Giancarlo Bigi, Marco Castellani,
Massimo Pappalardo, Mauro Passacantando

Nonlinear programming techniques for equilibria

Springer

Preface

In science the term “equilibrium” has been widely used in physics, chemistry, biology, engineering and economics, among others, within different frameworks. It generally refers to conditions or states of a system in which all competing influences are balanced. For instance, in physics the *mechanical equilibrium* is the state in which the sum of all the forces and torques on each particle of the system is zero, while a fluid is said to be in *hydrostatic equilibrium* when it is at rest, or when the flow velocity at each point is constant over time. In chemistry the *dynamic equilibrium* is the state of a reversible reaction where the forward reaction rate is equal to the reverse one. In biology the *genetic equilibrium* denotes a situation in which a genotype does not evolve any more in a population from generation to generation. In engineering the *traffic equilibrium* is the expected steady distribution of traffic over public roads or over computer and telecommunication networks. Even more, the well-known *equilibrium theory* is a fundamental branch of economics studying the dynamics of supply, demand, and prices in an economy within either one (*partial equilibrium*) or several (*general equilibrium*) markets: the basic model of supply and demand is an example of the former while the Arrow-Debreu and Radner models are examples of the latter.

Actually, the term “equilibrium” has always been very relevant also in mathematics, particularly in dynamical systems, partial differential equations and calculus of variations. After the breakthrough of game theory and the concept of Nash equilibrium, the term has been used in mathematics in much larger contexts involving relevant aspects of operations research and mathematical programming. Indeed, many “equilibrium problems”, including some of the above mentioned, can be modelled in this framework through different mathematical models such as optimization, complementarity, variational inequalities, multiobjective optimization, noncooperative games and inverse optimization among others. All these mathematical models share an underlying common structure that allows to conveniently formulate them in a unique format.

This book focuses on the analysis of this unifying format for equilibrium problems. Since it allows describing a large number of applications, many researchers devoted their efforts to study it and nowadays many results and algorithms are available: as optimization fits in this format, nonlinear programming techniques have often been the key tool of their work. The book aims at addressing in particular two core issues such as the existence and computation of equilibria. The first chapter illustrates a sample of applications, the second addresses the main theoretical issues, while the third introduces the main algorithms available for computing equilibria. A final chapter is devoted to quasi-equilibria, a more general format that is needed to cover more complex applications having additional features such as shared resources in noncooperative games. Finally, basic material on sets, functions and multivalued maps that are exploited throughout the book are summarized in the appendix. To make the book as readable as possible, examples and applications have been included. We hope that this book may serve as a basis for a second level academic course or a specialised course in a Ph.D. programme and stimulate further interest in equilibrium problems.

Pisa, Italy, Giancarlo Bigi
L'Aquila, Italy Marco Castellani
Pisa, Italy Massimo Pappalardo
Pisa, Italy Mauro Passacantando
July 2018

Contents

Chapter 1
Equilibrium models and applications

As already mentioned in the preface, the term “equilibrium” is widespread in science in the study of different phenomena. In this chapter a small selection of equilibrium problems from different areas is given, each leading to a different kind of mathematical model. The equilibrium position of an elastic string in presence of an obstacle, which is depicted in Section 1.1, coincides with the solution of a complementarity problem, the Nash equilibrium fits in well to model a power control multi-agent system described in Section 1.2, the steady distribution of traffic over a network is represented by a variational inequality in Section 1.3, the Markowitz portfolio theory is viewed as a multiobjective problem in Section 1.4, the shadow price theory is viewed as a saddle point problem for the nonlinear case in Section 1.5, the solution of the input-output model given in Section 1.6 is a fixed point and the quality control problem in a production system illustrated in Section 1.7 is an inverse optimization problem. Finally, the last section is devoted to show that all these mathematical models, which are apparently different, have a common structure that leads to a unified format: the Ky Fan inequality or the “equilibrium problem” using the “abstract” name introduced by Blum, Muu and Oettli to stress this unifying feature.

1.1 Obstacle problem

Consider an elastic one-dimensional string bounded on a plane. The endpoints of the string are kept fixed, and its natural position is a straight line between the endpoints. What shape does the string take at the equilibrium if an obstacle is inserted in between the two endpoints? The string stretches over the obstacle: it sticks to the obstacle somewhere while it remains stretched as a straight line elsewhere (see Figure 1.1).

Mathematically, the equilibrium position of the string may be described by a function $u : [0,1] \to \mathbb{R}$ with $u(0) = u(1) = 0$, with 0 and 1 denoting the x-coordinates of the two fixed endpoints. Analogously, the profile of the obstacle may be described

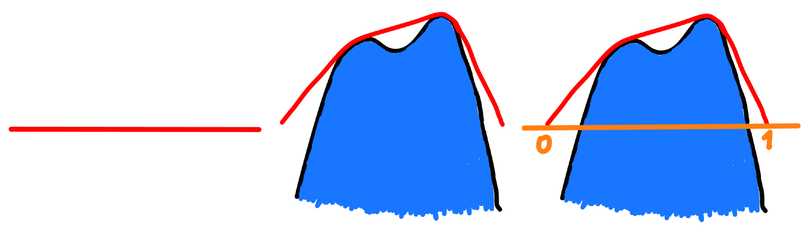

Fig. 1.1 The obstacle problem.

