_+\text{).} \tag{48}$$

Remark 49. In the control literature equicontinuity of a semiflow would correspond to 'semi-global non-asymptotic incremental stability', but we shall keep the term equicontinuity for brevity and consistency with [10].

Definition 50 (Uniformly almost periodic flow)**.** We say that a flow (ϕ, X, ρ) is *uniformly almost periodic* if for any $\varepsilon > 0$ there is a *syndetic* set $A \subseteq \mathbb{R}$, (i.e. $\mathbb{R} = A + B$ for some compact set $B \subseteq \mathbb{R}$), for which

$$\rho(\phi(t,x),x) \le \varepsilon \text{ for all } t \in A, x \in X. \tag{49}$$

For the readers convenience we reproduce the results, [10, Theorem 8] and [15, Proposition 4.4.], that we will use.

Theorem 51 ([10])**.** *Let (ϕ, X, ρ) be an equicontinuous semiflow and let X be either locally compact or complete. Let Ω be its ω-limit set. Then (ϕ, Ω, ρ) is an equicontinuous semiflow of homeomorphisms of Ω onto Ω. This generates an equicontinuous flow.*

The backwards flow given by Theorem 51 is only unique on Ω, (see Remark 15 which also applies here).

Proposition 52 ([15])**.** *Let (ϕ, X, ρ) be a flow, with X compact. Then the following are equivalent:*

(i) The flow is equicontinuous.
(ii) The flow is uniformly almost periodic.

In our case we focus on the non-expansive property which is a particular form of equicontinuity. We prove stronger results in this special case.

Proof of Proposition 14. By Theorem 51 (ϕ, Ω, d) is an equicontinuous flow with an equilibrium point $\bar{\mathbf{z}}$. Let $R > 0$ be arbitrary, and define

$$Y_R = \left\{ \mathbf{z}(0) \in \Omega : \sup_{t\in\mathbb{R}} d(\mathbf{z}(t), \bar{\mathbf{z}}) \le R \right\}. \tag{50}$$

As the flow is equicontinuous, Y_R is a closed bounded subset of $\mathbb{R}^{n+m}$ and hence compact, and moreover, the union of the sets Y_R over $R \ge 0$ is Ω. By Proposition 52 the flow (ϕ, Y_R, d) is uniformly almost periodic. By the non-expansive property, $d : Y_R \times Y_R \to \mathbb{R}$ is a non-increasing along the direct product flow, and is a continuous function on a compact set. Hence we have the inequality, for any two points $\mathbf{z}(0), \mathbf{z}'(0) \in Y_R$,

$$\begin{aligned} \lim_{t\to-\infty} d(\mathbf{z}(t), \mathbf{z}'(t)) &= \sup_{t\in\mathbb{R}} d(\mathbf{z}(t), \mathbf{z}'(t)) \\ &\ge \inf_{t\in\mathbb{R}} d(\mathbf{z}(t), \mathbf{z}'(t)) = \lim_{t\to\infty} d(\mathbf{z}(t), \mathbf{z}'(t)). \end{aligned} \tag{51}$$

We claim that the two limits are equal. Indeed, by uniform almost periodicity there are sequences $t_n \to \infty$ and $t'_n \to -\infty$ as $n \to \infty$ for which

$$0 = \lim_{n\to\infty} d(\mathbf{z}(t_n), \mathbf{z}(0)) = \lim_{n\to\infty} d(\mathbf{z}(t'_n), \mathbf{z}(0)) \tag{52}$$

and the analogous limits hold for $\mathbf{z}'$ for the same sequences t_n, t'_n. Hence, by continuity of d, we have

$$\lim_{t\to-\infty} d(\mathbf{z}(t), \mathbf{z}'(t)) = d(\mathbf{z}(0), \mathbf{z}'(0)) = \lim_{t\to\infty} d(\mathbf{z}(t), \mathbf{z}'(t)). \tag{53}$$

Hence $d(\mathbf{z}(t), \mathbf{z}'(t))$ is constant. By picking R big enough, this holds for any $\mathbf{z}(0), \mathbf{z}'(0) \in \Omega$, which completes the proof that the sub-semiflow generates a flow of isometries.

