

On the emergence of oscillations in distributed resource allocation

Thomas Holding^a, Ioannis Lestas^b

^a*Mathematics Institute, Zeeman Building, University of Warwick, Coventry, CV4 7AL, United Kingdom*

^b*Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom*

Abstract

We consider the problem of resource allocation in a decentralized market where users and suppliers trade for a single commodity. Due to the lack of strict concavity, convergence to the optimal solution by means of classical gradient type dynamics for the prices and demands, is not guaranteed. In the paper we explicitly characterize in this case the asymptotic behaviour of trajectories and provide an exact characterization of the limiting oscillatory solutions. Methods of modifying the dynamics are also given, such that convergence to an optimal solution is guaranteed, without requiring additional information exchange among the users.

1 Introduction

Problems of distributed resource allocation have been extensively studied by several scientific communities as a result of their significance in important applications such as allocation of resources in a communication networks (e.g. Internet congestion control, multipath routing), or market mechanisms in economic networks (e.g. [17,25,15,26]). A classical approach in this context is to consider the problem of maximizing an aggregate user utility by means of appropriate decentralized update schemes for the primal and dual variables, so as to reach a saddle point of the corresponding Lagrangian.

It is well known that if a concave-convex Lagrangian is strictly concave with respect to the primal variables, then the gradient method introduced by Arrow, Hurwicz and Uzawa [2] is guaranteed to converge to the optimal solution. The gradient method has been extensively used in network optimization problems due to the simplicity and decentralized form of the protocols it leads to; see e.g. the nice overview in [9], the convergence results in [6] where the non-smooth character of continuous time gradient dynamics is explicitly addressed, the recent discrete time results in [22], the significant works in [17,15], and the use of gradient dynamics in various important applications such as communication networks [7,19], distributed optimization [12,27], game

theory [11], and power networks [28,21]. In many cases, however, the underlying structure of the problem gives rise to a concave-convex Lagrangian which is not strictly concave, in which case convergence to the optimal solution, by means of classical gradient type dynamics, can be problematic. This oscillatory behaviour was noted in the early work of Arrow, Hurwicz and Uzawa in [2] and subsequent economic studies [15], but such oscillations due to a lack of strict concavity have also been observed in more recent works, see e.g. the discussion in [9] and the examples provided therein, and also analogous oscillations observed in [23] in the context of smart grids. Despite the significance of the gradient method, the problem of providing an *exact* characterization to its asymptotic behaviour in cases where it does not necessarily converge to an equilibrium point has not been addressed in the literature. This is, in fact, in general a non trivial problem due to the non-linearity of the dynamics.

We focus in this paper on an important such case where oscillations can emerge. In particular, we consider the problem of distributed resource allocation in an economic network where multiple consumers and suppliers trade for a single commodity with constraints in the supply. This is a classical problem in microeconomics [26], [2] which is known to be not strictly concave and hence conventional gradient based schemes for updating the supply and demand are not guaranteed to converge to the optimum. One of our main results is to fully characterize in this case the asymptotic behaviour of trajectories, giving an exact characterization to the limiting oscillatory solutions. This allows to quantify how severe

* This work was partly supported by an ERC starting grant.
Email addresses: t.holding@warwick.ac.uk (Thomas Holding), ic120@cam.ac.uk (Ioannis Lestas).

these oscillations can be in a general network (it is shown in the paper that these can be of arbitrarily large amplitude, and cannot be mitigated by the choice of utility functions) and also motivate ways of avoiding them.

We also discuss within the paper how appropriate modifications, which lead to higher order dynamics in the local update schemes, provide guarantees for convergence to the desired equilibrium point. A distinctive feature of the modifications proposed, is that apart from being distributed, they do not require any additional information transfer among users/suppliers, which can be important in a competitive economic setting. A comparison is also made with other distributed modifications, such as ones where the network constraints are incorporated in a modified Lagrangian. These go back to the early work of Arrow, Hurwicz and Uzawa [2] and have also been used in more recent studies, such as the continuous time modifications described in [9], which are also linked with corresponding discrete time methods such as the ADMM mechanism reviewed in [4]. These lead here, however, to schemes where each user needs to be aware of the demand of other users which use the same supplier, and thus require additional cooperative information exchange.

Also a side technical result in the paper for general gradient dynamics, is the fact that when the problem is not strictly concave, in which case trajectories are not guaranteed to converge to an equilibrium point, the gradient dynamics satisfy a notion of stability linked to incremental stability [1].

This paper is an extended version of the manuscript [13], where preliminary results have been stated. In a recent study [14] we have also shown that some of the features of the oscillatory solutions (such as the fact that they satisfy linear ODEs), are retained in the case where (sub)gradient methods are applied on general concave-convex functions. Nevertheless an exact characterization of the oscillations, as in this paper, is in general an involved problem due to the non-linear and switching nature of the dynamics, and requires an exploitation of the underlying structure.

The paper is structured as follows.

- In section 2 various preliminary results and definitions are provided that are necessary to formulate the problem. These include the notions of stability that will be used, and also results on the Arrow-Hurwicz-Uzawa gradient method for convergence to a saddle point of a concave-convex function.
- In section 3 we provide the problem formulation. The economic network that will be studied is described in detail and we consider non-linear ODEs that follow when classical gradient dynamics are used in this context.

- In section 4 we state the main results of the paper. In particular, we classify the asymptotic behaviour of the system described in section 3 and provide an exact characterization of the limiting oscillatory trajectories.
- Section 5 discusses methods of modifying the dynamics such that convergence to an optimal point can be guaranteed, focusing on schemes that do not introduce additional information transfer among users.
- Numerical examples are provided in section 6 and conclusions are drawn in section 7.
- The proofs of the results presented in the main text of the paper are given in appendix A. In appendix B an extended version of Theorem 17 is presented, where a technical assumption associated with the existence of a strictly positive saddle point is relaxed.

2 Preliminaries

2.1 Notation

Real numbers are denoted by \mathbb{R} , and the non-negative reals by \mathbb{R}_+ . For vectors $x, y \in \mathbb{R}^n$ the inequality $x < y$ means $x_i < y_i$ for all $i = 1, \dots, n$, $\|\cdot\|$ denotes a norm on \mathbb{R}^N and $|\cdot|$ denotes specifically the Euclidean norm.

For a matrix $A \in \mathbb{R}^{n \times m}$, the entries of A are denoted by A_{ij} and the kernel and transpose of A are denoted by $\ker(A)$ and A^T respectively. The matrix $\text{diag}(J_1, \dots, J_n)$ with J_1, \dots, J_n square matrices, of possibly different sizes, will denote the block diagonal matrix with diagonal blocks the matrices J_1, \dots, J_n .

The space of k times continuously differentiable functions is denoted by C^k . For a sufficiently differentiable function $f(x, y) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ we denote the partial derivatives with respect to each component of the vectors x or y by f_{x_i} and f_{y_i} respectively. f_x will denote the gradient of f with respect to x , i.e. the vector $(f_{x_1}, f_{x_2}, \dots, f_{x_n})$, respectively f_y . The Hessian matrices of f with respect to x and y are denoted by f_{xx} and f_{yy} respectively, while f_{xy} denotes the matrix of partial derivatives defined as $[f_{xy}]_{ij} := \frac{\partial^2 f}{\partial x_i \partial y_j}$.

The distance from a vector $x \in \mathbb{R}^n$ to a subset $E \subset \mathbb{R}^n$ will be denoted and defined by $\text{dist}(x, E) = \text{dist}(E, x) = \inf\{\|x - y\| : y \in E\}$.

We will use the notation (\bar{x}, \bar{y}) to denote a saddle point of a function $f(x, y)$ (see Definition 6 in section 2.3).

For two trajectories $x(t), y(t)$ of a dynamical system that take values in \mathbb{R}^n we say $x(t)$ converges to $y(t)$ if $\|x(t) - y(t)\| \rightarrow 0$ as $t \rightarrow \infty$.

2.2 Stability

For the purposes of this section we will look at the autonomous differential equation

$$\dot{x} = f(x) \quad (1)$$

where $x \in \mathbb{R}^N$, and $f : D \rightarrow \mathbb{R}^N$ is nice enough for the solution to exist and be unique for any initial condition $x(0)$.

A standard definition of stability of equilibrium points can be found in, for example, [18]. This extends to stability of positively invariant sets in the following manner.

Definition 1 (Stability of positively invariant sets)

A positively invariant set E is stable if for any $\epsilon > 0$ there is a $\delta > 0$ such that any solution $x(t)$ of (1) with $\text{dist}(x(0), E) < \delta$ has $\text{dist}(x(t), E) < \epsilon$ for all $t \geq 0$.

The notion of stability of an invariant set can be used to define stability of solutions $x(t)$ by considering stability of their positive orbit $\{x(t) : t \geq 0\}$; however, in the analysis that follows we are able to establish a stronger form of stability.

Definition 2 (Pathwise stability) We say that the flow generated by (1) is pathwise stable, if there is a norm $\|\cdot\|$ such that for any two solutions $x(t), y(t)$ of (1),

$$\sup_{t \geq 0} \|x(t) - y(t)\| \leq \|x(0) - y(0)\|. \quad (2)$$

Remark 3 This notion is similar to that of Incremental stability (see e.g. [1]), which is usually taken to mean that $\|x(t) - y(t)\|$ converges to zero.

Remark 4 Pathwise stability implies the stability (Definition 1) of any positively invariant set.

2.3 Arrow-Hurwicz-Uzawa gradient method

In this section we summarise the results of Arrow, Hurwicz and Uzawa [2]. The reader is encouraged to consult [9] for a more modern view with various applications.

Definition 5 (Concave-Convex function) We say that a function $g(x, y) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is (strictly) concave in x (respectively y) if for any fixed y (respectively x), $g(x, y)$ is (strictly) concave as a function of x , (respectively y). If g is concave in x and convex in y we call g concave-convex.

Concave programming is concerned with an optimisation problem of maximising a concave function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ subject to the restriction, $g(x) \geq 0$, where $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a concave function. Some well known

preliminary results will now be stated without proof (see for example [2] or the more recent [5]). We will be considering in the paper *primal* problems of the form

$$\max_{x \geq 0, g(x) \geq 0} f(x) \quad (3)$$

where $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, $g(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are concave functions.

The Lagrangian for (3) is

$$\varphi(x, y) = f(x) + y^T g(x) \quad (4)$$

where $y \in \mathbb{R}_+^m$ are the Lagrange multipliers.

Definition 6 (Saddle point) A saddle point of $\varphi(x, y) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a non-negative pair $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $\forall x, y \geq 0$,

$$\varphi(x, \bar{y}) \leq \varphi(\bar{x}, \bar{y}) \leq \varphi(\bar{x}, y). \quad (5)$$

Theorem 7 Let g be concave on $x \geq 0$, and Slater's condition hold, i.e.,

$$\exists x' > 0 \text{ with } g(x') > 0. \quad (6)$$

Then \bar{x} is an optimum of (3) iff $\exists \bar{y}$ with (\bar{x}, \bar{y}) a saddle point of (4).

The min max optimization problem associated with finding a saddle of (4) is the dual problem of (3).

