Mathematical Programming with LFS Functions

Faranak Sharifi Mokhtarian

Department of Mathematics and Statistics
McGill University, Montreal

July 1992

A Thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements of the degree of
Master of Science

Contents

Abstract

Differentiable functions with a locally flat surface (LFS) have been recently introduced and studied in convex optimization. Here we extend this notion in two directions: to non-smooth convex and smooth generalized convex functions. An important feature of these functions is that the Karush-Kuhn-Tucker condition is both necessary and sufficient for optimality. Then we use the properties of linear LFS functions and basic point-to-set topology to study the "inverse" programming problem. In this problem, a feasible, but nonoptimal, point is made optimal by stable perturbations of the parameters. The results are applied to a case study in optimal production planning.

Résumé

Récemment, on a introduit dans l'étude d'optimisation convexe les fonctions dérivables à surface localement plate (SLP). Ici, nous développons cette idée dans deux directions: fonctions convexes non-dérivables, et fonctions convexes généralisées dérivables. Une caractéristique importante de ces fonctions est le fait que la condition Karush-Kuhn-Tucker est non-seulement suffisante mais aussi nécessaire pour l'optimalité Ensuite, nous utilisons les propriétés des fonctions SLP linéaires et la topologie point-par-ensemble dans l'étude du problème de la programmation "inverse". Ici, un point réalisable, mais non-optimale, devient optimale par des perturbations stables des paramètres. On applique les résultats dans une étude de cas qui concerne la planifications de la production optimale.

Acknowledgements

I would like to express my sincere thanks to Professor Sanjo Zlobec, my thesis supervisor, for his guidence, counselling and understanding throughout my Master's program.

I would like to thank my husband, Farzad Mokhtarian, for his continued support and encouragement.

Thanks are also due to Xiaojie Zhou and Marchand Van Rooyen for their interest and our useful discussions.

I am grateful to Fonds FCAR and the Department of Mathematics and Statistics at McGill University for the financial support.

Introduction

Characterizations of optimal solutions of convex programs were formulated in mid- and late seventies. They are part of the theory formulated by Ben-Israel, Ben-Tal and Zlobec, and their colleagues (see, e.g., [4,6]). In the presence of extraneous conditions, their results recover the classical theory of Karush, Kuhn and Tucker [16]. On the other hand, characterizations of "optimal inputs"(optimal parameters) over "regions of stability" (i.e., regions of continuity of outputs) and thus characterizations of "optimal realizations" of linear and convex models are quite recent (see, e.g. [7, 24,26,25,32,34,37]). The concept of a differentiable convex function with a "locally flat surface" (LFS) was recently introduced in [31]. These functions enjoy some interesting and important properties. One such property is that for these functions the Karush-Kuhn-Tucker condition is both necessary and sufficient for optimality. If some constraints do not belong to this class, then the optimality conditions assume an asymptotic form.

The purpose of this thesis is twofold: The first objective is to study convex (not necessarily differentiable) as well as differentiable generalized convex functions with a locally flat surface. The second is to find a solution to the "inverse problem" for linear mathematical models.

Chapter 1 extends the concept of differentiable convex programs with LFS property to the non-smooth (i.e., not-necessarily-differentiable) convex case. Unlike the differentiable case, providing an algebraic characterization for LFS functions in the general case is not trivial. However, several results (specially Corollary 1.3) help us identify these functions. Given a mathematical program, we regroup the constraints to those that have LFS property at a given feasible point and to those that do not. Using the program rewritten in this form, we then take the approach by Ben-Israel, Ben-Tal and Zlobec [4,6] and characterize optimality in "primal", "dual" and "saddle-point" forms. The "primal" form (Theorem 1.6) involves only elements (solutions and directions) of the primal (decision variable) space. The "dual" forms (Theorems 1.7, 1.8, Corollaries 1.4, 1.5 and 1.6) involve elements (dual sets and Lagrange multipliers) of the dual space. The "saddle-point" form (Theorem 1.9) involves the Lagrangian function. Some applications of these LFS functions in single- and multi-objective programming conclude the chapter.