It remains to show that Ω is convex. To this end let $\mathbf{z}(t), \mathbf{z}'(t)$ be two trajectories of (ϕ, Ω, d). Let that $\lambda \in (0,1)$ and define $\mathbf{z}''(t) = \lambda\mathbf{z}(t) + (1-\lambda)\mathbf{z}'(t)$. By the same argument as used in the proof of [24, Proposition 34] we deduce that $\mathbf{z}''(t)$ is a trajectory of the original semiflow, but (as argued above) by uniform almost periodicity of (ϕ, Ω, d) we have a sequence of times $t_n \to \infty$ for which $d(\mathbf{z}(t_n), \mathbf{z}(0)) \to 0$ as $n \to \infty$ and the same limit for $\mathbf{z}'(t)$. Hence $d(\mathbf{z}''(t_n), \mathbf{z}''(0)) \to 0$ also, showing that $\mathbf{z}''(0)$ is in the ω-limit set. □

Proof of Corollary 16. From the text above the statement of the Corollary we have that all trajectories of the semiflow converge to its ω-limit set (denoted as Ω). Also from Proposition 14 we have that (ϕ, Ω, d) defines a flow of isometries.

In the remainder of the proof we strengthen the convergence to Ω to convergence to a trajectory of the flow (ϕ, Ω, d). Let $\mathbf{z}(t)$ be a trajectory of the flow. From the convergence to the set Ω there exist points $\mathbf{z}^{(n)} \in \Omega$ and times t_n such that,

$$|\mathbf{z}(t_n) - \mathbf{z}^{(n)}| \le 1/n. \tag{54}$$

We now consider the trajectories $\mathbf{z}^{(n)}(t)$ of the flow (ϕ, Ω, d) with $\mathbf{z}^{(n)}(t_n) = \mathbf{z}^{(n)}$ (note that $\mathbf{z}^{(n)}(t) \in \Omega$ for all $t \in \mathbb{R}$). From the non expansive property we have for all $t \ge t_n$,

$$|\mathbf{z}(t) - \mathbf{z}^{(n)}(t)| \le 1/n. \tag{55}$$

From the boundedness of $\mathbf{z}(t)$ and (55) the set $\{\mathbf{z}^{(n)} : n \in \mathbb{N}\}$ is relatively compact, and by the constant distance of each trajectory of the flow (ϕ, Ω, d) from any equilibrium point, the set of initial conditions $\{\mathbf{z}^{(n)}(0) : n \in \mathbb{N}\}$ is also relatively compact. There is hence a subsequence n_k for which $\mathbf{z}^{(n_k)}(0)$ tends to a point $\mathbf{z}'(0) \in \Omega$ as $k \to \infty$ (using also the fact that that Ω is closed [10, Theorem 5]). We claim that $|\mathbf{z}(t) - \mathbf{z}'(t)| \to 0$ as $t \to \infty$, where $\mathbf{z}'(t)$ is the trajectory of the flow (ϕ, Ω, d) that at time $t = 0$ is equal to $\mathbf{z}'(0)$. Indeed, for any $\epsilon > 0$ there exists a $k \in \mathbb{N}$ such that for all $t \ge t_{n_k}$, we have

$$|\mathbf{z}(t) - \mathbf{z}^{(n_k)}(t)| \le \varepsilon/2 \tag{56}$$

and also for all $t \ge 0$,

$$|\mathbf{z}'(t) - \mathbf{z}^{(n_k)}(t)| \le \varepsilon/2 \tag{57}$$

where in each case we have used the non-expansive property. The claim now follows from the triangle inequality, which completes the proof. □

B.2. Convergence to solutions of dynamics projected on an affine subspace

We now use the isometry property together with the geometry of the convex projection term to obtain the key result of this section, Theorem 18, which states that the limiting dynamics of a non-expansive ODE restricted to a convex set K have C^1 smooth vector field and lie inside one of the faces of K.

To prove the theorem we will make use of a simple lemma on faces of convex sets.