Suppose that φ is a Lagrangian such that:

$$\begin{aligned} \varphi(x, y) : \mathbb{R}^n \times \mathbb{R}^m &\rightarrow \mathbb{R}, \\ \varphi \in C^2, \varphi &\text{ is concave in } x \text{ and convex in } y. \end{aligned} \quad (7)$$

We wish to design a dynamical system that will enable us to converge to a saddle point. An obvious choice would be to send x in the direction of increasing φ , and y in the direction of decreasing φ . So we choose

$$\begin{aligned} \dot{x}_i &= \begin{cases} 0 & \text{if } x_i = 0 \text{ and } \varphi_{x_i} < 0 \\ \varphi_{x_i} & \text{otherwise} \end{cases} \\ \dot{y}_i &= \begin{cases} 0 & \text{if } y_i = 0 \text{ and } \varphi_{y_i} > 0 \\ -\varphi_{y_i} & \text{otherwise} \end{cases} \end{aligned} \quad (8)$$

where the cases above are used to keep $x, y \geq 0$. This is called the Arrow-Hurwicz-Uzawa gradient method.

The following definitions characterise solutions to the gradient method in terms of the switching at the boundary, and will be important in the analysis throughout the paper.

Definition 8 We say that a coordinate i of x in a solution pair (x, y) of (8) is,

- Active if either $x_i > 0$ or $\varphi_{x_i} > 0$.
- Semi-active if $x_i = 0$ and $\varphi_{x_i} = 0$.
- Inactive if $x_i = 0$ and $\varphi_{x_i} < 0$.

We use the same definitions for the components of y , but with $-\varphi_{y_i}$.

Note that x_i (or y_i) cannot be negative, so a coordinate is either active, inactive or semi-active.

Definition 9 We say a solution $(x(t), y(t))$ of the gradient method (8) applied to (7) is proper, iff no coordinate x_i or y_j ever becomes inactive. This is equivalent to $(x(t), y(t))$ solving

$$\dot{x} = \varphi_x, \quad \dot{y} = -\varphi_y, \quad (9)$$

which is (8) without the switching at the boundary.

In practical terms, we can just ignore the switching at the boundary for proper solutions.

The following is a known convergence result associated with the case that φ is strictly concave in x .

Theorem 10 (Arrow, Hurwicz and Uzawa[2]) Let (7) hold, and in addition $\varphi(x, y)$ be strictly concave in x . Let φ have a saddle point $(\bar{x}, \bar{y}) \geq 0$. Then \bar{x} is unique, and when the gradient method (8) is applied to φ , $x(t) \rightarrow \bar{x}$ for any initial condition $(x(0), y(0)) \geq 0$.

If φ is only concave-convex, then the gradient method may have solutions that do not converge to a saddle. However we can make some statements about such solutions.

Theorem 11 (Arrow, Hurwicz and Uzawa [2]) Let (7) hold, (\bar{x}, \bar{y}) be a saddle point and $(x(t), y(t))$ be a solution of (8). Then $(x(t), y(t))$ converges to the set of solutions $(x'(t), y'(t))$ of (8) which have the Euclidean norm $|(x'(t), y'(t)) - (\bar{x}, \bar{y})|$ constant.

3 Economic network

Throughout the paper we consider the problem of distributed resource allocation in an economic network where consumers and suppliers trade for a single commodity. As it will be discussed in this section, the underlying structure of the problem leads to a lack of strict concavity in the Lagrangian, hence classical pricing and user dynamics can fail to converge to the desired optimal solutions. One of our main results is to fully characterize in this case the asymptotic behaviour of trajectories.

We define a network $\mathcal{G} = (S, C, Z)$ of suppliers $S = (s_i)_1^m$, consumers $C = (c_i)_1^n$ and the links between them

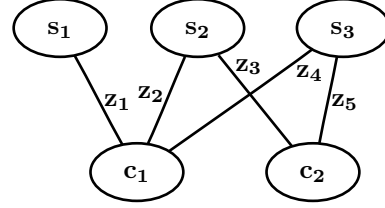


Fig. 1. An example network.

$Z = (z_l)_1^L$. The links determine the suppliers each consumer can trade with. We will always assume that the network is connected.

Notation. We write \sim to indicate connection between consumers, suppliers and links. So that $c_i \sim z_l$ iff c_i is connected to z_l , $s_j \sim z_l$ iff s_j is connected to z_l , and $s_j \sim c_i$ iff the supplier s_j is connected to the consumer c_i via some link z_l .

If c_i is specified then the sum $\sum_{z_l \sim c_i}$ will be over all links z_l that are connected to c_i , and similarly in other cases. We will overload z_l to be both the l th link, and the amount of goods being sent down the l th link. The set of L links forms a vector $z \in \mathbb{R}^L$.

Each consumer c_i has a strictly concave utility function $U_i : \mathbb{R}_+ \rightarrow \mathbb{R}$, which gives the utility of the consumer as a function and of his aggregate demand $\sum_{c_i \sim z_l} z_l$. Each supplier s_j has a maximum output $Y_j \geq 0$. We also denote $Y = (Y_1, Y_2, \dots, Y_n)$. The primal problem of maximising total utility is then

$$\begin{aligned} \mathbf{Primal:} \quad & \max_{z \geq 0} \sum_{c_i} U_i \left(\sum_{c_i \sim z_l} z_l \right) \\ & \text{s.t.} \quad \sum_{c_i \sim z_l} z_l \leq Y. \end{aligned} \quad (10)$$

It is easy to see that this problem is concave and it is also assumed that Slater's condition (6) is satisfied, so we formulate the dual problem using the Lagrangian

$$\varphi(z, p) = \sum_{c_i} U_i \left(\sum_{c_i \sim z_l} z_l \right) + \sum_{s_j} p_j \left(Y_j - \sum_{z_l \sim s_j} z_l \right) \quad (11)$$

where $p \in \mathbb{R}_+^m$ are Lagrange multipliers (or marginal prices in economic terms). The dual problem is then,

$$\mathbf{Dual:} \quad \text{Find a saddle point of } \varphi(z, p) \text{ in } z \geq 0, p \geq 0 \quad (12)$$

with $\varphi(z, p)$ as defined in (11).

Solving problem (10) in a decentralised manner is a classical problem in microeconomics [26]. A standard approach is to use the gradient method [2] to solve the

problem, which leads to decentralized update rules for the prices and demands.

In particular, the gradient method (8) applied to (11) yields the system:

$$\begin{aligned} \dot{z}_l &= M_{z_l} \left(U'_l \left(\sum_{c_i \sim z_l} z_l \right) - p_j \right) \quad \text{where } s_j \sim z_l \sim c_i \\ \dot{p}_j &= M_{p_j} \left(\sum_{z_l \sim s_j} z_l - Y_j \right) \end{aligned} \quad (13)$$

where U'_i is the derivative of U_i . The M_{z_l} and M_{p_j} are operators that act to keep the supply along the links and the prices non-negative and are defined by

$$M_{p_j} q = \begin{cases} 0 & \text{if } q < 0 \text{ and } p_j = 0 \\ q & \text{otherwise.} \end{cases} \quad (14)$$

In this paper we are concerned with the asymptotic behaviour of (13) and we will give a complete characterisation of its limiting solutions.

Remark 12 *The gradient dynamics (13) for the prices and demand can be given an economic interpretation as the action of consumers and suppliers in a decentralised market who aim to optimise their individual utility and profit, respectively, by means of only local information [15], [26]. More precisely, each consumer c_i is viewed as having the net utility function V_i given by*

$$V_i(z, p) = U_i \left(\sum_{c_i \sim z_l} z_l \right) - \sum_{c_i \sim z_l \sim s_j} p_j z_l \quad (15)$$

where U_i has been adjusted with the cost of purchasing the commodity at the current prices. The gradient dynamics are then interpreted as each consumer c_i performing gradient updates to maximise V_i , assuming, due to lack of global knowledge, that marginal prices will remain fixed. At the same time, the suppliers s_j adjust the prices p_j based on the aggregate demand of the consumers connected to them such that at equilibrium this will be equal to the supply Y_j (i.e. price updates are analogous to those in a classical tatonnement process [26]).

Remark 13 *It should be noted that, despite the strict concavity of the utility functions, the optimisation problem (10) and associated Lagrangian (11) are not strictly concave if any consumer is connected to more than one link. The lack of strict concavity means that the gradient method is not guaranteed to converge to an optimal solution to (10). The asymptotic behaviour is hence comprised of non-trivial trajectories and one of the main results in the paper is to provide an exact characterization*

of those, showing that these are solutions to explicit linear ODEs¹.

We will rewrite the system (13) in vector form, defining two matrices that determine the connections between suppliers, consumers and the links.

The matrices $H \in \mathbb{R}^{n \times L}$ and $A \in \mathbb{R}^{m \times L}$ are defined by

$$\begin{aligned} H_{il} &= 1 \text{ if } c_i \sim z_l, \text{ and } 0 \text{ otherwise.} \\ A_{jl} &= 1 \text{ if } s_j \sim z_l, \text{ and } 0 \text{ otherwise.} \end{aligned} \quad (16)$$

To illustrate this we give these in the case of the example in figure 1.

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \quad (17)$$

We will also consider the addition of gains to (13), introducing positive constants k_1, \dots, k_m and k'_1, \dots, k'_L . The final system is, (using the notation in (13)),

$$\begin{aligned} \dot{z} &= M_z K' (H^T U' (Hz) - A^T p) \\ \dot{p} &= M_p K (Az - Y) \end{aligned} \quad (18)$$

where M_p, M_z act diagonally by M_{p_j} and M_{z_l} respectively, $K = \text{diag}(k_1, \dots, k_m)$, $K' = \text{diag}(k'_1, \dots, k'_L)$, and U' is defined by

$$(U'(y))_i = U'_i(y_i) \quad \text{for } i = 1, \dots, n. \quad (19)$$

Remark 14 *As defined, M_z and M_p are operators not matrices, but for simplicity we will from now on view them as 0-1-matrices that are defined to depend on the expressions they are applied to in (18).*

Remark 15 *Using (11), (18) could also be formulated as*

$$\begin{aligned} \dot{z} &= M_z K' \varphi_z \\ \dot{p} &= M_p K (-\varphi_p). \end{aligned} \quad (20)$$

We will put a regularity requirement on the utility functions:

$$U_i : \mathbb{R}_+ \rightarrow \mathbb{R}, \quad U_i \in C^2, \quad (21)$$

U_i is strictly increasing and strictly concave.

It will be useful for the later analysis to single out a saddle point (\bar{z}, \bar{p}) . The results will not depend on the

¹ It should be noted that a such a complete characterisation of the asymptotic behaviour does not exist in the literature in either a discrete or continuous time setting.

choice of (\bar{z}, \bar{p}) from among the possibly multiple saddle points.

Finally, it will sometimes be more convenient to work under the assumption:

$$(\bar{z}, \bar{p}) \text{ is strictly positive.} \quad (22)$$

Results in section 4 make use of this assumption. When this assumption does not hold then these results are more technical and are presented in appendix B. Note that section 5, where modification methods are discussed, does not use this assumption.

4 Asymptotic behaviour

In this section we provide an exact characterization of the asymptotic behaviour of system (18). In particular, one of our main results is to show that despite the non-linearity of the dynamics, all solutions of (18) converge to trajectories that satisfy an explicit linear ODE. To state this result we will first need to define the following.

We define a matrix Q and an associated second order ODE by

$$Q = K' A^T K A, \quad \ddot{z}(t) + Q\tilde{z}(t) = 0 \quad (23)$$

and a subspace \mathcal{Q} by

$$\mathcal{Q} = \text{span}(\{v \in \ker(H) : v \text{ is an eigenvector of } Q\}). \quad (24)$$

With these definitions we can now define the set \mathcal{X} of limiting trajectories of (18).