In Chapter 2, the concept of differentiable convex functions with LFS property is further extended to the differentiable generalized convex case. These functions are then characterized geometrically (Theorem 2.1 and Corollary 2.1) and algebraically (Theorem 2.2). Again, we regroup the constraints to those with LFS property at a given point and to those without this property, in order to derive characterizations of optimality similar to what we derived in the convex case. It is interesting that the saddle-point-necessary optimality condition does not hold here even when all the functions are pseudoconvex and enjoy the LFS property at an optimal solution (Example 2.2). In the completely pseudoconvex case, if some of the constraints do not belong to the class of LFS functions, the optimality conditions again assume an asymptotic form (Theorem 2.7). The last section of the chapter extends the characterizations of differentiable convex multi-objective programs, which have appeared recently (see, e.g., [30,32]), to the pseudoconvex case. It further illustrates the role of differentiable quasiconvex functions with LFS property in multi-objective programming.

In Chapter 3, we focus our attention to mathematical models and input optimization. A new notion of stability is introduced and compared with the usual one (see, e.g., [32]) in input optimization and also with the "weak stability" introduced in [24]. We then study optimal inputs with respect to this new notion of stability and compare the results with the related ones from the literature. These results are then applied to the classical control problem of Zermelo, adapted from [27].

The most important subject of Chapter 3 is the inverse problem. Considering a mathematical model initially running with some input and a nonoptimal feasible point, our objective is to perturb the system in a stable way so as to reach an input at which the feasible point becomes optimal. This objective may not be achievable, in which case an alternative is to search for an optimal parameter at which the given feasible point becomes as close as possible to being optimal. For the sake of simplicity, we study only the inverse problem for linear models, i.e., an important subclass of LFS models. In particular, we give a characterization of optimal parameters for the inverse problem using the duality properties of linear programming (Theorem 3.15). We also study the relation between the stability of the primal and dual linear models. A numerical algorithm for finding a stable path that leads to an optimal parameter is adapted from [7]. The results are applied to a case study that was first studied, in a simpler form, in [15].

Chapter 1

LFS Functions in Convex Programming

We will first recall some basic notions in convex programming and then introduce a class of differentiable and nondifferentiable convex functions for which the Karush-Kuhn-Tucker condition is necessary for optimality. We regroup the constraints of a given convex program to those that belong to this class and those that do not. For this form of the convex program, we will then give primal, dual and a saddle-point characterization of optimality. We will further show that for the functions that belong to this class in multi-objective optimization, Pareto solutions coincide with strong Pareto solutions.

We recall that a function $f : R^n \to R$ is *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

for every x, $y \in R^n$ and every $0 \leq \lambda \leq 1$. We will study mathematical programs of the form

$$(P) \qquad \begin{array}{ll} \text{Min} & f^0(x) \\ \text{s.t.} & \\ & f^i(x) \leq 0, \quad i \in \mathcal{P} = \{1, \ldots, m\} \end{array}$$

where all the functions $f^j : R^n \to R$, $j \in \{0\} \cup \mathcal{P}$, are assumed to be continuous. Such programs are called *convex* if all the functions are further assumed to be convex. At a feasible point

$$x^* \in F = \{x \in R^n : f^i(x) \leq 0, \ i \in \mathcal{P}\},$$

with F referred to as the feasible set, we denote the set of feasible directions by

$$F(x^*) = \{d \in R^n : \exists\, T > 0 \ \ni \ x^* + td \in F \ \ \forall\, 0 \leq t \leq T\}$$

and the set of active constraints by

$$\mathcal{P}(x^*) = \{i \in \mathcal{P} : f^i(x^*) = 0\}.$$

Furthermore, we recall the set

$$\mathcal{D}(x^*) = \{d \in R^n : \nabla f^i(x^*)d \leq 0, \quad i \in \mathcal{P}(x^*)\}.$$

Associated with the convex program (P), is also the *minimal index set of active constraints* denoted by

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

We recall that a set C in R^n is called a cone if it is closed under multiplication by nonnegative scalars. It is a blunt cone if it is closed under multiplication by positive scalars. It is a polyhedral cone, if it can be written in the form

$$C = \{x \in R^n : Ax \leq 0\} \text{ for some } A \in R^{m \times n}.$$

We also recall that for any nonempty set S in R^n, the *polar* of S is the set

$$S^+ = \{u \in R^n : (u, x) \geq 0, \quad x \in S\},$$

and the *convex hull* of S, $conv\{S\}$, is the smallest convex set containing S.