Lemma 53. *Let $K \subseteq \mathbb{R}^n$ be non-empty closed and convex and $A \subseteq K$. Let F be the minimal face of K containing A, then* $\operatorname{relint}(F)$ *intersects* $\operatorname{Conv} A$.

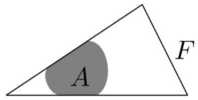

Fig. 4. This figure illustrates the claim of Lemma 53. The triangle F is the minimal face containing the convex set A (shaded region). If A intersects two subfaces of F, then, as shown, to be convex it must also intersect the relative interior of F.

The statement of this lemma and the idea behind its proof are illustrated by Figure 4.

Proof. As faces are convex, the minimal face containing A is the same as the minimal face containing $\operatorname{Conv} A$. So we are free to assume without loss of generality that A is convex. Assume for a contradiction that $A \cap \operatorname{relint}(F) = \emptyset$. Define the set $\mathcal{F}$ as

$$\{C : C \text{ is a proper face of } F \text{ and } A \cap (\operatorname{relint} C) \neq \emptyset\}.$$

Note that every point in the relative boundary of F lies in the relative interior of some proper face of F by property (e) below Definition 3. This implies that $\mathcal{F}$ is not empty. Now, either there is a face C in $\mathcal{F}$ that contains all other faces in $\mathcal{F}$, or there are two faces $F_1, F_2 \in \mathcal{F}$ such that there is no face $F_3 \in \mathcal{F}$ containing both F_1 and F_2. In the first case, C is a face containing A that is strictly contained in F, contradicting minimality of F. In the second case let $x_i \in (\operatorname{relint} F_i) \cap A$ for $i = 1, 2$, (note that $x_1 \neq x_2$ by property (e) of faces), and let x_3 be some point in the open line segment between x_1 and x_2. By convexity of A, $x_3 \in A$. Hence x_3 lies in $\operatorname{relint}(F_3)$ for some face F_3, and $F_3 \in \mathcal{F}$, as otherwise x_3 would lie in $\operatorname{relint}(F)$ contradicting the assumption that $(\operatorname{relint} F) \cap A = \emptyset$. We claim that F_3 contains both F_1 and F_2, a contradiction. Indeed, first we note that $x_1, x_2 \in F_3$ by property (ii) in Definition 3 as $x_3 \in F_3$. Then, as F_i is convex and $x_i \in \operatorname{relint}(F_i)$, F_i can be written as the union of line segments which have x_i as an interior point (i.e. not an end point). But each of these line segments touches F_3 at x_i, so by Definition 3(ii) each lies entirely within F_3. □

Proof of Theorem 18. **Step 1: Identification of the limiting equation.** First, by Lemma 17 and Proposition 14 (ϕ, Ω, d) is a flow of isometries. Now let F be the minimal face that contains Ω, i.e. the intersection of all faces that contain Ω, and N_F be its normal cone (in step 2 of the proof we will identify this face more precisely). We note that the vector field in (13) must be directed parallel to V, as otherwise trajectories would leave F, contradicting $\Omega \subseteq F$.

It is sufficient to show that if $\mathbf{z} = \mathbf{z}(0) \in \Omega$ with $\mathbf{n}(t) = \mathbf{P}_{N_K(\mathbf{z}(t))}(\mathbf{f}(\mathbf{z}(t)))$ then $\mathbf{n}(t)$ is orthogonal to F. If $\mathbf{z}(t) \in \operatorname{relint} K$ then $N_K(\mathbf{z}(t)) = N_F$ and the orthogonality holds. Otherwise $\mathbf{z}(t)$ lies in the relative boundary of F.

As each solution of the differential equation (13) holds only for almost all times t and we wish to consider an uncountably infinite family of solutions, we run the risk of taking an uncountable union of sets of measure zero, (which does not necessarily have zero measure). Avoiding this makes the proof technical. To better communicate the idea of the proof, we shall first give the proof that would work if the differential equations held for all times t.

Step 1.1: Heuristic (unrigorous) proof.

