Parametric programming: An illustrative mini encyclopedia*

SANJO ZLOBEC†

Abstract. *Parametric programming is one of the broadest areas of applied mathematics. Practical problems, that can be described by parametric programming, were recorded in the rock art about thirty millennia ago. As a scientific discipline, parametric programming began emerging only in the 1950's. In this tutorial we introduce, briefly study, and illustrate some of the elementary notions of parametric programming. This is done using a limited theory (mainly for linear and convex models) and by means of examples, figures, and solved real-life case studies.*

Among the topics discussed are stable and unstable models, such as a projectile motion model (maximizing the range of a projectile), bilevel decision making models and von Stackelberg games of market economy, law of refraction and Snell's law for the ray of light, duality, Zermelo's navigation problems under the water, restructuring in a textile mill, ranking of efficient DMU (university libraries) in DEA, minimal resistance to a gas flow, and semi-abstract parametric programming models. Some numerical methods of input optimization are mentioned and several open problems are posed.

Key words: *parametric programming, optimal parameter, convex programming, point-to-set mapping, stable model, unstable model, input optimization, optimal control, duality*

AMS subject classifications: 90C25, 90C30, 90C31, 90C46, 90C90

Received December 12, 1999

The tutorial is organized by topics as follows:

*This paper is based on a tutorial given for students at the Sixth International Conference on Parametric Optimization and Related Topics, Dubrovnik, Croatia, October 4–8, 1999.

†Mc Gill University, Department of Mathematics and Statistics, Burnside Hall, 805 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6, e-mail: zlobec@math.mcgill.ca

1. Introduction

Applied mathematics uses *mathematical models* to describe, and possibly solve, real-life problems. If a model describes an optimization problem, then it is called a *mathematical programming model* or an *optimization model*. These models typically contain two types of variables: those that can be changed, controlled or influenced by the decision maker are called *parameters* (*inputs*, *stimuli*), the remaining ones are *decision variables* (*outputs*, *instruments*). We denote the parameters by $\theta \in \mathbb{R}^p$ and the decision variables by $x \in \mathbb{R}^n$.

The two kinds of variables are often related by a system of equations and inequalities such as $x \in F(\theta) = \{x \in \mathbb{R}^n : g^i(x,\theta) \leq 0, i \in I, h^j(x,\theta) = 0, j \in J\}$ where $g^i, h^j : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, $i \in I$, $j \in J$ are some functions, and I and J are finite index sets. Models of the form

$$\min_{(x)} \ f(x,\theta) \qquad \text{subject to} \quad x \in F(\theta)$$

where θ is allowed to vary over some set $\mathbf{F}$ in $\mathbb{R}^p$, are termed *parametric programming models*. *Parametric programming* is the study of such models. Their local analysis, around a fixed θ, is referred to as *sensitivity analysis*. The classical sensitivity analysis studies how the values of a function (say, the minimum value function) change with small perturbations of the argument or the parameter.

Since every equality constraint can be replaced by two inequalities, one can assume that $J = \emptyset$. Then the mathematical programming model (abbreviated: *model*) is said to be *linear* (resp. *convex*) if the constraints $f(\cdot,\theta)$, $g^i(\cdot,\theta) : \mathbb{R}^n \to \mathbb{R}$, $i \in I$ are linear (resp. convex) for every $\theta \in \mathbb{R}^p$. The set $F(\theta)$ is in the space of decision variables and it is called the *feasible set for a given parameter* θ. We will also talk about *programs*. In our terminology, a program is a special case of a model when the parameter θ *is fixed, i.e., when* θ *is not allowed to vary.*

We assume that one can distinguish between the two types of variables: parameters and decision variables. This may not always be easy. Often it is not clear which variables should be "parameters" and which ones are "decision variables". The rule of thumb is to choose those variables that one can change, control or influence as parameters.

2. History

Some of the oldest and most common practical problems can be formulated as parametric programming models. One of these is to determine the maximal range of a projectile (stone, snowball, shot ball, arrow, spear, bullet, ski jumper) that is launched into the air and is then allowed to move freely. Rock paintings, some possibly 30 millennia old, illustrate the problem. (See *Figure 1* reproduced from a cave in Spain, same in Becker [3].)

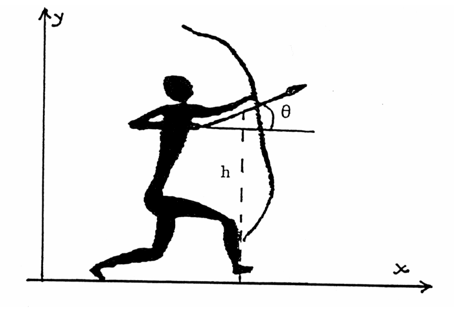

Figure 1. *Parametric programming in practice*

The motion of the projectile is complicated by air resistance and rotational and gravitational forces of the earth. If these are neglected, then the range depends on the initial speed v of the projectile and on an angle θ of the launch relative to the horizontal axis. Assuming that the initial speed is constant, the range still depends on θ.

A basic problem now is to *determine the angle of the launch that yields the maximal range*. It has taken many centuries to mathematically formulate and solve this problem. It appears that Galileo Galilei (1564–1642) was the first person who gave the modern, quantitative description of projectile motion. In particular, he was aware of the validity of treating the horizontal and vertical components of projectile's motion as independent motions. (See, e.g., Tipler [40, vol. 1, p.63] for a translation of Galileo's comment.)

Using this approach one can formulate the optimal range problem as a parametric programming model:

2.1. Projectile motion

A situation where a projectile (say, an arrow) is being launched into the air from an initial height of h meters with the initial velocity v and an angle of projection θ is depicted in *Figure 1.* The path of its centre of mass can be described in time t after applying the second law of motion of Isaac Newton (1642–1727) to the two components:

$$\frac{d^2x}{dt^2} = 0; \text{ no acceleration along the } x \text{ axis.}$$

$$\frac{d^2y}{dt^2} = -g; \text{ constant acceleration along the } y \text{ axis. (Here } g \approx 9.81\,\mathrm{m/s^2}.)$$

After solving these differential equations with the initial conditions, e.g., $x(0) = 0$, $y(0) = h$, and noting that the components of the initial velocity vector relative to the x and y axes are $v_x = v\cos\theta$ and $v_y = v\sin\theta$, the path of the projectile's centre of mass is described by

$$x = x(t) = vt\cos\theta$$
$$y = y(t) = -\tfrac{g}{2}t^2 + vt\sin\theta + h.$$

The angle θ can be considered as a parameter. Given θ, the projectile is at the level zero when $y(t) = 0$, i.e., the corresponding feasible set is

$$F(\theta) = \{t : -\frac{g}{2}t^2 + vt\sin\theta + h = 0\}.$$

It consists of the two roots of the equation $y(t) = 0$. The parametric programming model, describing the maximal range problem, can be formulated as

$$\begin{aligned} &\max_{(t)} f(t,\theta) = vt\cos\theta \qquad \text{subject to} \\ &-\tfrac{g}{2}t^2 + vt\sin\theta + h \le 0. \end{aligned} \tag{2.1}$$

Solution: In this situation we can determine the optimal parameter (angle). Indeed, for every θ, the optimal decision variable is the larger root $t = t^o(\theta) = \frac{2v}{g}\sin\theta$, when $h = 0$. Hence, after substitution in the objective function, the *optimal value function* is

$$f^o(\theta) = f(t^o(\theta), \theta) = vt^o(\theta)\cos\theta = \frac{v^2}{g}\sin 2\theta.$$

It assumes the maximal value when $\sin 2\theta = 1$. Hence the *optimal parameter* (i.e., the optimal projection angle of the launch) is $\theta^* = 45^o$; and the *optimal value of the model* (i.e., the corresponding maximal range) is $f^o(\theta^*) = \frac{v^2}{g}$.

The solution is slightly more complicated if the projectile is launched from a positive initial height $h > 0$. Studies of the best shot putters (with $h \approx 2m$) show that the maximum range occurs when the projection angle is about 42^o.

Remark 1. *The problem of optimizing the optimal value function is one of the two basic problems of parametric programming. The other one is the study of stability.*

Let us consider the situation when the projectile is launched from the height $h = 0$ and let us determine how the maximal range changes with small changes of the height $h = \epsilon \ge 0$. This illustrates "sensitivity analysis":

2.2. Sensitivity

The sensitivity information is obtained from the model

$$\begin{aligned} &\max_{(t)} f(t,\theta) = vt\cos\theta \qquad \text{subject to} \\ &t \in F(\theta,\epsilon) = \{t : \tfrac{g}{2}t^2 - vt\sin\theta \le \epsilon\}. \end{aligned}$$

After substituting the optimal larger root for $t = t(\epsilon)$ in the objective function, the sensitivity information is obtained from the optimal value function

$$f^o(\theta,\epsilon) = vt(\epsilon)\cos\theta = \frac{v}{g}\left[v\sin\theta + \sqrt{v^2\sin^2\theta - 2g\epsilon}\right]\cos\theta.$$

In particular, its derivative at $\epsilon = 0$ is $\frac{\partial f^o}{\partial \epsilon}(\theta, 0) = u(\theta) = \cot\theta$. Hence we conclude that the optimal range is *most sensitive* when the projectile is launched at a small angle $\theta > 0$, $\theta \approx 0$. It is *least sensitive* for the projection angle $\theta = \frac{\pi}{2}$; in this case $\frac{\partial f^o}{\partial \epsilon}(\theta, 0) = 0$. Indeed, for the projection angle $\theta = \frac{\pi}{2}$, the projectile lands at the same point where it is launched from, regardless of the height $\epsilon = 0$ or $\epsilon > 0$.

The sensitivity function $u(\theta)$ turns out to be an optimal solution of the corresponding "dual" model. (We will talk about this later. Since the variable t denotes time, we have opted for the notation t rather than x in this special case.)

Parametric programming models often appear in problems of static equilibria:

Example 1. (Romeo's problem) *Romeo is climbing on a ladder of negligible weight. He wants to know the maximal distance s he can climb before the ladder slips. Let L denote the length of the ladder and μ the coefficient of friction between the ladder and the floor. Assume that the ladder leans against a frictionless vertical floor at an angle $0 \le \theta \le \theta^* < \frac{\pi}{2}$. (See Figure 2.)*

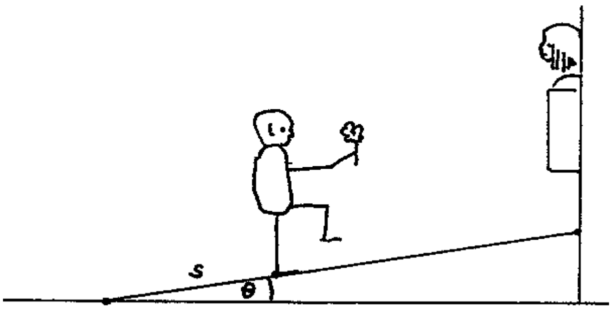

Figure 2. *Romeo's problem*

Then Romeo's problem can be described by the linear parametric model

$$\max_{(s)} s \qquad \text{subject to}$$

$$s - \mu L \tan\theta \le 0, \quad s \le L, \quad 0 \le \theta \le \theta^*.$$

The inequality involving the tangent function is obtained using the conditions for static equilibrium (e.g., Tipler [40, v.1, p.283]). The ladder will not slip as long as this inequality is satisfied. The sensitivity function, relative to the right–hand side perturbations of the first constraint is constant: $u = u(\theta) = 1$, $0 \le \theta < \frac{\pi}{2}$. One can also consider μ or L as parameters.

The above two models have important properties: *uniqueness of optimal solutions* and "*stability*". Indeed, for a fixed value of the parameter θ there is only one optimal decision variable. Also, continuous feasible perturbations of the parameter θ imply continuous changes of the feasible set $F(\theta)$. Many problems of physics, especially in mechanics, can be described by mathematical models having these two properties. They are said to be "well posed". In the late 1950's, parametric programming models appeared describing problems outside the world of physics. These models were expressed in terms of both equations and inequalities. In many of these models at least one of the two properties of well posedness was violated. New mathematical tools had to be developed to study these models. In particular, basic notions of applied mathematics, such as "*optimality*" and "*stability*", had to be re-examined and, often, redefined. *These are not uniquely defined, and one may depend on the other.* The role of "stability" in mathematical modelling has become increasingly important even to the point that *these days*, when confronted with the choice between optimality and stability, *many managers would rather live with non-optimal but stable plans*, e.g., Carlson et al. [79, p.755]. Indeed, *optimality and stability are the two basic notions of parametric programming.*

Parametric programming, as a scientific discipline, began to emerge in the early 1950's. The term "parametric linear programming" appears to be first used by Manne [26] in 1953. Presently there are at least 15 books devoted to parametric programming (with hundreds of references). There have been 20 annual symposia on mathematical programming with data perturbations organized by Tony Fiacco at George Washington University and 6 biannual symposia on parametric programming and related topics initiated by members of the "Berlin school of parametric programming" and František Nožička. (At least 29 students obtained doctorates under his supervision.)

3. Stability

Let us introduce a notion of stability. We study models

(P,θ)

$$\min_{(x)} f(x,\theta) \text{ subject to } x \in F(\theta) = \{x \in \mathbb{R}^n : f^i(x,\theta) \le 0,\ i \in P\}.$$

All functions $f, f^i : \mathbb{R}^{n+p} \to \mathbb{R}$, $i \in P$ in the model (P,θ) are assumed to be continuous. There are many notions of "stability". We will define *stability*, essentially, as *continuity of the feasible set mapping* $F : \theta \mapsto F(\theta)$. Following Berge [4], we define continuity of a *general* point-to-set mapping using the notions of closed and open mappings.