The cones of directions of *constancy, nonascent* and *descent* are respectively denoted by

$$\begin{aligned} D_f^=(x^*) &= \{d \in R^n : \exists \bar{\alpha} > 0 \ni f(x^* + \alpha d) = f(x^*), \quad 0 < \alpha < \bar{\alpha}\}, \\ D_f^\leq(x^*) &= \{d \in R^n : \exists \bar{\alpha} > 0 \ni f(x^* + \alpha d) \leq f(x^*), \quad 0 < \alpha < \bar{\alpha}\} \end{aligned}$$

and

$$D_f^<(x^*) = \{d \in R^n : \exists \bar{\alpha} > 0 \ni f(x^* + \alpha d) < f(x^*), \quad 0 < \alpha < \bar{\alpha}\}.$$

If $\{f^k : k \in \Omega\}$ is a set of functions indexed by a set Ω, then for each of the relations

$$\text{“relation”} \triangleq \text{“}=\text{”}, \text{“}\leq\text{”}, \text{“}<\text{”}, \text{“}\geq\text{”}, \text{ etc.}$$

we use the following abbreviation:

$$D_\Omega^{\text{relation}}(x^*) \triangleq \bigcap_{k \in \Omega} D_{f^k}^{\text{relation}}(x^*).$$

We will denote the domain of a convex function f by $dom f$ and the interior of the domain of f by $int(dom f)$. The *relative interior* of the domain of f, $ri(dom f)$, is defined as the interior which results when $dom f$ is regarded as a subset of its affine hull (the smallest affine set containing $dom f$). Unless otherwise specified, we always assume that a function f is of finite value at any point in its domain.

We now collect some useful properties in the following lemma. More details and proofs can be found in [4].

Lemma 1.1 *Let f be a real-valued convex function, and let $x^\star \in dom f$. Furthermore, let S and T be nonempty subsets of R^n. Then*

(a) $compD_f^{<}(x^\star) = D_f^{\geq}(x^\star)$ *where comp denotes the complement;*

(b) $D_f^{\leq}(x^\star) = D_f^{<}(x^\star) \cup D_f^{=}(x^\star)$;

(c) the blunt cone $D_f^{<}(x^\star)$ *is convex;*

(d) the cone $D_f^{\leq}(x^\star)$ *is convex;*

(e) $conv\{D_f^{=}(x^\star)\} \subset D_f^{\leq}(x^\star)$;

(f) $D_{\mathcal{P}^=}^{=}(x^\star)$ *is convex;*

(g) $\mathcal{P}^=$ *is the maximal subset* Ω *of* $\mathcal{P}(x)$ *with the property that, for every* $k \in \Omega$, $d \in F(x) \Rightarrow d \in D_\Omega^{=}(x)$;

(h) $S^+ = \{conv\{S\}\}^+ = \{clS\}^+$;

(i) if S and T are closed and convex cones, $\{S \cap T\}^+ = cl[S^+ + T^+]$. *Moreover, if S and T are polyhedral, the "closure" can be omitted.*

1.1 Differentiable LFS functions

For a differentiable function f, we denote the null space of its gradient (considered as a row vector) at a point $x^\star \in dom f$ by

$$N(\nabla f(x^\star)) = \{d \in R^n : \nabla f(x^\star)d = 0\}.$$

Definition 1.1 *A differentiable convex function f has a locally flat surface at $x^\star$ if*

$$N(\nabla f(x^\star)) = D_f^{=}(x^\star).$$

The abbreviation "LFS" is used for such functions.

