Journal of Global Optimization, **29:** 97–112, 2004.

Saddle-Point Optimality: A Look Beyond Convexity*

S. ZLOBEC
McGill University

(Received 27 August 2002; accepted 1 September 2003)

Abstract. The fact that two disjoint convex sets can be separated by a plane has a tremendous impact on optimization theory and its applications. We begin the paper by illustrating this fact in convex and partly convex programming. Then we look beyond convexity and study general nonlinear programs with twice continuously differentiable functions. Using a parametric extension of the Liu-Floudas transformation, we show that every such program can be identified as a relatively simple structurally stable convex model. This means that one can study general nonlinear programs with twice continuously differentiable functions using only linear programming, convex programming, and the inter-relationship between the two. In particular, it follows that globally optimal solutions of such general programs are the limit points of optimal solutions of convex programs.

Key words: Convex model, global optimum, Liu-Floudas transformation, structural stability.

1. Introduction

Consider mathematical programs of the form

$$\text{(NP)} \quad \begin{aligned} &\operatorname{Min} f(z) \\ &\text{s.t.} \\ &f^i(z) \leqslant 0, \quad i \in P = \{1, \ldots, m\} \end{aligned}$$

where the functions are defined on the entire space R^n and continuous. We denote the feasible set by

$$F = \{z : f^i(z) \leqslant 0, \ i \in P\}$$

and consider a point $z^* \in F$. The basic problem in optimization is to find conditions under which z^* locally or globally optimizes f on F. The idea is to use properties of the objective function, constraint functions and the feasible set.

It is important to describe optimal solutions for several reasons, including

(i) verification whether a numerical solution is optimal or 'close' to an optimum;
(ii) formulation of numerical methods;

---

* Research partly supported by an NSERC of Canada grant. This paper is based on a lecture given by the author on the occasion of Professor Lj. Martić's 80th birthday at KOI'02 in Trogir, Croatia.

(iii) economic interpretation of Lagrange multipliers (as economic indicators);
(iv) obtaining additional (often non-intuitive) information about the system;
(v) formulation of 'duality' theories.

Optimality of z^* can be fully described, in many situations, using the basic geometric fact that two disjoint convex sets can be separated by a plane, called 'hyperplane' in R^n, $n \geqslant 3$. This fact is known as the Hyperplane Separation Theorem. Moreover, the slopes of the hyperplane are known as 'Lagrange multipliers' and they typically have an economic interpretation. (In linear programming these are 'shadow prices', in convex programming these are 'values' of the constraints; e.g., [14].) By separating the two sets, a saddle point is created in a space of higher dimension that divides the problem into two parts: an optimization problem in the decision variables z for some 'slope' u and an optimization problem in terms of the slopes for a fixed optimal solution z^*. This leads to 'dual' formulations of the original problem.

The structure of the paper is as follows: First we will recall some basic results on saddle-point optimality from convex and partly convex programming. Although these results are known, they are recalled here to familiarize ourselves with the notation and the results that will also be used in the general case. We begin with convex programs in Section 2. The characterization of optimality given here has been around since at least mid-1970's, e.g., [2, 14]. 'Convexity' is a magic word in optimization, because local and global optima coincide, so one talks only about an 'optimal solution'. Convex programming is the most thoroughly studied area of nonlinear optimization. Therefore it is important to know if some properties of convex programs carry on to nonconvex programs. In Section 3 we consider partly convex (PC) programs and characterize their global and local optima. (PC) are programs that become convex after 'freezing' some variables. A global characterization of optimality is possible for these programs on the 'region of cooperation' of a feasible z^*, e.g., [13]. However, a characterization of local optimality requires that the feasible set mapping be open, e.g., [11, 14]. Every 'classical' result on optimality in (PC) programming, which is either necessary or sufficient, is a particular case of these results. In Section 4 we introduce one such result. It is a modification of a well-known necessary conditions for a local optimum proved in [9] by the Implicit Function Theorem and Farkas' Lemma. In Section 5 we look beyond convexity and partial convexity. We study the general (NP) and characterize its global optimum. This is possible using a parametric extension of the Liu-Floudas transformation [6]. Using this extension we show that general programs with twice continuously differentiable functions can be identified as partly convex programs in a space of twice-higher dimension. Hence the characterizations of optimality for general programs follow from those given for usual partly convex programs. In particular, a global optimum of a general program with twice continuously differentiable functions is obtained as the limit of optimal solutions of particular convex programs.

Another important property of the parametric Liu-Floudas transformation is that it always yields a structurally stable model! Hence this transformation can be used to stabilize an arbitrary mathematical program. This has been essentially an open problem even in linear programing; e.g., [14, 15]. It is well known that augmentations by artificial and slack variables do not make an unstable model stable. Let us recall that 'structurally stable' models are those for which the feasible set changes 'continuously' with data. Under continuity of the constraints assumption, this is equivalent to saying that the feasible set mapping is open (or lower semi-continuous); e.g. [11, 12, 14, 15], also [4] and Section 3 below. The saddle-point optimality conditions given hereby are full characterizations of optimality. They are not always simple and some extra work may be required to make them useful.