by a function $f : [0,1] \to \mathbb{R}$ with $f(0) < 0$ and $f(1) < 0$. At any point $x \in [0,1]$ the string is above the obstacle, that is $u(x) \geq f(x)$. Moreover, the string and the obstacle touch, that is $u(x) = f(x)$, or the string is a straight line near x, that is $u''(x) = 0$. Therefore, if f has second order derivatives, then the shape of the string u has second order derivatives as well. Moreover, notice that the absence of any other force beyond the pressure of the obstacle guarantees that u is a concave function, which means $u''(x) \leq 0$ for any $x \in (0,1)$. Summarising, the equilibrium position of the string is given by any u satisfying the conditions

$$\begin{cases} u(0) = u(1) = 0, \\ u(x) \geq f(x),\ u''(x) \leq 0 \qquad x \in (0,1), \\ (u(x) - f(x))u''(x) = 0 \qquad x \in (0,1). \end{cases} \tag{1.1}$$

This functional system can be turned into a system of inequalities and equalities in finite dimension through piecewise linear approximations of u and f given on a finite grid of points. Specifically, fix any $n \in \mathbb{N}$ and consider $t_i = i/(n+1)$, $u_i = u(t_i)$ and $f_i = f(t_i)$ for $i = 0, 1, \dots, n+1$. Exploiting finite differences, the second order derivative of u at t_i can be approximated through

$$u''(t_i) \approx (n+1)^2\,(u_{i+1} - 2u_i + u_{i-1}), \qquad i = 1, \dots, n.$$

Therefore, system (1.1) can be approximated by the following system

$$\begin{cases} u_0 = u_{n+1} = 0, \\ u_i - f_i \geq 0,\ u_{i+1} - 2u_i + u_{i-1} \leq 0 \qquad i = 1, \dots, n, \\ (u_i - f_i)(u_{i+1} - 2u_i + u_{i-1}) = 0 \qquad i = 1, \dots, n, \end{cases} \tag{1.2}$$

in the n variables $u_1, \dots, u_n$. In turn, (1.2) can be written in a compact form introducing the $n \times n$ tridiagonal matrix

$$M = \begin{bmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix}$$

and the vector $q \in \mathbb{R}^n$ with $q_i = 2f_i - f_{i+1} - f_{i-1}$. Indeed, $\bar{u} \in \mathbb{R}^n$ satisfies (1.2) if and only if $\bar{x} = \bar{u} - f$ satisfies

$$\bar{x} \geq 0, \qquad M\bar{x} + q \geq 0, \qquad \langle \bar{x}, M\bar{x} + q \rangle = 0, \tag{1.3}$$

where the inequality $\geq$ is meant componentwise. Notice that the nonnegative conditions imply either $\bar{x}_i = 0$ or $(M\bar{x} + q)_i = 0$ for $\langle \bar{x}, M\bar{x} + q \rangle = 0$ to hold. Systems like (1.3) are known as (linear) *complementarity problems* since they require that the product of nonnegative quantities should be zero.

1.2 Power control in wireless communications

A cellular network is designed to provide several users with access to wireless services over a large area that is divided into smaller areas called cells: each of them represents the area covered by a single base station which is often located at the center of the cell. In a code-division multiple-access (CDMA) system, mobile users operate using the same frequency and they need to adjust their transmit power to ensure a good performance (e.g., in terms of quality of service) by controlling the interference while minimizing the overall cost at the same time.

For the sake of simplicity, consider a single-cell CDMA system with N mobile users. Each user i has to select a value for the uplink transmit power $x_i \geq 0$ in order to minimize its own cost function

$$c_i(x_i, x_{-i}) = \lambda_i x_i - \alpha_i \log(1 + \gamma_i(x)), \tag{1.4}$$

where $x_{-i} = (x_j)_{j \neq i}$ is the transmit power of all users except i and γ_i is the Signal-to-Interference-plus-Noise Ratio (SINR) function

$$\gamma_i(x) = \frac{W}{R} \frac{h_i x_i}{\sum\limits_{j \neq i} h_j x_j + \sigma^2},$$

where W is the chip rate, R is the total rate, $h_j \in (0,1)$ is the channel gain from user j to the base station in the cell and σ^2 is the Gaussian noise variance. The objective function of each user is the difference between a pricing function (that assigns a price λ_i per power unit) and the gain obtained from a better SINR (α_i is called benefit parameter and represents the desired level of SINR). Therefore, an increase of the power level on one hand implies a benefit in terms of interference and on another hand a price in terms of the power consumed.

In this framework a vector of power levels chosen by the mobile users provides an equilibrium state if no user can decrease its own cost function by changing its power level unilaterally. More generally, a situation in which several selfish decision makers interact each other is known as *noncooperative game* and the equilibrium concept given above as *Nash equilibrium*.

