



Stability of primal–dual gradient dynamics and applications to network optimization[☆]

Diego Feijer, Fernando Paganini^{*}

Department of Electrical Engineering, Universidad ORT, Cuareim 1451, 11.100 Montevideo, Uruguay

ARTICLE INFO

Article history:

Received 18 August 2009

Received in revised form

3 May 2010

Accepted 22 July 2010

Available online 1 September 2010

Keywords:

Convex optimization

Dynamics

Lyapunov stability

Network resource allocation

ABSTRACT

This paper considers dynamic laws that seek a saddle point of a function of two vector variables, by moving each in the direction of the corresponding partial gradient. This method has old roots in the classical work of Arrow, Hurwicz and Uzawa on convex optimization, and has seen renewed interest with its recent application to resource allocation in communication networks. This paper brings other tools to bear on this problem, in particular Krasovskii's method to find Lyapunov functions, and recently obtained extensions of the LaSalle invariance principle for hybrid systems. These methods are used to obtain stability proofs of these primal–dual laws in different scenarios, and applications to cross-layer network optimization are exhibited.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Recent advances in the application of convex optimization to resource allocation in communication networks, originating in Kelly, Maulloo, and Tan (1998), have motivated a new look at classical optimization methods based on motion along the gradient. While gradient algorithms are not the most efficient for centralized computation, they are particularly well-suited for implementations across a distributed network.

We focus here on methods that seek a *saddle point* of a function $L(x, \lambda)$ of two vector variables (maximum in x , minimum in λ), each moving in the direction of the respective partial gradient, with appropriate signs:

$$\dot{x} = \frac{dx}{dt} = K \left[\frac{\partial L}{\partial x} \right]^T, \quad \dot{\lambda} = \frac{d\lambda}{dt} = \Gamma \left[-\frac{\partial L}{\partial \lambda} \right]^T. \quad (1)$$

In contrast to gradient maximization (or minimization) alone, where it is easy to show convergence since the objective function is monotonic along trajectories (see e.g. Khalil, 1996), the asymptotic behavior of the above dynamics is not immediate. Nevertheless, for $L(x, \lambda)$ concave in x and convex in λ , a quadratic form of the state around a saddle point is a Lyapunov function, as shown in the

classical study of Arrow, Hurwicz, and Uzawa (1958). Thus stability holds with great generality, and asymptotic stability under some restrictions (e.g. strict concavity in x). Other related work in the optimization literature can be found in Nedic and Ozdaglar (2009) and references therein.

Saddle points arise when $L(x, \lambda)$ is the Lagrangian of a convex optimization problem. Here relations (1) are modified to include projections that restrict the variable domains, and termed *primal–dual* (sub)gradient dynamics. These optimization problems have appeared recently in the area of Internet congestion control (see e.g. Srikant, 2004) and cross-layer network optimization: see Chiang, Low, Calderbank, and Doyle (2007) for a recent survey. It is noteworthy that, to some degree, this literature has rediscovered some of the classical results: e.g., the convergence proofs in Wen and Arcak (2003) for primal–dual laws are based on the same quadratic Lyapunov functions in Arrow et al. (1958).

In this paper we revisit these topics with a combination of old and new analytical tools, expanding on our conference paper (Feijer & Paganini, 2009). One tool is the use of Lyapunov functions of the Krasovskii type (Khalil, 1996; Krasovskii, 1959): a quadratic form of the *vector field*, which we show (1) always allows. It turns out that for unconstrained domains there are strong connections between the two Lyapunov functions, a fact we establish in Section 2. In optimization-based problems where there are switching projections in the dynamics, we show in Section 3 how to complete the stability analysis using the LaSalle invariance principle for hybrid systems (Lygeros, Johansson, Simic, Zhang, & Sastry, 2003). We also apply the theory to a cross-layer network resource allocation problem from Lee, Chiang, and Calderbank (2007, 2006) and Zhang, Zheng, and Chiang (2008).

[☆] Research partially supported by ANII-Uruguay, grants FCE 2007_265 and BE_INI_2008_70. The material in this paper was partially presented at the American Control Conference, June 10–12, 2009, St. Louis, MO, USA. This paper was recommended for publication in revised form by Associate Editor Masayuki Fujita under the direction of Editor Ian R. Petersen.

^{*} Corresponding author. Tel.: +598 2 9021505 1253; fax: +598 2 9081370.

E-mail addresses: feijer@ort.edu.uy (D. Feijer), paganini@ort.edu.uy (F. Paganini).

The work in Feijer and Paganini (2009) is also generalized in this paper to primal–dual gradient laws for non-strictly concave problems, a question that also goes back to Arrow et al. (1958). In these cases, the standard primal–dual dynamics can oscillate; we study two modifications that obtain stabilization, giving a stability proof through the Krasovskii method. These results are applied to the cross-layer network problem, and shown to expand the range of application of the method. Conclusions are given in Section 5.

2. Gradient laws for concave–convex saddle-point problems, and Krasovskii's method

Consider a function $L : X \times \Lambda \rightarrow \mathbb{R}$, where $X \subset \mathbb{R}^n$, $\Lambda \subset \mathbb{R}^m$. Throughout, $L(x, \lambda)$ is assumed concave in x and convex in λ . In particular, for L continuously differentiable ($L \in C^1$) we have the first order properties¹

$$L(\tilde{x}, \lambda) \leq L(x, \lambda) + \frac{\partial L}{\partial x}(x, \lambda) \cdot (\tilde{x} - x), \quad (2)$$

$$L(x, \tilde{\lambda}) \geq L(x, \lambda) + \frac{\partial L}{\partial \lambda}(x, \lambda) \cdot (\tilde{\lambda} - \lambda), \quad (3)$$

and for $L \in C^2$ the second order properties (see Boyd & Vandenberghe, 2004)

$$\frac{\partial^2 L}{\partial x^2} \leq 0, \quad \frac{\partial^2 L}{\partial \lambda^2} \geq 0. \quad (4)$$

We say $(\hat{x}, \hat{\lambda})$ is a saddle point of L if

$$L(x, \hat{\lambda}) \leq L(\hat{x}, \hat{\lambda}) \leq L(\hat{x}, \lambda) \quad \text{for all } x \in X, \lambda \in \Lambda.$$

For this initial section, suppose $X = \mathbb{R}^n$, $\Lambda = \mathbb{R}^m$, $L \in C^2$, and the gradient laws

$$\dot{x}_j = k_j \left[\frac{\partial L}{\partial x_j} \right], \quad \dot{\lambda}_i = \gamma_i \left[-\frac{\partial L}{\partial \lambda_i} \right]; \quad (5)$$

in vector notation, these have the form (1) except now $K = \text{diag}(k_j) > 0$ and $\Gamma = \text{diag}(\gamma_i) > 0$ are diagonal matrices. We will also refer to these as “gradient” laws, although \dot{x} and $[\frac{\partial L}{\partial x}]^T$ are only co-linear when $K = kI$.