2. Convex Programming

These are programs (NP) where all functions are assumed to be convex. In order to characterize optimality of a feasible point z^*, we use the index set

$$P^= = \{i \in P : z \in F \Rightarrow f^i(z) = 0\}$$

called the *minimal index set of active constraints.*

The results on optimality (and stability, e.g., [14]) are significantly simplified if Slater's condition holds. (It is the condition that there exists a point z' such that $f^i(z') < 0, i \in P$. This is true if, and only if, $P^= = \emptyset$.) The index set P can always be represented as the union of the two sets: $P^=$ and $P^< = P \backslash P^=$, i.e., $P = P^= \cup P^<$. When the index set $P^=$ is known, then one can introduce the set

$$F^= = \{z \in R^n : f^i(z) \leqslant 0, i \in P^=\}.$$

This is a convex set that contains the feasible set F. Since every constraint f^i, with $i \in P^=$, has the property that $f^i(z) = 0$ for every $z \in F$, one can replace $f^i(z) \leqslant 0$ by $f^i(z) = 0$, $i \in P^=$. Hence the inequalities in the definition of $F^=$ can be replaced by equations, i.e.,

$$F^= = \{z \in R^n : i \in P^=\}.$$

If Slater's condition holds, then $F^= = R^n$.

Remark. Situations where Slater's condition is not satisfied (i.e., when $P^= \neq \emptyset$) include programs with at least one linear equation as a constraint, programs describing bi-level decision making processes, such as von Stackelberg games of market economy, lexicographic problems, and some formulations of multi-objective problems.

In the formulation of optimality we will use the Lagrangian function

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

Cardinality of the index set $P^{<}$ is denoted by the letter c, i.e., c = card $P^{<}$, and the non-negative orthant in R^c is denoted by

$$R^c_+=\{u=(u_1,u_2,\ldots,u_c)^T\in R^c : u_i\geqslant 0,\ i=1,\ldots,c\}.$$

THEOREM 2.1 (Characterization of optimality in convex programming). *Consider the convex program* (NP). *A point $z^*\in F^=$ is an optimal solution if, and only if, there exists $u^*\in R^c_+$ such that*

$$\mathscr{L}^{<}(z^*,u)\leqslant\mathscr{L}^{<}(z^*,u^*)\leqslant\mathscr{L}^{<}(z,u^*)$$

for every $u\in R^c_+$ and every $z\in F^=$.

Proof. See, e.g., [14]. The two convex sets to be separated here are

$$C_1=\{y : y\geqslant[f(z),f^1(z),\ldots,f^c(z)]^T \text{ for at least one } z\in F^=\}$$

and

$$C_2=\{y : y<[f(z^*),0,\ldots,0]^T\}$$

in R^{c+1}. (The vector ordering is taken component-wise.) Since $F^=$ is a convex set, so is C_1. Convexity of C_2 is obvious. In the proof of necessity we note that $C_1\cap C_2=\varnothing$. (Otherwise z^* is not optimal.) After separation the leading coefficient is non-zero by the requirement $z\in F^=$ and all slopes are non-negative by the unboundedness of C_2. The sufficiency proof is straightforward.

Remark. The 'reduced' Lagrangian $\mathscr{L}^{<}(x,u)$ can be replaced by the 'classical' Lagrangian

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

However, the set $F^=$ must still be used!

When the constraints satisfy Slater's condition then one can drop the assumption $z^*\in F^=$ (i.e., one can replace it by $z^*\in R^n$). This result is equivalent to the well-known Karush-Kuhn-Tucker theorem. For the differentiable and subdifferentiable versions of these results see, e.g., [2, 14].

3. Partly Convex Programming

These are programs (NP) in which $z=(x,\theta)$ and we make a distinction between the 'state variable' x and 'control variable' (or 'parameter') θ. The program can be rewritten as

$$
\begin{array}{ll}
 & \operatorname{Min} f(x,\theta) \\
(\mathrm{PC},\theta) & \text{s.t.} \\
f^i(x,\theta)\leqslant 0, & i\in P=\{1,\dots,m\}.
\end{array}
$$

Minimization is carried out simultaneously relative to both variables x and θ. We say that the program is *partly convex* (PC) if the functions $f(\cdot,\theta)$, $f^i(\cdot,\theta)$: $R^n\to R, i\in P$ are convex for every $\theta\in R^p$.

There are many models that lead to (PC,θ) and a natural distinction between x and θ. They range from multi-stage heat exchanger problems in chemical engineering and pooling and blending in oil refineries to configuration of clusters of atoms and molecules and isoperimetry problems in geometry, e.g., [14]. In particular, finite-dimensional optimal control problems can be formulated as (PC,θ), e.g., [3, 9, 10].