1.3 Traffic network

Consider a traffic network represented by a set of nodes N, a set of arcs $A \subseteq N \times N$ and a set OD of pairs of nodes that represent the origin and the destination of paths. For each pair $s \in OD$ there is a known demand d_s representing the rate of traffic entering and exiting the network at the origin and the destination of s respectively. The demand d_s has to be distributed among a given set P_s of paths connecting the pair s and let x_p denote the portion of d_s routed on path p. Let P be the set of all the n paths, i.e., the union of all the sets P_s over all $s \in OD$, and $x = (x_p)_{p \in P}$ the vector of all path flows. The set of feasible path flows is hence given by

$$C = \left\{ x \in \mathbb{R}^n_+ : \quad \sum_{p \in P_s} x_p = d_s, \quad \forall\, s \in OD \right\}. \tag{1.5}$$

Since the flow z_a on each arc a is the sum of all flows on paths to which the arc belongs, the arc flow vector $z = (z_a)_{a \in A}$ is given by $z = \Delta\, x$, where Δ is the arc-path incidence matrix:

$$\Delta_{a,p} = \begin{cases} 1 \text{ if arc } a \in p, \\ 0 \text{ otherwise.} \end{cases}$$

A nonnegative arc cost function $t_a(z)$, which represents the travel time in traversing arc a and depends upon the whole arc flow vector z, is given for each arc a. Assuming that the path cost function is additive, the travel time $T_p(x)$ on path p is equal to the sum of the travel times on all the arcs of path p, that is

$$T_p(x) = \sum_{a \in p} t_a(\Delta\, x).$$

Therefore, the path cost map is $T : \mathbb{R}^n \to \mathbb{R}^n$ with $T(x) = \Delta^T t(\Delta\, x)$.

According to the *Wardrop equilibrium principle*, a vector $\bar{x} \in C$ is an equilibrium flow if it is positive only on paths with minimum cost, i.e., the following implication

$$\bar{x}_p > 0 \implies T_p(\bar{x}) = \min_{q \in P_s} T_q(\bar{x})$$

holds for any $s \in OD$ and $p \in P_s$.

It is possible to prove that a path flow $\bar{x} \in C$ is a Wardrop equilibrium if and only if the inequality

$$\langle T(\bar{x}), y - \bar{x} \rangle \geq 0, \qquad \forall\, y \in C, \tag{1.6}$$

holds. In fact, setting $\widetilde{T}_s = \min_{p\in P_s} T_p(\bar{x})$ for any $s \in OD$, if $\bar{x}$ is a Wardrop equilibrium, then any $y \in C$ satisfies

$$\begin{aligned}
\langle T(\bar{x}), y-\bar{x}\rangle &= \sum_{s\in OD}\sum_{p\in P_s} T_p(\bar{x})(y_p-\bar{x}_p) \\
&= \sum_{s\in OD}\left[\sum_{p\in P_s:\ \bar{x}_p>0} T_p(\bar{x})(y_p-\bar{x}_p) + \sum_{p\in P_s:\ \bar{x}_p=0} T_p(\bar{x})(y_p-\bar{x}_p)\right] \\
&= \sum_{s\in OD}\left[\sum_{p\in P_s:\ \bar{x}_p>0} \widetilde{T}_s(y_p-\bar{x}_p) + \sum_{p\in P_s:\ \bar{x}_p=0} T_p(\bar{x})y_p\right] \\
&\geq \sum_{s\in OD}\left[\sum_{p\in P_s:\ \bar{x}_p>0} \widetilde{T}_s(y_p-\bar{x}_p) + \sum_{p\in P_s:\ \bar{x}_p=0} \widetilde{T}_s y_p\right] \\
&= \sum_{s\in OD}\widetilde{T}_s\sum_{p\in P_s}(y_p-\bar{x}_p) \\
&= \sum_{s\in OD}\widetilde{T}_s(d_s-d_s)=0,
\end{aligned}$$

where the third equality follows from the definition of Wardrop equilibrium, the inequality from the definition of $\widetilde{T}_s$ and the fifth equality from the feasibility of y and $\bar{x}$. Thus, inequality (1.6) is satisfied. Conversely, if inequality (1.6) holds, then consider an arbitrary pair $s \in OD$, two paths $p, q \in P_s$ with $\bar{x}_p > 0$ and the path flow y defined as follows:

$$y_r = \begin{cases} \bar{x}_r & \text{if } r \neq p, q, \\ 0 & \text{if } r = p, \\ \bar{x}_p + \bar{x}_q & \text{if } r = q. \end{cases}$$

Then, it is clear that $y \in C$ and

$$0 \leq \langle T(\bar{x}), y-\bar{x}\rangle = T_p(\bar{x})(y_p-\bar{x}_p) + T_q(\bar{x})(y_q-\bar{x}_q) = \bar{x}_p(T_q(\bar{x})-T_p(\bar{x})),$$

hence $T_q(\bar{x}) \geq T_p(\bar{x})$. Since path $q \in P_s$ is arbitrary, $T_p(\bar{x}) = \min_{q\in P_s} T_q(\bar{x})$ holds and hence $\bar{x}$ is a Wardrop equilibrium.

Inequalities like (1.6) are known as *variational inequalities*.

1.4 Portfolio selection

Suppose there are n risky assets, where asset i gives the random return R_i. Recall that the return R of an asset is simply the percentage change in the value from one time to another; more precisely, the return at time t is defined by

$$R = \frac{V_t - V_{t-1}}{V_{t-1}},$$

where V_t is the total value of the asset at time t and V_{t-1} is the total value at an earlier time $t-1$. For the sake of simplicity, assume that R_i are n jointly distributed random variables with finite second moment.