Definition 1. *A point-to set mapping* $\Gamma : \mathbb{R}^p \to \mathbb{R}^n$ *is said to be* closed *at* $\theta^* \in \mathbb{R}^p$ *if, given any sequence* $\theta^k \to \theta^*$ *and a sequence* $x^k \in \Gamma(\theta^k)$*, such that* $x^k \to x^*$*, it follows that* $x^* \in \Gamma(\theta^*)$*.*

Fact 1. *The feasible set mapping* $F : \theta \mapsto F(\theta) = \{x \in \mathbb{R}^n : f^i(x,\theta) \le 0, i \in P\}$ *is closed. The claim is an immediate consequence of continuity of the constraint functions.*

Definition 2. *A point-to-set mapping* $\Gamma : \mathbb{R}^p \to \mathbb{R}^n$ *is* open *at* $\theta^* \in \mathbb{R}^p$ *if, given any sequence* $\theta^k \to \theta^*$ *and any point* $x^* \in \Gamma(\theta^*)$*, there is a sequence* $x^k \in \Gamma(\theta^k)$ *such that* $x^k \to x^*$*.*

Definition 3. *A point-to-set mapping* $\Gamma : \mathbb{R}^p \to \mathbb{R}^n$ *is* continuous *at* $\theta^* \in \mathbb{R}^p$ *if it is both* closed and open *at* θ^**.*

A closely related notion is the one of "lower semi-continuity":

Definition 4. *A point-to-set mapping* $\Gamma : \mathbb{R}^p \to \mathbb{R}^n$ *is* lower semicontinuous *at* $\theta^* \in \mathbb{R}^p$ *if, for each open set* $A \in \mathbb{R}^n$ *satisfying* $A \cap \Gamma(\theta^*) \neq \emptyset$*, there exists a neighbourhood* $N(\theta^*)$ *of* θ^* *such that* $A \cap \Gamma(\theta) \neq \emptyset$ *for each* $\theta \in N(\theta^*)$*.*

Fact 2. *The notions of open and lower semicontinuous mappings are equivalent.*

Remark 2. *Since the feasible set mapping* $F : \theta \mapsto F(\theta)$ *is closed, this mapping* is continuous if, and only if, it is lower semicontinuous *(i.e.,* open*).*

We will study mainly linear and convex models. They will often be studied locally around an arbitrary but fixed $\theta^* \in \mathcal{F} = \{\theta : F(\theta) \neq \emptyset\}$, i.e., we will perform a sensitivity analysis. In this case, we will denote converging sequences by $\theta \to \theta^*$, rather than $\theta^k \to \theta^*$, $k \to \infty$. Also, the results are simplified if one makes a weak technical assumption that the set of optimal solutions exists and that it is bounded at θ^*. The set of all optimal solutions of the program (P, θ^*) is denoted by $F^o(\theta^*)$.

Definition 5. *Consider the convex model* (P, θ) *around some* $\theta^* \in \mathcal{F}$*. The objective function* f *is said to be* realistic *at* θ^* *if* $F^o(\theta^*) \neq \emptyset$ *and bounded.*

Theorem 1. (Characterization of continuity of the feasible set mapping) *Consider the convex model* (P, θ) *around some* $\theta^* \in \mathcal{F}$*. The following statements are equivalent:*

(i) *The point-to-set mapping* $F : \theta \mapsto F(\theta)$ *is continuous at* θ^**.*

(ii) *For every realistic objective function* f *there exists a neighbourhood* $N(\theta^*)$ *of* θ^* *such that* $F^o(\theta) \neq \emptyset$ *and uniformly bounded for every* $\theta \in N(\theta^*)$*. Moreover, all limit points of the sequences of optimal solutions* $x^o(\theta) \in F^o(\theta)$*, as* $\theta \in N(\theta^*)$*,* $\theta \to \theta^*$ *are contained in* $F^o(\theta^*)$ *(i.e., the optimal solutions mapping is closed).*

(iii) *For every realistic objective function* f *there exists a neighbourhood* $N(\theta^*)$ *such that both* $F^o(\theta) \neq \emptyset$ *for every* $\theta \in N(\theta^*)$ *and* $\theta \to \theta^*$ *implies* $f^o(\theta) \to f^o(\theta^*)$*.*

We can now define local and global "stability" of a model.

Definition 7. (Local and global stability) *Consider a convex model* (P, θ) *around some* $\theta^* \in \mathcal{F}$*. We say that the* model is stable at θ^* *if the objective function is realistic at* θ^* *and if the feasible set mapping* $F : \theta \mapsto F(\theta)$ *is continuous at* θ^**. The model is* globally stable *if it is stable at every* $\theta \in \mathcal{F}$*.*

The following sufficient condition for local stability is simple and useful.

Theorem 2. *Consider the convex model* (P, θ) *around some* $\theta^* \in \mathcal{F}$*. Assume that the objective function is realistic at* θ^* *and that the constraints of the program* (P, θ^*) *satisfy* Slater's condition*, i.e.,*

$$\text{"there exists an } x \text{ such that } f^i(x, \theta^*) < 0,\ i \in P\text{".}$$

Then the model is stable at θ^**.*

Stability and instability are illustrated in *Figures 3* and *4*, respectively.

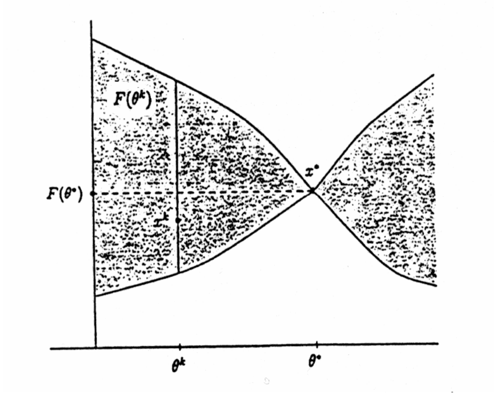

Figure 3. *Stable model at* θ^*

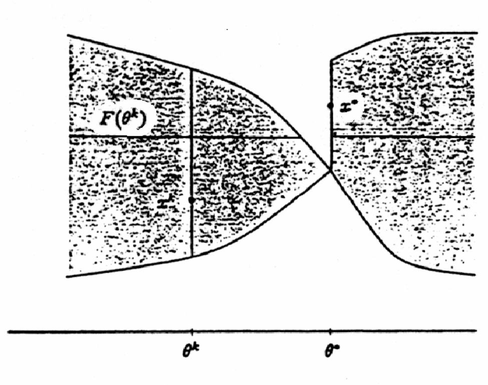

Figure 4. *Unstable model at* θ^*

A majority of models in applied mathematics, including parametric programming models that describe situations in physics and engineering, that are governed by the laws of Newton, appear to be globally stable relative to feasible perturbations of the parameters. In contrast, many parametric programming models outside the world of physics are unstable.

The "caveman' s problem" of determining a *projection angle, subject to Newton's law, that maximizes the arrow's range* is a globally stable parametric programming model. In a modern version of this problem, a businessman wants to determine *prices of products, subject to market constraints, to maximize profit.* The businessman's model, given below, *is globally stable relative to the feasible set*, but *there are some complications*! They are caused by the non-uniqueness of the set of optimal solutions.

Illustration 1. (A stable optimal pricing model) *Suppose, academically, that a businessman, the owner of a small corner store, wants to sell three products:* x_1,

x_2 and x_3 with some unit prices θ_1, θ_2 and θ_3, respectively. He assumes that he can sell any amount of products, provided that certain constraints are satisfied. Let us analyse his situation and determine a pricing policy that maximizes the shop owner's profit. Suppose that the "optimal pricing policy" model is

$$\max_{(x)} \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

subject to the constraints

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & \leq & 60 \\ 2x_1 & + & 4x_2 & + & x_3 & \leq & 80 \\ \theta_1 & + & 2\theta_2 & + & \theta_3 & \leq & 100 \\ \theta_1 & & & - & \theta_3 & \leq & 10 \\ & & & & \theta_3^2(x_1+1) & \leq & 1 \end{array}$$

$$x_i \geq 0, \quad i = 1,2,3; \quad \theta_i \geq 0, \quad i = 1,2,3.$$

The prices are considered as "parameters" and the products to be sold are the "decision variables". One can show (using e.g. Theorem 2.*) that this model is globally stable at every feasible $\theta = (\theta_i)$ with $\theta_3 \neq 1$ relative to feasible perturbations of parameters. Stability here implies, in particular, continuity of the optimal value function. However, it does not imply continuity of the set of optimal solutions. Let us analyse the model numerically.*

Suppose, initially, that the prices are set at, say, $\theta_1 = 10$, $\theta_2 = 20$, and $\theta_3 = 0.5$. After substituting these values in the model, and solving the corresponding linear program, one finds that a corresponding optimal sales profile is $x^* = (x_i^*) = (40,0,0)^T$. This means that the owner should sell only 40 units of item 1 to achieve the maximal profit, which is 400. Suppose that he wants to determine a pricing policy that yields a higher profit. Using "input optimization" (described in Section 8) one finds that the profit increases along a path such as:

$$\theta^*(t) = (\theta_i^*(t)) = (10,20,0.5)^T + t\left(\frac{1}{\sqrt{41}}, \frac{2}{\sqrt{41}}, \frac{1}{\sqrt{41}} - \frac{1}{2}\right)^T, \quad 0 \leq t \leq 1.$$

The path begins at $t = 0$ and ends at $t = 1$, connecting the initial policy $\theta^*(0) = (\theta_i^*(0)) = (10,20,0.5)^T$ with the new policy:

$$\theta^*(1) = (\theta_i^*(1)) = \left(10 + \frac{1}{\sqrt{41}}, 20 + \frac{2}{\sqrt{41}}, \frac{1}{\sqrt{41}}\right)^T.$$

The optimal sales profile $x^* = (40,0,0)^T$ remains the same on the entire path. The optimal profit function on this path is $f^o(\theta^*(t)) = 400 + \frac{40t}{\sqrt{41}}$ with the highest value $f^o(\theta^*(1)) = 400 + \frac{40}{\sqrt{41}} = 406.25$ which is achieved at the path's end $\theta^*(1)$.

Suppose that the store owner wants to increase his profit further by finding a "globally optimal" pricing policy. He finds that this policy, from the initial $\theta^* = \theta^*(1)$, is achieved at the end of the path

$$\theta^\dagger(t) = (\theta_i^\dagger t)) = \theta^* + t\left(-10 - \frac{1}{\sqrt{41}}, 30 - \frac{2}{\sqrt{41}}, -\frac{1}{\sqrt{41}}\right)^T, \quad 0 < t \leq 1.$$

It is $\theta^\dagger(1) = (0, 50, 0)^T$*. The optimal value function on this new path is* $f^o(\theta^\dagger(t)) = 400 + \frac{40}{\sqrt{41}} + 20t\left(30 - \frac{2}{\sqrt{41}}\right)$*,* $0 < t \le 1$*. Its graph on the two paths is depicted in Figure 5:*

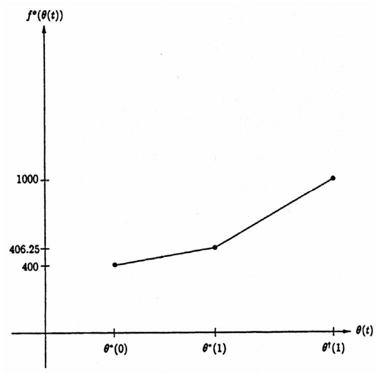

Figure 5. *Stable model: Continuity of the profit function*

However, for any pricing policy chosen on the second path $\theta^\dagger(t)$*,* $0 < t \le 1$*, the optimal solution (sales) profile is uniquely determined and essentially different from the previous one. It was* $x^* = (40, 0, 0)^T$*. Now it is* $x^\dagger = (0, 20, 0)^T$*! The owner may expect to achieve the maximal possible profit if he sells only* 20 *items* x_2 *at the price of* $\theta_2 = 50$ *per unit. Then his globally optimal profit would be* 1000*.*

Remark 3. The jump occurs because the set of optimal solutions in the decision variable x does not depend on the parameter θ continuously. *Indeed, the set of optimal solutions for perturbations along* $\theta^*(t)$*,* $0 \le t \le 1$ *is not unique. It is the set* $x_1 = 40(1-\lambda)$*,* $x_2 = 20\lambda$*,* $x_3 = 0$*, for all* $0 \le \lambda \le 1$*. This set includes both* $x^* = (x_i^*)$ *and* $x^\dagger = (x_i^\dagger)$ *as special cases of* λ *and it is the same set for every choice of the parameter taken on* $\theta^*(t)$*,* $0 \le t \le 1$*. However, there is only one optimal solution for the parameter chosen from the path* $\theta^\dagger(t)$*,* $0 < t \le 1$ *and it is* $x^\dagger = (x_i^\dagger)$*.*

Figure 6 below depicts discontinuity of the set of optimal solutions:

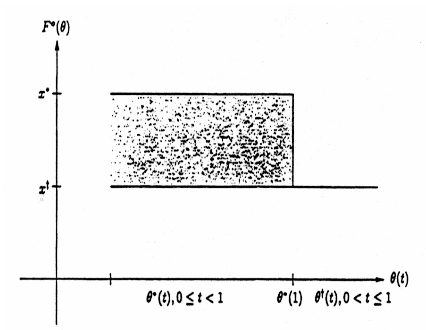

Figure 6. *Stable model: Discontinuity of optimal solutions*

The fact that the set of optimal solutions in x can be discontinuous in a stable model, as the parameter changes, is the cause of instability in the models with "enforced optima". These are models whose feasible sets are the sets of optimal solutions of some other models. A model with enforced optima follows.