Theorem 1.1 *Consider the convex function f, differentiable at $x^\star \in dom f$, such that $\nabla f(x^\star) \neq 0$. Then f has LFS propety at $x^\star$ if, and only if, the cone $D_f^{\leq}(x^\star)$ is polyhedral.*

Proof: (Sufficiency:) Since $\nabla f(x^\star) \neq 0$, it follows that $D_f^{<}(x^\star) \neq \emptyset$.

$$\begin{aligned} D_f^{\leq}(x^\star) &= \{d : \nabla f(x^\star)d \leq 0\} \\ &\quad \text{by polyhedrality and } \nabla f(x^\star) \neq 0 \text{ assumptions} \\ &= D_f^{<}(x^\star) \cup N(\nabla f(x^\star)). \end{aligned}$$

Also

$$D_f^{\leq}(x^\star) = D_f^{<}(x^\star) \cup D_f^{=}(x^\star).$$

In either case $D_f^{\leq}(x^\star)$ is represented as a union of two disjoint sets. Hence

$$D_f^{=}(x^\star) = N(\nabla f(x^\star)),$$

i.e., f has LFS property at $x^\star$.
(Necessity:) Assume that f has LFS property at $x^\star$. Then, by Definition 1.1,

$$D_f^{=}(x^\star) = \{d : \nabla f(x^\star)d = 0\}.$$

Therefore

$$\begin{aligned} D_f^{\leq}(x^\star) &= D_f^{<}(x^\star) \cup D_f^{=}(x^\star) \\ &= \{d : \nabla f(x^\star)d \leq 0\}, \end{aligned}$$

which is a polyhedral cone.

Remark: Note that the assumption $\nabla f(x^\star) \neq 0$ is not required for the necessity part of the proof. Therefore, if f has the LFS property at $x^\star$, then $D_f^{\leq}(x^\star)$ is polyhedral regardless of whether the gradient of f is zero at $x^\star$ or not. On the other hand, as the following example shows, if $\nabla f(x^\star) = 0$, then $D_f^{\leq}(x^\star)$ being polyhedral does not imply that f has the LFS property at $x^\star$.

Example 1.1 Consider $f(x) = x_1^2 + x_2^2$ at $x^\star = (0,0)^T$. Here
$D_f^{=}(x^\star) = \{0\}$, $D_f^{<}(x^\star) = \emptyset$ and $D_f^{\leq}(x^\star) = \{0\}$ (a polyhedral cone). While

$$N(\nabla f(x^\star)) = R^2 \text{ since } \nabla f(x^\star) = 0.$$

□

Theorem 1.1 may not help us identify an LFS function easily. For a faithfully convex function, i.e., a function of the form

$$f(x) = h(Ax + b) + a^T x + \alpha, \tag{1.1}$$

where $h : R^m \to R$ is strictly convex, $A \in R^{m \times n}$, a, $x \in R^n$, $b \in R^m$ and $\alpha \in R$, Zhou [30] proved the following result:

Theorem 1.2 *Let f be a differentiable convex function of the form (1.1) and $x^\star \in dom f$. If $\nabla f(x^\star) = 0$, then f has LFS property at $x^\star$ if, and only if, f is constant. If $\nabla f(x^\star) \neq 0$, then f has LFS property at $x^\star$ if, and only if,*

$$rank \begin{bmatrix} A \\ a^T \end{bmatrix} = 1. \tag{1.2}$$

Consider a program of the form (P) where all the functions are differentiable but not necessarily convex. We refer to the system

$$\nabla f^0(x^\star) + \sum_{i \in \mathcal{P}(x^\star)} \lambda_i f^i(x^\star) = 0$$

$$\lambda_i \geq 0, \;\; i \in \mathcal{P}(x^\star)$$

as the Karush-Kuhn-Tucker system or the KKT system for short. Conditions that guarantee consistency of the KKT system are known in literature as the *constraint qualifications*. An important application of the LFS functions then follows.

Theorem 1.3 *Consider the convex program (P) where all the functions are differentiable and the constraint functions have LFS property at a feasible point $x^\star$. Then $x^\star$ is optimal if, and only if, the KKT system is consistent.*

Proof: It is shown in [4] that $x^\star \in F$ is not optimal if, and only if, the system

$$\begin{array}{rcl} \nabla f^0(x^\star)d & < & 0 \\ \nabla f^i(x^\star)d & \leq & 0, \;\; i \in \mathcal{P}(x^\star) \\ \text{with} & \text{``=''} & \text{if, and only if, } d \in D^{=}_{f^i}(x^\star) \end{array}$$

is consistent. If the constraint functions have LFS property at $x^\star$, then the condition on "=" can be omitted. An application of the Farkas lemma (see, e.g., [17]) proves the theorem.