Now consider the (PC) program in the form (PC,θ) and its arbitrary feasible point $z^*=(x^*,\theta^*)$, where $x^*=x^o(\theta^*)$ is an optimal solution of the convex program (PC,θ^*). We wish to give a condition which is both necessary and sufficient for its *global* optimality. Following [13], we state the condition over the 'region of cooperation' of z^*. Let us recall how this region is defined. First, denote the feasible set of (PC,θ) by

$$Z=\{(x,\theta)\in R^N: f^i(x,\theta)\leqslant 0,\ i\in P\}.$$

For every fixed θ denote the feasible set in x by

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

Also

$$P^=(\theta)=\{i\in P: x\in F(\theta)\Rightarrow f^i(x,\theta)=0\}, \text{the 'minimal index set of active constraints'}$$

and

$$P^<(\theta)=P\backslash P^=(\theta).$$

Given a feasible $\theta^*\in F=\{\theta: F(\theta)\neq\varnothing\}$, denote

$$K(\theta^*)=\{\theta\in F: P^=(\theta)\subset P^=(\theta^*)\}$$

and, given the feasible point $z^*=(x^*,\theta^*)\in Z\subset R^N$, consider the set

$$\Omega(z^*)=\{z=(x,\theta): x\in F(\theta)\quad \text{for } \theta\in K(\theta^*)\}.$$

The above objects $F(\theta), P^{=}(\theta), P^{<}(\theta), K(\theta^*), \Omega(z^*)$ can be considered as point-to-set mappings, i.e., as F: $\theta \to F(\theta)$, etc. They are important in the study of 'structural stability' of parametric models; e.g., [8, 13–15].

The Lagrangian that is used in checking global optimality of $z^* = (x^*, \theta^*)$ is of the form

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

We study its behaviour on a set of (x, θ) in R^N, determined by $\theta \in K(\theta^*)$ and the corresponding $x = x(\theta) \in F_*^=(\theta)$, i.e., on

$$\{F_*^=(\theta), K(\theta^*)\}$$

where

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

Cardinality of the set $P^<(\theta^*)$ is denoted by the letter c and the non-negative orthant in R^c by R_+^c.

THEOREM 3.2 (Characterization of a global optimum in partly convex programming; e.g., [13, 14]). *Consider the partly convex program* (PC,θ) *and its feasible point* $z^* = (x^*, \theta^*)$. *Suppose that* x^* *is an optimal solution of the convex program* (PC,θ^*). *Then* z^* *is globally optimal on its region of cooperation* $\Omega(z^*)$ *if, and only if, there exists a vector function* U: $K(\theta^*) \to U(\theta) \in R_+^c$ *such that*

$$L_*^<(z^*,u) \leqslant L_*^<(z^*, U(\theta^*)) \leqslant L_*^<(z, U(\theta))$$

for every $z = (x, \theta) \in \{F_*^=(\theta), K(\theta^*)\}$ *and* $u \in R_+^c$.

In order to characterize *local* minima, we need two extra assumptions: uniqueness of the optimal solution x^* of the convex program (PC,θ^*) and openness of the feasible set mapping F: $\theta \to F(\theta)$. We recall that the point-to-set mapping F: $R^p \to R^n$ is open at $\theta^* \in R^p$ if, given any point $x^* \in F(\theta^*)$ and any sequence $\theta^k \to \theta^*$, there is a sequence $x^k \in F(\theta^k)$ such that $x^k \to x^*$.

THEOREM 3.3 (Characterization of a local optimum in partly convex programming; e.g., [13, 14]). *Consider the partly convex program* (PC,θ) *and its feasible point* $z^* = (x^*, \theta^*)$, *where* x^* *is a unique optimal solution of the convex program* (PC,θ^*). *Assume that the feasible set mapping* F *is open at* θ^* *relative to* F. *Then* z^* *is a local minimum if, and only if, there exists a vector function* U: $F \cap N(\theta^*) \to R_+^c$ *such that*

$$L_*^<(z^*,u) \leqslant L_*^<(z^*, U(\theta^*)) \leqslant L_*^<(z, U(\theta))$$

for every $z=(x,\theta)\in\{F_*^{=}(\theta), F\cap N(\theta^*)\}$ *and* $u\in R_+^c$, *where* $N(\theta^*)$ *is some neighbourhood of* θ^*.

Similarly to the convex case, let us observe that the saddle-point inequality in the above theorems is restricted to $x\in F_*^{=}(\theta)$ and not to $x\in R^n$.

Remark. A relationship between local optima and the feasible set mapping has been studied in, e.g., [14]. How to characterize local optima in PC programming when the feasible set mapping is not open appears to be an open question; see [8].

EXAMPLE 3.4. Consider a discrete optimal control problem, borrowed from [9] with a slight adjustment (an equation is replaced by inequality):

$$\begin{aligned}&\mathrm{Min} - x\theta_1\theta_2\\ &\quad x^2+2\theta_1^2+3\theta_2^2\leqslant 27 \quad x\geqslant 1,\ 1\leqslant\theta_1\leqslant 3/\sqrt{2},\ 1\leqslant\theta_2\leqslant 2.\end{aligned}$$

We wish to know whether $\theta_1^*=3/\sqrt{2}$, $\theta_2^*=\sqrt{3}$ is a locally or globally optimal control with the corresponding optimal state $x^*=3$? Since the program is PC, we can use Theorem 3.3 to check local optimality. According to this theorem, local optimality is equivalent to the existence of a non-negative function $U=U(\theta)$ such that

$$-27\leqslant Ux^2-\theta_1\theta_2 x+U(2\theta_1^2+3\theta_2^2-27)$$

for every x and all feasible θ's close to θ^*. (This U corresponds to the nonlinear constraint.) The right-hand side function is quadratic convex in x and its values are bigger than or equal to -27 if

$$U=\theta_1^2\theta_2^2/4(2\theta_1^2+3\theta_2^2).$$

This establishes local optimality of θ^*. Moreover, since $K(\theta^*)=\{\theta: 2\theta_1^2+3\theta_2^2<26\}$ and this set contains all $1\leqslant\theta_1\leqslant 3/\sqrt{2}, 1\leqslant\theta_2\leqslant 2$, Theorem 3.2 confirms also global optimality.