4. Instability

A simple unstable linear model in one scalar variable follows:

Illustration 2. *Consider the model*

$$\min x_1, \quad \textit{subject to} \quad \theta\, x_1 = 0, \ \ -1 \le x_1 \le 1.$$

If $\theta = 0$*, then the feasible set is* $F(0) = [-1, 1]$ *and the optimal solution is* $x_1^o = -1$*. For any perturbation* $\theta \neq 0$*,* $F(\theta) = \{0\}$ *and the optimal solution is* $x_1^o = 0$*.*

Examples of unstable linear models on the canonical form with a full row rank coefficient matrix are given and studied in, e.g., Zlobec [45]. An illustration of instability in a situation of "enforced" optima is given next.

Illustration 3. *Let us consider the optimal pricing problem from* Illustration 1..

The owner of the corner store knows the optimal choice of prices and what and how many unit products he should sell to maximize the profit. Now suppose that he reports his sales to a higher level authority, say, the government (internal revenue) for taxation. Suppose that the authority wants to encourage the store owner to achieve the highest possible profit. However the store owner will be taxed by the number of unit products than he can optimally sell (not by how many units he actually sells). If the tax per unit product x_1*,* x_2 *and* x_3 *is* 3, 2, 1 *monetary units, respectively, then the authority's objective is to maximize* $3x_1 + 2x_2 + x_3$ *and his model is*

$$\max 3x_1 + 2x_2 + x_3, \quad \textit{where} \ \ x = (x_i) \ \ \textit{is an optimal solution of the program}$$

$$\max_{(x)} \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & \le & 60 \\ 2x_1 & + & 4x_2 & + & x_3 & \le & 80 \\ \theta_1 & + & 2\theta_2 & + & \theta_3 & \le & 100 \\ \theta_1 & & & - & \theta_3 & \le & 10 \end{array}$$

$$\theta_3^2(x_1 + 1) \le 1, \ \ x_i \ge 0, \ \ i = 1, 2, 3; \ \ \theta_i \ge 0, \ \ i = 1, 2, 3.$$

In this situation, the feasible set of the higher-level authority is the shop owner's set of optimal solutions. Hence the higher-level authority's model is unstable whenever the shop owner's set of optimal solutions experiences a discontinuity. The instability at the higher level manifests itself in jumps of the optimal solutions and *the optimal value as the prices are perturbed. Let us identify one such "critical" set of prices. For any price from the path*

$$\theta^*(t) = (10, 20, 0.5)^T + t\left(\frac{1}{\sqrt{41}}, \frac{2}{\sqrt{41}}, \frac{1}{\sqrt{41}} - \frac{1}{2}\right)^T, \quad 0 \le t \le 1$$

the set $\{(40(1-\lambda), 20\lambda, 0)^T : 0 \le \lambda \le 1\}$ *represents the shop owner's optimal solutions; recall Figure 6. Hence the higher level's authority optimal solution uniquely is* $x^* = (40, 0, 0)^T$ *giving him the optimal value 120. However, on the continued path*

$$\theta^\dagger(t) = \theta^*(1) + t\left(-10 - \frac{1}{\sqrt{41}}, 30 - \frac{2}{\sqrt{41}}, -\frac{1}{\sqrt{41}}\right)^T, \quad 0 < t \le 1$$

the shop owner's optimal solution is unique: $x^\dagger = (0, 20, 0)^T$*. This is the only solution that he can offer to the authority; hence this is also the authority's optimal solution. The authority's optimal value now drops to 40. The authority's optimal value function is depicted in Figure 7.*

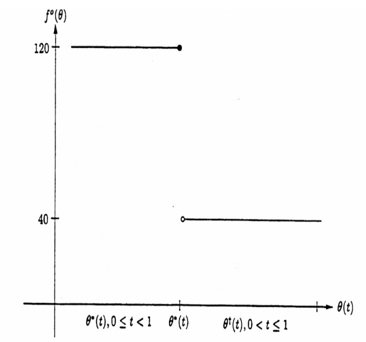

Figure 7. *Unstable model: Discontinuity of the revenue function*

The above is an example of a lexicographic programming model or a bi-level model. In these models a higher level decision maker finds its optimal solution on the set of optimal solutions of a lower level decision maker. These models typically appear in market economy, where they are also called *von Stackelberg games.* The identification of stable and unstable perturbations is a non-trivial task even in linear models.

5. Von Stackelberg games

These are bi-level (or multi-level) decision making processes. The upper level decision maker is referred to as the "*leader*" and the lower level decision maker is the "*follower*". The leader offers a set of rules (e.g., numerical values of the parameter θ) to the follower and requests from him to produce a *complete set of optimal solutions.* The follower obliges and *the leader then chooses a point x from this set that he thinks is locally best for his objective.* The leader wants to improve the current value of his objective. He offers another parameter to the follower and the process repeats until the leader finds an optimal parameter θ. Note that, for a specified value of the parameter, the *feasible set of the leader is the set of optimal solutions of the follower. Complications (instability) occur when the follower does not find a unique optimal solution.* Then we have situations depicted, essentially, in *Figure 6* and *Illustration 3.* Let us illustrate one such situation by a simple example.

Illustration 4. (The leader influences the feasible set of the follower) *Suppose that the leader wishes to maximize the objective* $\Phi^1(x, \theta) = \frac{x_1}{\theta}$ *and that the follower wishes to maximize his own objective* $\Phi^2(x, \theta) = x_1 + x_2$. *Let* $\theta > 0$ *and the feasible set of the follower be determined by* $x_1 + \theta x_2 \leq 1$, $x_1 \geq 0$, $x_2 \geq 0$. *The game begins by the leader offering the value of the parameter, say,* $\theta = 1$ *to the follower. The follower is required to respond by producing the set of all optimal solutions. He finds that this set is the* segment $\{[x_1 \ \ x_2] : x_1 + x_2 = 1, \ x_1 \geq 0, \ x_2 \geq 0\}$. The leader then chooses the best point for him from this set*, which is* $x^0(1) = [1, 0]^T$ *and the value of his objective is thus* $\frac{1}{1} = 1$. *He now realizes that this value might increase if the parameter* θ *is decreased, so he offers some* $\theta < 1$ *to the follower. But with this value the follower finds only one optimal solution:* $x^0(\theta) = \left[0, \frac{1}{\theta}\right]^T$. The leader's feasible set has shrunk. *He has no choice but, according to the rules of the game, to accept this point as an optimal solution.* The value of his objective now drops *to* $\frac{0}{\theta} = 0$*! (The situation is depicted by* Figure 8.*)*

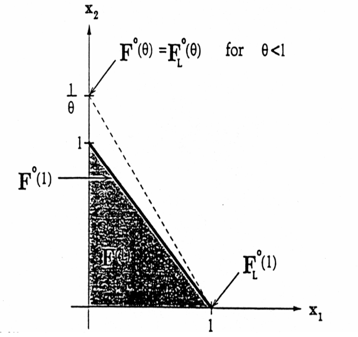

Figure 8. *Instability caused by discontinuity of the follower's set of optimal solutions.*

In some von Stackelberg games the leader influences the objective of the follower and not his feasible set. Interesting games are those where there is one leader (say, the central bank of a country) and several followers (other banks and companies) looking for equilibrium points (e.g., Pareto solutions).

6. Optimality conditions

Optimality conditions are mathematical statements that describe optimal states of a system. Optimality of a feasible decision variable x (for a given parameter θ) or, more generally, of a parameter θ, can be *fully characterized* (without "regularization assumptions") only for simple classes of programs and models (linear and *some* convex !). Otherwise one needs restrictive assumptions and uses first or second order optimality conditions to obtain either necessary or sufficient conditions for optimality. Before studying convex *models*, let us recall the more familiar situation of convex *programs* (i.e., the fixed parameter case):

(CP)

$$\min f(x), \qquad f^i(x) \le 0, \ \ i \in P = \{1, ..., m\}.$$

Here the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraints $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in P$ are convex functions. (Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if $f[\lambda x + (1-\lambda)y] \le \lambda f(x) + (1-\lambda)f(y)$ for every $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, and $0 \le \lambda \le 1$. In order to concentrate on the essentials, we assume that the functions are defined on the entire space $\mathbb{R}^n$.)

The following result is an extension of the classical method of Lagrange from equations to convex inequalities: (Notation: $P(x^*) = \{i \in P : f_i(x^*) = 0\}$ denotes *active constraints* at x^*.)

Theorem 3. (Karush-Kuhn-Tucker conditions) *Consider the convex program (CP) where all functions are assumed to be differentiable. Also assume that the constraints satisfy Slater's condition. Then a feasible point x^* is optimal if, and only if, the system*

$$\nabla f(x^*) + \sum_{i \in P(x^*)} u_i \nabla f^i(x^*) = 0, \ u_i \ge 0, \ i \in P(x^*)$$

is consistent.

The Slater condition assumption can be omitted in *Theorem 3. if the active constraints are "functions with a locally flat surface"* (abbreviated: LFS functions). These are possibly *the simplest nonlinear functions that retain many properties of linear functions*. We will introduce this class for differentiable functions. (Notation: $D_f^{=}(x^*) = \{d \in \mathbb{R}^n : f(x^* + \alpha d) = f(x^*), \ 0 < \alpha < \alpha'$ for some $\alpha' > 0$ denotes the cone of directions of constancy of f at x^*; $N(\nabla f(x^*))$ is the *null-space* of the gradient)

Definition 2. *A differentiable convex function $f : \mathbb{R}^n \to \mathbb{R}$ is said to have a* locally flat surface *at $x^* \in \mathbb{R}^n$ if $N(\nabla f(x^*)) = D_f^{=}(x^*)$.*

All linear functions are LFS at every $x^* \in \mathbb{R}^n$, so are many nonlinear functions, e.g., $f(t) = e^t$.

Sometimes optimality conditions have clear physical interpretations:

Physical interpretation of the KKT conditions (*G. I. Joe Minimizes travel time over two regions*) Suppose that G.I. Joe (or an object) can move freely in a plane between two fixed points A and B belonging to two different regions. Let his velocities in these regions be v_1 and v_2, respectively. We study the movement in the space R^2 using the Euclidean distance. Suppose that the coordinates of A and B in the (x_1, x_2)-plane are $(0, a)$ and $(b_1 + b_2, 0)$, respectively, and that the two regions are separated by a straight line parallel with the x_2 axis and passing through the point $(b_1, 0)$. (See *Figure 9*) An important question is: *How should G.I. Joe move from A to B in order to minimize his travel time?* Note that the variables a, b_1, b_2, v_1 and v_2 could be considered as "parameters", but we will not go into this here. The angles of incidence α and refraction β are "decision variables". Given some numerical values of the parameters, we wish to determine an optimal solution α^* and β^*.

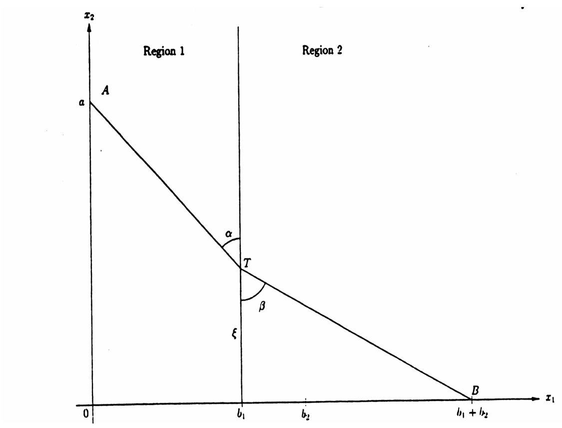

Figure 9. *Travelling over two regions*

This problem can be formulated as a convex program. Let us denote by T a point on the border between the two regions where the object crosses from one region into the other. Then denote by AT the distance between the points A and T and by TB the distance between T and B. Since $b_1 = AT \sin\alpha$ and $b_2 = TB \sin\beta$, the time of travel from A to B is

$$f(\alpha,\beta) = \frac{AT}{v_1} + \frac{TB}{v_2} = \frac{b_1}{v_1 \sin\alpha} + \frac{b_2}{v_2 \sin\beta}$$

subject to the constraint $b_1 \text{cot}\alpha + b_2 \text{cot}\beta = a$. Rather than working with trigonometric functions, we use the substitution $x_1 = \text{cot}\alpha$ and $x_2 = \text{cot}\beta$. Since $\sin^2\alpha = \frac{1}{1+x_1^2}$ and $sin^2\beta = \frac{1}{1+x_2^2}$, the problem is formulated as *the convex program*

$$\min f(x_1, x_2) = \frac{b_1}{v_1}\sqrt{1+x_1^2} + \frac{b_2}{v_2}\sqrt{1+x_2^2} \tag{6.1}$$

$$b_1 x_1 + b_2 x_2 = a.$$

For the sake of simplicity, we consider only the angles from the interval $0 < \alpha, \beta < \frac{\pi}{2}$; hence the non-negativity constraints on the variables x_1 and x_2 are omitted. Note that the constraint is linear, hence LFS. An optimal solution of the minimal travel time program, formulated above, is unique and it is characterized by the KKT conditions, which are here

$$\frac{b_1 x_1}{v_1\sqrt{1+x_1^2}} + \lambda b_1 = 0, \quad \frac{b_2 x_2}{v_2\sqrt{1+x_2^2}} + \lambda b_2 = 0$$

for some multiplier $\lambda \in \mathbb{R}$. The elimination of the multiplier yields the well-known *"law of refraction"*:

$$\frac{v_1}{v_2} = \frac{\cos\alpha}{\cos\beta}.$$

The solution of the above problem requires solving fourth degree polynomials.

Instead of a G.I. Joe, one can consider a ray of light passing between two media. Since the ray of light travels "optimally", by the "generalized" Fermat's principle, it must satisfy the above law of refraction. In this case the law is called the *Snell law.* (See also Collatz and Wetterling [75].)