Eqs. (5) define a dynamical system $\dot{z} = F(z)$ in the state vector $z = (x, \lambda)$, of which every saddle point \hat{z} is an equilibrium. In the work of Arrow et al. (1958), the dynamics are studied through a quadratic norm of $z - \hat{z}$, of the form

$$W(z) = (z - \hat{z})^T Q (z - \hat{z}), \quad (6)$$

$$\text{with } Q = \frac{1}{2} \begin{bmatrix} K^{-1} & 0 \\ 0 & \Gamma^{-1} \end{bmatrix}. \quad (7)$$

Following Arrow et al. (1958), we first show that W is non-increasing along trajectories of (5). For this purpose, write

$$\begin{aligned} \dot{W}(z) &= \frac{\partial L}{\partial x}(x, \lambda) \cdot (x - \hat{x}) - \frac{\partial L}{\partial \lambda}(x, \lambda) \cdot (\lambda - \hat{\lambda}) \\ &\leq L(x, \lambda) - L(\hat{x}, \lambda) + L(x, \hat{\lambda}) - L(x, \lambda) \\ &= [L(\hat{x}, \hat{\lambda}) - L(\hat{x}, \lambda)] + [L(x, \hat{\lambda}) - L(\hat{x}, \hat{\lambda})] \\ &\leq 0. \end{aligned} \quad (8)$$

Here the first step uses (2)–(3), the second the saddle point conditions. This has an important initial consequence: the solutions of (5) are *bounded*.

We introduce now an alternative Lyapunov function candidate from Krasovskii's method (see Khalil, 1996):

$$V(z) = \dot{z}^T Q \dot{z} = F(z)^T Q F(z), \quad (9)$$

which can be differentiated to give

$$\dot{V}(z) = \dot{z}^T \left\{ \left(\frac{\partial F}{\partial z} \right)^T Q + Q \left(\frac{\partial F}{\partial z} \right) \right\} \dot{z}.$$

If $Q > 0$ is found such that the term in braces is negative semidefinite at every z , then V is decreasing along trajectories. It turns out that Q from (7) satisfies this condition. Indeed, by differentiation we have

$$Q \left(\frac{\partial F}{\partial z} \right) = \frac{1}{2} \begin{bmatrix} \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 L}{\partial x \partial \lambda} \\ -\frac{\partial^2 L}{\partial \lambda \partial x} & -\frac{\partial^2 L}{\partial \lambda^2} \end{bmatrix}.$$

The equality of partial derivatives $\frac{\partial^2 L}{\partial x_j \partial \lambda_i}$ in both orders leads to

$$\left(\frac{\partial F}{\partial z} \right)^T Q + Q \left(\frac{\partial F}{\partial z} \right) = \begin{bmatrix} \frac{\partial^2 L}{\partial x^2} & 0 \\ 0 & -\frac{\partial^2 L}{\partial \lambda^2} \end{bmatrix} \leq 0,$$

invoking (4). Thus $\dot{V} \leq 0$.

So we have two functions which are non-increasing across trajectories. In particular, Lyapunov stability in the broad sense holds. To establish *asymptotic* stability requires further hypothesis and/or modifications of the dynamics, as will be discussed in future sections. In particular, we will use the LaSalle invariance principle (Khalil, 1996) and the function V .

$W(z)$ and $V(z)$ have similar structure: the same Q is used, the increment $(z - \hat{z})$ being replaced by $F(z)$; the fact that they both decrease is not coincidental. Indeed, we have the following relationship, that goes beyond the case of saddle point problems, and to our knowledge has not been made explicit before. The implication below also holds when inequalities are assumed strict for $z \neq \hat{z}$.

Proposition 1. Given a dynamical system $\dot{z} = F(z)$, $F \in C^1$, with equilibrium point \hat{z} . Suppose $Q = Q^T$ satisfies $(\frac{\partial F}{\partial z})^T Q + Q(\frac{\partial F}{\partial z}) \leq 0$ for every z . Then $W(z) = (z - \hat{z})^T Q (z - \hat{z})$ satisfies $\dot{W} \leq 0$ along trajectories.

Proof. Differentiation gives $\dot{W}(z) = 2(z - \hat{z})^T Q F(z)$; suppose we had $\dot{W}(z) > 0$ for some $z \neq \hat{z}$. Introduce

$$h(\theta) := 2(z - \hat{z})^T Q F(\hat{z} + \theta(z - \hat{z})), \quad \theta \in [0, 1].$$

Then, $h(0) = 0$ and $h(1) > 0$, so the mean value theorem gives $h'(\theta) > 0$ for some $\theta \in (0, 1)$. This yields

$$\begin{aligned} h'(\theta) &= 2(z - \hat{z})^T Q \frac{\partial F}{\partial z} \cdot (z - \hat{z}) \\ &= (z - \hat{z})^T \left\{ \left(\frac{\partial F}{\partial z} \right)^T Q + Q \left(\frac{\partial F}{\partial z} \right) \right\} (z - \hat{z}) > 0, \end{aligned}$$

with the Jacobian evaluated at $\hat{z} + \theta(z - \hat{z})$. This contradicts the hypothesis. \square

So we see that for dynamics characterized by a smooth vector field, whenever a Lyapunov function is found via Krasovskii's method, a quadratic Lyapunov function is also available. Still, Krasovskii's method may provide a useful way to search the space of Q 's, in particular it requires no explicit knowledge of the equilibrium \hat{z} . Both techniques carry through as well to differential equations with switching that arise in gradient methods for optimization under constraints, to be considered in the next section. In that case, however, the above proposition has no simple counterpart.

¹ Notational conventions: for a scalar L , $\frac{\partial L}{\partial x}$ is the gradient as a row vector. $\frac{\partial^2 L}{\partial x^2}$ is the Hessian. For a vector valued $F(z)$, $\frac{\partial F}{\partial z}$ is the Jacobian matrix.

3. Primal–dual gradient laws for Lagrangian optimization and application to networks

We study constrained optimization problems of the form

Problem 2.

maximize $U(x)$

subject to $g_i(x) \leq 0, \quad i = 1, \dots, m.$

We assume the functions $U(x)$ and $g_i(x)$ of $x \in X$ are in C^2 , concave and convex respectively. $g(x)$ is the column vector of components $g_i(x)$. To simplify our presentation we focus on the case $X = \mathbb{R}^n$; we remark below on the changes needed to address the other main case of interest, the non-negative orthant $X = \mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_j \geq 0, j = 1, \dots, n\}$.

Assumption 1. The problem has a finite optimum $\hat{U} = U(\hat{x})$, and Slater's condition holds: the constraints are strictly feasible.