Remark. Possibly a more natural formulation of optimal control (abbreviated: OC) problems uses the optimal value function $f^o(\theta)$ defined as

$$\begin{aligned}f^o(\theta)=\min f(x,\theta)&\\ \text{s.t.}&\\ x\in F(\theta).&\end{aligned}$$

In the study of $f^o(\theta)$ one is interested in calculating and characterizing its local and global optima on the set of feasible parameters $F=\{\theta : F(\theta)\neq\varnothing\}$, i.e.,

$$\text{(PP)}\quad \begin{array}{l}\min f^o(\theta)\\ \text{s.t.}\\ \theta\in F.\end{array}$$

The problems (PP) and (PC,θ) are closely related in PC programming. In fact, they are equivalent at the global optimality level, provided that the feasible set of (PC,θ) is compact. In this case, if θ^* globally minimizes $f^o(\theta)$, then θ^* and an optimal solution $x^o(\theta^*)$ of the convex program (PC,θ) with $\theta=\theta^*$ is a globally optimal solution of (PC,θ) in (x,u) and vice versa. The problem (PP) is a basic problem in parametric programming. It has been well studied in the context of convex parametric programming 'models'. (These are PC programs in which the variable θ is considered as a parameter and allowed to vary.) The optimization problem is then reduced to optimization of the optimal value function in (PP). Local and global optimality of the parameter have been characterized for convex models in many papers, see, e.g., [8, 11, 12, 14].

An advantage of formulating OC problems as (PP) is that the latter may lead to the discovery of perturbations that cause discontinuities of the feasible set mapping, as shown in 'input optimization' methods, e.g., [14]. These discontinuities generally occur even in linear models; e.g., [15]. However, they do *not* occur, e.g., if perturbations of θ continuously generate a unique feasible point $x\in F(\theta)$, which happens in many 'classical' OC problems (also in our preceding example). This might be the reason why discontinuities of the feasible set mapping are not usually studied in the classical optimal control. Non-uniqueness of feasible points typically occurs in control problems with inequality constraints, e.g., [7].

4. Particular Cases in PC Programming

Theorems 3.2 and 3.3 are full characterizations of optimality and hence they contain many special cases. For example, if θ is fixed, they yield Theorem 2.1. We will now look at a classical result that gives only a *necessary* condition for local optimality in NP and OC. This result will be reformulated here for partly convex programs. Being only a necessary condition for optimality (and not a characterization) it may pass some non-optimal points as candidates for optimality. The result is new and it is given here for the sake of comparison with the saddle-point approach.

Consider the partly convex (PC,θ) around a feasible point $z^*=(x^*,\theta^*)$. Here are the assumptions:

(A1) This is a new assumption: For every feasible θ in a neighbourhood of θ^*, the index set $P^=(\theta)$ is constant. If we denote this index by R, and denote

$f^i = h^i, i \in R$, then we know that $h^i(x) \leqslant 0$ can be replaced by $h^i(x,\theta) = 0, i \in R$. Now the PC program (PC,θ) can be rewritten as

$$\underset{(x,\theta)}{\mathrm{Min}} f(x,\theta)$$
$$h^i(x,\theta) = 0, \quad i \in R$$
$$f^j(x,\theta) \leqslant 0, \quad j \in Q = P \backslash R.$$

Note that the inequalities in R are replaced by equations. Further we assume that $R = \{1, \ldots, n\}$ where n is the dimension of x.

Other assumptions are standard and they are taken from the literature, e.g., [9]:

(A2) All functions $f, h^i, i \in R, f^j, j \in P \backslash R$ are twice continuously differentiable.

(A3) The $n \times n$ matrix evaluated at x^* and θ^*

$$M = \begin{bmatrix} \partial h^1/\partial x_1, \ldots, \partial h^n/\partial x_1 \\ \ldots \\ \partial h^1/\partial x_n, \ldots, \partial h^n/\partial x_n \end{bmatrix}$$

is non-singular. This assumption guarantees (by the Implicit Function Theorem) the existence of a continuous differentiable function $\xi : R^p \to R^n$ and a ball B around $\theta^* \in R^p$ such that for all $\theta \in$ B, the solution x to $h^i(x,\theta) = 0, i \in R$ is given by $x = \xi(\theta)$. We will now use this function instead of the equality constraints, i.e., the problem is reduced to

$$\underset{(\theta)}{\mathrm{Min}}\, \phi(\theta) = f(\xi(\theta), \theta)$$
$$\phi^j(\theta) = f^j(\xi(\theta), \theta) \leqslant 0, \quad j \in Q. \tag{4.1}$$