The conditions for globally optimal parameters are *significantly simplified* also for convex models *if the constraint functions are* LFS *functions* in the decision variable x. We consider a convex LFS model

(P, θ)

$$\min_{(x)} f(x, \theta) \qquad \text{subject to}$$

$$f^i(x, \theta) \leq 0, \;\; i \in P.$$

Here all *constraint* functions $f^i(\cdot, \theta) : \mathbb{R}^n \to \mathbb{R}$, $i \in P$ are assumed to be LFS at every $x \in F^o(\theta)$, $\theta \in F$. ($F^o(\theta)$ denotes the set of all optimal solutions of the program (P, θ) for a fixed θ from the feasible set of parameters F.) Optimality of parameters for such models can be described using the classical Lagrangian:

$$L(x, u; \theta) = f(x, \theta) + \sum_{i \in P} u_i f^i(x, \theta).$$

Theorem 4. *(***Characterizing *globally* optimal parameters for convex LFS models)** *Consider a convex LFS model (P, θ) around some θ^*. Assume that the optimal value function $f^o : \mathbb{R}^p \to \mathbb{R}$ exists on the entire feasible set F. Let x^* be an optimal solution of the program (P, θ^*). Then θ^* minimizes f^o on F if, and only if, there exists a non-negative vector function $u^* : F \to \mathbb{R}^m_+$ such that*

$$L(x^*, u; \theta^*) \leq L(x^*, u^*(\theta^*); \theta^*) \leq L(x, u^*(\theta); \theta)$$

for every $u \in \mathbb{R}^m_+$, every $x \in \mathbb{R}^n$, and every $\theta \in F$.

What does the above result give, when it is applied, say, to the projectile motion model?

Illustration 5. ("Caveman's rule") *Consider the projectile motion model (2.1) introduced in Section 2.* This is a convex LFS model! *For any $\theta > 0$, the optimal t is $t = \frac{2v}{g} \sin\theta$. An angle θ^* is globally optimal, according to the above theorem if, and only if, $g^2 \cos\theta \cdot t^2 - 2gv \sin 2\theta \cdot t + 2v^2 \sin\theta \cdot \sin 2\theta^* \geq 0$ for every $\theta > 0$ and every $t \in \mathbb{R}$. The values of this quadric are non-negative if, and only if, its minimal value is non-negative. This is the case if, and only if, $\sin 2\theta^* \geq \sin 2\theta$ for every $\theta > 0$. Hence the maximum range of the projectile, say, the caveman's arrow, launched from the height $h = 0$, is achieved if, and only if, θ^** maximizes the function *$f(\theta) = \sin 2\theta$.*

7. Duality

With every convex model (P, θ) one can associate one or more "dual" models. A dual model, essentially, is a statement that the saddle point of a Lagrangian can be reached by minimizing the Lagrangian relative to x, and then by maximizing it relative to a suitable u. First let us explain the idea for the convex program

(CP)

$$\min f(x) \qquad f^i(x) \le 0, \;\; i \in P.$$

We use the Lagrangian

$$\mathcal{L}^<(x,u) = f(x) + \sum_{i \in P \setminus P^=} u_i f^i(x).$$

Here $P^= = \{i \in P : x \in F \;\Rightarrow\; f^i(x) = 0\}$ is the *minimal index set of active constraints.* This index set generates the set of decision variables

$$F^= = \{x : f^i(x) \le 0, \; i \in P^=\} = \{x : f^i(x) = 0, \; i \in P^=\}.$$

For non-negative vectors $u \in \mathbb{R}_+^{\mathrm{card}(P \setminus P^=)}$, the minimization of the Lagrangian over $F^=$ defines the *subdual function*

$$\varphi_{P^=}(u) = \min_{x \in F^=} \mathcal{L}^<(x,u).$$

Let us consider only those u's for which an optimal solution $x = x^o(u) \in F^=$ exists. Then, with such u's, the "dual program" is

$$\max \varphi_{P^=}(u)$$

$$u \in \mathbb{R}_+^{\mathrm{card}(P \setminus P^=)}.$$

Since we know how to characterize optimality of $x^o(u)$ on the convex set $F^=$, the dual of the convex program can be written as follows:

(D)

$$\max\{f(x) + \sum_{i \in P \setminus P^=} u_i f^i(x)\}$$

$$u \in \mathbb{R}_+^{\mathrm{card}(P \setminus P^=)}$$

$$\nabla^T f(x) + \sum_{i \in P \setminus P^=} u_i \nabla^T f^i(x) \in \{\bigcup_{i \in P^=} D_i^=(x)\}^+$$

$$x \in F^=$$

(Here M^+ denotes the polar set of M.) If the constraints satisfy Slater's condition, then $P^= = \emptyset$, $F^= = \mathbb{R}^n$, and the dual is significantly simplified:

7.1. Dual in the presence of Slater's condition

If the constraints of the convex program (CP) satisfy Slater's condition then the dual is

$$\max\{f(x) + \sum_{i \in P} u_i f^i(x)\}$$

$$\nabla f(x) + \sum_{i \in P} u_i \nabla f^i(x) = 0, \;\; u \in \mathbb{R}_+^m.$$

In order to formulate the dual of the convex *model* (P,θ), one must resolve the obstacle presented by the variable index set $P^{=}(\theta)$ that appears in the Lagrangian. This set determines the number of constraint functions in the Lagrangian. One can proceed as follows: Denote all subsets of $P^{=}(\theta)$, obtained by varying θ, by Π:

$$\Pi = \{\Omega \subset P : \Omega = P^{=}(\theta) \text{ for some } \theta \in F\}.$$

Now, given an $\Omega \subset \Pi$, the set of all feasible parameters θ for which $P^{=}(\theta) = \Omega$ is denoted by $\mathbf{F}_\Omega = \{\theta \in F : P^{=}(\theta) = \Omega\}$. Note that $F = \cup_{\Omega\subset\Pi}\mathbf{F}_\Omega$.

The feasible set F is thus divided into disjoint regions $\mathbf{F}_\Omega$, each determined by an index set Ω. Each subset $\Omega \in \Pi$ generates a subdual function and a subdual, using the Lagrangian

$$L_\Omega(x,u;\theta) = f(x,\theta) + \sum_{i\in P\setminus\Omega} u_i f^i(x,\theta)$$

and a point-to-set mapping $F_\Omega : \mathbf{F}_\Omega \mapsto \mathbb{R}^n$ defined by $F_\Omega(\theta) = \{x : f^i(x,\theta) \le 0, \ i \in \Omega\}$. Fix an $\Omega \in \Pi$ and a $\theta \in \mathbf{F}_\Omega$.

Definition 6. *Consider a convex model* (P,θ) *and a subset* $\Omega \in \Pi$. *For* $\theta \in \mathbf{F}_\Omega$ *and a non-negative vector function* $u : \mathbf{F}_\Omega \to \mathbb{R}_+^{\mathrm{card}(P\setminus\Omega)}$, *the function*

$$\varphi_\Omega(u,\theta) = \inf_{x\in F_\Omega(\theta)} L_\Omega(x,u(\theta);\theta)$$

is called the Ω-subdual function. *The* dual, *determined by the same* $\Omega \subset \Pi$, *is*
$(D,\Omega;\theta)$

$$\sup \varphi_\Omega(u,\theta)$$
$$u : F_\Omega \to \mathbb{R}_+^{\mathrm{card}(P\setminus\Omega)}$$

where only those functions u are considered for which the above infimum exists (has finite value) for the Lagrangian in the variable x. We will use the terminology: Ω-dual. *The functions* $u = u(\theta) \in \mathbb{R}_+^{\mathrm{card}(P\setminus\Omega)}$, $\theta \in \mathbf{F}_\Omega$, *that solve* $(D,\Omega;\theta)$, *are called the* solutions of the Ω-dual. *These solutions have the number of components corresponding to the cardinality of the set* Ω. *The number of* Ω*-duals is the cardinality of the set* Π. *The collection of all these* Ω*-duals, i.e., the set* $\{(D,\Omega;\theta) : \Omega \subset \Pi\}$, *is called* the dual of the model (P,θ). *Note that the dual is a model because the parameter* θ *varies over* $\mathbf{F}_\Omega$ *and the feasible set* F.

Since we know how to characterize optimal solutions of the Lagrangian for a fixed θ, the Ω-dual for differentiable functions can be written as

$(D,\Omega;\theta)$

$$\max\{f(x,\theta) + \sum_{i\in P\setminus\Omega} u_i f^i(x,\theta)\}$$
$$u \in \mathbb{R}_+^{\mathrm{card}(P\setminus\Omega)}$$
$$\nabla^T f(x,\theta) + \sum_{i\in P\setminus\Omega} u_i \nabla^T f_i(x,\theta) \in \{\bigcap_{i\in\Omega} D_i^{=}(x,\omega)\}^{+}$$
$$x \in F_\Omega(\theta), \quad \theta \in \Omega.$$

The Ω-subduals are closely related to the original model (P, θ). In particular, one can estimate the optimal value of the model on the sets $\mathbf{F}_\Omega$ by feeding the minimum function with suitable non-negative functions $u : F_\Omega \to \mathbb{R}_+^{\text{card}(P\setminus\Omega)}$. One can also determine whether a given θ is globally optimal on $\mathbf{F}_\Omega$. These duals are useful in von Stackelberg games (and other models where Slater's condition is not or cannot be satisfied). The components of the dual solution are functions in θ; they provide sensitivity information and can be considered as the "values" (or "value functions") of the constraints as θ varies.

Let us illustrate the above ideas on the model introduced in *Section 2*.

Illustration 6. (The dual of the projectile motion model) *Consider the model (2.1). Since t denotes the time variable, we use the notation $x = t$. The constraint satisfies Slater's condition at every $t > 0$, so the set Π is just the empty set $\Omega = \emptyset$. Hence the dual model is*

$$\max_{u \geq 0} \; -vt\cos\theta + u[\frac{g}{2}t^2 - vt\sin\theta] \qquad \textit{subject to}$$

$$-v\cos\theta + u[gt - v\sin\theta] = 0.$$

The only u for which the subdual function has a minimum is $u = \frac{v\cos\theta}{gt - v\sin\theta}$. But t appearing in u minimizes the Lagrangian. Substituting u into the objective function, and setting the derivative with respect to the variable t to zero, yields $t = \frac{2v}{g}\sin\theta$. The back substitution gives the dual solution $u = \cot\theta$. (This solution was obtained earlier by sensitivity analysis).

Another illustration is taken from underwater navigation:

Illustration 7. (Zermelo's problem under the water) *The dynamics of an object (say, a torpedo) with a velocity of unit magnitude relative to a three-dimensional medium is described by a system of differential equations*

$$\begin{aligned} dz_1/dt &= u + \cos\Phi\cos\varphi \\ dz_2/dt &= v + \cos\Phi\sin\varphi \\ dz_3/dt &= w + \sin\Phi. \end{aligned}$$

Here u, v, w are components of the constant velocity vector of the medium; Φ denotes the angle between the velocity vector of the object and its projection on the (z_1, z_2) plane and φ is the angle between the projection and the z_3 axis. The problem of finding the steering angles Φ and φ that minimize the time of reaching a convex target can be formulated as a convex LFS model: Assume that the object, at time $t = 0$, is at the origin, that the components of the velocity vector of the medium are, say, $u = 2$, $v = 0$, $w = 0$ and that the target is the unit sphere

$$T = \{[z_1 z_2 z_3]^T : (z_1 - 10)^2 + (z_2 - 1)^2 + (z_3 - 2)^2 \leq 1\}.$$

Then, after solving the system with the initial condition and substitution, the minimal time problem is described by the convex model

$$\min_{(\Phi,\varphi)} t \qquad \textit{subject to}$$

$$(2t + t\cos\Phi\cos\varphi - 10)^2 + (t\cos\Phi\sin\varphi - 1)^2 + (t\sin\Phi - 2)^2 \leq 1. \qquad (7.1)$$

The dual solution (obtained from the KKT condition for fixed Φ and φ) is

$$U(\Phi,\varphi) = \tfrac{1}{2}\left[-116 - 16\cos\Phi\cos\varphi + 40\cos\Phi\sin\varphi + 80\sin\phi + 99\cos^2\Phi\cos^2\varphi \right.$$
$$\left. +10\cos^2\Phi\sin 2\varphi + 20\sin 2\Phi\cos\varphi - 3\cos^2\Phi + 2\sin 2\Phi\sin\varphi\right]^{-\frac{1}{2}}.$$

Its graph is depicted in Figure 10. *It describes sensitivity (rate of change, derivative) of the optimal sailing time to reach the target relative to small perturbations of the radius of the target. The points that minimize the dual function are the steering angles for which the optimal sailing times are least sensitive to small perturbations of the radius of the target, i.e., for these perturbations, these are the* most "robust" steering angles. *(One can verify the claims directly by performing sensitivity analysis on the right-hand side of (7.1).)*

Note: *The most robust steering angles obtained from duality generally are different from the angles that steer a torpedo to the target in least time. They are also different from the most robust steering angles for other kinds of perturbations of the target. Zermelo's problems are easily adjusted to situations when both the target and the object are moving.*

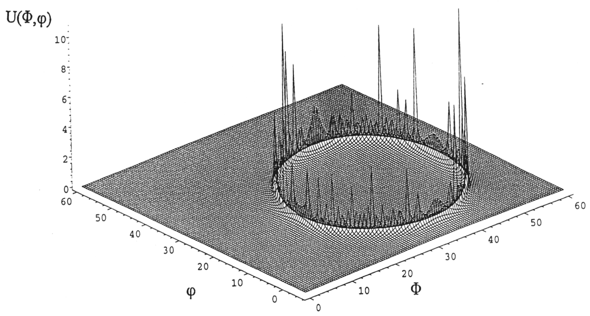

Figure 10. *Dual solution for Zermelo's problem*

8. Input optimization

Input optimization is a term used to describe a collection of numerical methods that *optimize the optimal value function using only stable perturbations* of the parameters. The methods are applied from some "initial" input $\theta = \theta^o$, hence they are designed to solve

$$\min f^o(\theta)$$
$$\theta \in \pi(\theta^o).$$

Here $f^o(\theta)$ is the optimal value function and $\pi(\theta^o)$ is a prescribed class of all "stable" paths emanating from θ^o, i.e., paths on which the point-to-set mapping $F : \theta \mapsto F(\theta)$ is lower semicontinuous at every point. A solution to an input optimization problem is both: *a stable path emanating from θ^o and its end point θ^* which locally optimizes the optimal value function relative to the region of stability at θ^*.* (This region is a collection of all continuous paths emanating from θ^* on which $F(\theta) \to F(\theta^*)$ as $\theta \to \theta^*$.) The solution generally depends on the initial choice θ^o and the prescribed class of paths. In particular, if there are two or more disjoint feasible or stable regions, then an input optimization process cannot leave the region where the iteration has begun from θ^o. (Such situations typically occur in, e.g., Zermelo's navigation problems. Depending on various parameters, such as the speed of water, torpedo or a boat, the feasible region consists of various disjoint sets.)