Let M be a given sum of money to be invested in the n different assets and let x_i denote the amount to be allocated to the asset i. The vector $x \in \mathbb{R}^n$ is called a *portfolio* if

$$x_1 + x_2 + \cdots + x_n \leq M.$$

Notice that the non-negativeness of the components x_i could not be required. A negative x_i represents a *short position* for the risky asset i: a short position is an investment strategy where the investor sells shares of borrowed stock in the open market. The expectation of the investor is that the price of the stock will decrease over time, at which point he will purchase the shares in the open market and return the shares to the broker which he borrowed them from.

Fixed a portfolio x, the net profit is described by the random variable

$$R = x_1 R_1 + x_2 R_2 + \cdots + x_n R_n.$$

A first simple method for establishing the goodness of the portfolio could be to maximize the expected return

$$g_1(x) = \mathrm{E}[x_1 R_1 + x_2 R_2 + \ldots + x_n R_n] = x_1 \mu_1 + x_2 \mu_2 + \ldots + x_n \mu_n,$$

where $\mu_i = \mathrm{E}[R_i]$, that is solving the following linear programming problem

$$\begin{cases} \max \ \mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n \\ x_1 + x_2 + \ldots + x_n \leq M \\ x_i \geq 0, \quad i = 1, \ldots, n. \end{cases}$$

The constraints $x_i \geq 0$ mean that only long positions are allowed, hence these conditions should be omitted to include possible short sales in the model.

However, this approach does not provide meaningful results. Indeed, if there exists an index j such that $\mu_j > 0$ and $\mu_j > \mu_i$ for each $i \neq j$, then the problem has a unique optimal solution $\bar{x}$ with $\bar{x}_j = M$ and all the other components equal to zero, which represent an undiversified portfolio and it could be very risky: don't put all your eggs in one basket!

Risk aversion is the behaviour of investors, when exposed to uncertainty, to prefer a bargain with a more certain, but possibly lower, expected payoff rather than another bargain with an uncertain payoff. One simple measure of financial risk is the variance of the net profit, i.e.

$$g_2(x) = \mathrm{Var}(x_1 R_1 + x_2 R_2 + \ldots + x_n R_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j,$$

where

$$\sigma_{ij} = \mathrm{Cov}(R_i, R_j) = \mathrm{E}[R_i R_j] - \mathrm{E}[R_i]\,\mathrm{E}[R_j]$$

is the covariance between R_i and R_j.

The investor would like to maximize the expected return g_1 and to minimize the risk measured by variance g_2. Since these two objectives are typically conflicting, more reasonable solutions are available by considering a trade-off between risk and return according to the Markowitz theory. One simple way to perform it is to consider portfolios such that no other portfolio provides a larger return paired with a lower risk. Formally, a portfolio x is *strictly dominated* by another portfolio x' if $g_1(x') > g_1(x)$ and $g_2(x') < g_2(x)$. Hence, the corresponding selection problem consists in finding a portfolio $\bar{x} \in \mathbb{R}^n_+$ which is not strictly dominated by any other portfolio. Such portfolios are called weak Pareto optimal.

1.5 Optimal production under restricted resources

A company produces a mix of n commodities aiming at maximizing the profit while also evaluating further investments in the m raw materials (or resources) that are needed. The profit is given by a function $f : \mathbb{R}^n \to \mathbb{R}$, where each variable x_j represents the quantity of commodity j that is produced. The use of each resource i depends on the total production $x = (x_1, \ldots, x_n)$ and is given by the function $g_i : \mathbb{R}^n \to \mathbb{R}$. Therefore, the maximization of the profit amounts to maximizing $f(x)$ subject to the constraints $g_i(x) \leq c_i$, where c_i denotes the available quantity of resource i.

The tool to evaluate further investments in resources are the so-called shadow prices, that is the marginal changes of the value of the optimal profit as the quantities c_i vary. According to the theory of Lagrange multipliers, under a standard constraint qualification, if a mix $\bar{x}$ is optimal then there exists a nonnegative vector of shadow prices $\bar{\lambda}$ satisfying the following conditions:

$$\sum_{i=1}^{m} \bar{\lambda}_i \frac{\partial g_i}{\partial x_j}(\bar{x}) \geq \frac{\partial f}{\partial x_j}(\bar{x}), \qquad j = 1, \ldots, n, \tag{1.7}$$

$$\bar{\lambda}_i (c_i - g_i(\bar{x})) = 0, \qquad i = 1, \ldots, m, \tag{1.8}$$

$$\bar{x}_j \left[\frac{\partial f}{\partial x_j}(\bar{x}) - \sum_{i=1}^{m} \bar{\lambda}_i \frac{\partial g_i}{\partial x_j}(\bar{x}) \right] = 0, \qquad j = 1, \ldots, n. \tag{1.9}$$

They yield the following economic interpretation: the left hand side of (1.7) is the marginal value of the amount of the resources needed to increase the production of commodity j while the right hand side is the marginal profit due to the increase; (1.8) means that if the resource i is not used completely then it must be free, i.e., a resource with zero shadow price, while if it is positive, then all of the available supply must be fully used; (1.9) asserts that if commodity j is produced, i.e., $\bar{x}_j > 0$, then the marginal quantities in (1.7) must be equal, while if they are not equal then commodity j cannot be produced.