Under this assumption, the problem can be studied through its Lagrangian dual, with no duality gap (see Boyd & Vandenberghe, 2004). The Lagrangian for Problem 2 is

$$L(x, \lambda) = U(x) - \sum_{i=1}^m \lambda_i g_i(x), \quad (10)$$

which is concave in x , and linear (thus convex) in $\lambda \in \mathbb{R}_+^m = \{\lambda \in \mathbb{R}^m, \lambda_i \geq 0, i = 1, \dots, m\}$. An optimum \hat{x} of the problem corresponds to a saddle point $(\hat{x}, \hat{\lambda})$ of the Lagrangian. The corresponding primal–dual gradient laws are given by

$$\dot{x} = K \left[\frac{\partial L}{\partial x} \right]^T = K \left[\frac{\partial U}{\partial x} - \sum_{i=1}^m \lambda_i \frac{\partial g_i}{\partial x} \right]^T, \quad (11)$$

$$\dot{\lambda} = \Gamma \left[- \left(\frac{\partial L}{\partial \lambda} \right)^T \right]_+^+ = \Gamma [g(x)]_+^+. \quad (12)$$

Here $K = \text{diag}(k_i) > 0$, $\Gamma = \text{diag}(\gamma_j) > 0$ as before, but we now include a *positive projection* to keep the multipliers non-negative, defined as follows. For scalars w, z , $[w]_z^+ := w$ if $z > 0$ or $w > 0$, and $[w]_z^+ := 0$ otherwise. The projection is said to be *active* in the second case. For vectors w, z of the same dimension, the projection is defined componentwise. We denote the set of active projection indices in (12) by $\sigma = \{i : \lambda_i = 0, g_i(x) < 0\}$. Thus (12) is equivalent to $\dot{\lambda}_i = \begin{cases} \gamma_i g_i(x) & \text{if } i \notin \sigma \\ 0 & \text{if } i \in \sigma \end{cases}$. The dynamics (11)–(12) are then the system $\dot{z} = F(z, \sigma)$, where $z = (x, \lambda)$ and σ is a discrete “state”. In this *switching* system σ has no memory, it merely partitions the state-space in z among different vector fields.²

We first show the dynamics (11)–(12) still present bounded trajectories.

Proposition 3. Under Assumption 1, the trajectories of (11)–(12) are bounded.

Proof. With W from (6) we have

$$\begin{aligned} \dot{W}(z) &= \dot{x}^T K^{-1} (x - \hat{x}) + \dot{\lambda}^T \Gamma^{-1} (\lambda - \hat{\lambda}) \\ &= \frac{\partial L}{\partial x} \cdot (x - \hat{x}) + \sum_i [g_i(x)]_{\lambda_i}^+ \cdot (\lambda_i - \hat{\lambda}_i) \\ &\leq \frac{\partial L}{\partial x} \cdot (x - \hat{x}) + g(x)^T \cdot (\lambda - \hat{\lambda}). \end{aligned}$$

For the last step, notice that $g_i(x)(\lambda_i - \hat{\lambda}_i) \geq 0$ for $i \in \sigma$, since $g_i(x) < 0$ and $\lambda_i = 0$. The last expression is now equal to (8), so $W \leq 0$ follows as before. Hence, balls of the quadratic norm W are invariant under the flow. \square

We now adapt the Krasovskii method of the previous section to the current dynamics with switching. Consider the Lyapunov function

$$V(z, \sigma) = \frac{1}{2} \dot{x}^T K^{-1} \dot{x} + \frac{1}{2} \sum_{i \notin \sigma} \frac{|\dot{\lambda}_i|^2}{\gamma_i}. \quad (13)$$

This has the same form as (9), only now the vector field is not smooth, it depends on the switching state σ . For a time interval with fixed σ , we can differentiate the above to obtain

$$\begin{aligned} \dot{V}(z, \sigma) &= \dot{x}^T K^{-1} \ddot{x} + \sum_{i \notin \sigma} \frac{\dot{\lambda}_i \ddot{\lambda}_i}{\gamma_i} \\ &= \dot{x}^T \left(\frac{\partial^2 U}{\partial x^2} - \sum_{i=1}^m \lambda_i \frac{\partial^2 g_i}{\partial x^2} \right) \dot{x} \\ &\quad - \dot{x}^T \sum_{i=1}^m \dot{\lambda}_i \left(\frac{\partial g_i}{\partial x} \right)^T + \sum_{i \notin \sigma} \dot{\lambda}_i \frac{\partial g_i}{\partial x} \dot{x} \\ &= \dot{x}^T \left(\frac{\partial^2 U}{\partial x^2} - \sum_{i=1}^m \lambda_i \frac{\partial^2 g_i}{\partial x^2} \right) \dot{x} \leq 0, \end{aligned} \quad (14)$$

using the concavity of U and the convexity of g_i . We must also study the behavior of $V(z, \sigma)$ when the set σ of active projections changes. Here, V may be discontinuous. Notice, however:

- The set σ will be *enlarged* if a constraint i with $g_i(t^-) \leq 0$ reaches $\lambda_i(t) = 0$. In that case, the sum in (13) loses a term, so $V(t^-) \geq V(t^+)$. V is discontinuous but in the decreasing direction.
- σ is *reduced* when a constraint i with active projection at t^- no longer has it at t^+ . This happens only if $g_i(x)$ went through zero, from negative to positive, at t . An extra term is added to (13), but this term is initially at zero. Hence there is no discontinuity of V .

The following statement summarizes our analysis of V .

Proposition 4. The Lyapunov function V defined in (13) is non-increasing along trajectories of (11)–(12).

To establish *asymptotic* stability of the saddle point of L as desired, some additional restrictions are needed, as illustrated by the following example.

Example. Consider the linear program

maximize $c^T x$

subject to $Gx \leq h$.

The corresponding primal–dual dynamics are

$$\begin{aligned} \dot{x} &= K[c - G^T \lambda], \\ \dot{\lambda} &= \Gamma[Gx - h]_+^+. \end{aligned}$$

These are linear except for the projection constraints. Take $K = I$, $\Gamma = I$, and a case where $\hat{\lambda} > 0$. In a neighborhood of the saddle point we have $\dot{z} = A(z - \hat{z})$, where $A = \begin{bmatrix} 0 & -G^T \\ G & 0 \end{bmatrix}$ is skew-symmetric, with purely imaginary eigenvalues. This gives rise to harmonic oscillations which do not converge to a saddle point.

The simplest way to obtain stronger conclusions is to require *strict* concavity of the function $U(x)$, as was done in Arrow et al. (1958). We now establish this through the Lyapunov function (13).

² To address the case $X = \mathbb{R}_+^n$, a positive projection is required on (11) as well. This leads to a discrete state $\sigma = (\sigma_x, \sigma_\lambda)$. We refer to Feijer and Paganini (2009) for details on this alternative.

Theorem 5. Under [Assumption 1](#), if $U(x)$ is strictly concave, all trajectories of (11)–(12) converge to a saddle point of L .

Proof. We apply a LaSalle invariance principle for hybrid systems, developed by [Lygeros et al. \(2003\)](#); we summarize the essence as follows: assume the dynamics have a compact, positively invariant set Ω (i.e., trajectories starting in Ω stay in Ω) and a function $V(z, \sigma)$ that decreases along trajectories in Ω . Then every trajectory in Ω converges to \mathcal{L} , the maximal positively invariant set within Ω with trajectories satisfying

- (i) $\frac{d}{dt}V(z(t), \sigma) \equiv 0$ in intervals of fixed σ , and
- (ii) $V(z(t), \sigma^-) = V(z(t), \sigma^+)$ if σ switches at time t between σ^- and σ^+ .

In our case, Ω is a ball as considered in [Proposition 3](#). Using V in (13), we show that any trajectory satisfying (i)–(ii) must be an equilibrium.