Definition 16 We define \mathcal{X} as trajectories $(z(t), p(t)) \geq 0$ of the form

$$(z(t), p(t)) = (\bar{z}, \bar{p}) + (\tilde{z}(t), \tilde{p}(t)) \quad (25)$$

where $(\bar{z}, \bar{p}) > 0$ is a saddle point of the Lagrangian φ in (11) and $\tilde{z}(t)$ solves the linear ODE (23) with initial conditions $\tilde{z}(0), \dot{\tilde{z}}(0)$ in subspace \mathcal{Q} defined in (24). $\tilde{p}(t)$ is recovered in terms of $\tilde{z}(t)$ from

$$\dot{\tilde{z}}(t) = -K' A^T \tilde{p}(t) \quad (26)$$

which is well defined as A^T has trivial kernel.

We now state the main result on the asymptotic behaviour of (18) which is proved in appendix A.

Theorem 17 Let (21),(22) hold. Then the solutions of (18) satisfy the following:

- (i) For all initial conditions, a solution of (18) converges to a proper solution in the set \mathcal{X} defined by Definition 16 or is of that form.

- (ii) The flow generated by (18) is pathwise stable.

Various remarks on this result are included below.

Remark 18 Despite the non-linearity of the dynamics, the set \mathcal{X} has an explicit characterization as sinusoidal oscillating solutions to a linear ODE, with an explicit set of initial conditions.

Remark 19 The form of the linear ODE (23) depends only on the network structure and gain parameters. In particular, the ODE does not depend on the choice of utility functions U_i ; the set \mathcal{X} depends on the utility functions U_i only through the location of the saddle point (\bar{z}, \bar{p}) .

Remark 20 The existence of such oscillations has been observed before in the literature in examples with specific utility functions and network topologies (see e.g. [2],[9] and the references therein). Theorem 17 provides an exact characterization of these oscillations in a general network, thus allowing to quantify how severe these oscillations can be. In particular, the Theorem shows that these can be of arbitrarily large size and cannot be mitigated by the choice of utility functions. Furthermore, the oscillations are not among the optimal points of problem (10) and can lead to arbitrarily large deviations from the optimum. It should also be noted that the exact structure of the limiting oscillatory solutions can give insight into designing schemes to avoid them (discussed in section 5).

Remark 21 It is possible for the network parameters to be such that the set \mathcal{X} of limiting solutions of (18) contains only the saddle points. In this case Theorem 17 establishes convergence of (18) to an optimum of the optimisation problem (10). However, in section 6 we illustrate by means of simulations that, even when the network parameters are such that convergence is achieved, prolonged transient oscillatory behaviour can be observed, if the network parameters are close to a set of parameters that give rise to permanent oscillations.

The explicit form of the linear ODE (23) in terms of the network structure allows us to compute the eigenvectors and eigenvalues of Q and solve (23).

Corollary 22 Let $(z(t), p(t)) = (\bar{z}, \bar{p}) + (\tilde{z}(t), \tilde{p}(t))$ be a solution in \mathcal{X} . Then $\tilde{z}(t)$ can be written in the form

$$\tilde{z}_l(t) = \tilde{z}_l^0 + \tilde{z}_l^1(0) \cos(\sqrt{\lambda_j}t) + \frac{\dot{\tilde{z}}_l^1(0)}{\sqrt{\lambda_j}} \sin(\sqrt{\lambda_j}t) \quad (27)$$

where $\tilde{z}^0 \in \ker(Q) \cap \mathcal{Q}$, $\tilde{z}^1(0) \in \mathcal{Q}$ with $K' \tilde{z}^1(0) \in \ker(Q)^\perp$ where Q and \mathcal{Q} are defined by (23) and (24) respectively. The constants λ_j are the (non-zero) eigenvalues of Q given by

$$\lambda_j = k_j(k'_l + \dots + k'_{l+l'}) \quad (28)$$

for $s_j \sim z_l$.

The proofs of Theorem 17 and Corollary 22 may be found in appendix A. These make use of some geometric properties of the general gradient method (8), while also exploiting the structure of the problem considered. A side result that follows from the derivations is also the fact that the pathwise stability property (Theorem 17(ii)) holds for general gradient dynamics (Corollary 29 in appendix A).

5 Modification methods for convergence

We will now look at ways of modifying the system to avoid the occurrence of the oscillatory behaviour and cause convergence to a saddle point. It should be noted that for centralized optimisation problems more advanced methods with improved convergence rates are feasible (e.g. [5]). Here we focus on modifications that preserve the distributed structure of the problem discussing also how these can lead to different degrees of localization.

One such modification method was introduced by Arrow, Hurwicz and Uzawa [2], and works by transforming the constraints to compensate for the lack of strict concavity of the original Lagrangian. More precisely, the Lagrangian of (3) is formulated as

$$\varphi(x, y) = f(x) + \sum_{j=1}^m y_j \rho_j(g_j(x)) \quad (29)$$

where $\rho_j : \mathbb{R} \rightarrow \mathbb{R}$ are strictly increasing and strictly concave functions. The resultant system of equations when the gradient method is applied has a distributed structure; however, in the case of the economic problem considered here they are not fully localized, as each user needs to also be aware of the demand of other users trading with the same supplier².

A second such method is to add penalty functions (see e.g. [9]) to the Lagrangian to penalise violated constraints. The new Lagrangian of (3) becomes,

$$\varphi(x, y) = f(x) + \sum_{j=1}^m y_j g_j(x) + \sum_{i=1}^m \psi_i(g_i(x)) \quad (30)$$

where $\psi_j : \mathbb{R} \rightarrow \mathbb{R}$ are C^2 functions such that $\psi_j(u) = 0$ for $u \geq 0$ and $\psi_i(u)$ is strictly increasing and strictly concave on $u \leq 0$. This corresponds to a modification of the utility function f . However, in the case of the economic problem the new update rules are again not fully localized as above, requiring additional information exchange among users.

² This is due to the non-linearity of the functions ρ_j and the form of the constraints g_j , which involve the demand of all users trading with the same supplier s_j .

Analogous methods exist in discrete time, with the additional complication of having to incorporate ordering of the updates and appropriate choice of step size³. Examples of such methods are the alternating direction method of multipliers (ADMM) [4], [10] and proximal point algorithms [24], [8], [16]. Within the paper we will restrict our analysis to continuous time and show below that a simple modification method motivated from the structure of the oscillations can be sufficient to obtain convergence.

In particular, for the remainder of this section we will consider schemes where the total utility function is modified by introducing local auxiliary variables; it will be seen that these can lead to schemes with guaranteed convergence to optimality without introducing additional information transfer. More precisely, we will look at modifications of the form

$$U_{new}(z, z') = U_{old}(z) + \gamma(Bz - z') \quad (31)$$

where $\gamma : \mathbb{R}^{L'} \rightarrow \mathbb{R}$ is a strictly concave function with a sole maximum $\gamma(0) = 0$, B is a constant matrix and $z' \in \mathbb{R}^{L'}$ is an additional vector to be maximised over. Observe that this has the same z maximisers, meaning that a maximiser \bar{z} of U_{old} , corresponds to a maximiser $(\bar{z}, B\bar{z})$ of U_{new} , because $\gamma(Bz - z') \leq 0$ and $= 0$ iff $z' = Bz$.

Note that, although the function γ is strictly concave, the function $\gamma(Bz - z')$ will not be strictly concave as a function of (z, z') , so the new utility function $U_{new}(z, z')$ will not be strictly concave.

The importance of this modification is that, despite the lack of strict concavity, an appropriate choice of B and γ can lead to distributed update schemes that are fully localized (discussed in the end of the section), and have guaranteed convergence to a saddle point.

The corresponding Lagrangian and assumptions for the analysis when (31) is used as an objective are

$$\begin{aligned} \varphi(z, z', p) &= \sum_{c_i} U_i(\sum_{z_l \sim c_i} z_l) + \\ &\quad + \gamma(Bz - z') + p^T(Y - Az) \\ \varphi : \mathbb{R}^L \times \mathbb{R}^{L'} \times \mathbb{R}^m &\rightarrow \mathbb{R}, \quad \gamma : \mathbb{R}^{L'} \rightarrow \mathbb{R} \quad (32) \\ \gamma \in C^2 &\text{ strictly concave with maximum } \gamma(0) = 0 \\ B \in \mathbb{R}^{L' \times L} &\text{ is a constant matrix, and (21) holds.} \end{aligned}$$

³ It should be noted that when a large step size is used discrete time gradient dynamics can exhibit oscillations even for strictly concave problems. Various methods such as diminishing step size and Armijo rule [3] can be used to avoid such oscillations.

We define the generalised gradient method on (32) by setting

$$\dot{z} = M_z K' \varphi_z \quad (33)$$

$$\dot{z}' = M_{z'} K'' \varphi_{z'} \quad (34)$$

$$\dot{p} = M_p K \varphi_p \quad (35)$$

in analogy with (20), where K'' is an additional diagonal matrix with strictly positive entries k_i'' .

The proposition below states that the generalised gradient method (33)-(35) is guaranteed to converge to a saddle point. Note that since the concavity of the Lagrangian $\varphi(z, z', p)$ defined by (32) is not strict, Theorem 10 cannot be used to deduce this result.

Proposition 23 *Let (32) hold and B have a trivial kernel, then the generalised gradient method (33)-(35) on the modified Lagrangian (32) converges to a maximum of the original problem given by (10).*

It is shown below that the addition of L extra variables, i.e. $z' \in \mathbb{R}^L$, is not always necessary. We only need to add enough to make up for the non-strict concavity of the sum of $U_i(\sum_{c_i \sim z_l} z_l)$ as a function of z . This leads to the following Theorem, which is a generalization of Proposition 23 and is proved in appendix A.

Theorem 24 *The generalised gradient method (33)-(35) applied to (32) will converge to a maximum of the original problem (10) if $\ker(B) \cap \ker(H) = \{0\}$.*

Remark 25 *The condition $\ker(B) \cap \ker(H) = \{0\}$ in Theorem 24 is related to the structure of the oscillations (Theorem 17) stated in section 4. In particular, the oscillations characterised in the set \mathcal{X} lie in the subspace⁴ \mathcal{Q} given by (24). As $\mathcal{Q} \subseteq \ker(H)$, the condition $\ker(B) \cap \ker(H) = \{0\}$ of Theorem 24 can be seen as adding sufficient strict concavity to the modified Lagrangian so as to disrupt these oscillations. Note, however, that the modified Lagrangian is still not strictly concave.*

For the update rules to be distributed when the gradient method is applied, we consider a class of functions γ which are of the form

$$\gamma(y) = \sum_{a=1}^{L'} \gamma_a(y_a) \quad (36)$$

for some strictly concave C^2 functions $\gamma_a : \mathbb{R} \rightarrow \mathbb{R}$. The additional information transfer in the gradient dynamics

⁴ More precisely, \tilde{z} (the deviation of z from the saddle point) satisfies $\tilde{z}(t) \in \mathcal{Q}$ for all $t \geq 0$. This is shown in the proof of Theorem 17 in appendix A.

due to the modification will be determined in this case by the sparsity structure of B .

In particular, an important class of matrices B is when the sparsity of the rows of B is chosen such that for all $a = 1, \dots, L'$, the a 'th element of Bz is equal to a linear combination of z_l that are connected to same consumer. This implies that when the gradient method is applied, the update rules for the users will be based on only local information, without requiring additional information from other users.