If f is concave and g_i are convex, then the existence of such shadow prices $\bar{\lambda}$ is also a sufficient optimality condition for the mix $\bar{x}$. Moreover, the couple $(\bar{x}, \bar{\lambda})$ satisfies the conditions (1.7)-(1.9) if and only if it is a saddle point of the function

$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i (c_i - g_i(x)), \tag{1.10}$$

which is called the Lagrangian function; this means

$$L(x,\bar{\lambda}) \leq L(\bar{x},\bar{\lambda}) \leq L(\bar{x},\lambda), \ \ \forall\, x, \lambda \geq 0,$$

or, equivalently, $\bar{x}$ maximizes $L(\cdot,\bar{\lambda})$ and $\bar{\lambda}$ minimizes $L(\bar{x},\cdot)$.

1.6 Input-output analysis in an economy

Input-output mathematical models for the economy of a country are based on its disaggregation into sectors. Suppose the economy consists of n interdependent sectors (or industries) $S_1, \ldots, S_n$ each of which produces a single kind of good that is traded, consumed and invested within the same economy. Let x_i denote the quantity of good i produced by the sector S_i. Each sector utilizes some of the goods produced by the other industries for the production of its own good. More precisely, suppose that the sector S_j must use y_{ij} units of the good i in order to produce x_j units of the good j. The proportionality is the main assumption of the original Leontief input-output model: the exploitation of the good i is directly proportional to the production of the good j. In other words, the ratio $a_{ij} = y_{ij}/x_j$, called input coefficient, is constant and represents the units of good i needed for producing one unit of good j. Clearly, all the coefficients a_{ij} are nonnegative and moreover the possibility that $a_{ii} > 0$ holds for some i is not ruled out (for instance, a power station may use some of its own electric power for the production). The quantity

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$$

is the *internal demand* of the good i. In addition to the internal demand, which models the flow of goods in between the industries, suppose the existence of other non-productive sectors of the economy (such as consumers and governments), that may be grouped into the so-called *open sector* not producing anything but consuming goods from all the sectors. Denote by d_i the demand of the open sector from the sector S_i, which is called *final demand*. Therefore, the total output of the sector S_i must be equal to the internal plus the final demand:

$$x_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + d_i.$$

The output levels required by all the n sectors in order to meet these demands are given by the system of n linear equations coupled with nonnegativity conditions

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + d_1 = x_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + d_2 = x_2 \\ \quad\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n + d_n = x_n \\ x_i \geq 0, \quad i = 1, \ldots, n \end{cases}$$

that can be written in matrix form as

$$\begin{cases} Ax + d = x \\ x \in \mathbb{R}^n_+ \end{cases} \tag{1.11}$$

where the *input-output matrix* $A = (a_{ij})$ describes the interdependence of the industries.

The linearity assumption on the relation between each x_j and the amount x_{ij} is a very strong assumption. The assumption of constant returns to scale is arguable on the grounds that functions more complex than simple proportions are needed to describe production processes realistically, particularly in industries where at least one large installation (such as railroad tracks, dams or telephone lines) must be provided before any output can be produced. For this reason some authors proposed a nonlinear input-output model replacing the linear production functions $y_{ij} = a_{ij}x_j$ with the nonlinear functions $y_{ij} = a_{ij}(x_j)$. The basic assumptions on the nonlinear functions are the following:

- $a_{ij}(\cdot)$ is defined and continuously differentiable on $\mathbb{R}_+$,
- $a_{ij}(0) = 0$,
- $a'_{ij}(t) \geq 0$ for all $t \geq 0$.

Therefore, given the map $A : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ whose component i is defined by $A_i(x) = a_{i1}(x_1) + a_{i2}(x_2) + \ldots + a_{in}(x_n)$, the solution of the nonlinear input-output model consists in finding $\bar{x} \in \mathbb{R}^n_+$ such that $A(\bar{x}) + d = \bar{x}$, that is $\bar{x}$ is a fixed point of the map $x \mapsto A(x) + d$.

1.7 Quality control in production systems

A manufacturer produces one commodity aiming at maximizing the profit while controlling the quality level of the production at its facility over a planning horizon of n time periods. Indeed, the quality level affects both the expected demand of the commodity and the cost of its production.

Let $x_i \in \mathbb{R}$ denote the quality level of the production during period i; it may be assumed $x_i \in [0,1]$ without any loss of generality. Suppose $f_i : \mathbb{R}_+ \to \mathbb{R}_+$ is the expected demand and $g_i : \mathbb{R}_+ \to \mathbb{R}_+$ the production cost as a function of the quality level x_i during period i. If $D > 0$ denotes the total budget for producing the commodity and c_i the unitary price for period i, the problem can be modeled as the following mathematical program

$$\max\left\{\sum_{i=1}^{n} c_i f_i(x_i) \;:\; \sum_{i=1}^{n} g_i(x_i) \leq D,\; \ell_i \leq x_i \leq u_i\right\} \tag{1.12}$$

for some given lower and upper bounds $0 \leq \ell_i \leq u_i \leq 1$ on the quality level. In some production systems $x_i \leq x_{i+1}$ (for $i = 1, \ldots, n-1$) are reasonable additional constraints: the quality level of the production does not get lower over time.

