The Gradient Method and Lyapunov Functions applied to Decentralized Optimization Problems

Evdokia Nikolova
Emmanuel College
Cambridge University

May 15, 2003

I declare that this essay is work done as part of the Part III Examination. It is the result of my own work, and except where stated otherwise, includes nothing which was performed in collaboration. No part of this essay has been submitted for a degree or any such qualification.

Signed Evdokia Nikolova

Kumanovo st, bl.209/A/7
Sofia 1233
Bulgaria

Contents

1 Introduction

This essay was motivated by work on Internet Congestion Control. I was very interested in the concept of decentralized optimization for finding optimal flow rates of users in a network, given the network contraints and each user's level of demand for unit flow. Although the decentralized model was essentially a dynamical system, I tried to avoid the control side of it in terms of proving stability of the solution, I thought I was only interested in the optimization side, of finding the actual solution. Consequently, I realized that I cannot separate the two, optimization and control. My attention was caught by the basic tools used in the proofs of stability and convergence of congestion control models—the Gradient Method and Lyapunov Functions. Since I kept seeing them used interchangeably, I thought they must be closely linked but was surprised not to find any explicit relationship drawn between them in the existing literature. I realised subsequently that this is due to the fact that the Gradient Method belongs strictly in the Optimization and Non-linear programming literature ([AHU58],[Ber99]), as an established method for finding optima of constrained and unconstrained problems, while Lyapunov Functions are a tool in Control Theory used to prove stability of equilibria of dynamical systems ([Gle94]).

Although a relation can easily be drawn between these two fields of Applied Mathematics, so far I have not found any book that treats them simultaneously. Consequently, in this essay I decided to explore the connection between the Gradient Method and Lyapunov Functions in greater detail, specifically as it relates to the problem that I was interested in originally, that of Internet Congestion Control and more generally, in the context of decentralised optimization.

Finding the optima of a function is a basic problem that is introduced as early as high-school. If we wish to find the minima and maxima of a function $F(x)$, we know that we need to set the function's first derivative to zero,

$$F'(x) = 0.$$

When we solve this equation for x, we know that we have found a local minimum or maximum if the second derivative of the function at the point in question is greater than or less than zero respectively. If the second derivative is zero, then we do not have an optimum but an inflection point instead.

In a control theoretic problem, we are generally given a dynamical system,

$$\frac{dx}{dt} = f(x(t))$$

or in short,

$$\dot{x} = f(x).$$

Here and elsewhere the dot over a variable will always stand for differentiation with respect to time. The equilibrium, or stationary point of that system is given by $\dot{x} = 0$ or equivalently, by $f(x) = 0$. One of the fundamental questions addressed by control theory is to find the equilibria of dynamical systems and classify them as stable or unstable, etc.

At first sight, the two problems posed above, that of finding the optimum of a function and classifying the equilibrium of a dynamical system, are very distinct. However, if in the control theoretic problem we replace $f(x)$ by the derivative of its integral function $F(x)$ (that is, define $F(x)$ to be such that $F'(x) = f(x)$), the differential equation above becomes

$$\dot{x} = F'(x) \tag{1}$$

and the stationary points are determined by

$$F'(x) = 0,$$

so the problem now coincides with the optimization problem above.

This reveals the link between the two types of problems at one level, finding the stationary points of the dynamical system and the potential optimal (not by chance also called stationary) points of the optimization problem. At a second level, classifying the equilibria as stable or unstable will also relate to classifying the optimal points as minima or maxima. The connection between the gradient method and Lyapunov functions comes in at this second level.

The necessity of the gradient method comes from the fact that it is often difficult to solve $F'(x) = 0$ in practice and find the optimal points analytically. The method arises from the special property of a local optimum that the first derivative is zero at the optimum while its sign changes from positive to negative or from negative to positive in the neighbourhoods before and after the optimum. This intuition explains why we should expect a system of the type (1), with reasonable conditions on F, to converge with time to a local optimum of F. Equation (1) thus stands as a definition for one of the most straight-forward gradient methods—the gradient descent method. Many variations and improvements from this basic notion of the gradient method have been studied, and have been treated in detail in [Ber99].

On the other hand, the stationary points of $F(x)$ give the equilibrium points of the dynamical system (1). The existence of a Lyapunov function, defined in Chapter 2, proves that an equilibrium point is stable and the optima of the Lyapunov function gives precisely the equilibrium point. We can see that stability of a stationary point $\bar{x}$, defined in essence by the tendency of trajectories $x(t)$ to it, is very closely related to convergence of these trajectories to $\bar{x}$ as required by the gradient method in the search of an optimum point $\bar{x}$.

With this in mind, after introducing the basic theory of Lyapunov functions and the gradient method, we look at examples where both methods can be used to the same purpose or more precisely, the problems in which the two methods are used are equivalent to each other and can in some sense be thought of as a primal and a dual problem. The dynamical systems we look at stem mostly from Internet Congestion Control models ([Kel00], [JoT01], [KMT98]) as well as similar decentralized optimization models from the economics literature, in particular ([AHU58]).

2 Lyapunov Functions

2.1 Stability

Consider the first-order differential equation

$$\dot{x} = f(x, t),\ x \in \mathbb{R}^n.$$

Here and throughout this work, a dot over a variable will stand for differentiation of this variable with respect to time. We want to study the stability of the equilibrium points of this equation. The equilibrium points are defined by $\dot{x} = 0$. The definition of stability however is not so straightforward. There are a number of different notions of stability, depending on the type of problem and what we want to achieve. The main notions involve the ideas that a point is stable if trajectories tend to it (asymptotic stability), or stay nearby if they start nearby (Lyapunov stability). A third common notion is that of structural stability, in which a system is stable if, when perturbed a little, it still has the same basic solution. The following definitions are based on [Gle94].

Definition 2.1 *A trajectory of the differential equation*

$$\frac{dx(t)}{dt} = f(x(t)),$$

starting at time t_0, is defined by the curve $x(t)$ with $t \geq t_0$, such that it satisfies the given differential equation. $x(t_0)$ is called the initial position of the trajectory; usually $t_0 = 0$.

Definition 2.2 *An equilibrium point is Lyapunov stable if all trajectories starting sufficiently close to the equilibrium point remain close to it for all time. More formally, x is Lyapunov stable iff for all $\epsilon \geq 0$ there exists $\delta \geq 0$ such that if $|x-y| < \delta$ then $|f(x,t)-f(y,t)| < \epsilon$ for all $t \geq 0$.*

Definition 2.3 *An equilibrium point is attracting or quasi-asymptotically stable if all trajectories that start near the equilibrium point, approach it as $t \to \infty$.*

Definition 2.4 *An equilibrium point is neutrally stable if it is Lyapunov stable but not attracting.*

Definition 2.5 *An equilibrium point which is both Lyapunov stable and attracting is said to be asymptotically stable. Similarly, a system is asymptotically stable in a region if all trajectories in the region converge to a single equilibrium point.*

Rather than convergence to a single point, we could as well talk about convergence to a cycle in which case we call it a stable limit cycle. Stability is *global* if the initial position of trajectories can be any point in the state space, and *local* if the initial position can only be a point in a neighbourhood of the stable point.