Imposing condition (i) we see from (14) and the strict concavity of U that $\dot{x} \equiv 0$ at every interval of fixed σ ; since x is continuous, it is then in equilibrium, say $x \equiv \hat{x}$. If $g_i(\hat{x}) > 0$ for some i , the corresponding multiplier λ_i would grow linearly, contradicting boundedness. If $g_i(\hat{x}) < 0$ for some i , the corresponding projection must be active ($\lambda_i = 0$); otherwise λ_i would converge linearly to zero, producing a discontinuity of V as the projection becomes active, violating (ii).

Therefore $\dot{\lambda}_i \equiv 0$ for all i , and $\lambda \equiv \hat{\lambda}$, constant. The invariant trajectory must be an equilibrium point of the dynamics, a saddle point of the Lagrangian. \square

3.1. Application: primal–dual congestion control

As a first application of the preceding result, consider the optimization-based approach to Internet congestion control introduced by [Kelly et al. \(1998\)](#). The network is modeled as a set of sources \mathcal{S} sending traffic through a set \mathcal{L} of links. Each source is assigned a concave, increasing utility function of its rate, $U_j(x_j)$, and the resource allocation must

$$\text{maximize } U(x) = \sum_{j \in \mathcal{S}} U_j(x_j),$$

subject to link capacity constraints

$$y_i := \sum_{j \in \mathcal{S}(i)} x_j \leq c_i, \quad i \in \mathcal{L}.$$

Here $\mathcal{S}(i) \subset \mathcal{S}$ is the set of sources that contribute traffic to link i , its total rate being y_i . This problem falls in the class of [Problem 2](#), with two special features: the $g_i(x)$ are linear, and $U(x)$ is a separable function of its arguments. This implies the primal–dual laws (11)–(12) take the decentralized form

$$\dot{x}_j = k_j \left[U'_j(x_j) - \sum_{i \in \mathcal{L}(j)} \lambda_i \right],$$

$$\dot{\lambda}_i = \gamma_i [y_i - c_i]_{\lambda_i}^+.$$

Here the multiplier (price) of link i is adapted to reflect local congestion, and each source j regulates its rate based on network feedback, the total price across the subset $\mathcal{L}(j) \subset \mathcal{L}$ of links in its route. These kinds of laws appear in [Srikant \(2004\)](#) and [Wen and Arcak \(2003\)](#), and they are proved to be stabilizing in the case of strictly concave utilities, using a quadratic Lyapunov function of the kind mentioned before. Here we have an alternate proof based on [Theorem 5](#).

3.2. Generalized network utility maximization

As a natural continuation of the congestion control studies, recent research has incorporated into the optimization framework other resources to be allocated in networks, involving other layers

of protocols. In particular, many researchers have tackled the control of routing, medium access control (MAC), or physical layer control, most prominently for the case of wireless networks. Recent surveys which cover part of the substantial literature are [Chiang et al. \(2007\)](#) and [Lin, Shroff, and Srikant \(2006\)](#). In particular, [Chiang et al. \(2007\)](#) advocates the network utility maximization (NUM) paradigm, and its decompositions, as the unifying paradigm over which network architectures can be designed and controlled.

We now formulate one such NUM problem, which generalizes the congestion control case considered above.

Problem 6.

$$\text{Maximize } \sum_j U_j(x_j)$$

$$\text{subject to } g_i(x) \leq c_i(p), \quad i = 1, \dots, m.$$

Here $x \in X = \mathbb{R}^n$, $p \in \mathcal{P}$, convex and compact subset of \mathbb{R}^q , and the functions $g(x)$ and $c(p)$ are in C^2 , convex and strictly concave respectively.

Through the variable $p \in \mathcal{P}$ above, we can represent the dependence of link capacity on parameters of the lower layers, e.g. the multiple access protocols or physical layer parameters. Letting x enter the constraint in a possibly nonlinear (convex) way, provides more generality for the choice of primal variables, other than link rates themselves. An example is described in [Section 3.3](#).

[Problem 6](#) would be a special case of [Problem 2](#), if we consider (x, p) as primal variables of optimization. In that case, however, we cannot have strict concavity of the objective since p does not appear in it. For this reason we find it more useful to consider a variant of the primal–dual dynamics, where p is updated instantaneously as a function of the current multipliers, and x controlled dynamically. To define this, introduce first the Lagrangian for this problem,

$$L(x, p, \lambda) = \sum_{j=1}^n U_j(x_j) - \sum_{i=1}^m \lambda_i (g_i(x) - c_i(p)).$$

The dual problem will be $\min_{\lambda \geq 0} D(\lambda)$, where the Lagrangian dual function can be written as

$$D(\lambda) = \max_{x \in X} \left(\sum_{j=1}^n U_j(x_j) - \lambda^T g(x) \right) + \max_{p \in \mathcal{P}} \lambda^T c(p). \quad (15)$$

Note the “dual decomposition” (between variables x and p) in the solution of (15). Define

$$\varphi(\lambda) := \max_{p \in \mathcal{P}} \lambda^T c(p). \quad (16)$$

The function $\varphi(\lambda)$ is convex in λ (maximum of linear functions). If the $c_i(p)$ are strictly concave and \mathcal{P} convex, there is a unique maximizing p for every $\lambda \neq 0$, which we denote by $\bar{p}(\lambda)$. For the following development we require an additional assumption.

Assumption 2. For every $\lambda \in \mathbb{R}_+^m$, $\lambda \neq 0$, the optimum in (16) is achieved at an interior point in \mathcal{P} .

Proposition 7. Under [Assumption 2](#), $\bar{p}(\lambda)$ is differentiable and $\varphi(\lambda)$ is twice differentiable in $\lambda \in \mathbb{R}_+^m$, $\lambda \neq 0$, and

$$\left[\frac{\partial \varphi}{\partial \lambda} \right]^T = c(\bar{p}(\lambda)). \quad (17)$$

Proof. Since the optimum \bar{p} of (16) is interior, the following first order condition must be satisfied at \bar{p} :

$$h(\lambda, p) := \sum_i \lambda_i \frac{\partial c_i}{\partial p} = 0. \quad (18)$$

This equation implicitly defines $\bar{p}(\lambda)$, with a Jacobian

$$\frac{\partial h}{\partial p} = \sum_i \lambda_i \frac{\partial^2 c_i}{\partial p^2}$$

which is negative definite for $\lambda \neq 0$, due to strict concavity of $c_i(p)$. It is thus nonsingular, and the implicit function theorem implies $\bar{p}(\lambda)$ is differentiable. Now, differentiating $\varphi(\lambda) = \lambda^T c(\bar{p}(\lambda))$ we obtain

$$\frac{\partial \varphi}{\partial \lambda} = [c(\bar{p}(\lambda))]^T + \left[\sum_i \lambda_i \frac{\partial c_i}{\partial p} \right]_{p=\bar{p}} \frac{\partial \bar{p}}{\partial \lambda};$$

the second term vanishes due to (18). \square

Returning now to (15), define

$$\begin{aligned} \bar{L}(x, \lambda) &= \max_{p \in \mathcal{P}} L(x, p, \lambda) \\ &= \sum_{j \in \mathcal{S}} U_j(x_j) - \sum_{i \in \mathcal{L}} \lambda_i g_i(x) + \varphi(\lambda). \end{aligned}$$

Proposition 8. Let $(\hat{x}, \hat{\lambda}, \hat{p})$ be a saddle point of $L(x, p, \lambda)$ (maximum in (x, p) , minimum in λ). Then $(\hat{x}, \hat{\lambda})$ is a saddle point of $\bar{L}(x, \lambda)$.