Actually, modifying the quality level may be a difficult task and may require a lot of time in some particular production systems. In such situations it may be convenient to adjust prices and it is clearly unreasonable to increase or decrease each price too much. Therefore, some $\bar{c} = (\bar{c}_1, \ldots, \bar{c}_n)$ sufficiently close to the current prices $c^* = (c_1^*, \ldots, c_n^*)$ is seeked such that the current quality levels $\bar{x} = (\bar{x}_1, \ldots., \bar{x}_n)$ provide an optimal solution. The problem can be formally stated as follows:

$$\text{given a feasible solution } \bar{x} \text{ of (1.12)}, c^* \in \mathbb{R}^n_+ \text{ and } \delta > 0,$$
$$\text{find } \; \bar{c} \in \mathbb{R}^n_+ \; \text{ s.t. } \; \|\bar{c} - c^*\|_\infty \leq \delta \text{ and } \bar{x} \text{ is an optimal solution of (1.12).}$$

Problems like the above one are called *inverse optimization problems* since they aim at determining whether a given feasible point can be made optimal by adjusting the values of some parameters within a given range.

1.8 Ky Fan inequalities: a unifying equilibrium model

In this section all the mathematical models of the problems described in the previous sections are recast as particular cases of a *Ky Fan inequality*, that is the following mathematical equilibrium model

$$\text{find } \bar{x} \in C \text{ such that } f(\bar{x}, y) \geq 0 \text{ for all } y \in C, \tag{EP(f,C)}$$

where $C \subseteq \mathbb{R}^n$ is a nonempty closed set and $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is an equilibrium bifunction, i.e., $f(x,x) = 0$ for all $x \in C$. Precisely, the section aims at showing how complementarity problems, Nash equilibrium problems, variational inequalities, weak Pareto optimization problems, saddle point problems, fixed point problems and inverse optimization problems can be all formulated in the above format through suitable choices of f and C.

Complementarity problems

Given a closed convex cone $C \subseteq \mathbb{R}^n$ and a map $F : \mathbb{R}^n \to \mathbb{R}^n$, the complementarity problem asks to

$$\text{find } \bar{x} \in C \text{ such that } F(\bar{x}) \in C^* \text{ and } \langle \bar{x}, F(\bar{x}) \rangle = 0, \tag{1.13}$$

where $C^* = \{y \in \mathbb{R}^n : \langle y, x\rangle \geq 0, \ \forall x \in C\}$ is the dual cone of C. Notice that the complementarity problem (1.3) described in Section 1.1 is a special case of (1.13) with $C = \mathbb{R}^n_+$ and $F(x) = Mx + q$.

Solving the complementarity problem amounts to solving EP(f,C) with

$$f(x,y) = \langle F(x), y - x\rangle.$$

Indeed, if $\bar{x}$ solves the complementarity problem, then

$$f(\bar{x},y) = \langle F(\bar{x}), y - \bar{x}\rangle = \langle F(\bar{x}), y\rangle \geq 0, \qquad \forall\, y \in C,$$

that is $\bar{x}$ solves (EP). Conversely, if $\bar{x}$ solves EP(f,C), then choosing $y = 2\bar{x}$ and $y = 0$ provides $\langle \bar{x}, F(\bar{x})\rangle = 0$ and thus $\langle F(\bar{x}), y\rangle = f(\bar{x},y) \geq 0$ holds for all $y \in C$, that is $\bar{x}$ is a solution of the complementarity problem.

Note that the system of equations $F(x) = 0$ is a special complementarity problem with $C = \mathbb{R}^n$.

Nash equilibrium problems

In a noncooperative game with N players, each player i has a set of possible strategies $C_i \subseteq \mathbb{R}^{n_i}$ and aims at minimizing a cost function $c_i : C \to \mathbb{R}$ with $C = C_1 \times \cdots \times C_N$. A Nash equilibrium is any $\bar{x} \in C$ such that no player can reduce its cost by unilaterally changing its strategy, that is any $\bar{x} \in C$ such that

$$c_i(\bar{x}_i, \bar{x}_{-i}) \leq c_i(y_i, \bar{x}_{-i})$$

holds for any $y_i \in C_i$ and any $i = 1, \ldots, N$, where $\bar{x}_{-i}$ denotes the vector of strategies of all players except i. Finding a Nash equilibrium amounts to solving EP(f,C) with the so-called Nikaido-Isoda bifunction, i.e.,

$$f(x,y) = \sum_{i=1}^{N} \left[c_i(y_i, x_{-i}) - c_i(x_i, x_{-i})\right]. \tag{1.14}$$

Indeed, if $\bar{x}$ is a Nash equilibrium, all the terms in (1.14) are nonnegative for any $y \in C$ and hence $\bar{x}$ solves the equilibrium problem. Conversely, let $\bar{x}$ be a solution of EP(f,C) and assume, by contradiction, there exist an index i and a strategy $y_i \in C_i$ such that $c_i(\bar{x}_i, \bar{x}_{-i}) > c_i(y_i, \bar{x}_{-i})$. Choosing $y_j = \bar{x}_j$ for all $j \neq i$ leads to the contradiction

$$f(\bar{x},y) = c_i(y_i, \bar{x}_{-i}) - c_i(\bar{x}_i, \bar{x}_{-i}) < 0.$$

The power control game described in Section 1.2 is a Nash equilibrium problem where the strategy sets are $C_i = [0, +\infty)$ and the cost functions are defined in (1.4).