(A4) The tangent cone to the feasible set of the new problem at θ^* is regular, i.e.,

$$T = \{e : \partial f^j(x^*, \theta^*)/\partial\theta \cdot e \leqslant 0, j \in Q(x^*, \theta^*)\}$$

where

$$Q(x^*, \theta^*) = \{j \in Q : f^j(x^*, \theta^*) = 0, j \in Q\}.$$

Now denote

$$v^T = \partial f(x^*, \theta^*)/\partial x \cdot M^{-1},$$

$$\rho^T = u^T \begin{bmatrix} \partial f^{n+1}/\partial x \\ \ldots \\ \partial f^N/\partial x \end{bmatrix} \cdot M^{-1}$$

evaluated at x^* and θ^*, and finally

$$J(x,\theta,v)=f(x,\theta)-\sum_{i\in R}v_i h^i(x,\theta)$$

$$V(x,\theta,\rho,u)=-\sum_{j\in Q}u_j f^j(x,\theta)+\sum_{i\in R}\rho_i h^i(x,\theta)$$

$$L(x,\theta,v,\rho,u)=J(x,\theta,v)-V(x,\theta,\rho,u).$$

THEOREM 4.1 (Necessary conditions for local optimality in partly convex programming). *Consider the partly convex program* (PC,θ) *around a parameter* θ^* *and a corresponding optimal solution* $x^*=x^0(\theta^*)$ *of the convex program* (PC,θ^*). *Suppose that the assumptions* (A1)–(A4) *are satisfied. If* θ^* *is a locally optimal parameter then there exists* $v\in R^n$, *determined by*

$$\partial J(x^*,\theta^*,v)/\partial x=0$$

and vectors $\rho\in R^n$ *and non-negative* $u\in R^n_+$ *such that*

$$\partial V(x^*,\theta^*,\rho,u)/\partial x=0$$

and

$$\partial L(x^*,\theta^*,v,\rho,u)/\partial\theta=0.$$

Proof. The idea is to work with the problem (4.1) in the variable θ while keeping in mind that $h^i(\xi(\theta),\theta)\equiv 0, i\in$ R The latter gives

$$\partial h(\xi(\theta^*),\theta^*)/\partial\theta=\partial h(x^*,\theta^*)/\partial x\cdot[\partial\xi(\theta^*)/\partial\theta]+\partial h(x^*,\theta^*)/\partial\theta=0.$$

Note that these are $n\times p$, $n\times n$, $n\times p$ and $n\times p$ matrices, respectively. (Here $\partial h(\xi(\theta^*),\theta^*)/\partial\theta=[\partial h^i(\xi(\theta^*),\theta^*)/\partial\theta_j]$, etc.) Hence

$$[\partial\xi(\theta^*)/\partial\theta]=-M^{-1}\cdot[\partial h^i(x^*,\theta^*)/\partial\theta]. \tag{4.2}$$

If θ^* is a local minimum then, by regularity of T, for every $e\in R^p$ such that

$\partial\phi^j(\theta^*)/\partial\theta\cdot e\geqslant 0, j\in Q^*$ we must have $\partial\phi(\theta^*)/\partial\theta\cdot e\geqslant 0$.

Now Farkas' lemma and chain rule diffferentiations with (4.2) yield relations that are equivalent to the statement of the theorem. These are known arguments and the reader is referred to [9, Section 4.2] for details.

EXAMPLE 4.2. Consider the control problem

$$\text{Min} -x_1/\theta$$
$$-x_1-\theta x_2+1\leqslant 0$$
$$-x_2\leqslant 0$$
$$x_1+x_2-1\leqslant 0$$
$$-x_1\leqslant 0$$
$$-\theta+1/2\leqslant 0.$$

Can $\theta^*=1/2$ be a locally optimal control with the corresponding optimal state $x^*=x^0(\theta^*)=(1,0)^T$? Here $P^=(\theta^*)=\{1,2,3,5\}$ and we can choose $R=\{1,2\}$. The assumptions (A1)–(A4) are satisfied with

$$M=\begin{bmatrix} 1 & 0 \\ \theta^* & 1 \end{bmatrix}$$

$Q(x^*,\theta^*)=\{5\}, T=R_+, v_1=-2, v_2=1, \rho_1=u_1, \rho_2=-1/2\rho_1,$ and finally $u_2=4$ from $\partial L/\partial\theta =0$. Hence the necessary condition for local optimality is satisfied.

5. Nonlinear Programming: A Look Beyond Partly Convex Programs

In the study of (PC) programs one uses the tools of convex programming and point-to-set topology. However, one should keep in mind that these are essentially nonconvex programs. Their feasible sets may be nonconvex and even disjoint. Moreover, local and global optima do not generally coincide. In this section we take a look beyond convex and (PC) programs and study global optimality for general n-dimensional NP problems. We will show how these problems can be formulated as particular (PC) programs in 2n+1 dimensions. For this reason, all the results given above are applicable in the new context. First, let us introduce the following PC formulation of (NP)

$$\text{Min} f(z)-M\theta^T z+Mz^T z$$
$$(\text{LF};\theta,\epsilon)$$
$$f^i(z)-M\theta^T z+Mz^T z\leqslant 0,\ i\in P,\ z\in C$$
$$\|z-\theta\|\leqslant\epsilon$$