Proof. By hypothesis we have

$$L(x, p, \hat{\lambda}) \leq L(\hat{x}, \hat{p}, \hat{\lambda}) \quad \forall x \in X, p \in \mathcal{P}, \quad (19)$$

$$L(\hat{x}, \hat{p}, \hat{\lambda}) \leq L(\hat{x}, \hat{p}, \lambda) \quad \forall \lambda \in \mathbb{R}_+^m. \quad (20)$$

First note that $\bar{L}(\hat{x}, \hat{\lambda}) = L(\hat{x}, \hat{p}, \hat{\lambda})$, which follows from restricting (19) to $x = \hat{x}$. In particular $\hat{p} = \bar{p}(\hat{\lambda})$. Also from (19) we have

$$\bar{L}(x, \hat{\lambda}) \leq L(\hat{x}, \hat{p}, \hat{\lambda}) = \bar{L}(\hat{x}, \hat{\lambda}). \quad (21)$$

Now consider the inequality

$$\begin{aligned} L(\hat{x}, p, \lambda) - L(\hat{x}, \hat{p}, \hat{\lambda}) &= L(\hat{x}, p, \lambda) - L(\hat{x}, \hat{p}, \lambda) \\ &\quad + L(\hat{x}, \hat{p}, \lambda) - L(\hat{x}, \hat{p}, \hat{\lambda}) \\ &\geq \sum_i \lambda_i (c_i(p) - c_i(\hat{p})), \end{aligned}$$

where we have used (20). Since the right-hand side is zero for $p = \hat{p}$, we conclude the maximum of the right-hand side over $p \in \mathcal{P}$ is non-negative. Therefore

$$\bar{L}(\hat{x}, \lambda) = \max_{p \in \mathcal{P}} L(\hat{x}, p, \lambda) \geq L(\hat{x}, \hat{p}, \hat{\lambda}) = \bar{L}(\hat{x}, \hat{\lambda}). \quad (22)$$

Combining (21) and (22) concludes the proof. \square

We are thus led to consider a primal–dual gradient law for the function \bar{L} , which has the form

$$\dot{x}_j = k_j \left[\frac{\partial \bar{L}}{\partial x_j} \right] = k_j \left[U'_j(x_j) - \sum_i \lambda_i \frac{\partial g_i}{\partial x_j} \right], \quad (23)$$

$$\dot{\lambda}_i = \gamma_i \left[-\frac{\partial \bar{L}}{\partial \lambda_i} \right]_{\lambda_i}^+ = \gamma [g_i(x) - c_i(\bar{p}(\lambda))]_{\lambda_i}^+, \quad (24)$$

where we have used (17). Note that the only discontinuous switches in these dynamics are due to the projections in λ . We state the following result.

Theorem 9. Consider Problem 6 with strictly concave $U_j(x_j)$, and satisfying Assumptions 1 and 2. Then the dynamics (23)–(24) converge to $(\hat{x}, \hat{\lambda})$, such that $(\hat{x}, \bar{p}(\hat{\lambda}))$ are the optimum of Problem 6.

Proof. We extend the argument in Theorem 5. A first observation is that since the dynamics is the primal–dual gradient law of \bar{L} , the same argument in Proposition 3 implies that trajectories are bounded.

Considering now the Lyapunov function $V(z, \sigma)$ in (13), it is still true as before that it can only have discontinuities in the decreasing direction. Differentiating in an interval of constant σ , the only difference is that \bar{L} is no longer linear in λ , so we have an additional term in \dot{V} :

$$\dot{V}(z, \sigma) = \dot{x}^T \left(\frac{\partial^2 U}{\partial x^2} - \sum_{i=1}^m \lambda_i \frac{\partial^2 g_i}{\partial x^2} \right) \dot{x} - \dot{\lambda}^T \frac{\partial^2 \varphi}{\partial \lambda^2} \dot{\lambda}.$$

Now $\varphi(\lambda)$ is convex, so the Hessian $\frac{\partial^2 \varphi}{\partial \lambda^2}$ is positive semidefinite and we still have $\dot{V} \leq 0$.

Invoking again the LaSalle principle in Lygeros et al. (2003), we characterize the invariant set under the hypothesis $\dot{V} \equiv 0$. In particular, we still have $\dot{x} \equiv 0$ from the strict concavity of U . Also, from the last term above we conclude that $\frac{\partial \varphi}{\partial \lambda}$ is constant in time, so $c(\bar{p}(\lambda))$ is constant, as well as $g(\hat{x})$. So the right-hand side of (24) is constant in time; from here on, the argument in Theorem 5 can be reproduced to conclude the $\lambda \equiv \hat{\lambda}$. \square

3.3. Cross-layer congestion and contention control

We briefly describe here an application of the formulation in Problem 6 to a problem originated in Lee et al. (2006). Consider a wireless network made up of a set of nodes, which use a random multiple access control. Each node k accesses the medium with probability P^k , and when transmitting it chooses one of its outgoing links in $\mathcal{L}_{out}(n)$ with probability p_i . Thus, we have the following convex constraints (which define \mathcal{P}) for every node:

$$p_i \geq 0, \quad \sum_{i \in \mathcal{L}_{out}(k)} p_i = P^k, \quad P^k \leq 1. \quad (25)$$

The transmission of a link is interfered with by another link, if the receiver of the former is within range of the transmitter of the latter; when such a collision occurs, we assume no useful information is transmitted. So the capacity of a link i depends on the probability of accessing the channel without presence of the interfering nodes; we write

$$C_i(p) := c_i p_i \prod_{k \in \mathcal{N}_I(i)} (1 - P^k),$$

where $\mathcal{N}_I(i)$ is the set of nodes that interfere with link i .

We wish to allocate the network degrees of freedom (rates and transmission probabilities) to optimize an overall network utility. The resulting optimal congestion and contention control problem is

$$\begin{aligned} &\text{maximize} \quad \sum_{j \in \mathcal{S}} U_j(x_j) \\ &\text{subject to} \quad (25) \text{ and } \sum_{j \in \mathcal{S}(i)} x_j \leq C_i(p), \quad \forall i. \end{aligned} \quad (26)$$

The difficulty with this problem is that the last constraint is non-convex; to overcome this, a change of variables was proposed in Lee et al. (2007). Take the logarithm on both sides of inequality (26), and define $\tilde{x}_j := \log x_j$, $\tilde{U}_j(\tilde{x}_j) := U_j(e^{\tilde{x}_j})$. The problem is then equivalent to

$$\begin{aligned} &\text{maximize} \quad \sum_{j \in \mathcal{S}} \tilde{U}_j(\tilde{x}_j) \\ &\text{subject to} \quad (25) \text{ and } \log \left(\sum_{j \in \mathcal{S}(i)} e^{\tilde{x}_j} \right) \leq \tilde{C}_i(p), \quad \forall i \end{aligned} \quad (27)$$

where $\tilde{C}_i(p) := \log c_i + \log p_i + \sum_{k \in \mathcal{N}_I(i)} \log(1 - P^k)$. Note that \tilde{C}_i is concave in the variables p_i, P^k . Also, the left-hand side of (27) is convex \tilde{x}_j (log-sum-exp function, see Boyd & Vandenberghe, 2004). Therefore, the new formulation falls in the class of Problem 6, provided that the new objective function \tilde{U}_j is concave

in \tilde{x}_j . This happens (see Lee et al., 2007) when the original utility functions $U_j(\cdot)$ satisfy

$$x_j U_j''(x_j) + U_j'(x_j) \leq 0. \quad (28)$$

In particular, for the class of so-called “ α -fair” utility functions from Mo and Walrand (2000), with

$$U_j'(x_j) = x_j^{-\alpha}, \quad (29)$$

the above condition holds if $\alpha \geq 1$.