Let $C = \operatorname{Conv} \Omega$, then, by the definition of a face, $\Omega \subseteq F$ implies that $C \subseteq F$. From Lemma 53 and the minimality of F we deduce that C must intersect $\operatorname{relint} F$. Thus there are $\mathbf{x}(0), \mathbf{y}(0) \in \Omega$ and $\lambda \in (0,1)$ with $\mathbf{w} = \lambda \mathbf{x}(0) + (1-\lambda)\mathbf{y}(0) \in \operatorname{relint} F$. Set $W = \frac{1}{2}|\mathbf{x}(t) - \mathbf{z}(t)|^2$. By the isometry property of the flow we know that $\dot{W} = 0$ at t. We also have,

$$\begin{aligned} \dot{W}(t) &= (\mathbf{x}(t) - \mathbf{z}(t))^T (\dot{\mathbf{x}}(t) - \dot{\mathbf{z}}(t)) \\ &= (\mathbf{x}(t) - \mathbf{z}(t))^T (\mathbf{f}(\mathbf{x}(t)) - \mathbf{f}(\mathbf{z}(t))) + \\ &\quad - (\mathbf{x}(t) - \mathbf{z}(t))^T \mathbf{P}_{N_K(\mathbf{x}(t))}(\mathbf{f}(\mathbf{x}(t))) + \\ &\quad + (\mathbf{x}(t) - \mathbf{z}(t))^T \mathbf{P}_{N_K(\mathbf{z}(t))}(\mathbf{f}(\mathbf{z}(t))). \end{aligned} \tag{58}$$

The first term in (58) is non-positive due to the assumption that the ODE satisfies (13). The other two terms are non-positive due to the definition of the normal cone. Hence $\dot{W} = 0$ implies that $(\mathbf{x}-\mathbf{z})^T \mathbf{n} = 0$. Similarly we obtain $(\mathbf{y}-\mathbf{z})^T \mathbf{n} = 0$. Taking a convex combination of these equalities, we obtain

$$(\mathbf{w}-\mathbf{z})^T \mathbf{n} = \lambda (\mathbf{x}-\mathbf{z})^T \mathbf{n} + (1-\lambda)(\mathbf{y}-\mathbf{z})^T \mathbf{n} = 0 + 0 = 0 \tag{59}$$

and as $\mathbf{w}$ is in the relative interior of F this implies that $\mathbf{n}$ is orthogonal to F.

Step 1.2: Rigorous proof. We now give the fully rigorous proof. We must show that the set of times t when $\mathbf{n}(t)$ is not orthogonal to F is of measure zero. Let Ω' be a countable dense subset of Ω that contains $\mathbf{z}(0)$. By invariance of Ω under the flow ϕ, the set $\phi(t, \Omega') = \{\phi(t, \mathbf{x}) : \mathbf{x} \in \Omega'\}$ is also dense in Ω for any $t \in \mathbb{R}$. Then the set

$$\begin{aligned} A = \{t \in [0, \infty) : \exists \mathbf{x}(0) \in \Omega' \text{ such that} \\ \dot{\mathbf{x}}(t) \neq \mathbf{f}(\mathbf{x}(t)) - \mathbf{P}_{N_K(\mathbf{x}(t))}(\mathbf{f}(\mathbf{x}(t)))\} \end{aligned} \tag{60}$$

is the countable union of measure zero sets, and is hence of measure zero. From the isometry property and by considering $W(t) = \frac{1}{2}|\mathbf{x}(t) - \mathbf{z}(t)|^2$ with $\mathbf{x}(0) \in \Omega'$, it follows that $(\mathbf{x}(t) - \mathbf{z}(t))^T \mathbf{n}(t) = 0$ for all $\mathbf{x}(0) \in \Omega'$ and $t \in [0, \infty) \setminus A$. Thus, for $t \in [0, \infty) \setminus A$, $(\mathbf{x} - \mathbf{z}(t))^T \mathbf{n}(t) = 0$ for all $\mathbf{x}$ in a dense subset of Ω, and hence for any $\mathbf{x} \in \Omega$. The proof now follows as step 1.1. above.