Remark: It is well known (see, e.g., [29]) that if all the functions in program (P) are differentiable, then the condition

$$clF(x^\star) = \mathcal{D}(x^\star) \tag{1.3}$$

is a constraint qualification. An indirect proof shows that the LFS property implies (1.3). Indeed, the implication is true, because

$$\begin{array}{rcl} F(x^\star) & = & \{d : \nabla f^i(x^\star)d < 0, \;\; i \in \mathcal{P}(x^\star)\} \cup D^{=}_{\mathcal{P}(x^\star)}(x^\star) \\ & = & \{d : \nabla f^i(x^\star)d \leq 0, \;\; i \in \mathcal{P}(x^\star)\} \;\; \text{for LFS functions} \\ & = & clF(x^\star) \\ & = & \mathcal{D}(x^\star). \end{array}$$

1.2 Nondifferentiable LFS Functions

Now consider convex, but not-necessarily-differentiable (non-smooth), functions. First we recall some concepts dealing with directional derivatives and subgradients.

Definition 1.2 *Let f be an arbitrary function from R^n to R, and let $x^\star \in dom f$. The directional derivative of f at $x^\star$ in the direction $d \in R^n$ is defined as*

$$f'(x^\star; d) = \lim_{\lambda \to 0^+} \frac{f(x^\star + \lambda d) - f(x^\star)}{\lambda}.$$

For a convex function f, the directional derivative exists for any direction d and at any point in the domain of f (see, e.g., [21]).

Definition 1.3 *A vector $h \in R^n$ (considered as a row vector), is a subgradient of the function $f : R^n \to R$ at $x^\star \in dom f$ if*

$$f(x) \geq f(x^\star) + h(x - x^\star) \quad \textit{for every } x \in R^n.$$

The set of all subgradients of f at $x^\star$ is denoted by $\partial f(x^\star)$ and is called the subdifferential of f at $x^\star$. If f is convex and differentiable, then $\partial f(x^\star) = \{\nabla f(x^\star)\}$. We now recall the following two useful lemmas from [21].

Lemma 1.2 *Let f be a convex function, and let $x^\star \in dom f$. Then h is a subgradient of f at $x^\star$ if, and only if,*

$$f'(x^\star; d) \geq (h, d) \tag{1.4}$$

for any $d \in R^n$.

Lemma 1.3 *If f is a convex function, then at each point $x^\star \in int(dom f)$ the subdifferential $\partial f(x^\star)$ is a nonempty, compact and convex set; furthermore,*

$$f'(x^\star; d) = \max\{(h, d) : h \in \partial f(x^\star)\}.$$

We can therefore conclude the following corollary.

Corollary 1.1 *Let f be a convex function, and let $x^\star \in int(dom f)$. Then*

$$\{d : f'(x^\star; d) \leq 0\}^+ = \{\lambda \partial f(x^\star) : \lambda \leq 0\},$$

where $\{\lambda \partial f(x^\star) : \lambda \leq 0\} = \{\lambda y : y \in \partial f(x^\star), \quad \lambda \leq 0\}$.

Proof: It follows from Lemma 1.3 that

$$\{d : f'(x^\star; d) \leq 0\} = \{d : (h, d) \leq 0, \quad h \in \partial f(x^\star)\}.$$

Since

$$\{d : (h, d) \leq 0, \quad h \in \partial f(x^\star)\}^+ = \{\lambda \partial f(x^\star) : \lambda \leq 0\}$$

the result immediately follows.

□

The following lemma is a well-known result (see [21, Chapter 23]).

Lemma 1.4 *Let f be a convex function, and let $x^\star \in ri(dom f)$ be a point where f does not achieve its minimum. Then*

$$\{D_f^{<}(x^\star)\}^+ = cl\{\lambda \partial f(x^\star) : \lambda \leq 0\}.$$

Furthermore, if $x^\star \in int(dom f)$, then

$$\{D_f^{<}(x^\star)\}^+ = \{\lambda \partial f(x^\star) : \lambda \leq 0\}.$$

Remark: The assumption that f does not achieve its minimum at $x^\star$ is the same as $D_f^{<}(x^\star) \neq \emptyset$ or $0 \notin \partial f(x^\star)$.