Variational inequality problems

Given a closed convex set $C \subseteq \mathbb{R}^n$ and a map $F : \mathbb{R}^n \to \mathbb{R}^n$, the variational inequality problem asks to

$$\text{find } \bar{x} \in C \text{ such that } \langle F(\bar{x}), y - \bar{x} \rangle \geq 0 \ \text{ for all } y \in C. \tag{1.15}$$

Solving this problem amounts to solving EP(f,C) with

$$f(x,y) = \langle F(x), y - x \rangle.$$

Notice that the variational inequality (1.6) which models the traffic equilibrium problem in Section 1.3 is a special case of (1.15) with F equal to the path cost map T.

More general formats of variational inequalities are included in the EP(f,C) format. For instance, if $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a multivalued map with compact values, then

$$\text{find } \bar{x} \in C \text{ and } \bar{u} \in F(\bar{x}) \text{ such that } \langle \bar{u}, y - \bar{x} \rangle \geq 0 \text{ for all } y \in C,$$

amounts to solving EP(f,C) with

$$f(x,y) = \max_{u \in F(x)} \langle u, y - x \rangle.$$

Given two maps $F, G : \mathbb{R}^n \to \mathbb{R}^n$ and a function $h : \mathbb{R}^n \to (-\infty, +\infty]$, another kind of generalized variational inequality problem asks to

$$\text{find } \bar{x} \in \mathbb{R}^n \text{ such that } \langle F(\bar{x}), y - G(\bar{x}) \rangle + h(y) - h(G(\bar{x})) \geq 0 \text{ for all y} \in \mathbb{R}^n.$$

Solving this problem amounts to solving EP(f,C) with $C = \mathbb{R}^n$ and

$$f(x,y) = \langle F(x), y - G(x) \rangle + h(y) - h(G(x)).$$

Notice that the presence of G and h does not allow formulating this problem in the standard format (1.15).

Weak Pareto optimization problems

Given m real-valued functions $\psi_i : \mathbb{R}^n \to \mathbb{R}$, $\bar{x} \in X$ is called a weak Pareto global minimum of the vector function $\psi = (\psi_1, \ldots, \psi_m)$ over a set $X \subseteq \mathbb{R}^n$ if there exists no element $y \in X$ such that $\psi_i(y) < \psi_i(\bar{x})$ for all $i = 1, \ldots, m$. Note that in the portfolio selection problem proposed in Section 1.4, the objective functions are $\psi_1(x) = -g_1(x)$ and $\psi_2(x) = g_2(x)$ with $X = \{x \in \mathbb{R}^n_+ : \ x_1 + \cdots + x_n \leq M\}$.

Finding a weak Pareto global minimum amounts to solving EP(f,C) with $C = X$ and

$$f(x,y) = \max_{i=1,\dots,m} [\psi_i(y) - \psi_i(x)].$$

Indeed $f(\bar{x},y) \geq 0$ for any $y \in X$ if and only if for any $y \in X$ there exists an index $i = 1,\dots,m$ such that $\psi_i(y) - \psi_i(\bar{x}) \geq 0$, that is the definition of weak Pareto global minimum.

Saddle point problems

Given two sets $C_1 \subseteq \mathbb{R}^{n_1}$ and $C_2 \subseteq \mathbb{R}^{n_2}$, a saddle point of a function $H : C_1 \times C_2 \to \mathbb{R}$ is any $\bar{x} = (\bar{x}_1, \bar{x}_2) \in C_1 \times C_2$ such that

$$H(\bar{x}_1, y_2) \leq H(\bar{x}_1, \bar{x}_2) \leq H(y_1, \bar{x}_2)$$

holds for any $y = (y_1, y_2) \in C_1 \times C_2$. The production problem described in Section 1.5 is a saddle point problem where H is the opposite of the Lagrangian function (1.10).

Finding a saddle point of H amounts to solving EP(f,C) with $C = C_1 \times C_2$ and

$$f((x_1,x_2),(y_1,y_2)) = H(y_1,x_2) - H(x_1,y_2).$$

Indeed, a saddle point of H is a Nash equilibrium in a two-person zero-sum game, that is a noncooperative game where the cost function of the first player is H and the cost function of the second player is $-H$ (what one player wins is exactly what the other player loses).

Fixed point problems

Given a set $C \subseteq \mathbb{R}^n$, a fixed point of a map $F : C \to C$ is any $\bar{x} \in C$ such that $\bar{x} = F(\bar{x})$. For instance, the solution of the nonlinear input-output model described in Section 1.6 is a fixed point of the map $F(x) = A(x) + d$ over the set $C = \mathbb{R}^n_+$.