2.2 Lyapunov's method

Lyapunov's method is used to study the stability of non-linear dynamical systems, without solving the differential equations involved. [Jac74] refers to it as the "second method", as opposed to the "first method" of linearizing about singularities and considering eigenvalues of the resulting linear equations. The linearization method can only prove local stability, thus the Lyapunov's method is preferred when we need global results.

The method is applied to systems of the form

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

which have an equilibrium point at the origin. It considers a function of the state variables x, called a Lyapunov function, whose values are a measure of distance to the equilibrium point.

Definition 2.6 *([Gle94]) Suppose that the origin, $x = 0$, is an equilibrium point for the differential equation $\dot{x} = f(x)$, $x \in \mathbb{R}^n$. Let G be an open neighbourhood of 0 and*

$$V : cl(G) \to \mathbb{R}$$

be a continuously differentiable function. Then we can define the derivative of V along trajectories by differentiating V with respect to time using the chain rule, so

$$\dot{V}(x(t)) = \frac{dV}{dt} = \dot{x} \cdot \nabla V = f \cdot \nabla V = \sum f_i(x) \frac{\partial V(x)}{\partial x_i}$$

where the subscripts denote the components of f and x. Then V is a Lyapunov function on G iff V is continuously differentiable on $cl(G)$ and
(i) $V(0) = 0$ and $V(x) > 0$ for all $x \in cl(G) \setminus \{0\}$;
(ii) $\dot{V} \leq 0$ for all $x \in G$.

A Lyapunov function resembles an energy function. The idea is that an equilibrium point will be stable if the distance from it to the trajectories of the dynamical system tends to decrease.

Lemma 2.1 *(Bounding Lemma, [Gle94]) Suppose that G is some open bounded domain in $\mathbb{R}^n$ with boundary ∂G, and that $V : cl(G) \to \mathbb{R}$ is a Lyapunov function. If there exists $x_0 \in G$ such that $V(x) > V(x_0)$ for all $x \in \partial G$ then*

$$S(x_0) = \{x \in cl(G) \mid V(x) \leq V(x_0)\}$$

is a bounded set in G and $f(x_0, t) \in S(x_0)$ for all $t \geq 0$.

Proof Sketch: $S(x_0)$ is bounded and it is in G by definition. Now $V(x)$ is decreasing with time hence a trajectory starting at x_0 will stay in $S(x_0)$. □

Theorem 2.2 *(Lyapunov's First Stability Theorem, [Gle94]) Suppose that a Lyapunov function can be defined on a neighbourhood of the origin, $x = 0$, which is a stationary point of the differential equation $\dot{x} = f(x)$. Then the origin is Lyapunov stable.*

The problem with Lyapunov functions is that there is no general method of finding them, for general non-linear systems. Also, they would rarely exist on the whole state space; most often we would only be able to find a Lyapunov function on a small neighbourhood of the equilibrium point.

Once we find a Lyapunov function, however, not only does it let us prove stability of an equilibrium point but it also helps find the largest invariant set to which trajectories tend to, namely this is the largest invariant subset of

$$E = \{x \in G \mid \dot{V}(x) = 0\},$$

the set of points where the time derivative of the Lyapunov function vanishes.

Theorem 2.3 *(La Salle's Invariance Principle [Gle94]) Suppose that $x = 0$ is a stationary point of $\dot{x} = f(x)$ and V is a Lyapunov function on some neighbourhood G of $x = 0$. If $x_0 \in G$ has its forward trajectory, $\gamma^+(x_0)$, bounded with limit points in G and M is the largest invariant subset of E, then*

$$f(x_0, t) \to M \text{ as } t \to \infty.$$

By the theorems above, asymptotic stability will be achieved if the time derivative of the Lyapunov function is strictly positive except at the stationary points.

Theorem 2.4 *(Lyapunov's Second Stability Theorem [Gle94]) Suppose $x = 0$ is a stationary point for $\dot{x} = f(x)$ and let V be a Lyapunov function on a neighbourhood G of $x = 0$. If $\dot{V}(x) < 0$ for all $x \in G \backslash \{0\}$, then $x = 0$ is asymptotically stable.*

The above discussion was mainly about first-order systems, *i.e.,* systems of the form $\frac{dx}{dt} = f(x)$, where x and $f(x)$ are vectors. It is naturally extended to second-order systems, the latter being defined as

$$\frac{dx}{dt} = f(x, y), \qquad \frac{dy}{dt} = g(x, y),$$

again with x, y, $f(x, y)$, $g(x, y)$ being vectors of appropriate dimensions.

2.3 A Special Class of Lyapunov Functions

As we noted above, it is in general difficult and there are no methods to find Lyapunov Functions. However, based on [KMT98], we have noted a special class of Lyapunov Functions which can be found in a straighforward manner. Let $F(x, y)$ be a function of x and y with F_x and F_y denoting its partial derivatives with respect to x and y accordingly. Consider a system of the form

$$\frac{dx}{dt} = F_x(x, y) \tag{2}$$

$$\frac{dy}{dt} = F_y(x, y). \tag{3}$$

Then $F(x, y)$ offers a possible Lyapunov function for the system (2)-(3) since

$$\frac{dF}{dt} = \frac{\partial F}{\partial x}\frac{dx}{dt} + \frac{\partial F}{\partial y}\frac{dy}{dt} = F_x^2 + F_y^2 \geq 0. \tag{4}$$

Note, in the definition of a Lyapunov function (see definition 2.6) we require that its time derivative be non-positive and the function have a global minimum at the equilibrium point. This is simply a matter of convention and we can easily invert the definition to require that the time derivative be non-negative and the function have a global maximum at the equilibrium point. While the former helps our intuition by associating Lyapunov functions with energy functions, the latter is more suited in an economics or optimization setting where we prefer to talk of maximizing profits and revenues. In light of this, we offer an alternate definition of a Lyapunov function.

Definition 2.7 *Suppose that $(x, y) = (a, b)$ is an equilibrium point for the system of differential equations*

$$\dot{x} = f(x, y), \;\; \dot{y} = g(x, y),$$

where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Let G be an open neighbourhood of (a, b) and $V : cl(G) \to \mathbb{R}$ be a continuously differentiable function. Then we can define the derivative of V along trajectories by differentiating V with respect to time using the chain rule, so

$$\dot{V}(x, y) = \frac{dV}{dt} = \sum f_i(x, y)\frac{\partial V}{\partial x_i} + \sum g_j(x, y)\frac{\partial V}{\partial y_j}$$

where the subscripts denote the components of x, y, f and g. Then V is a Lyapunov function on G iff V is continuously differentiable on $cl(G)$ and
(i) (a, b) is a unique maximum of V on G;
(ii) $\dot{V} \geq 0$ for all $(x, y) \in G$.