Step 2: Identification of the limiting face. Finally we will show that the face F defined above is in fact the minimal face F' containing the equilibrium points of the semiflow (ϕ, K, d). We argue by contradiction. If $F \neq F'$ then there must be some trajectory $\mathbf{z}(t)$ in Ω and a time t_0 with $\mathbf{z}(t_0) \in F \backslash F'$. For $T > 0$ we define $\mathbf{z}(t; T) = \frac{1}{2T} \int_{-T}^{T} \mathbf{z}(t+s)\, ds$. For any finite T this is a convex combination of trajectories in Ω, and as Ω is convex by Proposition 14, $t \mapsto \mathbf{z}(t; T)$ is a trajectory in Ω. Next, as the semiflow is uniformly almost periodic due to Proposition 52 the trajectory $\mathbf{z}(t)$ is an almost periodic function. Therefore, the limit $T \to \infty$ of $\mathbf{z}(t; T)$ exists (see e.g. [15]), and this limit is clearly a constant ($\mathbf{z}'$ say) independent of t. As Ω is closed, $\mathbf{z}' \in \Omega$ and being a constant, is an equilibrium point of the semiflow.

To obtain a contradiction we argue that $\mathbf{z}' \notin F'$ which is impossible as F' contains all equilibrium points. Indeed, this follows as the trajectory $\mathbf{z}(t)$, being almost periodic and passing through $\mathbf{z}(t_0) \in F \setminus F'$ spends a positive proportion of its time in $F \setminus F'$. Therefore, there is a $\delta > 0$ such that for any sufficiently large T, the average $\mathbf{z}(t;T)$ satisfies $d(\mathbf{z}(t;T), F) \geq \delta$ and this property carries over to the limit $\mathbf{z}'$. □

C. Subgradient method

In this section we give the proofs of the results of subsection IV-C.

Proof of Theorem 20. We apply Theorem 18, noting that $\mathbf{f}(\mathbf{z})$ in (7) satisfies the inequality in (13) [41], [20].

Case (i). This follows directly from Theorem 18.

Case (ii). As F must contain all K-restricted saddle points, it must contain a point in the interior of K. The only such face is K itself whose affine span is $\mathbb{R}^{n+m}$ (as K has non-empty interior) which has normal cone $\{\mathbf{0}\}$. Therefore in case (ii) (14) becomes the gradient method (6) and (16) holds, using also the isometry property of the flow on Ω.

The convexity and isometry properties of Ω stated in Theorem 20 follow from Proposition 14. □

D. A general convergence criterion

In this section we give the proofs of subsection IV-D.

Proof of Theorem 31. By Theorem 20(i) any solution $\mathbf{z}(t)$ in the ω-limit set of the subgradient method on K solves (15). By using $\mathbf{\Pi}$, the orthogonal projection matrix onto the orthogonal complement of N_V, the ODE (15) can be written as (10). Noting also the isometry property of the ω-limit set we have by Theorem 11, Corollary 12 (in section IV-A), that $\mathbf{z}(t)$ satisfies (34) and (35) for all $t \in \mathbb{R}$ and $r \in [0,1]$. Therefore, if there are no non-constant trajectories of the subgradient method on K satisfying these conditions then the ω-limit set consists only of equilibrium points and the subgradient method on K is globally convergent. □

Proof of Corollary 34. This follows from Theorem 20, and Theorem 11, Corollary 12 using the arguments in the proof of Theorem 31. □

Appendix B
Proofs of the results in section V

A. Modification methods

Proof of Theorem 41:

We prove convergence of each modification method in turn.

1) Auxiliary variables method:

Proposition 54. *Let* (36) *hold, and assume that there exists a K'-restricted saddle point. Then the subgradient method* (7) *on K' applied to φ' is globally convergent.*

Proof. We prove global convergence to an equilibrium point by making use of Theorem 31. In particular, we show that the only solutions of the subgradient method applied to ϕ', which satisfy both (34) and (35), are equilibrium points.

Without loss of generality, we assume, by a translation of coordinates, that $\bar{\mathbf{z}}' = (M\bar{x}, \bar{x}, \bar{y}) = \mathbf{0}$ is an equilibrium point. Since the auxiliary variables are unconstrained the orthogonal complement of N_V in Theorem 31 is a subspace of the form $\mathbb{R}^{n'} \times V'$ where $V' \subseteq \mathbb{R}^{n+m}$ is an affine subspace.