In order to simplify calculations, let us work only with the class of *linear perturbations*, i.e., perturbations of the form $\theta = \theta^k + \alpha d^k$, $\alpha \geq 0$, $k = 0, 1, 2, ...$ This restriction generally sacrifices optimal inputs (parameters) that can be reached only by nonlinear perturbations from θ^o. It may also increase the number of iterations. Lower semi-continuity of the mapping $F : \theta \mapsto F(\theta)$ on the *entire path* between two inputs, say, between an "initial" θ^o and a "final" θ^*, including continuity at θ^* "from the left", is guaranteed under rather restrictive assumptions.

Input optimization problems are solved iteratively. They are modelled after the feasible direction methods of mathematical programming: Given a feasible input approximation $\theta^k \in F$, $k = 0, 1, ...$; a new input θ^{k+1} is obtained in two stages. In the first stage, a *"stable improvable feasible direction generator"* is used to produce a direction d with the following local properties:

(i) $\theta^k + \alpha d \in F$, for every $0 \leq \alpha \leq \alpha'$, and some $\alpha' > 0$;

(ii) the feasible set mapping F is lower semicontinuous relative to F at every $\theta^k + \alpha d$, $0 \leq \alpha \leq \alpha'$;

(iii) the optimal value function is improvable, i.e., $f^o(\theta^k + \alpha d) < f^o(\theta^k)$ for $\alpha > 0$ close to 0.

The direction generator uses an appropriate *marginal value formula* like the one described below:

We study a convex model (P, θ) around some feasible θ^* with a realistic objective function at θ^*. Let us consider the region of stability S at θ^*. Along some path in S consider a sequence $\theta \in S$, $\theta \to \theta^*$, and then the sequence $f^o(\theta) \to f^o(\theta^*)$. A formula for the limit

$$\lim_{\theta \in S, \theta \to \theta^*} \frac{f^o(\theta) - f^o(\theta^*)}{\|\theta - \theta^*\|}$$

is called the *marginal value formula* at θ^*. The limit generally depends on the point θ^* and on the path. The marginal value formula at θ^* can be expressed in terms of the first derivative of the Lagrangian function

$$L_*^<(x, u; \theta) = f(x, \theta) + \sum_{i \in P \setminus P^=(\theta^*)} u_i f^i(x, \theta)$$

and the two limits:

$$s = \lim_{\theta \in S, \theta \to \theta^*} \frac{\theta - \theta^*}{\|\theta - \theta^*\|}$$

and

$$z = \lim_{\theta \in S, \theta \to \theta^*} \frac{x^o(\theta) - x^o(\theta^*)}{\|\theta - \theta^*\|}.$$

Here $x^o(\theta)$ is an optimal solution of the program (P, θ) for a fixed θ. Typically, the required assumptions are: continuous differentiability of functions in the model (P, θ), lower semi-continuity of the point-to-set mapping

$$F_*^= : \theta \mapsto F_*^=(\theta) = \{x : f^i(x, \theta) \leq 0, \ i \in P^=(\theta^*)\}$$

and uniqueness of the saddle point $\{x^o(\theta^*), U^o(\theta^*)\}$ for the program (P, θ^*). This saddle point is defined as

$$L_*^<(x^o(\theta^*), u; \theta^*) \leq L_*^<(x^o(\theta^*), U^o(\theta^*); \theta^*) \leq L_*^<(x, U^o(\theta^*); \theta^*)$$

for every $u \in \mathbb{R}_+^c$ and every $x \in F^=(\theta^*) = \{x : f^i(x, \theta) = 0, \ i \in P^=(\theta^*)\}$. The fact that lower semi-continuity of the mapping $F_*^=$ implies lower semi-continuity of F is used in the proof of the following the marginal value formula:

Theorem 5. (The basic marginal value formula) *Consider the convex model (P, θ) with a realistic objective function at some θ^*. Let us assume that the mapping $F_*^=$ is lower semicontinuous at θ^*, relative to a set S containing θ^*, and that the saddle point $\{x^o(\theta^*), U^o(\theta^*)\}$ is unique. Also suppose that the gradients $\nabla f(x, \theta)$, $\nabla f^i(x, \theta)$, $i \in P \backslash P^=(\theta^*)$ are continuous at $(x^o(\theta^*), \theta^*)$. Then for every sequence $\theta \in S$, $\theta \mapsto \theta^*$, and for every path $x^o(\theta) \to x^o(\theta^*)$, for which the limits s and z exist, we have*

$$\lim_{\theta \in S, \theta \to \theta^*} \frac{f^o(\theta) - f^o(\theta^*)}{\|\theta - \theta^*\|} = \nabla_x L_*^<(x^o(\theta^*), U^o(\theta^*); \theta^*) z + \nabla_\theta L_*^<(x^o(\theta^*), U^o(\theta^*); \theta^*) s.$$

Remark 4. *Under Slater's condition this formula uses the classical Lagrangian and the first term can be omitted. In this case one does not have to worry about the stability requirement and can pick any improvable direction for the step-size search. Such stable improvable feasible direction $d = (d_i)$ is*

$$d = -\nabla_\theta L(x^o(\theta^k), U^o(\theta^k); \theta^k).$$

In this case the optimal value function locally decreases along d in the direction $\alpha \geq 0$ and the limit s is $s = \frac{d}{\|d\|}$. This approach appears to work well in practice.

In the second stage, once d is determined, one solves the *"step-size problem"*, i.e., one determines a step-size $\alpha_k \geq 0$ that minimizes (or at least decreases) the optimal value function on a segment $\{\theta^k + \alpha d : 0 \leq \alpha \leq \alpha'\} \cap F$, for some $\alpha' > 0$. The substitution of $\theta = \theta^k + \alpha d = \theta(\alpha)$ into (P, θ), yields the problem in $n + 1$ variables (x and α):

(P, θ)

$$\min_{(x,\alpha)} f(x, \alpha)$$
$$f^i(x, \alpha) \leq 0, \ \ i \in P$$
$$\alpha \geq 0.$$

For every fixed α this is a convex program in the variable x. Denote its optimal value by $f^o(\alpha)$. Then the step-size problem is

$$\min_{(\alpha)} f^o(\alpha)$$
$$\theta^k + \alpha d \in F, \quad \alpha \geq 0.$$

One can approximate a minimum of $f^o(\theta)$ on the interval $\theta^k + \alpha d \in F$, $\alpha \geq 0$ using a search method, such as the Fibonacci Method or the Golden Section Search. For each α used in the search, one solves the convex program (P, θ) in the variable x in order to obtain an optimal solution $x^o = x^o(\alpha)$ and then $f^o(\alpha) = f(x^o(\alpha), \alpha)$. When an optimal solution (step-size) α_k is found, then the new approximation is $\theta^{k+1} = \theta^k + \alpha_k d \in F$. (Instead of insisting on optimality one is often satisfied with a feasible step size α_k for which $f^o(\alpha_k) < f^o(0)$.)

Let us demonstrate how input optimization works in practice. We have solved two real-life problems:

Case study 1: Restructuring in a textile mill. A case study of a textile mill has been described by Naylor et al. [27] using a linear program. We will reformulate their *program* as a linear *model*, by allowing three most sensitive matrix coefficients to vary. Then globally optimal parameters are determined by input optimization. The solution will require a restructuring of the work force in the mill leading to a significantly higher profit. Naylor et al. study a company that is purchasing rough cotton and, through a series of twelve operations, it produces seven styles of materials. The production rate of required operations (expressed in hundred meters per hour) for each of the seven styles, is given in the following table:

PROCESS	B	P_1	P_2	P_3	P_4	D_1	D_2	Available Hours
Singeing	90	60	90	70	80	90	80	150
Desizing	130	100	90	110	80	130	120	150
Kier Boiling	15	9	10	8	9	13	12	900
Bleaching	10	11	10.5	11	11	11	12	1500
Drying	130	100	100	120	110	110	120	140
Mercerizing	8	5.5	6	6.5	7	7	8	2490
Printing	–	3	3	2	2.5	–	–	1800
Aging	–	50	40	40	60	–	–	150
Dyeing (blue)	–	–	–	–	–	40	–	150
Dyeing (red)	–	–	–	–	–	–	3	5140
Starching	20	18	18	16	15	20	15	500
Calendering	40	50	30	25	40	32	35	450

Here the process B refers to the bleached style, P_k, $k = 1, ..., 4$ to the printed styles and D_1 and D_2 to the dyed styles. The available amount of hours for each operation (in hundreds of hours) is given in the last column. The amount of time available for mercerizing is divided equally between three styles: bleached, printed, and dyed (each with 830 hours). The estimated profit for the seven styles (per meter) is

B	0.40($)
P_1	0.60($)
P_2	0.80($)
P_3	1.00($)
P_4	1.25($)
D_1	1.20($)
D_2	1.30($)

A corresponding linear program for profit maximization is

$$\max 0.40B + 0.60P_1 + 0.80P_2 + 1.00P_3 + 1.25P_4 + 1.20D_1 + 1.30D_2$$

$$\begin{array}{llllllll}
\frac{B}{90} & +\frac{P_1}{60} & +\frac{P_2}{90} & +\frac{P_3}{70} & +\frac{P_4}{70} & +\frac{D_1}{90} & +\frac{D_2}{80} & \leq 150 \\
\frac{B}{130} & +\frac{P_1}{100} & +\frac{P_2}{90} & +\frac{P_3}{110} & +\frac{P_4}{80} & +\frac{D_1}{130} & +\frac{D_2}{120} & \leq 150 \\
\frac{B}{15} & +\frac{P_1}{9} & +\frac{P_2}{10} & +\frac{P_3}{8} & +\frac{P_4}{9} & +\frac{D_1}{13} & +\frac{D_2}{12} & \leq 900 \\
\frac{B}{10} & +\frac{P_1}{11} & +\frac{P_2}{10.5} & +\frac{P_3}{11} & +\frac{P_4}{11} & +\frac{D_1}{11} & +\frac{D_2}{12} & \leq 1500 \\
\frac{B}{130} & +\frac{P_1}{100} & +\frac{P_2}{100} & +\frac{P_3}{120} & +\frac{P_4}{110} & +\frac{D_1}{110} & +\frac{D_2}{120} & \leq 140 \\
\frac{B}{8} & & & & & & & \leq 830 \\
 & \frac{P_1}{5.5} & +\frac{P_2}{6} & +\frac{P_3}{6.5} & +\frac{P_4}{7} & & & \leq 830 \\
 & & & & & \frac{D_1}{7} & +\frac{D_2}{8} & \leq 830 \\
 & \frac{P_1}{3} & +\frac{P_2}{3} & +\frac{P_3}{2} & +\frac{P_4}{2.5} & & & \leq 1800 \\
 & \frac{P_1}{50} & +\frac{P_2}{40} & +\frac{P_3}{40} & +\frac{P_4}{60} & & & \leq 150 \\
 & & & & & \frac{D_1}{40} & & \leq 150 \\
 & & & & & & \frac{D_2}{35} & \leq 140 \\
\frac{B}{20} & +\frac{P_1}{18} & +\frac{P_2}{18} & +\frac{P_3}{16} & +\frac{P_4}{15} & +\frac{D_1}{20} & +\frac{D_2}{15} & \leq 500 \\
\frac{B}{40} & +\frac{P_1}{5} & +\frac{P_2}{30} & +\frac{P_3}{25} & +\frac{P_4}{40} & +\frac{D_1}{32} & +\frac{D_2}{35} & \leq 450.
\end{array}$$

There are also the demand restraints $P_1 \geq 50$, $P_2 \geq 50$, $50 \leq P_3 \leq 1000$, $50 \leq P_4 \leq 500$ and, of course, $B \geq 0$, $D_1 \geq 0$, $D_2 \geq 0$. The sixth to eighth inequalities describe the constraints in the mercerizing department for each of the three styles (bleached, printed and dyed). The eleventh and the twelfth inequalities describe the constraints in the dyeing department (one for blue and one for red dyed style). Using shadow prices one finds that a *"bottle-neck"* of production (most sensitive technological coefficient) for the profit occurs in the mercerizing department where the blue dyed style D_1 has been processed. It also occurs in the starching department in the production of the printed style P_2 and the blue dyed style D_1.