Note, we stated this definition with respect to a second-order system of differential equations but it can readily be adapted to refer to a first or higher order system. Now we can prove the result suggested above, with appropriate constraints on $F(x, y)$.

Theorem 2.5 *Let $F : cl(G) \to \mathbb{R}$ where $G \in \mathbb{R}^n \times \mathbb{R}^m$ is some open ball, be a continuously differentiable function, with a unique maximum at $(\bar{x}, \bar{y}) \in G$. Then F is a Lyapunov function for the system of differential equations (2)-(3) and its maximizing value is the equilibrium point of the system.*

Proof: Condition (i) of definition 2.7 is automatically satisfied. Condition (ii) follows from what we established above, namely that

$$\frac{dF}{dt} = F_x^2 + F_y^2 \geq 0.$$

Therefore, $F(x, y)$ is a Lyapunov function for the system (2)-(3). Its maximizing value $(\bar{x}, \bar{y})$ is obtained by setting $F_{x_i} = \frac{\partial F}{\partial x_i}$ and $F_{y_j} = \frac{\partial F}{\partial y_j}$ to zero. This gives precisely the equilibrium point of the system. $\square$

Naturally, any condition on $F(x, y)$ above which guarantees the unique maximum of F can replace the latter. Often, decentralized optimization problems will involve strictly concave functions on the non-negative reals with interior maxima.

3 The Gradient Method

3.1 Basic Concept

The gradient method is a way of finding the local optimum of a function assuming that the function gradient can be computed. In simplest terms, the method solves

$$\dot{x} = f'(x)$$

in the case of a local maximum, and

$$\dot{x} = -f'(x)$$

in the case of a local minimum, where $\dot{x}$ is differentiation with respect to time. In the discrete time case, the method consists of iterating

$$x_{t+1} = x_t + f'(x)$$

until the sequence x_t converges and the point of convergence gives the desired optimum. We will only work in the continuous time case here.

3.2 The Gradient Method for Concave Programming

In resource allocation models, we are often interested in finding the maximum of a function $f(x)$, say utility, production or profit, subject to constraints $g(x) \geq 0$. We solve the problem by considering the Lagrangian

$$L(x, y) = f(x) + y^T g(x)$$

where y^T, the Lagrangian multipliers, find the natural interpretation of shadow prices of the resources to be allocated. Here, $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_m)$. $f(x)$ will usually be concave in x, and thus $L(x, y)$ is concave in x and linear or convex in y. We are thus interested particularly in gradient method theorems for finding saddle points of concave-convex functions $L(x, y)$, restricted to $x \geq 0$ and $y \geq 0$. The gradient method to problems of that type is given by the following system of differential equations.

$$\begin{array}{ll} \dot{x}_i = 0 & \text{if } L_{x_i} < 0 \text{ and } x_i = 0 \\ \dot{x}_i = L_{x_i} & \text{otherwise} \\ \dot{y}_j = 0 & \text{if } L_{y_j} < 0 \text{ and } y_j = 0 \\ \dot{y}_j = L_{y_j} & \text{otherwise} \end{array}$$

where $L_{x_i} = \frac{\partial L}{\partial x_i}$ and $L_{y_j} = \frac{\partial L}{\partial y_j}$. We can rewrite the above system in a more concise form as follows.

$$\dot{x}_i = \delta_{x_i} L_{x_i} \tag{5}$$

$$\dot{y}_j = \delta_{y_j} L_{y_j} \tag{6}$$

Here, $\delta_{x_i} = 0$ whenever $x_i = 0$ and $L_{x_i} < 0$ and $\delta_{x_i} = 1$ otherwise. Similarly for δ_{y_j}.

We call a solution $(x(t), y(t))$ of a system *regular* ([AHU58]) if whenever $x_i(t_\mu) = 0$ (or $y_j(t_\mu) = 0$) for some sequence $\{t_\mu\}_{\mu \in \mathbb{N}}$ such that $t_\mu \geq 0$ and $\lim_{v \to \infty} t_\mu = 0$, then there is a positive number $\bar{t} \geq 0$ such that $x_i(t) = 0$ (or respectively $y_j(t) = 0$) for $0 < t < \bar{t}$.

The main theorem proved in [AHU58] gives global stability of the system (5)-(6) in the strictly concave case.

Theorem 3.1 *[AHU58] Let $L(x, y)$ be strictly concave and continuously twice-differentiable in n-vector $x \geq 0$, convex and continuously twice-differentiable in m-vector $y \geq 0$, such that the system (5)-(6) has a regular solution with respect to the initial position $x^0 \geq 0$, $y^0 \geq 0$. Then there is a unique regular solution $(x(t), y(t); x^0, y^0)$ of with any initial position $(x^0, y^0) \geq 0$.*

Furthermore, if $L(x, y)$ possesses a saddle-point $(\bar{x}, \bar{y})$ under the constraints $x \geq 0$ and $y \geq 0$, the x-component $\bar{x}$ of any saddle point is uniquely determined and the x-component $x(t)$ of the solution of (5)-(6) with an arbitrary initial position $(x^0, y^0) \geq 0$ converges to $\bar{x}$.

Arrow and Hurwicz subsequently extend Uzawa's results in Theorem 3.1 to include convergence in the second variable.

Theorem 3.2 *[AHU58] If $L(x, y)$ is strictly concave in x for each y and convex in y for each x, then the gradient process converges in both x and y.*

3.3 The Gradient Method for a Concave-Concave Function

Suppose we were given a function $L(x, y)$ of two variable vectors which is concave in both. We will prove a similar theorem to theorem 3.1 for the gradient method in this case. First, let us define the gradient method. Similarly to the sadde point case, we aim to find a maximum for L, on $x \geq 0$, $y \geq 0$. Suppose the maximum is at $(\bar{x}, \bar{y})$, with $\bar{x} \geq 0$, $\bar{y} \geq 0$, i.e. it is a solution of

$$\bar{x} \geq 0, \, \bar{y} \geq 0 \tag{7}$$

$$L(\bar{x}, \bar{y}) \geq L(x, y) \text{ for all } x \geq 0, \, y \geq 0. \tag{8}$$

If $L(x, y)$ is concave in both variables and continuously differentiable with respect to both, condition (8) becomes equivalent to conditions (9) and (10) below.

$$\bar{L}_{x_i} \leq 0 \tag{9}$$

with equality for i whenever $\bar{x}_i > 0$ and where $\bar{L}_{x_i}$ stands for $\frac{\partial L}{\partial x_i}$, evaluated at $\bar{x}_i$.

$$\bar{L}_{y_j} \leq 0 \tag{10}$$

with equality for j whenever $\bar{y}_j > 0$.