Indeed, the additional term in the dynamics for a single link z_l connected to consumer c_i is given by

$$\dot{z}_l = k_l' M_{z_l} \left(\dots + \sum_a B_{al} \gamma_a' \left(\sum_d B_{ad} z_d - z_a' \right) \right) \quad (37)$$

Hence if B_{al} are chosen to be zero unless z_l and z_a are connected for all a to the same consumer c_i , all the terms in (37) are local to this consumer. Explicitly the sums reduce to

$$\sum_{a: z_a \sim c_i \sim z_l} B_{al} \gamma_a' \left(\sum_{d: z_d \sim c_i \sim z_a} B_{ad} z_d - z_a' \right) \quad (38)$$

A similar computation can be done for the dynamics of z_a' , and the price dynamics are the same as in the unmodified gradient method.

An obvious special case is when B is the identity matrix. In this case the term in (38) reduces to the simple form

$$\gamma_l'(z_l - z_l') \quad (39)$$

and the dynamics for z' simplify to⁵

$$\dot{z}_l' = -k_l'' M_{z_l'} \gamma_l'(z_l - z_l'). \quad (40)$$

We can also exploit Theorem 24 to reduce the number of additional variables z_l' , as in the following corollary (proved in appendix A).

Corollary 26 *Let (32) hold. Split the indices of the links up into sets $Z_i = \{1 \leq l \leq L : c_i \sim z_l\}$ of those connected to each consumer. Then pick one index b_i arbitrarily from each set Z_i and form $B \in \mathbb{R}^{(L-n) \times L}$ by taking the $L \times L$ identity matrix and removing the b_i th row for each $i = 1, \dots, n$. With this B and γ of the form (36) the generalised gradient method (33)-(35) applied to (32) will converge to a maximum of the original problem (10).*

⁵ Note that when $\gamma_l(\cdot)$ is a quadratic function, then (40) becomes analogous to a low pass filtering of z_l .

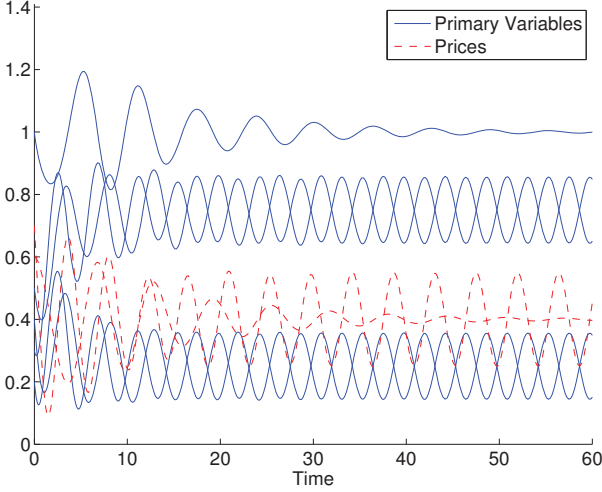


Fig. 2. Oscillations of the gradient method (18) applied to the network in Figure 1, with k_j, k'_l, Y_j, k''_i all set to 1. In this system \mathcal{Q} is non-trivial and this is evident in the presence of oscillatory modes that do not decay.

6 Numerical Examples

In this section we present some numerical examples that demonstrate the results. We will use the network in Figure 1, which can exhibit both convergent and oscillatory behaviour. The utility functions are set as $U_i(x) = \log(1+x)$ and the constants $k_j, k'_l, k''_3, k''_5, Y_j, k''_i$ are all set to 1, with $k'_2, k'_3 = 1$ in Figures 2 and 3, but equal to 1.1 in Figures 4 and 5. The trajectories in Figures 2 and 4 are generated by means of the unmodified gradient method (18). The modified gradient method (32)-(36) is used in Figures 3 and 5 with the functions $\gamma_l(\cdot)$ set to $\gamma_l(u) = -\frac{1}{2}u^2$ for all l , and the matrix B set to the identity matrix of the appropriate size. The initial conditions are the same for all Figures.

The form of \mathcal{Q} (defined in (35)) gives rise to permanent oscillating modes in the unmodified gradient method and these are demonstrated in Figure 2, where we have convergence to such an oscillating trajectory. In Figure 3 we see that the modified gradient method gives convergence to a saddle point with the same initial conditions.

In Figures 4 and 5 the different values of k'_2, k'_3 result in a different \mathcal{Q} which no longer gives rise to any permanent oscillating modes in the unmodified gradient method. Figure 4 shows the predicted convergence of the unmodified gradient method to a saddle point. However the closeness of the parameters k_j, k'_l to those in Figures 2 and 3, where there are permanent oscillations, causes prolonged transient oscillatory behaviour. In Figure 5 we see that the modified gradient method gives more rapid convergence in this case, with behaviour very similar to that in Figure 3.

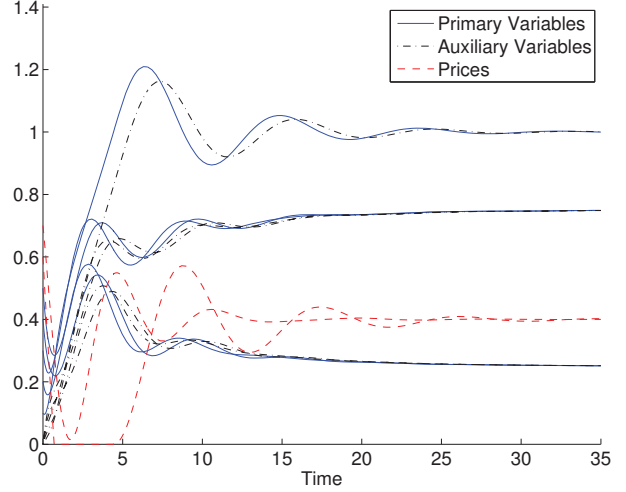


Fig. 3. Convergence of the modified gradient method (32)-(36) applied to the network in Figure 1, with k_j, k'_l, Y_j, k''_i all set to 1, the functions $\gamma_l(\cdot)$ defined by $\gamma_l(u) = -\frac{1}{2}u^2$, B set as the identity matrix, and the same initial conditions as in Figure 2. The modification always converges, and this is evident in the rapid decay of oscillations.

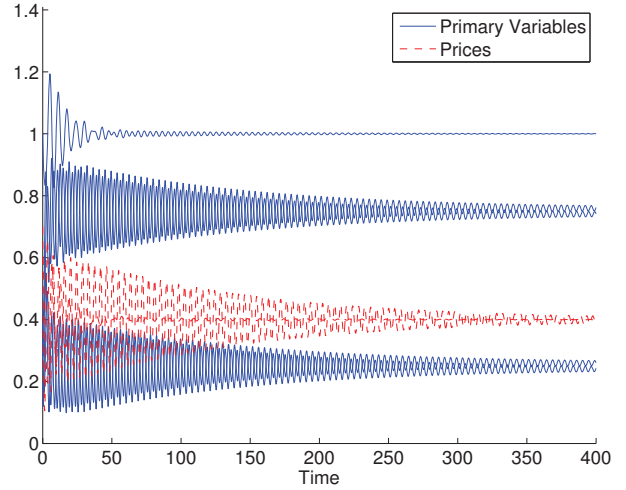


Fig. 4. Oscillations of the gradient method (18) applied to the network in Figure 1 with $k'_2, k'_3 = 1.1$, the rest of k_j, k'_l, Y_j, k''_i set to 1, and the same initial conditions as in Figure 2. In this case \mathcal{Q} , which depends on the k_j, k'_l , does not cause any permanent oscillating modes. Instead all oscillatory modes are damped and the system converges as predicted. However the closeness of the k_j, k'_l parameters to the case in Figures 2 and 3 where there are permanent oscillations is shown in the slow rate of decay of these oscillating modes.

7 Conclusions

We have considered the problem of distributed resource allocation in an economic network where users and suppliers trade for a single commodity. It has been discussed that due to the underlying structure of the problem the

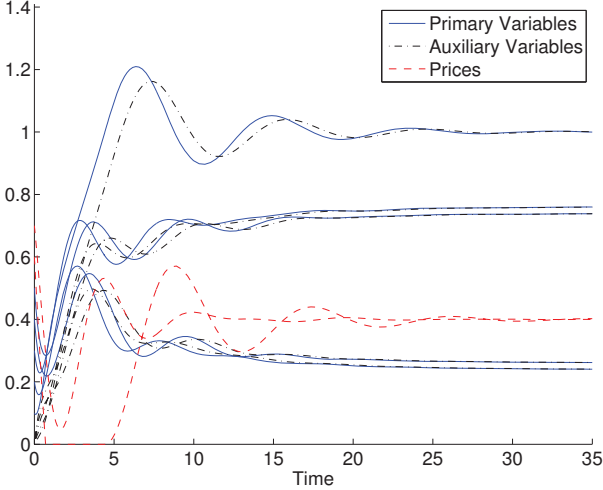


Fig. 5. Convergence of the modified gradient method (32)-(36) applied to the network in Figure 1, with $k'_2, k'_3 = 1.1$, the rest of k_j, k'_j, Y_j, k''_i set to 1, the functions $\gamma_l(\cdot)$ defined by $\gamma_l(u) = -\frac{1}{2}u^2$, B set as the identity matrix, and the same initial conditions as in Figure 2. The convergence behaviour of the modified gradient dynamics is almost identical to that in figure 3 where $k'_2, k'_3 = 1$ and all oscillatory modes are rapidly damped.

corresponding Lagrangian is not strictly concave-convex and hence convergence to an optimal solution by means of gradient type dynamics is not guaranteed. We have provided in this case a characterization of the asymptotic behaviour of trajectories, and provided an explicit description of the limiting solutions. Modifications have also been discussed such that convergence to an optimal solution can be guaranteed. Directions for future research include applications of these results and analysis tools in specific examples, such as analysis of electricity markets and improved schemes for multipath routing in communication networks.

It should also be noted that the derivations in the paper are of independent technical interest as they can be used as a basis for various extensions. For example, the pathwise stability property proved in this paper for general gradient dynamics, has been exploited in [14] to prove linearity of the asymptotic behaviour for general unconstrained gradient dynamics and subgradient dynamics with an internal saddle point. Extensions to general classes of subgradient dynamics will be addressed in future work.

Appendix A

In this appendix we provide the proof of the main results. For the convenience of the reader each of the proofs is presented in a separate subsection. The proof of Theorem 17 is given in section A.1. In section A.2 the proof of Corollary 22 is provided. Finally, in section A.3, the proofs of Theorem 24 and Corollary 26 are presented.

A.1 Proof of Theorem 17

In this subsection we prove Theorem 17. For the readers benefit we provide first an outline of the proof of the theorem. The proof is split into four steps.

- In step 1 we prove the pathwise stability of (18) which forms Theorem 17(ii). To do so, we define a new norm (denoted $\|\cdot\|_{\mathcal{W}}$) and show that the distance between any two solutions is non-increasing.
- In step 2 we prove that any solution of (18) converges to the orbit of a solution of (18) that lies a constant \mathcal{W} -distance from any saddle point.
- In step 3 we classify the solutions to (18) that lie a constant \mathcal{W} -distance from any saddle point, and show that the set of such solutions is the set \mathcal{X} given by Definition 16.
- In step 4 we combine steps 2 and 3 to establish convergence of solutions of (18) to the set \mathcal{X} . We then use the pathwise stability property proved in step 1 to upgrade a weaker notion of convergence that is used in step 2, to the stronger convergence property claimed by the theorem.

Step 1. In this step will prove that trajectories of (18) are pathwise stable, which will establish Theorem 17(ii).

We begin by defining a norm to take account of the matrices K and K' .