The management has decided to modernize the production. In particular, it wants to purchase new, more efficient machines for the two bottle-neck departments. Several different types of machines are available with different production rates. Which of these should be purchased? Should one invest only in the most efficient (most expensive) machines? Let us formulate this problem as a linear programming *model*. First, we denote the increased production rates in the two departments for the three processes by θ_1, θ_2 and θ_3, respectively. The new matrix coefficients, at the bottle-neck of production level, thus assume the form $a_{8,6} = (7 + \theta_1)^{-1}$, $a_{13,3} = (18 + \theta_2)^{-1}$, $a_{13,6} = (20 + \theta_3)^{-1}$. It is found that the available machines can operate with rates $0 \leq \theta_1 \leq 4$, $0 \leq \theta_2 \leq 7$, $0 \leq \theta_3 \leq 10$. The introduction of the parameters θ_1, θ_2 and θ_3 is a departure from a linear *program* to a linear *model*. Since the constraints satisfy Slater's condition (one can verify this with, e.g., $B = 1$, $P_1 = P_2 = P_3 = P_4 = 50.01$, $D_1 = D_2 = 1$) and the feasible set $F(\theta)$ is bounded for every θ from the above feasible box, the *model is globally stable relative to the feasible set*, and an input optimization method is applicable. Input optimization starts here from the origin $\theta = 0 \in \mathbb{R}^3$, and only linear paths are used. The step-size problems are solved using only three (!) searches per iteration. After 14 iterations, the *globally optimal input* $\theta_1^* = 4$, $\theta_2^* = 6.572$, $\theta_3^* = 6.572$ is found. Hence every piecewise linear path leading from the origin to θ^*, and remaining entirely in the above stable box, is a global solution of the input optimization problem. For example, one can first improve the rate of production for the mercerizing operation by 4 units (400 meters) and than improve the two starching operations by 6.572 units (657.2 meters) each.

An "old" linear programming solution and a "new" solution that corresponds to the optimal parameter are compared below:

"Old"	"New"
$B^o = 0$	$B^* = 0$
$P_1^o = 50$	$P_1^* = 50$
$P_2^o = 1996$	$P_2^* = 1814$
$P_3^o = 1000$	$P_3^* = 50$
$P_4^o = 500$	$P_4^* = 500$
$D_1^o = 5810$	$D_1^* = 6000$
$D_2^o = 0$	$D_2^* = 2276$.

The corresponding optimal values are $f^o = 1\,022\,380$ and $f^o(\theta^*) = 1\,231\,550$, which represents an improvement of about 20% (before deducting the cost of improving the efficiency, e.g. the cost of the new machines).

The input optimization solution suggests that the production of the printed style P_3 be reduced *from 1000 units to only 50* and that the mill should *start producing 2276 units of the dyed style* D_2. (With the old machines, the optimal profile of production did not include that particular style.) *The mill should restructure the work force to meet the new optimal requirements.* The workers who work in the printed style 3 department, should be trained to work on the dyed style 2. The result obtained by input optimization shows that *one does not have to purchase the most efficient (expensive) machines to achieve optimal results.* Only one "best" machine (for mercerizing) available on the market should be purchased, the other two (for starching) could be "sub-optimal". The restructuring can be done in several stages. It is a stable process as long as the parameters "move" within the feasible box.

When the model does not satisfy Slater's condition (e.g., if the model is bilevel), then the marginal value formula has to be modified. This was done in solving the next real-life problem. (It is assumed that the reader is familiar with the basic idea of data envelopment analysis: DEA).

8.1. Input optimization in DEA

Consider N decision making units each with m "inputs" $X^i \in \mathbb{R}^m$ and s outputs $Y^i \in \mathbb{R}^s$, $i = 1, ..., N$. The basic objective of DEA is to estimate the "efficiency" of a given decision making unit relative to the set of all decision making units. This can be done by estimating a meaningful "ratio" of outputs over inputs which leads to the Charnes, Cooper and Rhodes tests:

(CCR, k)

$$\max_{(x,y)} (y, Y^k)$$
$$(y, Y^j) \leq (x, X^j), \quad j = 1, ..., N$$
$$(x, X^k) = 1$$
$$x \geq 0, \quad y \geq 0; \quad k = 1, ..., N.$$

We use the notation $(u, v) = u^T v$ for the inner product. For the sake of simplicity we omit a non-Archimedean quantity in the tests. This may lead to more efficient decision making units than obtained by the original formulation. The optimal value of (CCR, k) is called the *efficiency ratio* of the decision making unit DMUk. If the ratio is equal to one, then DMUk is efficient. One can show that *these tests are globally stable* relative to positive input and output data. One of the difficulties with the CCR tests is that they may identify too many units as efficient. In that case one may wish to rank the efficient units by some other criterion, such as "rigidity to data". The idea is, for each efficient DMUk, $k = 1, ..., K$, to solve an optimization problem such as

(k, θ)

$$\max_{(x,y,\theta)} \|\theta\|$$
$$(Y^j(\theta), y) \leq (X^j(\theta), x), \quad j = 1, ..., N, \quad j \neq k$$
$$(X^k, x) = 1, \quad (Y^k, y) = 1, \quad x \geq 0, \quad y \geq 0.$$

Here $\theta \in \mathbb{R}^p$ is considered to be a parameter such that, at $\theta = 0$, (k, θ) is the unperturbed program (CCR, k). The most interesting perturbations are

$$[X^j(\theta)]_i = [X^j]_i - \theta_i, \quad j = 1, ..., N, \quad i = 1, ..., m,$$

$$[Y^j(\theta)]_l = [Y^j]_l + \theta_l, \quad j = 1, ..., N, \quad l = 1, ..., s,$$

for non-negative $\theta_i \geq 0$, $\theta_l \geq 0$. (Here the remaining $N-1$ units, including possibly the inefficient ones, are "attempting" to improve their efficiency standing.) In order to make the results meaningful, the perturbations are required to preserve positivity of the inputs. This is achieved by fixing some positive lower bounds. Also, for the sake of comparison, each input and output is scaled down to the range of numbers between 0 and 1. In the program (k, θ) one is looking for uniformly largest perturbations in θ of all remaining $N-1$ units that preserve the efficiency of DMUk. (Note that the efficiency of DMUk is guaranteed by the constraint $(Y^k, y) = 1$.) For the sake of simplicity, the norm $\|\theta\|$ can be chosen to be l_1 or l_∞,, in which case (k, θ) can be written as a linear program for every fixed θ. For any choice of the norm, (k, θ) is a "partly linear program". The optimal value of the program (k, θ)

is called the *radius of rigidity* of the efficient DMUk. One can calculate this radius by input optimization. Then, instead of solving the nonlinear program in (x, y, θ), one maximizes the optimal value function in θ using an appropriate marginal value formula. A major difficulty with this approach is that Slater's condition is not satisfied for any feasible fixed θ (because of the equality constraints). However, one can derive an appropriate marginal value formula that is suitable for models occurring in DEA. We will derive it for models of the form

(L, θ)

$$\begin{aligned} &\min_{(x)} f(x,\theta) \\ &f^i(x,\theta) \le 0, \ \ i \in P \\ &A(\theta)x = b. \end{aligned}$$

Here we assume that f, $f^i : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $i \in P$ are continuous functions, also $f(\cdot,\theta)$, $f^i(\cdot,\theta) : \mathbb{R}^n \to \mathbb{R}$, $i \in P$ are convex functions for every fixed $\theta \in \mathbb{R}^p$; $A : \mathbb{R}^p \to \mathbb{R}^{m\times n}$ is a continuous matrix function, while $b \in \mathbb{R}^m$ is fixed. For every $\theta \in \mathbb{R}^p$, we use the familiar notation: the feasible set is denoted by $F(\theta) = \{x : f^i(x,\theta) \le 0, i \in P, A(\theta)x = b]\}$ and the set of all optimal solutions $x^o(\theta)$ by $F^o(\theta) = \{x^o(\theta)\}$. Also, for every feasible $\theta \in F = \{\theta : F(\theta) \ne \emptyset\}$,

$$P^<(\theta) = \{i \in P : f^i(x',\theta) < 0, A(\theta)x' = b \quad \text{for some} \quad x'\} \ \text{ and}$$

$$F^=(\theta) = \{x \in \mathbb{R}^n : f^i(x,\theta) = 0, \ \ i \in P\backslash P^<(\theta)\} \cap \{x : A(\theta)x = b\}.$$

We study the behaviour of the optimal value function $f^o(\theta) = f(x^o(\theta),\theta)$ of the model (L,θ) around an arbitrary but fixed $\theta^* \in F$. We will study perturbations in the set

$$S = \{\theta : F^o(\theta^*) \subset F^=(\theta), \ F^o(\theta) \subset F^=(\theta^*)\}.$$

Also, we will use the Lagrangian

$$L_*^<(x,u;\theta) = f(x,\theta) + \sum_{i\in P^<(\theta^*)} u_i f^i(x,\theta).$$

and its saddle point $\{x^o(\theta^*), U^o(\theta^*)\}$. This is a point satisfying

$$L_*^<(x^o(\theta^*),u;\theta^*) \le L_*^<(x^o(\theta^*),U^o(\theta^*);\theta^*) \le L_*^<(x,U^o(\theta^*);\theta^*)$$

for every $u \in \mathbb{R}^c$, $u \ge 0$, where $c = \text{card}P^<(\theta^*)$, and for every $x \in F^=(\theta^*)$.

If the constraints satisfy the generalized Slater condition at θ^* (i.e., if there is an x such that $A(\theta^*)x = b$ and $f^i(x,\theta^*) < 0$, $i \in P$) and if the matrix $A(\theta^*)$ has the full row rank, then the two mappings F and $F^=$ are lower semicontinuous at θ^* relative to the set $S \cap F$. The optimal value function $f^o(\theta)$ can be locally decreased from an arbitrary, but fixed, θ^* along a path for which the limit $s = \lim_{\theta\in S,\theta\to\theta^*} \frac{\theta-\theta^*}{\|\theta-\theta^*\|}$ exists if, along this path, $\nabla_\theta L_*^<(x^o(\theta^*),U^o(\theta^*);\theta^*)s < 0$. This follows from the following theorem:

Theorem 6. (Simplified marginal value formula for input optimization in DEA) *Consider the convex model (L,θ) around some $\theta^* \in F$. Assume that the objective function is realistic at θ^* and that the saddle point $\{x^o(\theta^*), U^o(\theta^*)\}$*

is unique. Also assume that the gradients $\nabla_\theta f(x,\theta)$, $\nabla_\theta f^i(x,\theta)$, $i \in P$ *exist and that they are continuous at* $\{x^o(\theta^*), U^o(\theta^*)\}$. *Finally, assume that the constraints of the program* (L,θ^*) *satisfy the generalized Slater condition and that the matrix* $A(\theta^*)$ *has a full row rank. Then, for every sequence* $\theta \in S \cap F$, $\theta \to \theta^*$, *for which the limit* s *exists, we have the marginal value formula*

$$\lim_{\theta \in S\cap F, \theta\to\theta^*} \frac{f^o(\theta) - f^o(\theta^*)}{\|\theta - \theta^*\|} = \nabla_\theta L_*^<(x^o(\theta^*), U^o(\theta^*); \theta^*)s.$$

A *bad news* is that, when input optimization is applied, the generalized Slater condition assumption may be satisfied at some "old" iteration θ^i, but not at a "new" one θ^{i+1}. Another difficulty with applying it is that, typically, the saddle point is not unique.

Assume that the above programs (k,θ), $k = 1,...,K$, for the K efficient decision making units, are solved by input optimization. The K globally optimal values of these programs are their radii of rigidity. The efficient units can now be ranked by these radii: a unit with the largest radius of rigidity is ranked first, the one with the smallest radius of rigidity is ranked last. Note that the unit with the largest radius of rigidity will keep its efficiency under the largest uniform perturbations of all other $N-1$ units. The ranking depends on a particular type of perturbations used in the programs (k,θ), $k = 1,...,K$. Also, for different types of perturbations one generally obtains different rankings.

Instead of perturbing several or all parameters at the *same* time, one can simplify the numerical effort by perturbing the *same* type of inputs or outputs for every DMU, except the efficient one under the consideration. In particular, one can focus on *only one* specific input or output for every unit. In this case, there is only one direction (non-negative) of improvement in θ of the optimal value function (which is now a function of a scalar variable).

Comment 1. (Cooper's comment) *It was pointed out by* Cooper *[11] that the classification with respect to robustness is "worthwhile because the usual CCR ... models do not admit ranking for either or both of the following reasons:*

(1) *the measure may be incomplete because of the omission of inefficiencies represented by the non-zero slacks and/or*

(2) *the evaluations may be coming from different facets. This means that the evaluations are being affected by reference to different peer groups. For example, a DMU which is* 75% *efficient relative to one peer group is not necessarily less efficient than another DMU which is rated at* 80% *as efficient as another peer group.*

These points were first noted explicitly in Charnes et al. [8]. In a situation described in that paper, the unit costs of excessive inputs and unit prices of revenues losses led to the use of total opportunity cost as a basis for ranking to guide the choice of decision making units which were to be subjected to efficiency audits by the Texas Public Utility Commission. However, such costs prices are not always available, especially in not-for-profit entities, in which case an approach like the robustness tests could be found useful.