Then the maximum point $(\bar{x}, \bar{y})$ is a singular point of the following system of differential equations which we call the gradient method:

$$\dot{x}_i = \delta_{x_i} L_{x_i} \tag{11}$$

$$\dot{y_j} = \delta_{y_j} L_{y_j} \tag{12}$$

where $\delta_{x_i} = 0$ for $x_i = 0$ and $L_{x_i} < 0$ and it is 1 otherwise, similarly for δ_{y_j}.

Theorem 3.3 *Let $L(x, y)$ be strictly concave and continuously twice-differentiable in n-vector $x \geq 0$, concave and continuously twice-differentiable in m-vector $y \geq 0$, such that the system (11)-(12) has a regular solution with respect to the initial position $x^0 \geq 0$, $y^0 \geq 0$. Then there is a unique regular solution $(x(t), y(t); x^0, y^0)$ of (11)-(12) with any initial position $(x^0, y^0) \geq 0$.*

Furthermore, if $L(x, y)$ possesses a unique global maximum $(\bar{x}, \bar{y})$ under the constraints $x \geq 0$ and $y \geq 0$, the x-component $x(t)$ of the solution of (11)-(12) with an arbitrary initial position $(x^0, y^0) \geq 0$ converges to $\bar{x}$.

Remark 3.4 *Note, unlike the original theorem for saddle points, this one requires a necessary condition that the global maximum of $L(x, y)$ be unique. To see why, consider for example the function $f(x, y) = -(x - y)^2$, which is strictly concave in both x and y and is bounded above by 0. This function has infinitely many global maxima, all along the diagonal $x = y$.*

Proof: This proof is a modified version of the one given in [AHU58] to Theorem 3.1, which considers the case of a saddle point.

There are three parts to the proof. The first one establishes existence, uniqueness and continuity of the solution. The second part shows that the distance from the solution to the maximum point strictly decreases with time except when we have reached the maximum. The third part of the proof establishes the convergence of the solution to the maximum, for any initial position (x^0, y^0).

(a) This part is exactly the same as in [AHU58] since it is not related to the concavity properties of $L(x, y)$. We include the argument for completeness.

Let $z = (z_1, z_2, ..., z_l) = (x_1, ..., x_n, y_1, ..., y_m)$, with $l = n + m$. Let also

$$L_i(z) = L_{x_i}, \text{ for } 1 \leq i \leq n$$

$$L_j(z) = L_{y_j}, \text{ for } 1 \leq j \leq m.$$

Divide the indices 1, 2, ..., l into three sets:

$$S^0 = i\text{: } z_i^0 > 0, \text{ or } z_i^0 = 0 \text{ and } L_i(z^0) > 0$$

$$T^0 = i\text{: } z_i^0 = 0 \text{ and } L_i(z^0) = 0$$

$$R^0 = i\text{: } z_i^0 = 0 \text{ and } L_i(z^0) < 0$$

Define the (S^0, T^0, R^0)-system to be the following system of differential equations:

$$\dot{z}_i = \begin{cases} L_i(z) & \text{for } i \in S^0 \\ \max[0, L_i(z)] & \text{for } i \in T^0 \\ 0 & \text{for } i \in R^0 \end{cases} \tag{13}$$

If $L_1(z)$, ... $L_l(z)$ are continuously differentiable, then (13) has a solution $z(t; z^0)$ which is unique and continuous with respect to any initial condition $z^0 \geq 0$. Let t^1 be the smallest upper bound of t such that for all $0 < \sigma < t$,

$$\begin{array}{ll} z_i(\sigma) > 0 & \text{for } i \in S^0 \\ z_i(\sigma) = 0 \text{ or } L_i(z(\sigma)) \text{ is always positive} & \text{for } i \in T^0 \\ L_i(z(\sigma)) < 0 & \text{for } i \in R^0. \end{array}$$

From the regularity condition on the solution, $t^1 > 0$. Hence $z(t; z^0)$ is a solution of the system (11)-(12) in the interval $[0, t^1]$. On the other hand, if (11)-(12) has a regular solution with initial condition z^0, this solution must coincide with the solution of the (S^0, T^0, R^0)-system in $[0, t^1]$. Since the latter is unique and continuous, the solution to (11)-(12) is also unique and continuous with respect to z^0 in $[0, t^1]$.

Define similarly an (S^1, T^1, R^1)-system with initial condition $z^1 = z(t^1)$. Denote the solution $z(t; z^1)$. Define also t^2 for $z(t; z^1)$ similarly to the way t^1 was defined for $z(t; z^0)$. The curve

$$z(t; z^0) = \begin{cases} z(t; z^0) & \text{for } 0 \leq t < t^1 \\ z(t - t^1; z^1) & \text{for } t^1 \leq t < t^1 + t^2 \end{cases}$$

is a solution of the system (11)-(12) in $[0, t^1 + t^2]$. Repeating in this way, we obtain a sequence of times t^1, t^2, ..., $t^v > 0$ and a curve $z(t; z^0)$ which is a solution to the system (11)-(12) in $[0, t^1 + ... + t^v]$ for $v = 1, 2, ...$

Next, denote

$$t^* = \lim_{v \to \infty} t^1 + ... + t^v.$$

We will show that t^* cannot be finite. Suppose the contrary, that $t^* < \infty$. Then the set z^v ; $v = 1, 2, ...$ is bounded, where $z^v = z(t^v; z^0)$. Also there exists a constant C such that

$$L_i(z) < C \qquad \text{for all } z = z(t), \qquad 0 < t < t^*.$$

Then

$$|z_i(t) - z_i(t')| = |\int_{t'}^{t} \dot{z}_i dt| \leq \int_{t'}^{t} |L_i(z)| dt \leq C|t - t'|.$$

Thus, there exists z^* with

$$\lim_{t \to t^*} z(t) = z^*. \tag{14}$$

Define

$$S^* = i\text{: } z_i^* > 0, \text{ or } z_i^* = 0 \text{ and } L_i(z^*) > 0$$

$$T^* = i\text{: } z_i^* = 0 \text{ and } L_i(z^*) = 0$$

$$R^* = i\text{: } z_i^* = 0 \text{ and } L_i(z^*) < 0.$$

Then by (14) we cannot have the case $z_i^* = 0$ and $L_i(z^*) > 0$. Therefore we can find v such that for all τ,

$$t^1 + \cdots + t^v < \tau < t^*$$

$$z_i(\tau) > 0 \qquad \text{for } i \in S^*$$

$$z_i(\tau) = 0 \text{ or } z_i(\tau) > 0 \text{ and } L_i(z(\tau)) < 0 \qquad \text{for } i \in T^*$$

$$L_i(z(\tau)) < 0 \qquad \text{for } i \in R^*.$$

Constructing the (S^v, T^v, R^v)-system at z^v yields

$$t^{v+1} > t^* - (t^1 + \cdots + t^v)$$

which is a contradiction with the fact that $t^1 + \cdots + t^{v+1} < t^*$. Therefore $\lim_{v\to\infty} t^1 + ... + t^v$ is infinite, that is the solution of (11)-(12) is unique and continuous over the entire positive time axis.