where M is a positive scalar, C is a nonempty convex compact set, $\|\cdot\|$ is an arbitrary norm, and $\epsilon\geqslant 0$ is a scalar parameter. Recall that for a special choice of the norm, say, l_1 and l_∞, the constraint $\|z-\theta\|\leqslant\epsilon$ can be replaced by a system of linear inequalities. Note that for every fixed $\epsilon\geqslant 0$ and θ, (LF; θ,ϵ) is a *convex* *program* for all sufficiently large $M>0$, provided that f and $f^i, i \in P$ are twice continuously differentiable. This can be verified using the fact that the Hessian matrices in z, of the objective functions and the constraints, are positive definite on the compact set C for sufficiently large $M>0$. Also (LF; θ, ϵ) is *partly convex* in z and θ for every $\epsilon \geqslant 0$. Being linear in θ and convex in z, we can say that (LF; θ, ϵ) is a *partly linear-convex program*. If the vector θ and the scalar ϵ are jointly considered as a 'control' vector, then (LF; θ, ϵ) is a *convex* optimal control problem in the 'state' variable z! Using the terminology from, e.g., [14], (LF; θ, ϵ) is a *convex model* with the parameters θ and ϵ.

The above model is an extension of the program introduced by Liu and Floudas in [6] from $\epsilon=0$ to $\epsilon \geqslant 0$. These authors have observed that a feasible z^* of (NP) is its globally optimal solution if, and only if, the same z^* and $\theta^*=z^*$ is a globally optimal solution of (LF; θ, 0). This is a crucial observation for our study.

Let us first show that the model (LF; θ, ϵ) is 'structurally stable' for feasible perturbations in (θ, ϵ) at an optimal $\theta^*=z^*$ and $\epsilon^*=0$. This means that the feasible set mapping

$$F:\ (\epsilon, \theta) \to F(\epsilon, \theta) = \{z : f^i(z) - M\theta^T z + \mathrm{M} z^T z \leqslant 0, i \in P, z \in C, \|z-\theta\| \leqslant \epsilon\}$$

is open at $\epsilon^*=0$ and $\theta^*=z^*$ relative to its feasible perturbations.

THEOREM 5.1 (Stability of the Liu-Floudas model). *Consider* (NP), *where all functions are assumed to be twice continuously differentiable, with a unique globally optimal solution z^*, and the corresponding* (LF; θ, ϵ) *with M sufficiently large. Then the feasible set mapping F: $(\epsilon, \theta) \to F(\epsilon, \theta)$ is open at $\epsilon^*=0$ and $\theta^*=z^*$ relative to its feasible perturbations.*

Proof. Take the globally optimal solution z^* of (NP). Then $z^*=\theta^*$ and θ^* is a globally optimal solution of (LF; θ, 0). We have to show that, for an arbitrary feasible sequence $(\epsilon^k, \theta^k) \to (\epsilon^*, \theta^*), \epsilon^*=0$, there is a sequence $z^k \in F(\epsilon^k, \theta^k)$ converging to $z^* \in F(\epsilon^*, \theta^*)$. Indeed, for every element in such sequence $\{\epsilon^k, \theta^k\}$, there exists an $z^k \in F(\epsilon^k, \theta^k)$, by the feasibility assumption. Since the set C is compact, the sequence $\{z^k\}$ contains a subsequence converging to some point $w^* \in F(\epsilon^*, \theta^*)$. But the feasible set mapping F: $(\epsilon, \theta) \to F(\epsilon, \theta)$ is closed. (This is a consequence of continuity of the constraint functions.) Therefore $w^* \in F(\epsilon^*, \theta^*)$. Since $\epsilon^*=0$, this means that $w^*=\theta^*$. But we know that $z^*=\theta^*$. Hence $w^*=z^*$.

Remark. The above result says that (LF; θ, ϵ) is a 'stable' convex model at the optimal parameter even if the original (NP) is unstable. The 'stabilization' is important because, using this result, one can now study many multi-level and multi-objective optimization problems; see [1, 2, 4, 14, 15]. The models describing these problems are typically unstable.

An illustration of the Liu-Floudas stabilization follows:

EXAMPLE 5.2. Consider

$$\begin{aligned}&\text{Max } z\\&\text{s.t.}\\&\alpha z \leqslant 0, \quad 0 \leqslant z \leqslant 1\end{aligned}$$

where $\alpha \geqslant 0$ is a parameter. One can think of α as a small positive error around $\alpha = 0$. For any $\alpha > 0$, a unique global solution is $z^o(\alpha) = 0$ and the optimal value is $f^0(\alpha) = 0$. Since

$$\lim_{\alpha \to 0} z^o(\alpha) = 0 \quad \text{and} \quad \lim_{\alpha \to 0} f^o(\alpha) = 0$$

one might be tempted to conclude that the optimal value is 0 also for $\alpha = 0$. But this is not true, because $z^o(0) = 1$ and $f^o(0) = 1$. We see that every (arbitrarily small) perturbation of α at $\alpha = 0$ always gives results that are 'far' from the results of the unperturbed program. This kind of instability is not present in the parametric Liu-Floudas model. This model is here (after specifying $M = 1$)

$$\begin{aligned}&\text{Max } z + \theta z - z^2\\&\text{s.t.}\\&\alpha z - \theta z + z^2 \leqslant 0, \quad -\epsilon \leqslant z - \theta \leqslant \epsilon, \quad 0 \leqslant z \leqslant 1.\end{aligned}$$

Fix $\epsilon > 0$. For every $0 < \alpha \leqslant \epsilon$, an optimal solution is $z^o(\alpha) = 1, \theta^o(\alpha) = 1 + \alpha$ and the optimal value is $f^o(\alpha) = 1 + \alpha$. This time

$$\lim_{\alpha \to 0} z^o(\alpha) = 1 \text{ and } \lim_{\alpha \to 0} f^o(\alpha) = 1.$$

Continuity of the model is preserved. More numerical effort (better approximation of $\alpha = 0$) has produced more accurate results and the correct results in the limit.