Let $\mathbf{\Pi}$ be the orthogonal projection matrix onto the subspace $\mathbb{R}^{n'} \times V'$. We decompose $\mathbf{\Pi}$ on $\mathbb{R}^{n'} \times \mathbb{R}^{n+m}$ as

$$\mathbf{\Pi} = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi}' \end{bmatrix}. \tag{61}$$

Now let $\mathbf{z}(t) = (x'(t), x(t), y(t))$ be a solution of the modified subgradient method that satisfies (34) and (35), and let $(\tilde{x}(t), \tilde{y}(t)) = \mathbf{\Pi}'(x(t), y(t))$. The remainder of the proof is carried out in three steps.

Step 1: $x'(t)$ is constant. By the form of $\mathbf{A}(\mathbf{0})$ in (34) we deduce that $\dot{x}'(t) = 0$.

Step 2: $\tilde{x}(t)$ and $\tilde{y}(t)$ are constant. From the condition (35) that $\mathbf{\Pi}\mathbf{B}(r\mathbf{z})\mathbf{\Pi}\mathbf{z} = 0$ for $r \in [0,1]$, we have that

$$0 = \mathbf{z}^T\mathbf{\Pi}\mathbf{B}(r\mathbf{z})\mathbf{\Pi}\mathbf{z} = u^T\psi_{uu}u + \tilde{x}^T\varphi_{xx}\tilde{x} - \tilde{y}^T\varphi_{yy}\tilde{y} \tag{62}$$

where ψ_{uu} is the Hessian matrix of ψ evaluated at $u = M\tilde{x} - x'$. As each term is non-positive and ψ is strictly concave we deduce that $M\tilde{x} - x' = 0$ and $\tilde{x} \in \ker(\varphi_{xx}(\mathbf{0}))$. Thus $M\tilde{x}(t)$ is constant. By the condition that $\ker(M) \cap \ker(\varphi_{xx}) = \{0\}$ we deduce that $\tilde{x}(t)$ is constant. Then the form of $\mathbf{A}(\mathbf{0})$ allows us to deduce that $\tilde{y}(t)$ is also constant.

Step 3: $x(t)$ and $y(t)$ are constant. The vector field in (34) is orthogonal to $\ker(\mathbf{\Pi})$, so that $(\tilde{x}(t), \tilde{y}(t))$ being constant implies that $(x(t), y(t))$ are constant.

This completes the proof of convergence to an equilibrium point of the subgradient method applied to ϕ' . □

2) Penalty function method:

Proposition 55. *Let $K \subseteq \mathbb{R}^{n+m}$ be non-empty closed and convex as in* (1)*. Let* (37)*,* (38) *hold, and assume that there exists a K-restricted saddle point. Then the subgradient method* (7) *on K applied to φ' is globally convergent.*

Proof. Without loss of generality, we may assume by a translation of coordinates that $\mathbf{0}$ is a K-restricted saddle point. We apply Theorem 31 and let $F, V, \mathbf{\Pi}$ be as in Theorem 31 and $\mathbf{z}(t) = (x(t), y(t))$ be a trajectory of the subgradient method on K satisfying (34) and (35) for all $t \in \mathbb{R}$ and $r \in [0,1]$. Define $(\tilde{x}(t), \tilde{y}(t)) = \tilde{\mathbf{z}}(t) = \mathbf{\Pi}\mathbf{z}(t)$. We compute that

$$\mathbf{A}(\mathbf{0}) = \begin{bmatrix} 0 & g_x(0)^T \\ -g_x(0) & 0 \end{bmatrix}. \tag{63}$$

Step 1: $g_x(0)\tilde{x}(t) = 0$.