(b) Next, let $z(t; z^0)$ be a solution with any initial condition z^0. Define the distance from the solution to the maximum point $\bar{z} = (\bar{x}, \bar{y})$ to be

$$D(t) = \frac{1}{2}|z(t) - \bar{z}|^2.$$

We will show that

$$\dot{D}(t) \leq 0 \qquad \text{for any } z(t)$$

where equality is achieved only for $z(t)$ such that $x(t) = \bar{x}$.

$$\begin{aligned} \dot{D} &= \dot{z}(t) \cdot (z(t) - \bar{z}) \\ &= L_x \cdot \delta_x \cdot (x - \bar{x}) + L_y \cdot \delta_y \cdot (y - \bar{y}) \\ &= L_x \cdot (x_I - \bar{x}_I) + L_y \cdot (y_I - \bar{y}_I). \end{aligned} \tag{15}$$

The variables appearing above are defined as

$$L_x = (L_{x_1}, ..., L_{x_n})^T, \qquad L_y = (L_{y_1}, ..., L_{y_m})^T;$$

$$\delta_x = \begin{pmatrix} \delta_{x_1} & 0 & \cdots & 0 \\ 0 & \delta_{x_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_{x_n} \end{pmatrix}, \qquad \delta_y = \begin{pmatrix} \delta_{y_1} & 0 & \cdots & 0 \\ 0 & \delta_{y_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_{y_m} \end{pmatrix}$$

$$x_I = \delta_x \cdot x = (\delta_{x_1} x_1, ..., \delta_{x_n} x_n), \qquad x_{II} = x - x_I$$

$$\bar{x}_I = \delta_x \cdot \bar{x} = (\delta_{x_1} \bar{x}_1, ..., \delta_{x_n} \bar{x}_n), \qquad \bar{x}_{II} = \bar{x} - \bar{x}_I.$$

Note that $\bar{x}_I$ and $\bar{x}_{II}$ are dependent on t since δ_x depends on t. We define y_I, y_{II}, $\bar{y}_I$ and $\bar{y}_{II}$ similarly to the corresponding x-variables.

By strict concavity of $L(x, y)$ in x and non-strict concavity in y,

$$L_x \cdot (\bar{x} - x) > 0 \qquad \text{for } x \neq \bar{x}$$

and

$$L_y \cdot (\bar{y} - y) \geq 0 \qquad \text{for } y \neq \bar{y}$$

Therefore

$$L_x \cdot (\bar{x} - x) + L_y \cdot (\bar{y} - y) > 0.$$

Substituting $x_I = x - x_{II}$ and $y_I = y - y_{II}$ in (15), we obtain

$$\begin{aligned}\dot{D} &= L_x \cdot (x - \bar{x}) + L_y \cdot (y - \bar{y}) - L_x \cdot (x_{II} - \bar{x}_{II}) - L_y \cdot (y_{II} - \bar{y}_{II}) \\ &< L_x \cdot \bar{x}_{II} + L_y \cdot \bar{y}_{II} \leq 0\end{aligned}$$

for (x, y) with $x \neq \bar{x}$. This establishes what we wanted to show, that

$$\dot{D}(t) \leq 0 \qquad \text{for any } z(t)$$

with equality only for $z(t)$ such that $x(t) = \bar{x}$.

(c) This last part of the proof of the theorem shows convergence of $x(t)$ to $\bar{x}$. It does not use the concavity properties of L and is the same as in [AHU58]. Again, we include it for completeness.

Since $\dot{D} \leq 0$, $D^* = \lim_{t\to\infty} D(t)$ exists. On the other hand, the set $(x(t^v; z^0), v = 1, 2, ...)$ is bounded for all sequences t^v with $\lim_{t\to\infty} t^v = \infty$. Hence this set has an accumulation point x^* say and further we can find a subsequence $t^{v_k} : k = 1, 2, ...$ of t^v, such that

$$\lim_{k\to\infty} z(t^{v_k}; z^0) = Z^*,$$

where $z^* = (x^*, y^*)$ for an appropriate y^*. By (a) there exists a unique solution for any initial condition and in particular $z^v = z(t; z^v)$ is a unique solution with the initial condition $z^v = z(t^v; z^0)$. Therefore, $z(t; z^v) = z(t + t^v; z^0)$. Then

$$D^*(t) = D(t; z^*) = \lim_{k\to\infty} D(t; z^{v_k}) = \lim_{k\to\infty} D(t + t^{v_k}; z^0) = D^*,$$

since z and D are continuous with respect to t. Therefore

$$\dot{D} = 0 \qquad \text{at } z = z^*.$$

By (c), $x^* = \bar{x}$. Therefore $x(t^v)$ converges to $\bar{x}$, for any sequence t^v which tends to ∞. Hence

$$\lim_{t\to\infty} x(t) = \bar{x}.$$

□

Note, the rather involved proof in part (a) of the theorem is only required because of the constraints $x \geq 0$, $y \geq 0$. If these restrictions were removed, the proof of the theorem would proceed much like the above, only part (a) would become unnecessary and part (b) would become much simpler as well. In addition, by a symmetry argument we can extend the convergence of the gradient method to y as well if the function $L(x, y)$ is strictly concave in y as it is in x.

Corollary 3.5 *If $L(x, y)$ is strictly concave in both x and y, then the gradient process converges in both x and y.*

3.4 Application of the Gradient Method to a Resource Allocation Model

We follow the description of a resource allocation model in [AHU58]. Let x be a vector of activity levels, $g(x)$ a vector of net output of desired goods and $h(x)$ a vector of primary goods. Let y be a vector of the desired goods consumption, and $f(y)$ be the corresponding utility function of that consumption.