Case study 2: Ranking efficiently administered university libraries Let us illustrate the radius-of-rigidity method on a set of 15 university libraries listed below. First, CCR tests are applied with two input (staff and expenditures) and four output (volumes, volumes added, serials, microforms) data for the academic year 1994–95. Five libraries are found to be efficiently administered in that set (Alberta, California at L.A., Hawaii, Illinois at Urbana-Champaign, and Louisiana State.) We wish to perform a post-optimality analysis and rank these five libraries by their rigidity to data. First, each input and output data for all fifteen libraries is normalized to one. Then, for each efficient library , we have found simplified (one parameter) radii of rigidity relative to non-negative perturbations of each input and output. Bounds were imposed to ensure positivity of the input data. Thus, for input X_1 (number of staff), we set $0 \leq \theta \leq 0.3302$, and for input X_2 (expenditure), $0 \leq \theta \leq 0.1914$. For the outputs, upper bounds of 100 were imposed on θ. (This was deemed sufficiently large, given that the data was normalized. When this bound was achieved, we denote it by the sign ∞ in the table.) The radius of rigidity often attained these prescribed bounds. It is interesting to note that situations with zero radii of rigidity also occur. The numerical results are given in *Table 1* below. They are borrowed from the master's thesis Mann [24].

	Outputs				Inputs		Rank
	Y_1	Y_2	Y_3	Y_4	X_1	X_2	
Alberta				Inefficient			
B.C.	∞	∞	∞	∞	0.33	0.191	1-2
Brown				Inefficient			
California, L.A.	∞	∞	0	0.1482	0	0.191	5
California, San Diego				Inefficient			
Connecticut				Inefficient			
Duke				Inefficient			
Guelph				Inefficient			
Hawaii	∞	∞	∞	1.2521	0.33	0.191	3
Illinois, Urbana	∞	∞	∞	∞	0.33	0.191	1-2
John Hopkins				Inefficient			
Laval				Inefficient			
Louisiana State	∞	∞	∞	0.5601	0.33	0.0705	4
McGill				Inefficient			
McMaster				Inefficient			

Table 1. Ranking of efficiently administered libraries

After determining the radius of rigidity for each variable in turn, the following ranking of the five efficient libraries is produced: The universities of Illinois at

Urbana-Champaign and British Columbia are ranked tops, since the radius of rigidity reaches the imposed bounds for all variables. Hawaii would rank below these two, since the bounds were reached in all cases except one and, in that case, the radius of rigidity was better than in all other instances. Louisiana State is next, followed by California since two radii of rigidity of 0 would suggest a precarious efficiency evaluation. The zero radius for University of California library in the third output and the first input means that this library would lose its efficient status if *every other library* in the group of 15 libraries increases its serials or if every library decreases its staff by *any* number. On the other hand its efficiency would not change if the number of volumes or volumes added changes in any other library. No other efficient library in this group is so highly sensitive to perturbations of data.

The efficiency testing and ranking by rigidity of the university libraries could possibly be made more realistic if one distinguished between universities with and without medical schools. One could also include data such as the number of books circulated and the number of students who actually use the library. (These suggestions were made by Ms. V. Blažina, a professional librarian at Universit de Montréal.) The above ideas can also be applied internally, within a university, to its smaller departmental or faculty libraries. The library statistics is readily available from compilations such as the one by Kyrillidou et al. [20]. Studies of libraries using factor analysis use other criteria and generally produces different rankings.

9. Semi-abstract parametric programming

Problems of the form $\min_{(x)} f(x,\theta)$ subject to $x \in F(\theta)$ where $f : X \times \mathbb{R}^p \to \mathbb{R}$ is some continuous function, X is a normed linear vector space, and $\theta \in \mathbb{R}^p$ is allowed to vary over some set $F \in \mathbb{R}^p$, are called *semi-abstract parametric programming models. Semi-abstract parametric programming* is the study of these models.

Illustration 8. (Minimal resistance to a gas flow) *A well-known problem (see, e.g., the text by Krasnov et al. [19]) is to determine the shape of a solid (one can think of the nose of an airplane) that makes the least resistance to a gas flow. Assuming that the solid is obtained by rotation of some shape $x = x(t)$ around the t-axis (see* Figure 11*) one can show, under "idealized" assumptions, that the total force acting on the solid in the positive direction t is*

$$f = 4\pi\rho v^2 \int_0^a \left(\frac{dx}{dt}\right)^3 x dt.$$

Here ρ is the gas density and v is the velocity of the gas relative to the solid.

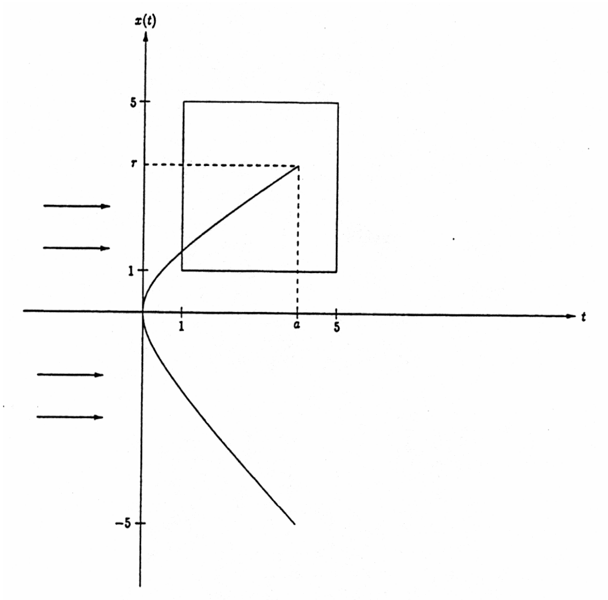

Figure 11. *Optimal shape problem*

We are imposing an extra requirement: The radius r of the solid and its length a should satisfy the conditions $1 \le a \le 5$, $1 \le r \le 5$. *Consider* $\theta = (a, r)^T \in \mathbb{R}^2$ *as the parameter and* $x = x(t)$ *as the "decision variable". The optimal shape problem is the semi-abstract parametric model:*

$$\min_{(x)} f(x, \theta) = 4\pi\rho v^2 \int_0^a \left(\frac{dx}{dt}\right)^3 x dt$$

$$x \in F(\theta) = \{x(t) : \left(\frac{dx}{dt}\right)^3 - 3 \cdot \frac{d}{dt}\left[x \cdot \left(\frac{dx}{dt}\right)^2\right] = 0,\ x(0) = 0,\ x(a) = r\}.$$

The differential equation constraint is the Euler-Lagrange equation from calculus of variations, applied to the functional $f(x, \theta)$. *This equation, under weak assumptions, is a necessary condition for optimality of a function defined on an abstract space.*

A globally optimal parameter θ^* *can be found here in two stages. First, for a fixed* θ, *the differential equation is solved and one obtains an optimal decision variable (function)* $x^o = x^o(t) = r\left(\frac{t}{a}\right)^{\frac{3}{4}}$. *After substituting this solution into the objective function f and integration, we find that the optimal value function is* $f^o(a, r) = \frac{27}{16}\pi\rho v^2 \frac{r^4}{a^2}$. *Minimization of this function on the constrained set yields the* optimal parameters $a^* = 5$ *and* $r^* = 1$. *From here, after a back-substitution, one finds that* $x^* = x^*(t) = \left(\frac{t}{5}\right)^{\frac{3}{4}}$ *is the* optimal shape *of the solid satisfying the constraints and that the least resistance to the gas flow, i.e., the optimal value of the model, is* $f^* = f^o(a^*, r^*) = \frac{27}{400}\pi\rho v^2$.

10. General applications

There are two kinds of general applications of parametric programming. Given an applied mathematics model, parametric programming provides a methodology to

(1) search for an optimal parameter, i.e., to *"optimize"* the model and/or

(2) study *stability (reliability)* of the model. (An objective function may not be required).

The two objectives ideally can be combined into *"stable parametric programming"*. Here one optimizes the optimal value function by stable perturbations of the parameter. An optimal parameter depends on the initial condition (starting parameter) and on a particular class of perturbations used.

Concrete applications of parametric programming include problems from

- physics, especially mechanics and equilibria problems. Classical problems, such as the least gas or water resistance achieved by varying the shape of an object, or finding the shortest time of descent subject to an obstruction, are usually studied in calculus of variations. Many of these can be formulated as semi-abstract parametric programming.

- chemistry. Problems here involve multi-stage heat exchanger designs (often formulated as partly linear or partly convex programs, e.g., Avriel and Williams [1], Mustapić et al. [28]) and configuration of clusters of atoms and molecules (minimization of Lennard-Jones interaction potential among spherical particles, e.g., Maranas and Floudas [27].)

- ill-posed problems in the sense of Hadamard (including problems with differential and integral equations). It has been observed by Tikhonov and Arsenin [39] p.xi that there are many such important problems. We quote the authors: " ...we shall show in the present book that the class of ill-posed problems includes many classical mathematical problems and, most significantly, that such problems have important applications..". In the introduction of their book Fritz John says: "One might say that the majority of applied problems are, and always have been, ill-posed, particularly when they require numerical answers." We note that ill-posed problems are described by models that are unstable or do not have unique (optimal) solutions. In the latter case, a higher level model is generally unstable.

- linear and nonlinear programming. This is a traditional area of parametric programming. Sample topics include: sensitivity analysis, shadow prices, DEA, degenerate linear programs, problems of simultaneously finding optimal prices *and* decision variables, pooling and blending in oil refineries (these are partly linear programs), understanding "ambiguities" when the optimal solutions jump but the optimal values remain relatively close, etc. (The latter is a consequence of the fact that the optimal solutions mapping is closed, but not necessarily continuous. These phenomena were reported in solving power system problems in electrical engineering, also in Tikhonov and Arsenin [39], Chapter vii.)

- zero-one linear programs. Mixed zero-one linear programs can be formulated as convex models.

- convergence of numerical algorithms. The *rate of convergence* of various algorithms can be determined using continuity properties of point-to-set mappings. Also, one can study convergence of discretization schemes for general abstract problems.

- parameter identification problems. This is another huge area that spans from parameter identification problems in differential equations, e.g., Scitovski and Jukić [34] to determination of kinetic constants in biological sciences, e.g., Dikšić [12], and study of mathematical models in hemodialysis , e.g., Sotirov et al. [37]. The area also includes generalized best approximation problems, e.g., Golub and Van Loan [15] and Jukić et al. [18]. Here the "parameters" (data) are optimized within specific bounds to find a best fit.

- economics, finance, management. These are "standard territories" for applications of parametric programming. Applications include the study of multilevel models, including von Stackelberg games, Leontiev models, e.g., Jemrić [17], and methods in portfolio optimization , e.g., Dupačova [13]. In particular, one could use stability aspects of parametric programming to understand and decrease "nervousness" of dynamic models used in dynamic lot-size algorithms where forecasts of future parameter values are frequently updated, e.g., Charlson et al. [6].

11. Some open problems

There are many open problems and territories yet to be studied in parametric programming. We are going to mention only some of those that are closely related to the topics introduced in this tutorial.

- Connection with nonlinear programs. It is shown by Liu and Floudas [23] that every nonlinear program with twice continuously differentiable functions can be transformed into a partly convex program. These programs are closely related to convex models. This means that *by studying convex models one can actually study a large class of general optimization problems.* The constructive aspects of the link (optimality, stability, numerical methods) are still mainly unexplored. (Unfortunately, some nice and useful properties of convex models are lost in the transformation.)

The next two problems are related to linear programming and, so far, they have been only partly solved in a constructive way:

- Basic problem I: Given a linear program in the canonical form with a full row rank coefficient matrix. Choose coefficients to be perturbed (e.g., several technological coefficients in the matrix). It is required to construct perturbations for which the feasible set mapping is (not) lower semicontinuous.

Remark 5. *Robinson [32] considers simultaneous perturbations of all coefficients and shows that, essentially, the existence of a positive $x > 0$ feasible point (i.e., the* *generalized Slater's condition)* is necessary and sufficient *for stability (also called* "regularity"*). However, his result is only sufficient for lower semi-continuity of the feasible set mapping for perturbations of specific coefficients. A stronger sufficient condition for lower semi-continuity (that implies regularity) is the existence of* at least one non-degenerate basic feasible solution. *A constructive* necessary *condition for lower semi-continuity for perturbations of specific coefficients can be given in terms of a subset of the index set of the decision variable, called the* "minimal index set of active variables"*. This result appears to have a potential for developing numerical methods for identifying unstable perturbations of specific coefficients. (This author gave a talk on this topic at the* Sixth International Conference on Parametric Optimization and Related Topics*, Dubrovnik, Croatia, October 4-8, 1999.).*

The second problem is related to the "partly" linear program

(PL, θ)

$$\min_{(x,\theta)} f(x,\theta) \qquad \text{subject to}$$
$$f^i(x,\theta) \le 0, \;\; i \in P = \{1, ..., m\}.$$

Here $f(\cdot,\theta)$, $f^i(\cdot,\theta) : \mathbb{R}^p \to \mathbb{R}$, $i \in P$, are linear functions for every $\theta \in \mathbb{R}^p$.

- Basic problem II: Given an optimal solution (x^*, θ^*) of the partly linear program (PL, θ). Specify coefficients to be perturbed. For what perturbations of $\theta \in \mathbb{R}^p$, emanating from θ^*, the point (x^*, θ^*) remains globally optimal for the program (PL, θ)?

- Stable planning is a term used to describe a particular type of stable parametric programming: Consider a convex model (P, θ) running with some fixed parameter θ'. Suppose that one wishes to achieve a prescribed goal (e.g., a prescribed plan or a profile of production) x^* which is *not* in the feasible set of the program (P, θ'), i.e., $x^* \notin F(\theta')$. One wishes to achieve the goal by perturbing the parameters. The *stable planning problem* is the problem of *how to change prescribed coefficients in a stable way* from the initial θ' to some θ^*, *so that x^* become feasible: $x^* \in F(\theta^*)$.*

  The problem can be formulated as a stable parametric model of the form

  (SP, θ)

  $$\min_{(x)} \|x - x^*\|$$
  $$f^i(x,\theta) \le 0, \;\; i \in P.$$

  Here the objective function is a norm that measures the distance between the prescribed point x^* and the feasible set $F(\theta)$. If, at a globally optimal input θ^*, the optimal value function has a positive value, then the goal x^* cannot be reached by a stable path emanating from θ', i.e., $x^* \notin F(\theta^*)$. A corresponding optimal solution $x^o = x^o(\theta^*) \in F(\theta^*)$ of the program (SP, θ^*) then is a *"best approximate solution"* (relative to the initial θ', the norm, and the class of stable perturbations used). If the optimal value is zero at θ^*, then $x^o(\theta^*) = x^* \in F(\theta^*)$ and θ^*, together with the path $\theta' \to \theta^*$, is an optimal solution of the stable planning problem. (Some case studies have been solved, e.g., "stable planning of university admission"; e.g., Leger [21].)