The condition (35) implies that the following expression is zero for all $s \in [0,1]$,

$$\tilde{\mathbf{z}}^T\mathbf{B}(s\mathbf{z})\tilde{\mathbf{z}} = \tilde{x}^T\varphi_{xx}\tilde{x} + [g_x\tilde{x}]^T\psi_{uu}[g_x\tilde{x}] + \psi_u(\tilde{x}^T g_{xx}\tilde{x}) \tag{64}$$

where φ_{xx} is evaluated at $s\mathbf{z}$, with g_x, g_{xx} at sx, and ψ_{uu}, ψ_{u^k} at $u = g(sx)$, and where $x^T g_{xx} x$ is the vector with ith component $x^T g^i_{xx} x$ where $g = [g^1, \ldots, g^m]^T$. All the terms are non-positive by the assumptions on ψ and φ. Strict concavity of ψ and that (64) vanishes for all $s \in [0,1]$ implies that $g_x(sx)\tilde{x} = 0$ for all $s \in [0,1]$. In particular $g_x(0)\tilde{x}(t) = 0$.

Step 2: $\tilde{x}(t)$ is constant.

Let $\mathbf{\Pi}$ be decomposed on $\mathbb{R}^n \times \mathbb{R}^m$ as

$$\mathbf{\Pi} = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix}. \tag{65}$$

Then $\tilde{x}, \tilde{y}$ satisfy

$$\dot{\tilde{x}} = \Pi_{11} g_x(0)^T \tilde{y} \qquad \dot{\tilde{y}} = -\Pi_{21} g_x(0)^T \tilde{y}. \tag{66}$$

Taking the time derivative of $g_x(0)\tilde{x} = 0$ we obtain $g_x(0)\Pi_{11} g_x(0)^T \tilde{y} = 0$. As Π_{11} is positive semi-definite, $\ker(g_x(0)\Pi_{11} g_x(0)^T) = \ker(\Pi_{11} g_x(0)^T)$, and hence $\dot{\tilde{x}} = \Pi_{11} g_x(0)^T \tilde{y} = 0$ and $\tilde{x}(t)$ is constant.

Step 3: $\tilde{y}(t)$ is constant.

The relation $\mathbf{\Pi}\dot{\tilde{\mathbf{z}}} = \dot{\tilde{\mathbf{z}}}$ implies that $\Pi_{11}\dot{\tilde{x}} + \Pi_{12}\dot{\tilde{y}} = \dot{\tilde{x}} = 0$ and $0 = \Pi_{12}\dot{\tilde{y}} = -\Pi_{12}\Pi_{21} g_x(0)^T \tilde{y}$. Therefore, again, as $\Pi_{12}\Pi_{21}$ is positive semi-definite we have $\tilde{y}^T g_x(0)\Pi_{12}\Pi_{21} g_x(0)^T \tilde{y} = 0$ and $\Pi_{21} g_x(0)^T \tilde{y} = 0 = -\dot{\tilde{y}}$, which implies $\tilde{y}$ is constant[7].

The fact that $x(t)$, $y(t)$ are constant can be deduced as in Step 3 of the proof of Proposition 54. □

3) Constraint modification method: We first consider the case without constraints. The proof below shows that the method works by disrupting the linear structure of the oscillating solutions by changing $\mathbf{A}(\mathbf{z})$ to ensure it is not equal to $\mathbf{A}(\mathbf{0})$, (where $\mathbf{0}$ is a saddle).

Proposition 56. *Let* (39) *hold and $\bar{\mathcal{S}} \neq \emptyset$. Then $\mathcal{S} = \bar{\mathcal{S}}$ and the gradient method* (6) *applied to φ' is globally convergent.*

Proof. Without loss of generality we may assume that $\mathbf{0}$ is a saddle point of φ. We use the classification of $\mathcal{S}$ given by [24, Theorem 13] and use the notation therein. We first compute,

$$\mathbf{A}(\mathbf{z}) = \begin{bmatrix} 0 & (\psi_g g_x)^T \\ -\psi_g g_x & 0 \end{bmatrix}. \tag{67}$$

Let $\mathbf{z}(t) = (x(t), y(t)) \in \mathcal{S}$ then we have

$$0 = \frac{d}{ds}[(\psi_g^i(g(sx))^T g_x(sx)x]_{s=0} \text{ for } i = 1, \dots, m \tag{68}$$