We assume that $f(y)$ is a strictly concave and strictly increasing function of y and that $g(x)$ and $h(x)$ are concave functions of x. We then need the following constraints:

$$g(x) - y \geq 0, \;\; h(x) \geq 0.$$

We would like to maximize utility subject to these constraints:

$$\max f(y) \text{ subject to } \quad \begin{aligned} & g(x) - y \geq 0, \\ & h(x) \geq 0, \\ & x \geq 0, \\ & y \geq 0. \end{aligned}$$

The Lagrangian for this problem is

$$L(x, y; p, q) = f(y) + p'[g(x) - y] + q'h(x).$$

Theorem 3.6 *([AHU58]) Suppose $f(y)$ is strictly concave and strictly increasing, $g(x)$ and $h(x)$ concave, and $(\bar{x}, \bar{y})$ maximizes $f(y)$ subject to the constraints $g(x) - y \geq 0$, $h(x) \geq 0$. Let p be the vector of Lagrangian multipliers associated with the constraints $g(x) - y \geq 0$. Then the solution of the gradient method has the properties:*

(a) $\lim_{t \to \infty} y(t) = \bar{y}$;

(b) if $\bar{y}_i > 0$, *then* $\lim_{t \to \infty} p_i(t) = \bar{p}_i$ *and* $\lim_{t \to \infty} g_i[x(t)] = \bar{y}_i$.

4 Use of the Gradient Method vs Lyapunov Functions

4.1 Proving Stability in a Simple Example

Consider a simple example of a second-order system from [Dra92]

$$\frac{dx}{dt} = ax - cy \tag{16}$$

$$\frac{dy}{dt} = cx + ay \tag{17}$$

where $a < 0$.

We can prove that $(x = 0, y = 0)$ is a stable equilibrium point of the system both through the use of the gradient method and of Lyapunov functions.

When we want to show asymptotic stability, we look for a Lyapunov function $L(x,y)$ which is strictly concave and has a positive time derivative except at the equilibrium point where the time derivative is zero. Since

$$\frac{d}{dt}L(x,y) = \frac{\partial L}{\partial x}\frac{dx}{dt} + \frac{\partial L}{\partial y}\frac{dy}{dt},$$

a general strategy is to try and find L such that $\frac{\partial L}{\partial x}$ is proportional to $\frac{dx}{dt}$, and similarly for y since then $\frac{d}{dt}L(x,y)$ will be a sum of squares and hence positive as necessary.

In our concrete example, we set $L(x,y)$ to be the sum of the integrated right-hand-sides of (16)-(17), with respect to x and y:

Claim 4.1 $L(x,y) = \frac{1}{2}ax^2 - cxy + cxy + \frac{1}{2}ay^2 = \frac{a}{2}(x^2+y^2)$ *provides a Lyapunov function for the system (16)-(17).*

Proof: Since $a < 0$, $L(x,y)$ is strictly concave in both x and y. Next,

$$\frac{d}{dt}L(x,y) = \frac{\partial L}{\partial x}\frac{dx}{dt} + \frac{\partial L}{\partial y}\frac{dy}{dt} = ax(ax-cy) + ay(cx+ay) = a^2(x^2+y^2).$$

So $\frac{dL}{dt} \geq 0$ as desired and this concludes the proof. Moreover, equality is achieved precisely at $(0,0)$ hence this is the unique asymptotically stable equilibrium point of the system. □

Next, we turn to the gradient method. We would like to use Theorem 3.1 to show convergence of the gradient method to a stable equilibrium. Thus, we would like to fit our system (16)-(17) to the form

$$\frac{dx}{dt} = \frac{\partial}{\partial x}\mathscr{F}(x,y) \tag{18}$$

$$\frac{dy}{dt} = -\frac{\partial}{\partial y}\mathscr{F}(x,y) \tag{19}$$

where $\mathscr{F}(x,y)$ is a function which is strictly concave in x and convex in y.

Claim 4.2 *There is a unique regular solution $(x(t),y(t);x^0,y^0)$ to (16)-(17), with any initial position (x^0,y^0). Furthermore, this solution converges to $(0,0)$ in both x and y.*

Proof: The above discussion suggests us to consider the function $\mathscr{F}(x,y) = \frac{a}{2}x^2 - cxy - \frac{a}{2}y^2$. The gradient method (18)-(19) is then precisely our original system (16)-(17). Since $a < 0$, $\mathscr{F}$ is strictly concave in x and strictly convex in y and it is continuously-twice differentiable in both variables, with a unique saddle point $(0,0)$. Applying Theorem 3.1 and Theorem 3.2 about the gradient method in the case of saddle points gives the result. □

It is worth noting that in this example the Lyapunov function and the Gradient method function are not the same. One may want to explore in what class of problems the two coincide. In such case we can define a primal and a dual problem where the primal optimizes a given function while the dual looks at the induced system of differential equations, trying to answer questions about its equilibrity points and stability.

Consider a modified version of system (16)-(17):

$$\frac{dx}{dt} = ax + cy \tag{20}$$

$$\frac{dy}{dt} = cx + ay \tag{21}$$

where $a < c < 0$.

Here, for the gradient method we need to take the function

$$\mathscr{V}(x, y) = \frac{a}{2}x^2 + cxy + \frac{a}{2}y^2.$$

This function is strictly concave both in x and y and it is continuously twice-differentiable in both variables. Also, it has a unique global maximum (0, 0) which we can see by writing $\mathscr{G}$ in the form

$$\mathscr{V}(x, y) = \frac{a-c}{2}x^2 + \frac{a-c}{2}y^2 + \frac{c}{2}(x+y)^2.$$

Applying Theorem 3.3 and Corollary 3.5 then shows that the gradient method defined by (20)-(21) converges to the global maximum. Hence $(0, 0)$ is a stable point of the system (20)-(21).

Further, note that $\mathscr{V}$ satisfies the conditions of a Lyapunov function by Definition 2.7, so this provides a second proof for the stability of the unique equilibrium point $(0, 0)$. The relationship between $\mathscr{V}$ and the system (20)-(21) which enables the equivalent use of the two methods is that

$$\frac{dx}{dt} = \frac{\partial \mathscr{V}}{\partial x} \tag{22}$$

$$\frac{dy}{dt} = \frac{\partial \mathscr{V}}{\partial y} \tag{23}$$

This last system defines a class of problems where the gradient method coincides with the use of Lyapunov functions.

4.2 Duality of the Gradient Method and Lyapunov Functions

Suppose we are given a function $V(x, y)$ which takes values in the real numbers and is strictly concave in both $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. We wish to find the maximum of that function on $x \geq 0$, $y \geq 0$. This formulates a primal problem:

PRIMAL

$$\begin{array}{ll} \text{max} & V(x, y) \\ \text{subject to} & x \geq 0 \\ & y \geq 0. \end{array} \tag{24}$$

If we know that $V(x,y)$ has a unique global maximum then we can use the gradient method to find that maximum. The gradient method induces a system of differential equations which we will call the DUAL problem.