For strict concavity of \tilde{U}_j , the inequality (28) must be strict (in the special case (29), $\alpha > 1$). If this happens, the problem falls in the class of Theorem 9 and thus the primal–dual dynamics (23)–(24) are globally stabilizing.

Remark.

- The primal–dual gradient dynamics for this problem were studied in Zhang et al. (2008) in discrete-time, together with stochastic issues associated with estimation of gradients. A stability result is given for the strictly concave case, using quadratic Lyapunov tools, and claimed to hold for $\alpha \geq 1$ in the family (29). However the $\alpha = 1$ case does not fall in this class, and indeed convergence can fail, as shown in the following section.
- The main advantage of the primal–dual formulation is that it allows for decentralized implementations. While the left-hand side of (27) is not separable between primal variables, computing the gradient can be distributed across a network. In contrast, a purely “dual” solution that instantaneously maximizes over x is not easy to decentralize, see Yu and Giannakis (2008) for related work. Note also that as shown in Lee et al. (2006), the maximization (16) can be computed with local information: the optimal p for a node depends on prices of interfering links.

4. Modified primal–dual gradient methods for non-strictly concave problems

As we saw in the previous sections, primal–dual gradient laws yield convergent trajectories when the objective is strictly concave, but can oscillate if concavity is not strict. In this section we study variations of the basic primal–dual law that can yield convergence in the latter case. For other recent work on this topic, see Nedic and Ozdaglar (2009).

4.1. Method based on modified constraints

We begin with a method first studied in Arrow et al. (1958). Namely, consider the following optimization problem

Problem 10 (Modified Constraints).

Maximize $U(x)$

subject to $\phi_i(g_i(x)) \leq 0, \quad i = 1, \dots, m,$

where each $\phi_i : \mathbb{R} \rightarrow \mathbb{R}$ is a twice continuously differentiable, strictly convex and strictly increasing function, such that $\phi_i(0) = 0$ and $\phi_i''(u) > 0$ for all $u \in \mathbb{R}$. Note that we could use the same function ϕ_i for all the constraints, for example, $\phi_i(u) = e^u - 1$. Now, since $\phi_i(u) \leq 0$ if and only if $u \leq 0$, the feasible sets of both Problems 2 and 10 are the same, thus the optimal points are identical. The Lagrangian for Problem 10 is given by

$$\tilde{L}(x, \lambda) = U(x) - \sum_{i=1}^m \lambda_i \phi_i(g_i(x)). \quad (30)$$

The idea is to use gradient laws for this modified Lagrangian to seek convergence to a saddle point. The corresponding primal–dual

dynamics are

$$\dot{x} = K \left[\frac{\partial \tilde{L}}{\partial x} \right] = K \left[\frac{\partial U}{\partial x} - \sum_{i=1}^m \lambda_i \phi_i'(g_i(x)) \frac{\partial g_i}{\partial x} \right]^T, \quad (31)$$

$$\dot{\lambda} = \Gamma \left[- \left(\frac{\partial \tilde{L}}{\partial \lambda} \right)^T \right]_{\lambda}^+ = \Gamma [\phi(g(x))]_{\lambda}^+, \quad (32)$$

where $\phi(g(x))$ is the column vector of $\phi_i(g_i(x))$. As the following result shows this algorithm can be used to solve Problem 2.

Theorem 11. Under Assumption 1, the dynamics (31)–(32) converge to a saddle point $(\hat{x}, \hat{\lambda})$ of (30), where \hat{x} is optimal for Problem 2.

Proof. Note first that Proposition 3 still applies here, so we know the trajectories are bounded. Consider now the Lyapunov function of the form (13), reproduced here for convenience:

$$V(z, \sigma) = \frac{1}{2} \dot{x} K^{-1} \dot{x} + \frac{1}{2} \sum_{i \notin \sigma} \frac{|\dot{\lambda}_i|^2}{\gamma_i}. \quad (33)$$

By an analogous argument to the one in Section 3, V will either be continuous or jump in the decreasing direction when the set of active projections σ changes. For an interval with fixed σ , we have

$$\begin{aligned} \dot{V}(z, \sigma) &= \dot{x}^T \frac{\partial^2 U}{\partial x^2} \dot{x} - \dot{x}^T \sum_{i=1}^m \lambda_i \phi_i''(g_i(x)) \frac{\partial g_i}{\partial x} \frac{\partial g_i}{\partial x} \dot{x} \\ &\quad - \dot{x}^T \sum_{i=1}^m \lambda_i \phi_i'(g_i(x)) \frac{\partial^2 g_i}{\partial x^2} \dot{x} \\ &\quad - \dot{x}^T \sum_{i=1}^m \dot{\lambda}_i \phi_i'(g_i(x)) \frac{\partial g_i}{\partial x} + \sum_{i \notin \sigma} \dot{\lambda}_i \overline{\phi_i(g_i(x))} \\ &= \dot{x}^T \frac{\partial^2 U}{\partial x^2} \dot{x} - \sum_{i=1}^m \lambda_i \phi_i''(g_i(x)) \left(\frac{\partial g_i}{\partial x} \dot{x} \right)^2 - \sum_{i=1}^m \lambda_i \phi_i'(g_i(x)) \dot{x}^T \frac{\partial^2 g_i}{\partial x^2} \dot{x}. \end{aligned} \quad (34)$$

Since ϕ_i is strictly increasing and convex, $\phi_i'(g_i(x)) > 0$ and $\phi_i''(g_i(x)) > 0$. Hence, the concavity of U and the convexity of g_i imply that $\dot{V} \leq 0$.

Therefore V decreases along trajectories of the switched system, and we are in a position to apply the invariance principle of Lygeros et al. (2003). It thus suffices to show that if a trajectory satisfies $\dot{V} \equiv 0$ (and, in particular, V has no discontinuous switches), then x and λ must be at an equilibrium of the dynamics, and thus a saddle point of the modified Lagrangian.