The main result on optimality follows:

THEOREM 5.3 (Characterization of global optimality for general nonlinear programs). *Consider* (NP), *where all functions are assumed to be twice continuously differentiable, its arbitrary feasible point* z^*, *and the model* (LF; θ, ϵ). *Assume that* (NP) *has a unique globally optimal solution. Then* z^* *is a globally optimal solution of* (NP), *if and only if, for all sufficiently large M, if, and only if,* z^* *and* $\theta^* = z^*$ *is the limit point of optimal solutions* $z^o = z^o(\epsilon), \theta^o = \theta^o(\epsilon)$ *of* (LF; θ, ϵ), *respectively, as* $\epsilon \to 0$.

Proof. Suppose that z^* is a unique globally optimal solution of (NP). Consider a sequence $\epsilon > 0, \epsilon \to 0$. For every ϵ of the sequence, there exists an optimal solution $z^o = z^o(\epsilon), \theta^o = \theta^o(\epsilon)$ of (LF; θ, ϵ). This is true because C is a compact set and the feasible set of (LF; θ, ϵ) is not empty, always containing at least z^*

and $\theta^* = z^*$. As $\epsilon \to 0$ the optimal solutions $z^o(\epsilon) \to w^*$ and $\theta^o(\epsilon) \to w^*$ for some w^*. Since the constraints are continuous, it follows that w^* and $\theta^* = w^*$ are a feasible point of (LF; $\theta, 0$). This means that w^* is a globally optimal solution of (NP), by the Liu-Floudas result. Since there is only one globally optimal solution of (NP), by assumption, it follows that $w^* = z^*$. The sufficiency part of the proof is straightforward: If z^* and $\theta^* = z^*$ is the limit point of optimal solutions $z^o = z^o(\epsilon), \theta^o = \theta^o(\epsilon)$ of (LF; θ, ϵ), respectively, as $\epsilon \to 0$, then z^* and $\theta^* = z^*$ is a feasible point of (LF; $\theta, 0$). But this means that z^* is a globally optimal solution of (NP).

Remark. A unique global optimum z^* of (NP) is the limit point of optimal solutions $z^o(\epsilon)$ of the convex programs (LF; $\theta^o(\epsilon), \epsilon$)

$$\begin{aligned}
&\mathrm{Min}\, f(z) - M\theta^o(\epsilon)^T z + \mathrm{M} z^T z \\
&f^i(z) - M\theta^o(\epsilon)^T z + \mathrm{M} z^T z \leqslant 0, \quad i \in P, \quad z \in C \\
&\|z - \theta^o(\epsilon)\| \leqslant \epsilon
\end{aligned}$$

as $\epsilon > 0, \epsilon \to 0$. Here $\theta^o(\epsilon)$ is a globally optimal solution of the convex model

$$\begin{aligned}
&\mathrm{Min}\, f^o(\theta) \\
&\theta \in F(\epsilon) = \{\theta : F(\epsilon, \theta) \neq \varnothing\}
\end{aligned}$$

where $f^o(\theta)$ is the optimal value function

$$f^o(\theta) = f(z^o(\theta)) - M\theta^T z^o(\theta) + M z^o(\theta)^T z^o(\theta).$$

The point $z^o(\theta)$ is an optimal solution, for fixed θ and ϵ, of the convex program (LF; θ, ϵ).

EXAMPLE 5.4. Consider the strictly *concave* program

$$\begin{aligned}
&\mathrm{Min}\, \sin z \\
&\text{s.t.} \\
&0 \leqslant z \leqslant 1.
\end{aligned}$$

One can choose $M = 1$, and the (LF; θ, ϵ) model is

$$\begin{aligned}
&\mathrm{Min}\, \sin z - \theta^T z + z^T z \\
&(\mathrm{LF}; \theta, \epsilon) \\
&z \in C = [0, 1] \\
&|z - \theta| \leqslant \epsilon.
\end{aligned}$$

For the sequence, e.g., $\epsilon^k = 1/k \to 0$ one can choose $\theta^o = \theta^o(\epsilon^k) = 1/(2k) \to 0$ and $z^o = z^o(\epsilon^k) = 0$. Hence $z^* = 0$ is a global minimum. Note that the sequence of *convex* programs to be solved in z is

$$\begin{aligned}&\operatorname{Min} \sin z - 1/(2k)z + z^2\\ &\text{s.t.}\\ &0 \leqslant z \leqslant 1, \quad |z - 1/(2k)| \leqslant 1/k, \quad k = 1, 2, \ldots\end{aligned}$$