Finding a fixed point amounts to solving EP(f,C) with

$$f(x,y) = \langle x - F(x), y - x \rangle.$$

In fact, if $\bar{x}$ is a fixed point of F, then it obviously solves EP(f,C). Conversely, if $\bar{x}$ solves EP(f,C), then choosing $y = F(\bar{x}) \in C$ provides

$$0 \leq f(\bar{x}, F(\bar{x})) = -\|\bar{x} - F(\bar{x})\|^2,$$

hence $\bar{x} = F(\bar{x})$.

If the set C is also convex, a further equivalent reformulation is available, that is the fixed point problem amounts to solving EP(f,C) with

$$f(x,y) = \langle y - F(x), y - x \rangle.$$

In fact, if $\bar{x}$ is a fixed point of F, then $f(\bar{x},y) = \|y-\bar{x}\|^2 \geq 0$ for any $y \in C$, i.e., $\bar{x}$ solves EP(f,C). Vice versa, if $\bar{x}$ is a solution to EP(f,C), then the convexity of C guarantees that $y = (\bar{x}+F(\bar{x}))/2 \in C$ and hence

$$0 \leq f(\bar{x},y) = \left\langle \frac{\bar{x}-F(\bar{x})}{2}, \frac{F(\bar{x})-\bar{x}}{2} \right\rangle = -\frac{1}{4}\|\bar{x}-F(\bar{x})\|^2,$$

thus $\bar{x} = F(\bar{x})$. Notice that the bifunctions defined in the two different reformulations satisfy different requirements: for instance, the former f is linear with respect to y while the latter is a strongly convex quadratic function in y.

Moreover, the EP(f,C) format includes also the fixed point problem when the map is multivalued. Indeed if $F : C \rightrightarrows C$ is a multivalued map with compact values, then finding $\bar{x} \in C$ such that $\bar{x} \in F(\bar{x})$ amounts to solving EP(f,C) with

$$f(x,y) = \max_{u \in F(x)} \langle x-u, y-x \rangle.$$

Inverse optimization problems

Given two closed sets $B \subseteq \mathbb{R}^n$ and $C \subseteq \mathbb{R}^m_+$, m functions $f_i : \mathbb{R}^n \to \mathbb{R}$ and p functions $h_j : \mathbb{R}^n \to \mathbb{R}$, the inverse optimization problem consists in determining a parameter $\bar{c} \in C$ such that at least one optimal solution of the maximization problem

$$\max \left\{ \sum_{i=1}^{m} \bar{c}_i f_i(x) \ : \ x \in B \right\}$$

satisfies the constraints $h_j(x) \leq 0$ for all j. The problem of Section 1.7 can be formulated in this fashion where $m = n$, $p = 1$, B is the feasible region of (1.12), C is the set of vectors $c \in \mathbb{R}^n_+$ such that $\|c-c^*\|_\infty \leq \delta$ and $h_1(x) = \|x-\bar{x}\|$.

Actually, this inverse optimization problem is equivalent to a noncooperative game with three players. The first player controls the variable $x \in B$ and aims at solving

$$\max \left\{ \sum_{i=1}^{m} c_i f_i(x) \ : \ x \in B \right\};$$

the second player controls the auxiliary variable $y \in \mathbb{R}^p_+$ and aims at solving

$$\max \left\{ \sum_{j=1}^{p} h_j(x) y_j \ : \ y \geq 0 \right\};$$

the third player controls the parameter $c \in C$ and aims at maximizing constant objective function, in other words he simply chooses the parameter c. Therefore, also this inverse optimization problem can be formulated in the EP(f,C) format via the Nikaido-Isoda bifunction (1.14).

1.9 Notes and references

Almost every paper on Ky Fan inequalities states that this model provides a general format that subsumes many other models so that a very large number of applications may be formulated in a unique fashion. The aim of this chapter is to corroborate this statement through a small selection of applications and the corresponding mathematical models. Section 1.1 examines the problem of determining the shape of an elastic string stretched over a body creating an obstacle [37]. This infinite dimensional problem can be numerically approximated by a complementarity system through standard discretization techniques. Section 1.2 describes the celebrated traffic network problem [110], which is based on the Wardrop equilibrium principle [128]. Its reformulation as a variational inequality was established in [54, 122]. Section 1.3 deals with a basic version of power control problems in wireless communications [69] that can be viewed as a noncooperative game and therefore turned into a system of inequalities through the Nikaido-Isoda aggregate bifunction [106]. Pareto optimization is the core of Section 1.4, that describes a simple portfolio selection problem relying on Markowitz's original approach [90]. Section 1.5 explains how optimal investment of resources, shadow prices and production of commodities are mixed in a mathematical model [18] that actually amounts to finding a saddle point of a suitable Lagrangian function. Section 1.6 describes the Leontief input-output model [89], which is probably the most well-known static model of the structure of a national economy. Indeed, Leontief received the Nobel Price in Economics "for the development of the input-output method and for its application to important economic problems" in 1973. The nonlinear version provided at the end of the section has been developed in [117]. The input-output analysis leads to a fixed point problem. Finally, inverse optimization (see [12]) allows modeling the problem of quality control in production systems that is addressed in Section 1.7 [133].

The solution set of each of the above problems coincides with the solution set of a Ky Fan inequality built by choosing a suitable feasible set C and a suitable bifunction f. This unified mathematical format has been explicitly proposed in [38, 99], following in the footsteps of the minimax inequality by Ky Fan [66].