DUAL

$$\frac{dx_i}{dt} = \delta_{x_i} V_{x_i}(x,y) \qquad (i = 1, ..., n)$$
$$\frac{dy_j}{dt} = \delta_{y_j} V_{y_j}(x,y) \qquad (j = 1, ..., m) \tag{25}$$

where $V_{x_i} = \frac{\partial V}{\partial x_i}$ and $V_{y_j} = \frac{\partial V}{\partial y_j}$. Also

$$\delta_{x_i} = \begin{cases} 0 & \text{for } x_i = 0 \text{ and } V_{x_i} < 0, \\ 1 & \text{otherwise.} \end{cases}$$

and

$$\delta_{y_j} = \begin{cases} 0 & \text{for } y_j = 0 \text{ and } V_{y_j} < 0, \\ 1 & \text{otherwise.} \end{cases}$$

The goal of the DUAL problem is to find the equilibrium points of the system (25) and to determine their stability.

Theorem 4.3 *Let $V(x,y) : \mathbb{R}^n \times \mathbb{R}^m \to \infty$ be twice continuously-differentiable and strictly concave in x and y, such that it has a unique global maximum $(\bar{x}, \bar{y})$ over $x \geq 0$, $y \geq 0$. Then*

(a) the maximum of V which is the solution to PRIMAL, coincides with the unique equilibrium point of DUAL and can be found by the gradient method;

(b) V provides a Lyapunov function for DUAL and hence the equilibrium of DUAL is stable.

Proof: (a) The maximum $(\bar{x}, \bar{y})$ of V is the unique solution to the system

$$\delta_{x_i} V_{x_i} = 0, \qquad \delta_{y_j} V_{y_j} = 0$$

and so is the equilibrium point of DUAL. Hence $(\bar{x}, \bar{y})$ is precisely the equilibrium of DUAL. By Theorem 3.3 and Corollary 3.5, the gradient method converges to the maximum $(\bar{x}, \bar{y})$.

(b)

$$\frac{dV}{dt} = \sum_i \frac{\partial V}{\partial x_i}\frac{dx}{dt} + \sum_j \frac{\partial V}{\partial y_j}\frac{dy}{dt} = \sum_i V_{x_i}^2 + \sum_j V_{y_j}^2 \geq 0$$

Therefore, since $(\bar{x}, \bar{y})$ is the unique maximum of V, V is a Lyapunov function for DUAL. Hence the equilibrium of DUAL is stable. $\square$

4.3 Application to Internet Resource Allocation and Congestion Control

4.3.1 Basic Model

We describe the mathematical model of the Internet after [KMT98], [Kel00] and [LPD01]. Let J be a set of resources or links, R a set of routes or users. A route r is a subset of links, $r \in J$. Let A be the link-route incidence matrix, that is

$$A_{jr} = \begin{cases} 1 & \text{if } j \in r \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

For a link $j \in J$, define its capacity by C_j. As we stated above, we can use the terms routes and users interchangeably since we assume each route to be associated with a unique user and vice versa. A rate x_r on route r induces utility $U_r(x_r)$ to the corresponding user. We assume utility functions to be increasing, strictly concave and continuously differentiable on $[0, \infty)$. For convenience, denote $C = (C_j, j \in J)$, the vector of capacities and also, $x = (x_r, r \in R)$, the vector of all users.

The problems in resource allocation can be posed at several levels. At a general system level, the problem is

$$\begin{array}{ll} \text{SYSTEM PROBLEM} & (U, A, C) \\ & \\ \max & \displaystyle\sum_{r \in R} U_r(x_r) \\ \text{subject to} & Ax \leq C \\ \text{over} & x \geq 0 \end{array} \tag{27}$$

The problem with this formulation is that the network does not always know the utility functions of the users and there is no central server which knows the link-route matrix A. Thus, we need to decentralize the problem. Consequently, we can formulate an equivalent problem at the user level:

$$\begin{array}{ll} \text{USER}_r \text{ PROBLEM} & (U_r; \lambda_r) \\ & \\ \max & U_r(\dfrac{w_r}{\lambda_r}) - w_r \\ \text{over} & w_r \geq 0 \end{array} \tag{28}$$

Here, w_r is the price that user r is willing to pay per unit time, to receive a flow of $x_r = \frac{w_r}{\lambda_r}$ for some constant of proportionality λ_r. We can also define a third problem at a network level which is similar to the system problem, only here the utility functions are assumed to be logarithmic with the price w_r as above input by each user.

$$\begin{array}{ll} \text{NETWORK PROBLEM} & (A, C; w) \\ \max & \sum_{r \in R} w_r \log x_r \\ \text{subject to} & Ax \le C \\ \text{over} & x \ge 0 \end{array} \tag{29}$$

Kelly proves in [Kel97] that there exist vectors $\lambda = (\lambda_r,\ r \in R)$, $w = (w_r,\ r \in R)$ and $x = (x_r,\ r \in R)$ such that $w_r = \lambda_r x_r$, w_r solves $\text{USER}_r(U_r, \lambda_r)$ and x solves the NETWORK problem and consequently the SYSTEM problem if $U_r(x_r) = w_r \log x_r$.

The proof finds the optimal rate x by defining the Lagrangian for the NETWORK problem:

$$L_{\text{network}} = \sum_{r \in R} w_r \log x_r + \mu^T (C - Ax - z) \tag{30}$$

where z is a vector of slack variables and μ is a vector of Lagrangian multipliers or shadow prices on the resources. The first order conditions

$$\frac{\partial}{\partial x_r} L_{\text{network}}(x, z; \mu) = \frac{w_r}{x_r} - \sum_{j \in r} \mu_j$$

$$\frac{\partial}{\partial z_j} L_{\text{network}}(x, z; \mu) = -\mu_j$$

determine the optimal rate:

$$x_r = \frac{w_r}{\sum_{j \in r} \mu_j},$$

where $\mu \ge 0$, $Ax \le C$ and $\mu^T(C - Ax) = 0$. Therefore the maximized value of the Lagrangian is

$$\max_{x,z \ge 0} L_{\text{network}}(x, z; \mu) = \sum_{r \in R} w_r \log \frac{w_r}{\sum_{j \in r} \mu_j} + \mu^T C,$$

from where we infer the dual to the NETWORK problem:

$$\begin{array}{ll} \text{DUAL} & \text{NETWORK } (A, C; w) \\ \max & \mathscr{U}(\mu) = \sum_{r \in R} w_r \log \sum_{j \in r} \mu_j - \sum_{j \in J} \mu_j C_j \\ \text{over} & \mu \ge 0. \end{array} \tag{31}$$

4.3.2 Congestion Control

The NETWORK and DUAL NETWORK problems above give a way of fairly ([Kel00]) allocating the network resources to all users based on user needs and network capacities. However, a centralised solution may be difficult and is undesirable. A decentralised algorithm for the DUAL NETWORK problem is given by the first-order system (slightly modified from the *dual algorithm* in [KMT98], to exclude stochastic perturbations)

$$\frac{d}{dt}\mu_j(t) = \kappa_j(\sum_{s:j\in s} x_s(t) - c_j) \tag{32}$$

$$x_r(t) = \frac{w_r}{\sum_{i\in r}\mu_i(t)} \tag{33}$$

where κ_j are positive constants to regulate the rate of convergence. This algorithm converges to a unique stable equilibrium point as established in the following theorem.