Imposing equality to zero in (34) we conclude that $\frac{\partial^2 U}{\partial x^2} \dot{x} \equiv 0$ and thus $\frac{\partial U}{\partial x}$ is constant, and for every constraint with $\lambda_i > 0$,

$$\begin{aligned} \frac{\partial^2 g_i}{\partial x^2} \dot{x} \equiv 0 &\implies \frac{\partial g_i}{\partial x} \text{ constant,} \\ \frac{\partial g_i}{\partial x} \dot{x} \equiv 0 &\implies g_i(x) \text{ constant.} \end{aligned}$$

In particular, while $\lambda_i > 0$ we have $\dot{\lambda}_i = \gamma_i \phi_i(g_i(x))$ constant; if this constant were strictly negative, λ_i would reach zero in finite time, and activate a projection that produces a discontinuity in $\dot{\lambda}_i$, hence in V ; this cannot occur in the invariant set under study. We also rule out $\phi_i(g_i(x)) > 0$, otherwise λ_i would grow without bound, violating boundedness. Therefore $\dot{\lambda}_i \equiv 0$, hence λ is at equilibrium. Returning now to (31) we see that \dot{x} is constant in the trajectory; due to boundedness of $x(t)$ we must have $\dot{x} \equiv 0$, so x must also be in equilibrium. \square

Remarks.

- The fact that the objective function is concave but not in the strict sense makes it possible to have more than one optimum for Problem 2. Theorem 11 states that the dynamics (31)–(32) will converge to one of these points.
- As mentioned, this approach was proposed by Arrow et al. (1958) who proved stability under the assumption that $U(x)$ and $g(x)$ are analytic. The preceding proof applies to functions in C^2 .

4.2. Method based on penalty function

Now we contemplate the possibility of modifying the objective of the problem while leaving the constraints unchanged. To this end, consider the dynamics given by

$$\dot{x} = K \left[\frac{\partial U}{\partial x} - \sum_{i=1}^m (\lambda_i + \psi'_i(g_i(x))) \frac{\partial g_i}{\partial x} \right]^T, \quad (35)$$

$$\dot{\lambda} = \Gamma [g(x)]_{\lambda}^+, \quad (36)$$

where $\psi_i : \mathbb{R} \rightarrow \mathbb{R}$ are C^2 functions such that:

$$u \leq 0 \implies \psi_i(u) = 0, \quad (37)$$

$$u \geq 0 \implies \psi_i \text{ is strictly convex and increasing.} \quad (38)$$

These are called *penalty functions* (see e.g. Freund, 2004) because they assign a penalty to the violated constraints. For example, functions of the form

$$\psi_i(u) = \begin{cases} v_i u^n, & u \geq 0, \\ 0, & u \leq 0, \end{cases}$$

verify the requirements for any integer $n > 1$.

In particular in the case $n = 2$, note that the additional term $\psi'_i(g_i(x))$ is proportional to λ_i , when it is positive. This term can be interpreted as adding “derivative action” to the control of x , a standard method of damping oscillatory systems.

The algorithm (35)–(36) is a primal–dual law that seeks a saddle point of

$$L(x, \lambda) = U(x) - \sum_{i=1}^m \psi_i(g_i(x)) - \sum_{i=1}^m \lambda_i g_i(x),$$

concave in x and convex in λ , Lagrangian of the following optimization problem.

Problem 12 (Modified Objective).

$$\text{Maximize } U(x) - \sum_{i=1}^m \psi_i(g_i(x))$$

$$\text{subject to } g_i(x) \leq 0, \quad i = 1, \dots, m.$$

From (37) and (38) we see that $g_i(x) \leq 0$ imply that $\psi_i(g_i(x)) = 0$, thus Problem 12 is equivalent to Problem 2. Hence, their optimal points (equilibria of (35)–(36)), must be identical. Note that the modified objective is still not strictly concave inside the constraint set; nevertheless, stabilization is obtained.

Theorem 13. Under Assumption 1, assuming the functions ψ_i satisfy (37)–(38), the dynamics (35)–(36) converge to a point $(\hat{x}, \hat{\lambda})$, where \hat{x} is an optimal of Problem 2.



Fig. 1. Simple wireless network of three nodes.

Proof. The proof has similarities with the previous cases. Again we have boundedness of trajectories, and we use the Lyapunov function $V(z, \sigma)$ in (33) to apply the LaSalle principle. For an interval with fixed σ , we obtain here

$$\begin{aligned} \dot{V}(z, \sigma) &= \dot{x}^T \frac{\partial^2 U}{\partial x^2} \dot{x} - \sum_{i=1}^m [\lambda_i + \psi'_i(g_i(x))] \dot{x}^T \frac{\partial^2 g_i}{\partial x^2} \dot{x} \\ &\quad - \sum_{i=1}^m \psi''_i(g_i(x)) \left(\frac{\partial g_i}{\partial x} \dot{x} \right)^2. \end{aligned} \quad (39)$$

Again we conclude that $\dot{V} \leq 0$, and invoking Lygeros et al. (2003) we reduce the problem to showing that the only trajectories that satisfy $\dot{V} \equiv 0$ are equilibrium points.

Imposing $\dot{V} \equiv 0$ in (39), we conclude the following:

- (i) $\frac{\partial^2 U}{\partial x^2} \dot{x} \equiv 0$, then $\frac{\partial U}{\partial x}$ remains constant;
- (ii) if $\lambda_i > 0$ or $g_i(x) > 0$, then $\frac{\partial^2 g_i}{\partial x^2} \dot{x} \equiv 0$, hence $\frac{\partial g_i}{\partial x}$ remains constant;
- (iii) if $g_i(x) > 0$, then $\frac{\partial g_i}{\partial x} \dot{x} \equiv 0$, thus $g_i(x)$ is constant.

Note that for the second cases in both (ii) and (iii) we have invoked (38). From (iii), $g_i(x)$ is constant whenever it is positive; if this could occur it would give a linear growth of the corresponding λ_i , contradicting boundedness. So $g_i(x(t)) \leq 0$ for all i ; in particular, $\lambda_i(t)$ is monotonically decreasing in the invariant trajectory, and non-negative, so it must converge to $\hat{\lambda}_i$.

Note also from (37) that $g_i(x(t)) \leq 0$ implies the penalty term disappears from the dynamics of x , it becomes

$$\dot{x} = K \left[\frac{\partial U}{\partial x} - \sum_{i: \lambda_i > 0} \lambda_i \frac{\partial g_i}{\partial x} \right]^T. \quad (40)$$

The right-hand side converges to a finite limit as $t \rightarrow \infty$, since the multipliers converge and gradients are constant, using (i) and (ii) above. Due to boundedness of $x(t)$, the limit must be zero.

Now observe that for $i \notin \sigma$, $\lambda_i = \frac{\partial g_i}{\partial x} \dot{x} \rightarrow 0$ as $t \rightarrow \infty$. So the function $\dot{\lambda}_i \leq 0$ is integrable and its derivative goes to zero, this implies $\dot{\lambda}_i \rightarrow 0$. We conclude that the Lyapunov function (33) goes to zero as $t \rightarrow \infty$, but this function was constant in the invariant trajectory. Therefore, we have $\dot{x} \equiv 0$, $\dot{\lambda} \equiv 0$. The invariant trajectory is an equilibrium of the primal–dual dynamics, hence a saddle point of the optimization problem. \square

4.3. Application example

We consider the wireless network depicted in Fig. 1, where node 0 sends data to node 2 through node 1. We assume that every node $n \in \{0, 1, 2\}$ accesses the medium randomly with probability P^n . Hence the effective capacities of links (0, 1) and (1, 2) are given by $C_{(0,1)} = c(P^0(1 - P^1))$ and $C_{(1,2)} = cP^1$, respectively, where c denotes the physical-layer transmission rate (e.g., $c = 11$ Mbps, as in the IEEE 802.11 standard).