Then by applying the chain rule we obtain

$$0 = [g_x(0)x]^T \psi_{gg}^i(0)[g_x(0)x] + \psi_g^i(0)^T(x^T g_{xx}(0)x), \tag{69}$$

where $x^T g_{xx}(0)x$ is the vector with components $x^T g_{xx}^i x$ where $g = [g^1, \dots g^m]^T$. All the terms are non-positive due to the assumptions on ψ and g. As $\psi_{gg}^i < 0$ we have $g_x(0)x = 0$. Hence $\dot{y} = 0$ and therefore y is constant. As $|x|^2 + |y|^2$ is also constant this means that $\dot{x}$ is zero. Therefore $\mathcal{S} = \bar{\mathcal{S}}$ and the gradient method is globally convergent. □

Now we extend the stability to the subgradient method on sets which have a product structure, by making use of Corollary 28.

Corollary 57. *Let $K \subseteq \mathbb{R}^{n+m}$ be non-empty closed and convex as in* (1). *Let* (39) *hold and there be a K-restricted saddle point. Then the subgradient method* (7) *on K applied to φ' is globally convergent.*

Proof. By Corollary 28 it suffices to prove that the subgradient method on $\text{aff}(F)$ is globally convergent, where F is an arbitrary face of K that contains a K-restricted saddle point $\bar{\mathbf{z}}$. By translation of coordinates we may assume that $\bar{\mathbf{z}} = \mathbf{0}$. By the product structure of K, $V = \text{aff}(F)$ must also decompose into $V = V_x \times V_y$ with $V_x \subseteq \mathbb{R}^n$ and $V_y \subseteq \mathbb{R}^m$ affine subspaces. Let the orthogonal projection matrices onto V_x, V_y, which exist as $(0,0) \in V_x \times V_y$, be P, Q respectively. Then the subgradient method on V, satisfies, for $(x,y) \in V$,

$$\dot{x} = P\varphi'_x = \varphi_x^V, \quad \dot{y} = -Q\varphi'_y = -\varphi_y^V \tag{70}$$

where $\varphi^V(x,y) := \varphi(Px, Qy)$. By a rotation[8] of coordinate bases we may assume that $V_x = \mathbb{R}^{n'} \times \{0\}$ and $V_y = \mathbb{R}^{m'} \times \{0\}$ for some $n' \leq n$ and $m' \leq m$. Then $\varphi^V : \mathbb{R}^{n'} \times \mathbb{R}^{m'} \to \mathbb{R}$ is of the form (39) and Proposition 56 gives convergence. □

B. Multi-path congestion control

Proof of Proposition 44. The *if* claim follows directly from the discussion preceding the proposition. For the *only if* we explicitly construct a trajectory that does not converge. Let u satisfy (43), then it can be directly verified that

$$\mathbf{z}(t) = \bar{\mathbf{z}} + ce^{t\mathbf{A}(\bar{\mathbf{z}})} \begin{bmatrix} u \\ -Au \end{bmatrix}$$

is a solution (for any $c > 0$) of the unconstrained gradient method (6) applied to φ. By taking c small enough using the fact that $\bar{\mathbf{z}} > 0$ (and the skew-symmetry of $\mathbf{A}(\bar{\mathbf{z}})$) we can ensure that $\mathbf{z}(t) > 0$ for all $t \in \mathbb{R}$, and hence $\mathbf{z}(t)$ is also a solution of the subgradient dynamics (42). □

[7]Note that step 3 could also be proved from the fact that the product structure of K implies that $V = \text{aff}(F)$ must also decompose into $V = V_x \times V_y$ with $V_x \subseteq \mathbb{R}^n$, $V_y \subseteq \mathbb{R}^m$ affine subspaces, thus implying $\Pi_{12} = \Pi_{21} = 0$ (this structure of V is used in the proof of Corollary 57).

[8]Note that a rotation of coordinates will transform ϕ' in (39) to a function that is still of the form specified in (39), i.e. in the new coordinates ϕ' can be written in terms of functions U, g, ψ that satisfy the conditions in (39).