Definition 27 We define the \mathcal{W} -norm on $\mathbb{R}^L \times \mathbb{R}^n$ as

$$\|(z, p)\|_{\mathcal{W}}^2 = z^T K'^{-1} z + p^T K^{-1} p. \quad (\text{A.1})$$

For $z \in \mathbb{R}^L$ this is defined as $\|z\|_{\mathcal{W}} = \|(z, 0)\|_{\mathcal{W}}$.

Pathwise stability is a consequence of the following lemma.

Lemma 28 Let (21) hold, and let $(z'(t), p'(t))$ and $(z(t), p(t))$ be two solutions of (18). Then

$$\frac{d}{dt} \|(z'(t) - z(t), p'(t) - p(t))\|_{\mathcal{W}}^2 \leq 0. \quad (\text{A.2})$$

Equality in (A.2) holds if and only if for each index l, j ,

- φ is a linear function on the line from (z, p) to (z', p') .
- If z_l (respectively p_j) is active then z'_l (respectively p'_j) is active or semi-active.
- If z_l (respectively p_j) is inactive then z'_l (respectively p'_j) is inactive or semi-active.

PROOF. We start with the inequality, and first prove the result for proper solutions. Let $(z'(t), p'(t))$ and $(z(t), p(t))$ be two proper solutions of (18) and define

$W(t) = \frac{1}{2} \|(z'(t), p'(t)) - (z(t), p(t))\|_{\mathcal{W}}^2$. Using the formulation of (18) in Remark 15,

$$\begin{aligned} \dot{W} &= (z' - z)^T K'^{-1} (\dot{z}' - \dot{z}) + (p' - p)^T K^{-1} (\dot{p}' - \dot{p}) \\ &= (z' - z)^T (\varphi'_z - \varphi_z) - (p' - p)^T (\varphi'_p - \varphi_p) \\ &= \int_0^1 \frac{d}{ds} \{ (z' - z)^T \varphi_z \circ \gamma(s) - (p' - p)^T \varphi_p \circ \gamma(s) \} ds \end{aligned} \quad (\text{A.3})$$

where φ'_z denotes φ_z at (z', p') , and $\gamma(s) = ((z' - z)s + z, (p' - p)s + p)$ traverses the line from (z, p) to (z', p') linearly. Note that only the partial derivatives of φ depend on s . Continuing, letting $\hat{z} = z' - z$ and $\hat{p} = p' - p$,

$$\begin{aligned} \dot{W} &= \int_0^1 \hat{z}^T \varphi_{zz} \circ \gamma(s) \hat{z} ds + \int_0^1 \hat{z}^T \varphi_{zp} \circ \gamma(s) \hat{p} ds + \\ &\quad - \int_0^1 \hat{p}^T \varphi_{pp} \circ \gamma(s) \hat{p} ds - \int_0^1 \hat{p}^T \varphi_{pz} \circ \gamma(s) \hat{z} ds \\ &= \int_0^1 \hat{z}^T \varphi_{zz} \circ \gamma(s) \hat{z} ds - \int_0^1 \hat{p}^T \varphi_{pp} \circ \gamma(s) \hat{p} ds. \end{aligned} \quad (\text{A.4})$$

By concavity/convexity we have that $\varphi_{zz}, \varphi_{pp}$ are negative/positive semi-definite which shows that $\dot{W} \leq 0$ with equality if and only if they vanish along the line, which proves claim (a).

Now we consider fully general solutions. This has the effect of replacing $\dot{z}_l = k'_l \varphi_{z_l}$ with $k'_l \max(0, \varphi_{z_l})$ if $z_l = 0$. We will compare this modification with the computation above and show that the modification can never make \dot{W} positive. We will look at the z coordinates. The p coordinates are similar. The first line of (A.3) above shows that \dot{W} is the sum of contributions from each index l . For each l there are three possibilities:

- (i) z_l, z'_l are both active or semi-active. Then there is no modification and the contribution to \dot{W} is non-positive as shown before.
- (ii) z_l, z'_l are both inactive or semi-active. Then $z_l = z'_l = 0$ and the contribution is zero.
- (iii) One of z_l, z'_l is active and the other inactive. By symmetry we need only consider z'_l active and z_l inactive. Then

$$\begin{aligned} (z'_l - z_l) k'_l{}^{-1} (\dot{z}'_l - \dot{z}_l) &= \\ &= (z'_l - z_l) (\varphi'_{z_l} - \max(\varphi_{z_l}, 0)) \\ &= (z'_l - z_l) (\varphi'_{z_l} - \varphi_{z_l}) + (z'_l - z_l) \min(0, \varphi_{z_l}). \end{aligned} \quad (\text{A.5})$$

The first term is the same as in the case (i), so is non-positive. The second term is strictly negative as $z'_l > z_l = 0$ and $\varphi_{z_l} < 0$ as z_l is inactive, which proves (b) and (c). \square

An immediate consequence of this proof is the fact that Lemma 28 also holds for the general gradient method

(7), (8), without this being restricted to the economic network. We hence have the following corollary.

Corollary 29 *Let (7) hold. Then the flow generated by the gradient method (8) is pathwise stable.*

Step 2. In this step we will prove that any solution of (18) converges to the orbit (see Definition 30 below) of a solution of (18) that lies a constant distance from any saddle point.

A simple consequence of Lemma 28 from step 1, is that the \mathcal{W} -distance of a solution of (18) to any saddle point is non-increasing in time. It seems natural, therefore, to apply LaSalle's theorem with the \mathcal{W} -distance to the saddle point as the Lyapunov like function. However, due to the switching in (18), an application of LaSalle's theorem in its classical form (as in e.g. [18]) is not possible. We will instead provide a direct proof that relies on pathwise stability⁶. To this end we need to define a form of convergence of trajectories.

Definition 30 *Let $(z(t), p(t))$ and $(z'(t), p'(t))$ be two solutions of (18). We say that $(z(t), p(t))$ converges to the orbit of $(z'(t), p'(t))$ if, as $t \rightarrow \infty$,*

$$\text{dist}((z(t), p(t)), \{(z'(t), p'(t)) : t \geq 0\}) \rightarrow 0. \quad (\text{A.6})$$

Lemma 31 *Let (21) hold and $(z(t), p(t))$ be a solution of (18). Then $(z(t), p(t))$ converges to the orbit of a solution $(z'(t), p'(t))$ of (18) which lies a constant \mathcal{W} -distance from any saddle point.*

PROOF. We will first construct a solution $(z'(t), p'(t))$ and show that $(z(t), p(t))$ converges to its orbit. Then we will establish that the constructed solution $(z'(t), p'(t))$ is a constant \mathcal{W} -distance from any saddle point.

By Lemma 28 $(z(t), p(t))$ is bounded as $t \rightarrow \infty$ so it has an ω -limit point. We define $(z'(0), p'(0))$ to be such a limit point, then there is a sequence $T_n \rightarrow \infty$ as $n \rightarrow \infty$ such that $\|(z', p')(0) - (z, p)(T_n)\|_{\mathcal{W}} < 1/n$. By Lemma 28 this implies that $\|(z', p')(t) - (z, p)(t + T_n)\|_{\mathcal{W}} < 1/n$ for all $t \geq 0$, which establishes the convergence of $(z(t), p(t))$ to the orbit of $(z'(t), p'(t))$.

To show that the distance to any saddle point is constant, we will first prove that $(z'(t), p'(t))$ is almost periodic in the sense that there is a sequence $\tau_n \rightarrow \infty$ as $n \rightarrow \infty$ for which, as $n \rightarrow \infty$,

$$\sup_{t \geq 0} \|(z', p')(t) - (z', p')(t + \tau_n)\|_{\mathcal{W}} \rightarrow 0. \quad (\text{A.7})$$

⁶ We would like to thank A. Cherukuri, E. Mallada and J. Cortés for kindly pointing out that the invariance principle for hybrid automata [20] also does not apply to the system (18).

We claim that $\tau_k = T_{n_k} - T_{n'_k}$, with $n_k, n'_k \rightarrow \infty$ as $k \rightarrow \infty$ chosen so that $n_k > n'_k$ and $\tau_k \rightarrow \infty$ as $k \rightarrow \infty$, is such a sequence. Indeed, for any $t \geq 0$ we have,

$$\begin{aligned} & \|(z', p')(t) - (z', p')(t + \tau_k)\|_{\mathcal{W}} \leq \\ & \left\| (z', p')(t) - (z, p)(t + T_{n'_k}) \right\|_{\mathcal{W}} + \\ & \left\| (z', p')(t + \tau_k) - (z, p)(t + T_{n'_k} + \tau_k) \right\|_{\mathcal{W}} + \\ & \left\| (z, p)(t + T_{n'_k}) - (z, p)(t + T_{n'_k} + \tau_k) \right\|_{\mathcal{W}} \end{aligned} \quad (\text{A.8})$$

The first two terms on the right hand side are each less than $1/n'_k$. The final term is less than $\frac{1}{n'_k} + \frac{1}{n_k}$ by the triangle inequality and the definition of T_n . Thus as $k \rightarrow \infty$ all three terms go to zero proving the claim.

Now let (\bar{z}, \bar{p}) be an arbitrary saddle point. By Lemma 28, the distance $W(t) = \|(z'(t), p'(t)) - (\bar{z}, \bar{p})\|_{\mathcal{W}}$ is non-increasing in time t . By the almost periodicity shown above, $W(t)$ takes values arbitrarily close to $W(0)$ at a sequence of times $t_n \rightarrow \infty$ as $n \rightarrow \infty$. Therefore it must be constant. \square

Step 3. In this step we will characterise the solutions of (18) that lie a constant \mathcal{W} -distance to any saddle point, and show that the set of such solutions is given by the set \mathcal{X} (Definition 16). To do so, we shall choose a particular saddle point (\bar{z}, \bar{p}) and find all solutions which have constant \mathcal{W} -distance to this particular saddle point. Then, a posteriori, we deduce that these solutions are also a constant \mathcal{W} -distance to *all* saddle points.

From assumption (22) the saddle point (\bar{z}, \bar{p}) is strictly positive. We begin with the following lemma, which allows us to ignore the switching at the boundary.

Lemma 32 *Let (21) hold and $(\bar{z}, \bar{p}) > 0$ be a saddle point. Then all solutions of (18) which have constant \mathcal{W} -distance to this saddle point are proper.*

PROOF. This is immediate from Lemma 28(b),(c) as all coordinates of the saddle point are active. \square

When we express solutions of (18) as $(\bar{z} + \tilde{z}(t), \bar{p} + \tilde{p}(t))$, it can be directly verified from (18) that the contribution of the non-linearity coming from the utility functions is zero if and only if $\tilde{z}(t) \in \ker(H)$. Indeed, for proper solutions we may ignore the switching at the boundary and convert to the second order differential equation,

$$\ddot{\tilde{z}} = K'(H^T(U'(H(\tilde{z} + \bar{z})))' - A^T K A(\tilde{z} + \bar{z}) + A^T Y) \quad (\text{A.9})$$

where $(U'(Hz))'$ is $\frac{d}{dt}(U'(Hz))$. If $\tilde{z} \in \ker(H)$ then the right hand side reduces to $-K'A^T K A \tilde{z}$ as all other terms

cancel because (\bar{z}, \bar{p}) is a saddle point. This discussion leads to the following lemma.

Lemma 33 *Let (21),(22) hold. Suppose that $(\tilde{z}(t) + \bar{z}, \tilde{p}(t) + \bar{p})$ is a proper solution of (18). Then if for some time t_0 we have $\tilde{z}(t_0) \in \ker(H)$ then for that time,*

$$\ddot{\tilde{z}}(t_0) + Q\tilde{z}(t_0) = 0, \quad (\text{A.10})$$

where Q is defined by (23).