The formulation (LF; θ, ϵ) is more suitable for numerical purpose than (LF; θ, 0) since an optimal solution of (NP) is obtained as the solution of a structurally stable model (without complications illustrated in Example 5.2). In order to find a global optimum z^* of a general, possibly nonconvex and unstable, (NP) one first transforms (NP) into (LF; θ, ϵ). For a fixed $\epsilon > 0$ one finds an optimal solution $z^o = z^o(\epsilon)$ and $\theta^o = \theta^o(\epsilon)$ of the corresponding PC program. There are at least two general and distinct ways how this can be done. One can use a GOP method; e.g., [5, 6]. These methods are specifically designed to solve PC programs. An alternative approach is to use a method of input optimization. These methods have been applied to a variety of convex models in, e.g., [14]. They are parametric in nature and minimize the optimal value function $f^o(\theta)$, using an appropriate marginal value formula, to find its minimizing point $\theta^o(\epsilon)$. In either case, as $\epsilon \to 0, \theta^o = \theta^o(\epsilon) \to z^*$ and $z^o = z^o(\epsilon) \to z^*$. It is too early to comment on comparison and numerical efficiency of these approaches. Their testing is a time-consuming process currently under way. In order to get a full picture we also plan to test unstable programs. Finding efficient numerical methods for (LF; θ, ϵ), $\epsilon \to 0$ may be a potentially important problem in numerical optimization. Although we do not have a computer code at this time, several nontrivial problems have already been solved using the Liu-Floudas parametric model. For example, the model has been recently used in [16] to solve a system of 3 nonlinear equations in 3 unknowns. The system was first written as an unconstrained problem and then its (LF; θ, ϵ), $\epsilon \to 0$ was solved by input optimization.

If a globally optimal solution is not unique then a characterization of optimality can be given based on the property that the optimal solutions mapping is closed, i.e., one uses a weak convergence of the optimal value function. However, it is not clear how to relax the condition that the functions in (NP) be twice continuously differentiable so that (LF; θ, ϵ) be convex. In particular, for what $f(z)$ is the function $f(z) - M\theta^T z + Mz^T z$ convex over a compact set C for every θ and all sufficiently large $M > 0$? We know that it is not enough to assume that the function $f(z)$ is just twice differentiable. Theorems 2.1 and 3.2 have been extended to abstract settings in [1]. They are applicable to convex OC problems. The Liu-Floudas parametric model has not yet been formulated in an abstract setting. If such model existed, it could be used to study nonconvex infinite-dimensional optimization and OC problems.

Acknowledgment

The author is indebted to C. Floudas for drawing his attention to reference [6], which has inspired this research. He is also indebted to the referees for their constructive criticism and to R. Trujillo-Cortez for his comments on the material from Section 5.

References

1. Asgharian, M. and Zlobec, S. (2002), Abstract parametric programming, *Optimization*, 51, 841–861.
2. Ben-Israel, A., Ben-Tal, A. and Zlobec, S. (1981), *Optimality in Nonlinear Programming: A Feasible Directions Approach*, Wiley-Interscience, New York.
3. Canon, M.D., Cullum Jr., C.D. and Polak, E. (1970), *Theory of Optimal Control and Mathematical Programming*, McGraw-Hill Series in Systems Science, New York.
4. Floudas, C. (2000), *Deterministic Global Optimization*, Kluwer Academic.
5. Floudas, C.A. and Zlobec, S. (1998), Optimality and duality in parametric convex lexicographic programming. in Migdalas et al. (eds.), *Multilevel Optimization: Algorithms and Applications*, Kluwer Academic, pp. 359–379.
6. Liu, W.B. and Floudas, C.A. (1993), A remark on the GOP algorithm for global optimization, *J. Global Optimization*, 3, 519–521.
7. Sethi, S.P. A survey of management science applications of the deterministic maximum principle, *TIMS Studies in the Management Sciences*, North-Holland v. 9 (1978), 33–67.
8. Trujillo-Cortez, R. and Zlobec, S. (2001), Stability and optimality in convex parametric programming, *Mathematical Communications*, 6, 107–121.
9. Vincent, T.L. and Grantham, W.J. (1981), *Optimality in Parametric Systems*, Wiley Interscience, New York.
10. Zangwill, W.I. (1969), *Nonlinear Programming*, Prentice-Hall Inc., Englewood Cliffs, N.J.
11. Zlobec, S. (1983) Characterizing an optimal input in perturbed convex programming, *Mathematical Programming*, 25, 109–121; Corrigendum: Ibid 35 (1986) 368–371.
12. Zlobec, S. (1985) Input optimization: I. Optimal realizations of mathematical models, *Mathematical Programming*, 31, 245–268.
13. Zlobec, S. (1995), Partly convex programming and Zermelo's navigation problems, *J. Global Optimization*, 7, 229–259.
14. Zlobec, S. (2001), *Stable Parametric Programming*, Kluwer Academic; series: Applied Optimization, 57.
15. Zlobec, S. (2001), Stability in linear programming: An index set approach, *Annals of Operations Research*, 101, 363–382.
16. Zlobec, S. (2003), *Estimating Convexifiers in Continuous Optimization*, Mathematical Communications (forthcoming).