- Inverse stable planning: Consider a convex model (P, θ) running with some parameter θ'. Let x' be its fixed feasible decision variable. Assume that x' is a point which is *not* an optimal solution of the program (P, θ'), but it is a "desirable" point which one wants to make optimal by changing parameters. In "stable inverse programming" one attempts to determine a stable path, leading from θ' to some θ^*, such that x' become an optimal solution of the program (P, θ^*).

  This problem can be formulated as a stable programming problem and solved by input optimization. One can use the model

$$\begin{aligned}&\min_{(d,\delta)} \nabla f(x', \theta)d \\ &\nabla f^i(x', \theta)d + \epsilon\|d - \delta^i\| \leq 0, \quad i \in P(x', \theta) \\ &\|d\| \leq 1, \quad \delta^i \in D_i^=(x', \theta), \quad \|\delta^i\| \leq 1, \quad i \in P(x', \theta).\end{aligned}$$

  Here $\epsilon > 0$ is an auxiliary scalar parameter, $P(x', \theta)$ is the set of active constraints at x' of (P, θ), and $D_i^=(x', \theta)$ is the cone of directions of constancy of $f^i(x, \theta)$ at x', $i \in P(x', \theta)$; all for a fixed θ. We know that, for a fixed θ, the point x' is an optimal solution of the program (P, θ) if, and only if, there exists a positive scalar $\epsilon^* > 0$ such that the optimal value of the program $(P, \theta; \epsilon)$ is zero for every $0 < \epsilon \leq \epsilon^*$. Hence, in order to make x' optimal, *one should bring the optimal value of the model to zero for all sufficiently small* $\epsilon > 0$*, by stable perturbations of* θ, starting from, say, θ'. If the constraints are LFS (in particular linear) in the x variable, then the problem is significantly simplified. In this case, instead of the above model, one can use the model

$$\begin{aligned}&\min_{(d)} \nabla f(x', \theta)d \\ &\nabla f^i(x', \theta) \leq 0, \quad i \in P(x', \theta), \quad \|d\| \leq 1.\end{aligned}$$

  Now one has to find a path connecting the initial θ' to some θ^* at which the optimal value of the program

$$\begin{aligned}&\min_{(d)} \nabla f(x', \theta^*)d \\ &\nabla f^i(x', \theta^*) \leq 0, \quad i \in P(x', \theta^*), \quad \|d\| \leq 1\end{aligned}$$

  is equal to zero. Then $x' \in F(\theta^*)$ is an optimal solution of the program (P, θ^*).

- Efficient input optimization methods, especially for bilevel models (von Stackelberg games) and multi-objective multilevel stable models, are yet to be developed.

  Remark 6. *Path-following methods, based on* nonlinear programming optimality conditions*, are described in, e.g., Guddat et al. [16].*

- Inverse multi-objective multilevel stable models; this is, essentially, terra incognita even for linear and convex models.

- Constructive study of optimality and stability in abstract formulations of parametric programming. Some results do exist for abstract von Stackelberg games; e.g., Lignola and Morgan [22].

- Controls and differential games. An important problem (the feedback law) can be formulated as follows: Given $T \in \mathbb{R}$ and $(\tau, \alpha) \in (-\infty, T] \times \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^{n+1} \to \mathbb{R}^n$, solve

$$\begin{aligned} &\min_{s.t.} f(x(T)) \\ &\tfrac{dx}{dt} \in G(t, x(t)), \quad x(\tau) = \alpha. \end{aligned}$$

  This is, essentially, a semi-abstract parametric programming model where the initial condition $(\tau, \alpha) = \theta$ can be considered as a parameter. Very little is known about the optimal value function and stability (in the sense of *Definition 7.*) and virtually nothing about optimal parameters; e.g., Clarke et al. [9]. General optimal control problems can be formulated as abstract parametric programming models. Then an optimal control is identified as an optimal parameter. Stable abstract parametric programming recovers " stable optimal control theory". This approach does not yet seem to have been studied.

- Study of LFS functions. LFS functions are all linear and many nonlinear convex functions. These are important functions for optimization, because optimality conditions in convex modelling with LFS constraints do not require a "regularization condition". Many of these functions f have a property that $\nabla f(x) = \alpha(x)c$, where $\alpha(x)$ is a scalar function and c is a constant vector depending only on f. LFS functions are not yet fully explored; e.g., Sharifi Mokhtarian and Zlobec [36], Neralić and Zlobec [31], and Trujillo-Cortez [41].

References

[1] M. Avriel, A. C. Williams, *An extension of geometric programming with applications in engineering optimization*, Journal of Engineering Mathematics **5**(1971), 187–194.

[2] B. Bank, J. Guddat, D. Klatte, B. Kummer, K. Tammer, *Nonlinear Parametric Optimization*, Akademie–Verlag, Berlin, 1982.

[3] P. Becker, *Peoples of Southern Africa*, The Star, Johannesburg, (A collection of 45 newspaper articles), 1971.

[4] C. Berge, *Topological Spaces*, Oliver and Boyd, London, 1963.

[5] M. P. Brunet, *Numerical Experimentations with Input Optimization*, M. Sc. Thesis, Department of Mathematics and Statistics, McGill University, 1989.

[6] R. C. Carlson, J. V. Jucker, D. H. Kropp, *Less nervous MRP systems: a dynamic economic lot-sizing approach*, Management Science **25**(1979), 745–761.

[7] A. CHARNES, W. W. COOPER, M. SEARS, S. ZLOBEC, *Efficiency evaluations in perturbed data envelopment analysis*, in: Parametric Optimization and Related Topics II, (J. Guddat et al. Eds.), Akademie Verlag Series: Mathematical Research **62**(1991), 38–49.

[8] A. CHARNES, W. W. COOPER, D. DIVINE, T. W. RUEFFI, D. THOMAS, *Comparison of DEA and egzisting ratio and regression systems for effecting efficiency eisluations of regulated cooperatives in Texas*, Research in Governmountal and Bonprofit Accounting **5**(1989), 187–211.

[9] F. H. CLARKE, YU. S. LEDYAEV, R. J. STERN, P. R. WOLENSKI, *Nonsmooth Analysis and Control Theory*, Graduate Texts in Mathematics, Springer Verlag, Vol. **178**, 1998.

[10] L. COLLATZ, W. WETTERLING, *Optimization Problems*, Springer-Verlag, Applied Mathematical Sciences, Vol. 17, New York, 1975.

[11] W. W. COOPER, *Private correspondence*, March, 1998.

[12] M. DIKŠIĆ, *The use of optimization procedures in estimation of parameters of the brain biological systems*, lecture presented at the Sixth International Conference on Parametric Optimization and Related Topics, Dubrovnik, Croatia, October 4-8, 1999.

[13] J. DUPAČOVA, *Portfolio optimization via stochastic programming: methods of output analysis*, two tutorial lectures presented at the Sixth International Conference on Parametric Optimization and Related Topics, Dubrovnik, Croatia, October 4-8, 1999. (Published in Mathematical Methods of Operations Research **50**(1999), 245–270.)

[14] I. I. EREMIN, N. N. ASTAFIEV, *Introduction to the Theory of Linear and Convex Programming*, Nauka, Moscow, 1976. (In Russian)

[15] G. H. GOLUB, C. F. VAN LOAN, *An analysis of the total least squares problem*, SIAM Journal of Numerical Analysis **17**(1980), 883–893.

[16] J. GUDDAT, F. GUERRA VASQUEZ, H. TH. JONGEN, *Parametric Optimization: Singularities, Pathfollowing and Jumps*, Teubner, Stuttgart and John Wiley, New York, 1990.

[17] I. JEMRIĆ, *Stable planning in Leontiev's models*, Proceedings of SYM-OP-IS' 88, Brioni, Croatia, 1988., pp. 139–142.

[18] D. JUKIĆ, R. SCITOVSKI, Š. UNGAR, *The best total least squares line in* $\mathbb{R}^3$, in: Proceedings of the 7-th International Conference on Operational Research - KOI'98, (I. Aganović, T. Hunjak and R. Scitovski, Eds.), Croatian Operational Research Society, Osijek, 1999, 311–316.

[19] M. L. KRASNOV, G. I. MAKARENKO, A. I. KISELEV, *Problems and Exercises in the Calculus of Variations*, Mir Publishers, Moscow, 1975.

[20] M. Kyrillidou, K. Maxwell, K. Stubbs(Editors), *ARL Statistics 1994-1995: A Computation of Statistics from the One Hundered and Nineteen Members of the Association of Research Libraries*, Association of Research Libraries, Washington, D.C., 1996.

[21] A. Leger, *Stability of a Mathematical Model for Admissions Planning at University*, M. Sc. Thesis, McGill University, 1983.

[22] M. B. Lignola, J. Morgan, *Topological existence and stability for Stackelberg problems*, Journal of Optimization Theory and Applications **84**(1995), 145–169.

[23] W. B. Liu, C. A. Floudas, *A remark on the GOP algorithm for global optimization*, Journal of Global Optimization **3**(1993), 519–521.

[24] G. M. Mann, *Efficiency Evaluations of North American University Libraries*, M. Sc. thesis, McGill University, Montreal, 1997.

[25] G. Mann, S. Zlobec, *Ranking efficiency by robustness in data envelopment analysis*, paper presented at the 20th International Symposium on Mathematical Programming with Data Perturbations, Washington, D.C., May 1998.

[26] A. S. Manne, *"Notes on parametric linear programming,"* RAND Report P-468, The Rand Corporation, Santa Monica, California, 1953.

[27] C. D. Maranas, C. A. Floudas, *A global optimization approach for Lennard-Jones microclusters*, Journal of Chemical Physics **97**(1992), 7667–7678.

[28] N. Mustapić, N. Šakić, Ž. Slavica, *Application of mathematical programming in the synthesis and analysis of process plant heat recovery network*, in: Proceedings of the 7-th International Conference on Operational Research - KOI'98, (I. Aganović, T. Hunjak and R. Scitovski, Eds.), Croatian Operational Research Society, Osijek 1999, 145–158.

[29] T. H. Naylor, E. T. Byrne, J. M. Vernon, *Introduction to Linear Programming: Methods and Cases*, Wadsworth, Belmont, California, 1971.

[30] L. Neralić, *Sensitivity in data envelopment analysis for arbitrary perturbations of data*, Glasnik Matematički. **32**(1997) (52), 315–335.

[31] L. Neralić, S. Zlobec, *LFS functions in multi-objective programming*, Applications of Mathematics **41**(1996), 347–366.

[32] S. Robinson, *Stability theory for systems of inequalities. Part I: Linear systems*, SIAM J. Numerical Analysis **12**(1975), 754–769.

[33] S. Robinson, *Stability theory for systems of inequalities. Part II: Differentiable nonlinear systems*, SIAM J. Numerical Analysis **13**(1976), 497–513.

[34] R. Scitovski, D. Jukić, *A method for solving the parameter identification problem for ordinary differential equations of the second order*, Appl. Math. Comput. **74**(1996), 273–291.

[35] L. Seiford, *A Bibliography of Data Envelopment Analysis 1978-1996*, Department of IEOP, The University of Massachusetts, Amherst, 1996.

[36] F. Sharifi Mokhtarian, S. Zlobec, *LFS functions in nonsmooth programming*, Utilitas Mathematica **45**(1994), 3–15.

[37] R. Sotirov, M. Jakić, M. Pavleković, *Mathematical models of hemodialysis*, Proceedings of the 7-th International Conference on Operational Research - KOI'98, (I. Aganović, T. Hunjak and R. Scitovski, Eds.), Croatian Operational Research Society, Osijek 1999, 209–218.

[38] von H. Stackelberg, *Marktform und Gleichgewicht*, Springer, Berlin, 1934. (Translation: *The Theory of the Market Economy*, Oxford University Press, Oxford, 1982.)

[39] A. N. Tikhonov, V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston & Sons, Washington, D.C., 1977.

[40] P. A. Tipler, *Physics for Scientists and Engineers (Volumes 1 and 2)*, Worth Publ., New York, 1991.

[41] R. Trujillo-Cortez, *LFS Functions in Stable Bilevel Programming*, M.Sc. Thesis, McGill University, Department of Mathematics and Statistics, Montreal, 1997.

[42] E. Zermelo, *Über das Navigationsproblem bei ruhender oder veranderlicher Windverteilung*, Zeitschrift für Angewandte Mathematik und Mechanik **11**(1931), 114–124.

[43] S. Zlobec, *Economizing under the conditions of restructuring*, in: Proceedings of the International Conference: Restructuring Transitional Economies (S. Sharma and P. Sikavica, Eds.), Faculty of Economics, University of Zagreb, v. **I**(1995), 453–467.

[44] S. Zlobec, *Stable parametric programming*, Optimization **45**(1999), 387–416. Augmented version of this paper to be published as a textbook.

[45] S. Zlobec, *Nondifferentiable optimization: Parametric programming*, in: Encyclopaedia of Optimization, (C. A. Floudas and P. M. Pardalos, Eds.), Kluwer Academic, 1999. (Forthcoming)

The reader is referred to the last two references for more literature on the subject.