When this condition does not hold, Lemma 28 shows that the \mathcal{W} -distance from the saddle point is decreasing. Indeed, we have slightly more.

Lemma 34 *Let (21) hold, then for any two proper solutions $(z'(t), p'(t)), (z(t), p(t))$ of (18), define $W(t) = \|(z' - z, p' - p)\|_{\mathcal{W}}^2$. Then $\dot{W} = 0$ at time t iff $z'(t) - z(t) \in \ker(H)$.*

PROOF. By Lemma 28(a) it is sufficient to show that φ is linear on the line between $(z'(t), p'(t))$ and $(z(t), p(t))$ if and only if $z'(t) - z(t) \in \ker(H)$. As φ is a linear function of p , this follows by the strict concavity of the utility functions U_i . \square

Thus all solutions which are at a constant \mathcal{W} -distance to the saddle point will have $\tilde{z}(t) \in \ker(H)$ for all times t , and hence will, by Lemma 33, satisfy the linear ODE (23). It is however possible to find solutions which have $\tilde{z}(t_0) \in \ker(H)$ for some time t_0 , but which are not at a constant \mathcal{W} -distance to the saddle point. The exact condition for the \mathcal{W} -distance to the saddle point to remain constant is for \tilde{z} to lie in the subspace defined below. This is an alternate definition to that given by (24). The equivalence between the two definitions will be proved below in Lemma 38.

Definition 35 (Subspace \mathcal{Q}) *We define $\mathcal{Q} = K'A^T K A$ and \mathcal{Q} to be the largest subspace of $\ker(H)$ which is invariant under the action of Q . By which we mean that if $v \in \mathcal{Q}$ then $Qv \in \mathcal{Q}$.*

Remark 36 *It is easy to see that \mathcal{Q} can equivalently be defined as*

$$\mathcal{Q} = \bigcap_{i=0}^{\infty} Q^i \ker(H). \quad (\text{A.11})$$

The contrapositive of this is that

$$v \notin \mathcal{Q} \implies \exists i \in \mathbb{N}, \text{ such that } Q^i v \notin \ker(H). \quad (\text{A.12})$$

Remark 37 *There is no guarantee that \mathcal{Q} is non trivial.*

Lemma 38 *Definition 35 and Equation (24) define the same set \mathcal{Q} .*

PROOF. First suppose that v is in the span defining \mathcal{Q} in (24). As both definitions define a linear space and \mathcal{Q} is symmetric with respect to the inner product associated with the \mathcal{W} -norm on \mathbb{R}^L , we may without loss of generality take v to be an eigenvector of Q with some eigenvalue λ . Then clearly $Q^i v = \lambda^i v \in \ker(H)$ for all positive integer q , so $v \in \mathcal{Q}$ as defined by Definition 35.

Second suppose that $v \in \mathcal{Q}$ as defined by Definition 35. As before, we express v as a linear combination of eigenvectors of Q . If $v \in \ker(A)$ then certainly v is an eigenvector of Q so it is in the span defining \mathcal{Q} in (24). By linearity it is sufficient to also consider the case where v is \mathcal{W} -orthogonal to $\ker(A)$. We express v as a linear combination of the eigenvectors of Q , $v = \sum_{j=1}^m a_j v^j$. Let the corresponding eigenvalues of Q be $\lambda_1, \dots, \lambda_m$. Take the set of eigenvalues with maximum magnitude, i.e. λ_j for which $|\lambda_j| = \max_{k=1, \dots, m} |\lambda_k|$. This is associated with a subset of the eigenvectors, which without loss of generality we take to be v^1, \dots, v^s . We compute for integer i ,

$$|\lambda_1|^{-2i} Q^{2i} v = \sum_{j=1}^s a_j v^j + \sum_{j=s+1}^m a_j v^j \left(\frac{\lambda_j}{|\lambda_1|} \right)^{2i}. \quad (\text{A.13})$$

When we take $i \rightarrow \infty$ the second term on the right vanishes, and as the left hand side is in $\ker(H)$ the first term on the right is also. Hence this term belongs to the span. We then subtract this term from v and argue inductively on the difference to show that v is in the span defining \mathcal{Q} in (24). \square

To prove that $\tilde{z}(t) \in \mathcal{Q}$ for some time t if and only if the solution $(z(t), p(t))$ is a constant \mathcal{W} -distance from the saddle point, we first show (Lemma 39) that if we start a solution to (18) with initial condition $\tilde{z}(0), \dot{\tilde{z}}(0) \in \mathcal{Q}$ then it obeys the *linear* equation (23).

Solutions to (23), with this restriction on the initial conditions, will remain in \mathcal{Q} for all time, and, by Lemma 34, will remain a constant \mathcal{W} -distance to the saddle point.

Lemma 39 *Let (21),(22) hold. Suppose a proper solution of (18), $(\tilde{z}(t) + \bar{z}, \tilde{p}(t) + \bar{p})$, has $\tilde{z}(t_0), \dot{\tilde{z}}(t_0) \in \mathcal{Q}$ for some t_0 , then $\tilde{z}(t), \dot{\tilde{z}}(t) \in \mathcal{Q}$ for all $t \geq t_0$, and $\tilde{z}(t)$ satisfies (23).*

PROOF. A solution of the linear equation (23) exists and is proper. For this solution the non-linear terms in equation (18) are zero as $\tilde{z}(t) \in \mathcal{Q} \subset \ker(H)$. Hence it is a solution of the full non-linear equation (18) and by uniqueness is the solution of (18). \square

Solutions in \mathcal{X} satisfy $\tilde{z}(t) \in \mathcal{Q} \subset \ker(H)$ for all times t and are hence a constant \mathcal{W} -distance from the saddle point (Lemma 34). To prove that these exhaust all proper solutions which lie a constant \mathcal{W} -distance to the saddle point we consider the case where a solution has $\tilde{z}(t) \in \ker(H)$ for all times t (a necessary and sufficient condition for the distance from the saddle point to be constant by Lemma 34), and deduce that $\tilde{z}(t) \in \mathcal{Q}$.

Lemma 40 *Assume (21),(22) and let $(z(t), p(t))$ be a proper solution of (18) then if $z(t) - \bar{z} \in \ker(H)$ for all time t , then $z(t) - \bar{z} \in \mathcal{Q}$ for all time t .*

PROOF. As the solution is proper and $\tilde{z}(t) = z(t) - \bar{z} \in \ker(H)$ for all times t , Lemma 33 implies that \tilde{z} solves the linear ODE (23). Any solution of this is analytic, and in particular we can expand \tilde{z} as a power series in t about any t_0 to arbitrarily high order. Suppose $\tilde{z} \in \ker(H)$ for all time t , as assumed in the lemma, but there is a t_0 for which $\tilde{z}(t_0) \notin \mathcal{Q}$. Then by Remark 36 there is smallest integer $q \geq 0$ such that $HQ^q \tilde{z} \neq 0$, i.e. $Q^q \tilde{z} \notin \ker(H)$. Expanding to twice this order,

$$\begin{aligned} \tilde{z}(t_0 + t) &= \tilde{z}(t_0) + \dot{\tilde{z}}(t_0)t + \frac{1}{2}\ddot{\tilde{z}}(t_0)t^2 + \dots \\ &\dots + \frac{1}{(2q)!} \frac{d^{2q}\tilde{z}}{dt^{2q}}(t_0)t^{2q} + O(t^{2q+1}) \end{aligned} \quad (\text{A.14})$$

From the equation for \tilde{z} we can compute these derivatives,

$$\begin{aligned} \frac{d^{2i}\tilde{z}}{dt^{2i}}(t_0) &= (-Q)^i \tilde{z}(t_0) \\ \frac{d^{2i+1}\tilde{z}}{dt^{2i+1}}(t_0) &= (-Q)^i \dot{\tilde{z}}(t_0) \end{aligned} \quad (\text{A.15})$$

for integer $i \geq 0$. By assumption the first $2q - 1$ derivatives lie in $\ker(H)$, but the $2q$ th derivative does not. By Taylor's theorem, there is some interval for which $\tilde{z}(t)$ does not lie in $\ker(H)$, which is a contradiction. \square

In summary, we have shown that a solution of (18) lies a constant \mathcal{W} -distance from the saddle point (\bar{z}, \bar{p}) if and only if it is in the set \mathcal{X} . It remains to show that solutions in \mathcal{X} are a constant distance from all saddle points. It can be seen from the form of the linear ODE (23), (or from the explicit solutions given by Corollary 22), that solutions in \mathcal{X} return to any neighbourhood of their initial condition, and as Lemma 28 implies that they cannot become further away from any saddle point, this implies that they also cannot become closer. This leads to the proposition below.

Proposition 41 *Assume that (21),(22) hold. Then the set of solutions of (18) that lie a constant \mathcal{W} -distance to any saddle point is exactly the set \mathcal{X} given by Definition 16.*

Step 4. In this step we will combine the results of the previous steps and complete the proof of Theorem 17(i), that solutions of (18) converge to solutions in \mathcal{X} .

By Proposition 41 the set \mathcal{X} is exactly the set of solutions to (18) that lie a constant distance to any saddle point, therefore Lemma 31 implies convergence to the orbit of a solution in \mathcal{X} . However, to obtain the full result we must upgrade the weaker convergence to the orbit of a solution in \mathcal{X} given by Lemma 31 to the stronger convergence to a solution in \mathcal{X} .

Proposition 42 *Let (21),(22) hold and $(z(t), p(t))$ be a solution of (18). Then, $(z(t), p(t))$ converges to a solution in \mathcal{X} (defined in Definition 16).*

PROOF of Proposition 42. Let $(z(t), p(t))$ be a solution of (18). By Lemma 31 $(z(t), p(t))$ converges to the orbit of a solution $(z'(t), p'(t))$ of (18) that lies a constant \mathcal{W} distance from any saddle point. By Proposition 41, the set of all solutions that lies a constant \mathcal{W} distance from any saddle point is exactly \mathcal{X} . In particular, $(z(t), p(t))$ converges to the set \mathcal{X} .

It now remains to upgrade the convergence to the set \mathcal{X} to convergence to a solution. To this end we use the convergence to the set \mathcal{X} and Lemma 28 to obtain points $(z^{(n)}, p^{(n)}) \in \mathcal{X}$ and times t_n such that,

$$\left\| (z(t_n) - z^{(n)}, p(t_n) - p^{(n)}) \right\|_{\mathcal{W}} \leq 1/n. \quad (\text{A.16})$$

By the form of \mathcal{X} there are solutions $(z^{(n)}(t), p^{(n)}(t)) \in \mathcal{X}$ with $(z^{(n)}(t_n), p^{(n)}(t_n)) = (z^{(n)}, p^{(n)})$, and by an application of Lemma 28, we have for all $t \geq t_n$,

$$\left\| (z(t) - z^{(n)}(t), p(t) - p^{(n)}(t)) \right\|_{\mathcal{W}} \leq 1/n. \quad (\text{A.17})$$

By the boundedness of trajectories, the set of initial conditions of these solutions $\{(z^{(n)}(0), p^{(n)}(0)) : n \in \mathbb{N}\}$ is relatively compact, and we can pass to a subsequence n_k for which $(z^{(n_k)}(0), p^{(n_k)}(0)) \rightarrow (z'(0), p'(0)) \in \mathcal{X}$ as $k \rightarrow \infty$. We claim that $\|(z(t), p(t)) - (z'(t), p'(t))\|_{\mathcal{W}} \rightarrow 0$ as $t \rightarrow \infty$. Indeed, for any $\epsilon > 0$ there is a $k \in \mathbb{N}$ such that for all $t \geq t_{n_k}$, we have

$$\left\| (z(t) - z^{(n_k)}(t), p(t) - p^{(n_k)}(t)) \right\|_{\mathcal{W}} \leq \epsilon/2 \quad (\text{A.18})$$

and also for all $t \geq 0$,

$$\left\| (z'(t) - z^{(n_k)}(t), p'(t) - p^{(n_k)}(t)) \right\|_{\mathcal{W}} \leq \epsilon/2 \quad (\text{A.19})$$

where in each case we have used Lemma 28 to extend to later times. The claim now follows from the

triangle inequality, which completes the proof of the proposition. \square

This concludes the proof of Theorem 17.