Theorem 4.4 *The function* $\mathscr{U}(\mu) = \sum_{s\in R} w_s \log(\sum_{i\in s}\mu_i) - \sum_{j\in J} c_j\mu_j$ *is a Lyapunov function for the system of equations (32)-(33).*

Proof: The function $\mathscr{U}$ is strictly concave on $\mu \geq 0$. This, together with the assumptions on $w_s > 0$ and $c_j > 0$ ensures that it has a unique interior maximum on the state space. The partial derivatives are

$$\frac{\partial}{\partial\mu_j}\mathscr{U}(\mu) = \sum_{s:j\in s}\frac{w_s}{\sum_{i\in s}\mu_i} - c_j \tag{34}$$

hence

$$\begin{aligned}
\tfrac{d}{dt}\mathscr{U}(\mu(t)) &= \sum_{j\in J}\frac{\partial\mathscr{U}}{\partial\mu_j}\frac{d\mu_j}{dt}\\
&= \sum_{j\in J}(\sum_{s:j\in s}\frac{w_s}{\sum_{i\in s}\mu_i} - c_j)\kappa_j(\sum_{s:j\in s}x_s(t) - c_j)\\
&= \sum_{j\in J}\kappa_j(\sum_{s:j\in s}\frac{w_s}{\sum_{i\in s}\mu_i} - c_j)^2.
\end{aligned}$$

So $\frac{d}{dt}\mathscr{U}(\mu(t)) \geq 0$, with equality at the unique maximizing point $\bar{\mu}$ attained by setting the partial derivatives (34) to 0. Hence $\mathscr{U}(\mu(t))$ is strictly increasing with time except at $\bar{\mu}$. Thus $\mathscr{U}(\mu)$ is a Lyapunov function for the system (32)-(33), which concludes the proof. □

Note, the $\mathscr{U}(\mu)$ function from Theorem 4.4 is precisely the same as the function we wanted to maximize in the DUAL NETWORK problem (31). This is not surprising from the duality between the gradient method and Lyapunov functions which we established in Section 4.2. Indeed, the algorithm given by the system (32)-(33) is the gradient method applied to the function $\mathscr{U}(\mu)$ which we want to maximize in the DUAL NETWORK problem. While this algorithm is essentially a congestion control algorithm, by solving DUAL NETWORK it is also a fair resource allocation algorithm. Thus we see that the problems of congestion control and decentralized resource allocation coincide.

5 Conclusion

The recent boom in Internet Congestion Control has provided a bridge between two mature fields, Optimization and Control theory. Although the concepts and tools used to study Internet Congestion Control are not novel and to a large extent this new area of research can be viewed as an instance of the general area of Resource Allocation in Economics, there is a novel link which necessarily emerges from the requirement that the algorithms searching for optimal allocations of flow rates must be decentralized and stable.

Remarkably, the idea that a complex multidimensional optimization problem may be solved more easily if it is decentralized, dates back at least to 1950-ies studies of Resource Allocation for Economic policy ([AHU58]). In a sense, it was coincidentally discovered that the gradient method—the main computational mechanism to achieve optimum allocation of scarce resources, was offering "for free" a decentralized solution. In those early Economic studies emphasis was placed on theoretical solutions and not on practical implementation, and so the decentralized solution was perhaps not so highly appreciated and the problems remained strictly within the Optimization area of research.

The decentralized feature of the gradient method however has come as a blessing in today's pressing problems on fair allocation of network flow since the immensity of network data and its partial, decentralized knowledge makes a central computation impossible. Another lucky coincidence has been that the problems of network flow allocation and congestion control have turned out equivalent. The fast and stable algorithms required to avoid congestion are essentially gradient algorithms which solve the network optimization problems. This essay has provided an insight as to why these lucky coincidences as not so coincidental, and the key to this insight is the discovered duality between optimization and control.

6 Acknowledgements

I had dwelt for a while on the duality between the convergence of the gradient method and the stability provided by Lyapunov functions, however it first stroke me explicitly when Dimitri Bertsekas put it in simple terms that, not surprisingly, the convergence of trajectories (under the gradient method) is in essence the same concept as stability of dynamical systems. I thank him for this insight. I am grateful to Asuman Ozdaglar for helpful comments and suggestions for further development. My special thanks goes to Frank Kelly as the first person who explained to me how to initiate research; I thank him for his patience and continual encouragement in pursuing my ideas.

References

[AHU58] K.J. Arrow, L. Hurwicz and H. Uzawa (1958), *Studies in Linear and Non-linear Programming.* Stanford University Press, Stanford, California.

[Ber99] D.P. Bertsekas (1999), *Non-linear Programming.* Athena Scientific, Belmont, MA.

[Dra92] P. G. Drazin (1992), *Nonlinear systems.* Cambridge University Press.

[Gle94] P. Glendinning (1994), *Stability, instability and chaos: an introduction to the theory of nonlinear differential equations.* Cambridge University Press.

[Jac74] O.L.R. Jacobs (1974), *Introduction to Control Theory.* Clarendon Press, Oxford.

[Jac88] V. Jacobson (1988), *Congestion avoidance and control.* In Proc. ACM SIGCOMM '88, pp. 314-329.

[JoT01] R. Johari and D. Tan (2001), *End-to-end Congestion Control for the Internet: Delays and Stability.* IEEE/ACM Transactions on Networking 9, 818-832.

[Kel97] F. Kelly (1997), *Charging and Rate Control for Elastic Traffic.* Eur Trans on Telecommun 8, 33-37.

[Kel00] F. Kelly (2000), *Models for a self-managed Internet.* Philosophical Transactions of the Royal Society A358, 2335-2348. http://www.statslab.cam.ac.uk/ frank/smi.html.

[KMT98] F.P. Kelly, A.K. Maulloo and D.K.H. Tan (1998), *Rate control in communication networks: shadow prices, proportional fairness and stability.* Journal of the Operational Research Society 49, 237-252.

[LPD01] S.H. Low, F. Paganini and J.C. Doyle (2001), *Internet Congestion Control.* IEEE Control Systems Magazine, Feb. 2002.

[Mas02] L. Massoulie (2002), *Stability of distributed congestion control with heterogeneous feedback delays.* IEEE Transactions on Automatic Control 47, 895-902.

[Vin00] G. Vinnicombe (2000), *On the stability of end-to-end congestion control for the Internet.* http://www-control.eng.cam.ac.uk/gv/internet.