Intuitively, the solution to the formulation in Section 3.3 is for node 0 to transmit always, i.e., $\hat{P}^0 = 1$; and for node 1 to access the medium with probability $\hat{P}^1 = 1/2$. This way, node 1 spends half of the time receiving data from node 0, and the other half sending that information to its final destination, node 2. The optimum end-to-end rate for source 0 is $\hat{x}_0 = \frac{c}{2}$, regardless of the choice of utility function.

Now if we consider $U(x) = \log(x)$, a Matlab simulation of the primal–dual control algorithm (23)–(24) leads to oscillations,

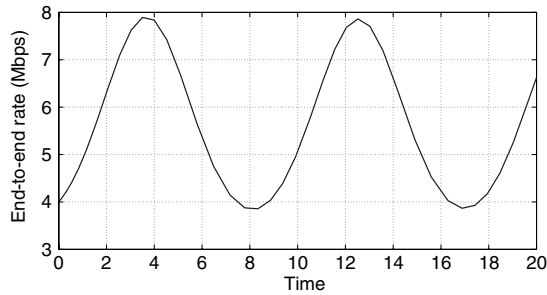


Fig. 2. Rate evolution under primal-dual dynamics.

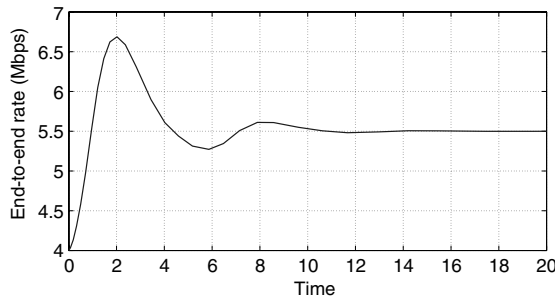


Fig. 3. Rate evolution under modified primal-dual dynamics.

as shown in Fig. 2 for the traffic rate x_0 of source 0. This lack of convergence to the equilibrium is not surprising since $\tilde{U}(\tilde{x}) := U(e^{\tilde{x}}) = \tilde{x}$ is not strictly concave (corresponds to $\alpha = 1$ in the family (29)).

As an alternative, we could apply a modified primal-dual gradient law of the form (35)–(36). With the introduction of the penalty function $\psi(u) = \max\{0, u^2\}$, the oscillations vanish (Fig. 3) and x_0 settles at the aforementioned optimum value. We note that this kind of control algorithm allows for distributed implementations.

5. Conclusion

We have studied stability of gradient laws for saddle point problems, motivated by primal-dual Lagrangian optimization. Much of our paper has involved revisiting, with more modern or alternative tools, the very rich theory already present in the classical monograph (Arrow et al., 1958). Along the way we have found the Krasovskii method as an alternative proof technique for these problems, with close connections to the quadratic Lyapunov functions; we have also obtained new extensions to non-strictly concave problems, and brought these tools to bear on distributed cross-layer network optimization.

References

Arrow, K., Hurwicz, L., & Uzawa, H. (1958). *Studies in linear and non-linear programming*. California, Stanford: Stanford University Press.

- Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
- Chiang, M., Low, S. H., Calderbank, A. R., & Doyle, J. C. (2007). Layering as optimization decomposition: a mathematical theory of network architectures. *Proceedings of the IEEE*, 95(January), 255–312.
- Feijer, D., & Paganini, F. (2009). Krasovskii's method in the stability of network control. In *Proceedings of the American control conference*.
- Freund, R. M. (2004). Penalty and barrier methods for constrained optimization. In *Lecture notes for nonlinear programming*. MIT.
- Kelly, F. P., Maulloo, A., & Tan, D. (1998). Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operations Research Society*, 49(3), 237–252.
- Khalil, H. K. (1996). *Nonlinear systems*. Prentice-Hall.
- Krasovskii, N. N. (1963). *Problems of the theory of stability of motion*. Stanford University Press (in Russian). English Translation.
- Lee, J.-W., Chiang, M., & Calderbank, A. R. (2007). Utility-optimal random-access control. *IEEE Transactions on Wireless Communications*, 6(7), 2741–2751.
- Lee, J.-W., Chiang, M., & Calderbank, A. R. (2006). Jointly optimal congestion and contention control based on network utility maximization. *IEEE Communication Letters*, 10(3), 216–218.
- Lin, X., Shroff, N., & Srikant, R. (2006). A tutorial on cross-layer optimization in wireless networks. *IEEE Journal on Selected Areas in Communications*, (Aug), 1452–1463.
- Lygeros, J., Johansson, K. H., Simic, S. N., Zhang, J., & Sastry, S. S. (2003). Dynamical properties of hybrid automata. *IEEE Transactions on Automatic Control*, 48, 2–17.
- Mo, J., & Walrand, J. (2000). Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5), 556–567.
- Nedic, A., & Ozdaglar, A. (2009). Subgradient methods for saddle point problems. *Journal of Optimization Theory and Applications*, 142(1), 205–228.
- Srikant, R. (2004). *The mathematics of internet congestion control*. Birkhauser.
- Wen, J.T., & Arcak, M. (2003). A unifying passivity framework for network flow control. In *IEEE Infocom*.
- Yu, Y., & Giannakis, G. B. (2008). Cross-layer congestion and contention control for wireless ad hoc networks. *IEEE Transactions on Wireless Communications*, 7(1).
- Zhang, J., Zheng, D., & Chiang, M. (2008). The impact of stochastic noisy feedback on distributed network utility maximization. *IEEE Transactions on Information Theory*, 54(2), 645–665.



Diego Feijer received his Electronics Engineering degree from Universidad ORT, Uruguay in 2009. From 2007 to 2009 he was Research Assistant in the Mathematics Applied to Telecommunications (MATE) research group at Universidad ORT. He is currently a Ph.D. student in the Laboratory for Information and Decision Systems at MIT. His research interests are networks, distributed optimization, control and dynamical systems.



Fernando Paganini received his Electrical Engineering and Mathematics degrees from the Universidad de la Republica, Montevideo, Uruguay, in 1990, and his M.S. and Ph.D. degrees in Electrical Engineering from the California Institute of Technology, Pasadena, in 1992 and 1996 respectively. From 1996 to 1997 he was a postdoctoral associate at MIT. Between 1997 and 2005 he was a faculty member at the Electrical Engineering Department at UCLA, reaching the rank of Associate Professor. Since 2005 he has been Professor of Electrical and Telecommunications Engineering at Universidad ORT, Uruguay. Dr. Paganini is the recipient of the 1994 O. Hugo Schuck best paper award at the American Control Conference, the Wilts and Clauser Prizes for outstanding Ph.D. thesis at Caltech in 1996, the 1999 NSF CAREER Award and the 1999 Packard Fellowship, and the 2004 George S. Axelby Award from the IEEE Control Systems Society. His research interests are control and networks.