A.2 Proof of Corollary 22

To prove Corollary 22 we will calculate the eigenvectors and eigenvalues of the matrix Q . We begin by splitting Q into blocks that each correspond to a single supplier.

Lemma 43 *Q has the structure (up to a reordering of the coordinates of \tilde{z}):*

$$Q = \text{diag}(J_1, \dots, J_m)$$

$$J_j = k_j \begin{bmatrix} k'_l & \dots & k'_l \\ k'_{l+1} & \dots & k'_{l+1} \\ \vdots & \vdots & \vdots \\ k'_{l+l'} & \dots & k'_{l+l'} \end{bmatrix} \quad (\text{A.20})$$

where $z_l, \dots, z_{l+l'}$ are all the links connected to the supplier s_j .

PROOF. We can compute the entries of Q explicitly from its definition.

$$Q_{ab} = \sum_{i,j,k} K'_{ai} A_{ij}^T K_{jk} A_{kb}$$

$$= \sum_j k'_a A_{ja} k_j A_{jb}$$

$$= \begin{cases} k'_a k_j & \text{if } z_a \sim s_j \sim z_b \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.21})$$

The stated form then follows immediately when we reorder the coordinates into blocks of links connected to the same supplier. \square

The block diagonal structure of Q means that the eigenvalues of Q are completely determined by the eigenvalues of each J_j .

Lemma 44 *The contribution to the eigensystem of Q from each J_j defined by (A.20) is l' zero eigenvalues with eigenvectors in $\ker(A)$, and a single eigenvalue/vector pair:*

$$v^j = [k'_l, \dots, k'_{l+l'}]^T$$

$$\lambda_j = k_j (k'_l + \dots + k'_{l+l'}) \quad (\text{A.22})$$

where the vector v^j is written out in its form as an eigenvector of J_j .

PROOF. By inspection, a basis of \mathbb{R}^L for the eigenvectors with eigenvalue 0 is,

$$\{e_l - e_{l+a} : 1 \leq a \leq l'\} \quad (\text{A.23})$$

where e_i is the standard Euclidean basis. These all lie in $\ker(A)$, and the dimension of the span is l' . The final eigenvector can be verified again by inspection, and this makes exactly $l' + 1$ eigenvectors, which is the size of J_j , so it is all of them. \square

Proving Corollary 22 is now a simple computation which we leave to the reader.

A.3 Convergence of the modified dynamics

In this section we present the proof of the convergence of the modified dynamics (33)-(35) to an optimum point of (10).

PROOF of Theorem 24. It should be noted that the \mathcal{W} -norm (Definition 27) and the proof of Lemma 28 easily extends to the generalised gradient method (33)-(35), and this allows Lemma 31 to also be extended to apply to solutions of (33)-(35). Therefore, any solution of (33)-(35) will converge to the set of solutions to (33)-(35) that lie a constant \mathcal{W} -distance to any saddle point. It hence suffices to show that the only solutions that have this distance constant are saddle points.

Let $((z(t), z'(t)), p(t))$ be a solution at constant \mathcal{W} -distance from a saddle point $((\bar{z}, \bar{z}'), \bar{p})$. By Lemma 28(a), using the fact that φ is always a linear function of p and also the strict concavity of both the utility functions and γ , we deduce that $Bz - z'$ and $H_z z$ are constant on the line from the saddle point to $((z(t), z'(t)), p(t))$.

By evaluating $Bz - z'$ and $H_z z$ at the saddle point we deduce that $Bz(t) - z'(t) = 0$ and $H_z z(t) = H\bar{z}$. Next we use the properties of the saddle point and (32) to express $\varphi_{z(t)} = -A^T(p(t) - \bar{p}) + B^T((\nabla\gamma)(Bz(t) - z'(t)))$ and $\varphi_{z'(t)} = -(\nabla\gamma)(Bz(t) - z'(t))$. Using that $Bz(t) - z'(t) = 0$ and γ has a maximum at 0, we deduce that $\varphi_{z(t)} = -A^T(p(t) - \bar{p})$ and $\varphi_{z'(t)} = 0$. By taking the time derivative of the relation $Bz(t) - z'(t) = 0$ we have $BM_z K' \varphi_{z(t)} = M_{z'} K'' \varphi_{z'(t)} = 0$, and the condition $\ker(B) \cap \ker(H) = \{0\}$ gives that either $M_z K' \varphi_{z(t)} = 0$ or $M_z K' \varphi_{z(t)} \notin \ker(H)$. The second possibility is ruled out by the relation $H_z z(t) = H\bar{z} = \text{constant}$, which when time differentiated gives $HM_z K' \varphi_{z(t)} = 0$. Therefore we have $\dot{z} = 0$ and z is constant. This means that \dot{p} is also constant, and by the boundedness of trajectories implied by the fact that the \mathcal{W} -norm distance to the saddle point is non-increasing, p must be constant also. Therefore all the time derivatives are zero and $((z(t), z'(t)), p(t))$ is a saddle point. \square

PROOF of Proposition 23 The proposition follows directly from Theorem 24.

PROOF of Corollary 26 The corollary follows from Theorem 24, and by noting that non-trivial elements of $\ker(B)$ are not in $\ker(H)$. \square

Appendix B

Here we will present an extended version of Theorem 17 that holds without the assumption of (22). In this case the boundary switching plays a key role, and the limiting trajectories can lie on the boundary. To relax (22), we introduce the concept of a maximally active saddle point, which will be used to define the set of limiting trajectories.

Definition 45 Given two saddle points (\bar{z}, \bar{p}) and (\bar{z}', \bar{p}') we say that

- (\bar{z}, \bar{p}) is not less active than (\bar{z}', \bar{p}') , which is denoted $(\bar{z}', \bar{p}') \preceq (\bar{z}, \bar{p})$, if, for any index l , $\bar{z}'_l > 0$ implies that $\bar{z}_l > 0$.
- (\bar{z}, \bar{p}) is more active than (\bar{z}', \bar{p}') , which is denoted $(\bar{z}', \bar{p}') \prec (\bar{z}, \bar{p})$, if both $(\bar{z}', \bar{p}') \preceq (\bar{z}, \bar{p})$ and there is an index l for which $\bar{z}'_l = 0$ and $\bar{z}_l > 0$.

We say that (\bar{z}, \bar{p}) is maximally active, if for any other saddle point (\bar{z}', \bar{p}') we have $(\bar{z}', \bar{p}') \preceq (\bar{z}, \bar{p})$.

It will be shown later in the appendix (Lemma 51) that a maximally active saddle is always guaranteed to exist.

The following proposition is an extension of Theorem 17 where (22) is not required to hold. As in the main text, we must first define the set of limiting solutions \mathcal{X}' .

Definition 46 Let (\bar{z}, \bar{p}) be a maximally active saddle point. We define $\hat{\mathcal{G}}$ to be the network obtained by removing the links z_l , that have $\bar{z}_l = 0$, from the original network \mathcal{G} . We denote the corresponding version of the set \mathcal{X} computed from $\hat{\mathcal{G}}$ as $\hat{\mathcal{X}}$.

Given a trajectory (\hat{z}, \hat{p}) in $\hat{\mathcal{X}}$ we extend it to $(z(t), p(t)) \in \mathbb{R}^{L+n}$ by putting $z_l = 0$ for each of the removed links. The set \mathcal{X}' is then defined those $(z(t), p(t))$ obtained in this way, that additionally satisfy $\varphi_{z_l(t)} = 0$ for each of the previously removed links.

Proposition 47 Let (21) hold. Then the solutions (18) satisfy the following:

- (i) For all initial conditions, a solution of (18) converges to a solution in the set \mathcal{X}' defined by Definition 46, or is of that form.

(ii) The flow generated by (18) is pathwise stable.

Remark 48 When (22) holds, the sets \mathcal{X} and \mathcal{X}' are equal and Proposition 47 reduces to Theorem 17.

The remainder of this section will consist of a proof of this Proposition.

PROOF of Proposition 47(ii). We simply note that the reasoning used in the proof of Theorem 17(ii) still holds as Lemma 28 does not depend on (22). \square

PROOF of Proposition 47(i).

To prove Proposition 47(i), and motivate why one should expect it to be true, we will first consider the more straightforward case where there is a saddle point which has all its coordinates either active or inactive, but not semi-active.⁷ The remaining case, where there are semi-active coordinates at the saddle point, is substantially longer and more complex, and will be considered later. During this first part of the proof we shall single out a saddle point (\bar{z}, \bar{p}) with this property.⁸

We will follow the same basic strategy as in the proof of Theorem 17. However, in this case we cannot assume that the trajectories that are a constant \mathcal{W} -distance to the saddle point are proper. Instead, we will use the properties of the chosen saddle point (\bar{z}, \bar{p}) to appropriately transform the problem.

Lemma 28(b)-(c) implies that trajectories $(z(t), p(t))$ that are a constant \mathcal{W} -distance from the saddle point (\bar{z}, \bar{p}) have z_l always inactive or semi-active if \bar{z}_l is inactive. Thus these coordinates are always zero and the trajectories are also solutions to the reduced system where we remove links z_l that are inactive at the saddle point.

This leads us to the following lemma.

Lemma 49 *Let (21) hold and let (\bar{z}, \bar{p}) be a maximally active saddle point with no semi-active coordinates. Then the set \mathcal{X}' obtained from (\bar{z}, \bar{p}) is the set of solutions of (18) that have the \mathcal{W} -distance to this saddle point constant.*

PROOF. The above discussion together with Theorem 17 applied to the reduced system \mathcal{G} establishes that any

⁷ Note that any saddle point with this property will be maximally active by Lemma 28(b)-(c), so the notion of maximally active will not be needed directly in this case, but will become important in the second part of the proof at the end of the appendix.

⁸ As in section 4, the results obtained are independent of this choice of saddle point.

solution of (18) which is a constant \mathcal{W} -distance to (\bar{z}, \bar{p}) will lie in \mathcal{X}' . Next we observe that the condition that $\varphi_{z_l(t)} \leq 0$ in the definition of \mathcal{X}' is necessary to ensure that these trajectories are also solutions to the non-reduced system. Conversely, any trajectory in \mathcal{X}' is a solution to (18) that is a constant \mathcal{W} -distance to the saddle point. \square

To prove convergence to these solutions we use the same strategy as in the proof of Theorem 17.

Lemma 50 *Let (21) and $(z(t), p(t))$ be a solution of (18) and (\bar{z}, \bar{p}) be a maximally active saddle point with no semi-active coordinates. Then, $(z(t), p(t))$ converges to a solution in \mathcal{X}' (defined in Definition 46).*

PROOF. This is identical to that of Proposition 42 but using Lemma 49 and set \mathcal{X}' instead of set \mathcal{X} . \square

This completes the proof of Proposition 47 when a maximally active saddle point has no semi-active coordinates.

We now move on to the remaining case, where any maximally active saddle point has at least one semi-active coordinate. We first show that a maximally active saddle point must exist.

Lemma 51 *Let (21) hold. Then there exists a maximally active saddle point.*

PROOF. It is sufficient to show that for any two saddle points (\bar{z}, \bar{p}) and (\bar{z}', \bar{p}') , there is a saddle point (\bar{z}'', \bar{p}'') with $(\bar{z}, \bar{p}) \preceq (\bar{z}'', \bar{p}'')$ and $(\bar{z}', \bar{p}') \preceq (\bar{z}'', \bar{p}'')$. We claim that $(\bar{z}'', \bar{p}'') = \frac{1}{2}(\bar{z} + \bar{z}', \bar{p} + \bar{p}')$ is such a saddle point.

To show this is indeed a saddle point we use that Lemma 28 implies that (\bar{z}'', \bar{p}'') is an equilibrium point of (18). Indeed, by Lemma 28 the solution of (18) starting at (\bar{z}'', \bar{p}'') is constrained to not become further in \mathcal{W} -distance from either (\bar{z}, \bar{p}) or (\bar{z}', \bar{p}') as time increases. By strict convexity of the \mathcal{W} -norm the only point obeying both conditions is (\bar{z}'', \bar{p}'') itself.

That $(\bar{z}, \bar{p}) \preceq (\bar{z}'', \bar{p}'')$ and $(\bar{z}', \bar{p}') \preceq (\bar{z}'', \bar{p}'')$ is now a simple computation. \square

We now state a lemma that will allow us to prove Proposition 47(i) in the remaining case. Recall that the notion of convergence to the orbit of a solution was defined in Definition 30.

Lemma 52 *Let (21) hold, and $(z(t), p(t))$ be a solution of (18). Let (\bar{z}, \bar{p}) be a maximally active saddle point. Then $(z(t), p(t))$ converges to the orbit of a solution $(z'(t), p'(t))$ of (18) that is a constant \mathcal{W} -distance to (\bar{z}, \bar{p}) and that satisfies, for each index l , $z'_l(t) = 0$ if $\bar{z}_l = 0$.*

With this lemma, the remaining case of Proposition 47 may be proved as follows. By construction, the set \mathcal{X}' consists of exactly those solutions of (18) that are both a constant \mathcal{W} -distance to the saddle point, and have z_l vanishing for l which have $\bar{z}_l = 0$. Thus, the above lemma gives convergence of an arbitrary solution to the orbit of a solution in \mathcal{X}' . Once we have this weaker form of convergence, the convergence can be upgraded as in the proof of Proposition 42.

The remainder of this section will be devoted to proving Lemma 52. To begin, we will extend Lemma 31 with an additional invariance property.

Lemma 53 *Let (21) hold and $(z(t), p(t))$ be a solution of (18). Then $(z(t), p(t))$ converges to the orbit of a solution $(z'(t), p'(t))$ that satisfies the following:*

(i) *For any $\tau \geq 0$, the distance*

$$\|(z'(t), p'(t)) - (z'(t+\tau), p'(t+\tau))\|_{\mathcal{W}} \quad (\text{B.1})$$

does not depend on t .

(ii) *$(z'(t), p'(t))$ is a constant \mathcal{W} -distance from any saddle point.*

PROOF. Lemma 31 implies both the convergence to the orbit of a solution $(z'(t), p'(t))$ and property (ii). (i) follows from the same arguments that are used to establish (ii) in the proof of Lemma 31. \square

The above properties in particular allows us to rule out convergence to solutions that have switching.

Corollary 54 *Let (21) hold and $(z(t), p(t))$ be a solution to (18) which satisfies Lemma 53 (i). Then for any index l , one of the following holds:*

- (a) z_l is identically zero.
- (b) z_l is never inactive.

and similarly for p_j .

PROOF. We will consider the links z_l , the prices p_j are similar. Suppose, for a contradiction, that there is an index l for which neither of (a) or (b) hold. Then there are times t_1, t_2 for which $z_l(t_1)$ is inactive and $z_l(t_2)$ is active. Suppose $t_2 > t_1$, the reverse case is similar. Let

$W(t) = \frac{1}{2} \|(z, p)(t) - (z, p)(t + (t_2 - t_1))\|_{\mathcal{W}}^2$, then, by Lemma 28, W is non-increasing and strictly decreasing at $t = t_1$, which contradicts Lemma 53(i). \square

The only way our constructed limiting solution $(z'(t), p'(t))$ could not satisfy the claim of Lemma 52 is if, for some index l , $\bar{z}_l = 0$ and case (b) of Corollary 54 holds, while case (a) does not. We shall show that this case would imply that (\bar{z}, \bar{p}) is not maximally active, which is a contradiction.

PROOF of Lemma 52. Lemma 53 gives convergence of $(z(t), p(t))$ to the orbit of a solution $(z'(t), p'(t))$ that satisfies the two conditions stated in the lemma. As we do not need to refer to the original solution $(z(t), p(t))$ again, we will drop the primes from $(z'(t), p'(t))$ to ease notation.

We will argue by contradiction. Assume that there is an index l for which $\bar{z}_l = 0$ but $z_l(t)$ is not identically zero. We will prove that this implies that (\bar{z}, \bar{p}) is not maximally active. To do this we will exhibit a saddle point that is more active than (\bar{z}, \bar{p}) .

Corollary 54 implies that for any index l' , $z_{l'}$ is either identically zero or never inactive. Let $\hat{\mathcal{G}}$ be the network obtained from \mathcal{G} by removing links $z_{l'}$ that have $z_{l'}(t)$ identically zero. Denote the version of \mathcal{X} obtained from $\hat{\mathcal{G}}$ as $\hat{\mathcal{X}}$, the vector $z(t)$ with these links removed as $\hat{z}(t)$ and similarly for other quantities.

$(\hat{z}(t), \hat{p}(t))$ is then a solution to the reduced system, and as, by Corollary 54, the remaining links $z_{l'}$ and prices p_j are never inactive, it is a proper solution. Because $(\hat{z}(t), \hat{p}(t))$ is proper, it is also a solution to (B.2), i.e. the system without the switching at the boundary.

$$\begin{aligned} \dot{\hat{z}} &= \hat{K}' \hat{\varphi}_{\hat{z}} \\ \dot{\hat{p}} &= -\hat{K} \hat{\varphi}_{\hat{p}}. \end{aligned} \quad (\text{B.2})$$

By Lemma 28(b)-(c) and Lemma 53(ii), (\hat{z}, \hat{p}) has no inactive coordinates and is an equilibrium point of (B.2). Lemma 53(ii) and the arguments⁹ from section 4 show that $(\hat{z}(t), \hat{p}(t)) \in \hat{\mathcal{X}}$, and is of the form,

$$(\hat{z}(t), \hat{p}(t)) = (\hat{z}, \hat{p}) + (\hat{z}^0, 0) + (\hat{z}^1(t), \hat{p}(t)) \quad (\text{B.3})$$

and $\hat{z}^0, \hat{z}^1(t), \hat{p}(t)$ obey the constraints given in the definition of \mathcal{X} . In particular, $(\hat{z} + \hat{z}^0, \hat{p})$ is a saddle point of

⁹ In section 4 assumption (22) was used to justify neglecting the switching at the boundary. In equation (B.2) this switching is neglected, so the analysis in section 4 can still be used to show that $(\hat{z}(t), \hat{p}(t))$ is of the form given in the definition of \mathcal{X} .

$\hat{\varphi}$. We wish to prove that $(\hat{\mathbf{z}} + \hat{z}^0, \hat{\mathbf{p}})$ is more active than $(\bar{\mathbf{z}}, \bar{\mathbf{p}})$, for which, as $\hat{\mathbf{z}}_l = 0$, it is sufficient to show that $\hat{z}_l^0 > 0$. This follows from (B.3), as $\hat{z}_l(t)$ is non-negative and not identically zero, $\hat{z}_l^1(t)$ oscillates around 0, and $\hat{\mathbf{z}}_l = 0$.

Finally, we must show that $(\hat{\mathbf{z}} + \hat{z}^0, \hat{\mathbf{p}})$ induces a saddle point $(\bar{\mathbf{z}} + \hat{z}^0, \bar{\mathbf{p}}) \geq 0$ of φ . This follows from $\hat{z}^0 \geq 0$, which holds because $(\hat{z}(t), \hat{p}(t)) \geq 0$, and the fact that extending a vector in $\ker(\hat{H})$ or $\ker(\hat{A})$ with zeros for the removed links yields a vector in $\ker(H)$ or $\ker(A)$ respectively. \square

This completes the proof of Proposition 47.

References

- [1] D. Angeli. A Lyapunov approach to incremental stability properties. *IEEE Transactions on Automatic Control*, 47(3):410–421, Mar 2002.
- [2] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in linear and non-linear programming*. Stanford University Press, 1958.
- [3] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.
- [4] 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.
- [5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
- [6] Ashish Cherukuri, Enrique Mallada, and Jorge Cortés. Asymptotic convergence of constrained primal–dual dynamics. *Systems & Control Letters*, 87:10–15, 2016.
- [7] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255–312, 2007.
- [8] J. Eckstein and D. P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(3):293–318, June 1992.
- [9] D. Feijer and F. Paganini. Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica*, 46(12):1974–1981, 2010.
- [10] 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.
- [11] B. Gharesifard and J. Cortés. Distributed convergence to Nash equilibria in two-network zero-sum games. *Automatica*, 49(6):1683–1692, 2013.
- [12] B. Gharesifard and J. Cortés. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3):781–786, 2014.
- [13] T. Holding and I. Lestas. On the emergence of oscillations in distributed resource allocation. In *52nd IEEE Conference on Decision and Control*, December 2013.
- [14] T. Holding and I. Lestas. On the convergence to saddle points of concave-convex functions, the gradient method and emergence of oscillations. In *53rd IEEE Conference on Decision and Control*, December 2014.
- [15] L. Hurwicz. The design of mechanisms for resource allocation. *The American Economic Review*, 63(2):1–30, May 1973.
- [16] A. N. Iusem. Augmented lagrangian methods and proximal point methods for convex optimization. *Investigaci'on Operativa*, pages 11–50, 1999.
- [17] 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.
- [18] H. K. Khalil. *Nonlinear Systems*. Prentice Hall, 2002.
- [19] Steven H Low and David E Lapsley. Optimization flow control: basic algorithm and convergence. *IEEE/ACM Transactions on Networking (TON)*, 7(6):861–874, 1999.
- [20] J. Lygeros, K. H. Johansson, S. N. Simić, J. Zhang, and S. S. Sastry. Dynamical properties of hybrid automata. *IEEE Transactions on Automatic Control*, 48(1):2–17, 2003.
- [21] E. Mallada, C. Zhao, and S. H. Low. Optimal load-side control for frequency regulation in smart grids. In *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 731–738, 2014.
- [22] A. Nedić and A. Ozdaglar. Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142(1):205–228, 2009.
- [23] D. Papadaskalopoulos and G. Strbac. Decentralized participation of flexible demand in electricity markets-part I: Market mechanism. *IEEE Transactions on Power Systems*, 28:3658–3666, 2013.
- [24] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, August 1976.
- [25] R. Srikant. *The mathematics of Internet congestion control*. Birkhauser, 2004.
- [26] H. R. Varian. *Microeconomic Analysis*. W.W.Norton & Company, 1996.
- [27] J. Wang and N. Elia. A control perspective for centralized and distributed convex optimization. In *50th IEEE Conference on Decision and Control and European Control Conference*, pages 3800–3805, 2011.
- [28] 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.