Definition 1.4 *Consider the convex function f and assume that $x^\star$ is a point in the domain of f. Then we say that f has generalized LFS property at $x^\star$ if*

$$\{d : f'(x^\star; d) = 0\} = D_f^{=}(x^\star).$$

The abbreviation "GLFS" is used for such functions. A graphic illustration of a GLFS function is given in Figure 1.1. This is the graph of the function $f(x) = |x_1| + |x_2|$

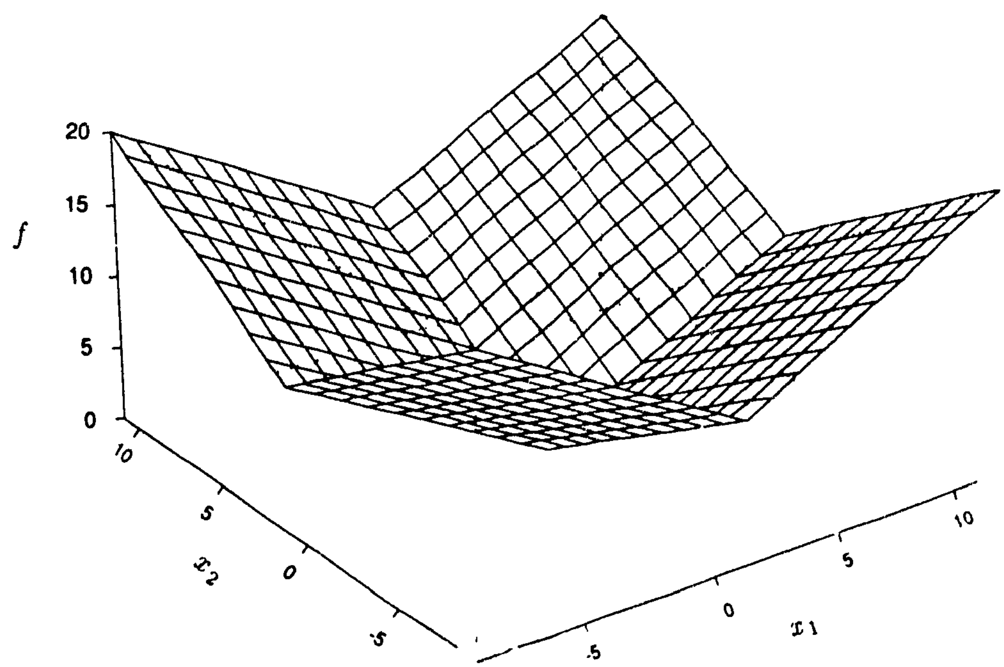

Figure 1.1: Graph of the GLFS function $f(x) = |x_1| + |x_2|$.

which has the GLFS property at any $x^\star$ in its domain. At any point on the surface, the directions along which the directional derivatives vanish are exactly the directions of constancy. We now give two important properties of these functions.

Corollary 1.2 *Consider a convex function f. Assume that f has GLFS property at $x^\star \in int(domf)$ where it achieves its minimum, that is $D_f^<(x^\star) = \emptyset$. Then*

$$\{D_f^=(x^\star)\}^+ = \{\lambda\partial f(x^\star) : \lambda \leq 0\}.$$

Proof: Since $D_f^<(x^\star) = \emptyset$, then

$$\{d : f'(x^\star; d) = 0\} = \{d : f'(x^\star; d) \leq 0\}.$$

This together with the GLFS property of f imply that

$$\begin{aligned}\{D_f^=(x^\star)\}^+ &= \{d : f'(x^\star; d) = 0\}^+ \\ &= \{d : (h, d) \leq 0, \quad h \in \partial f(x^\star)\}^+ \\ &= \{\lambda\partial f(x^\star) : \lambda \leq 0\}.\end{aligned}$$

Theorem 1.4 *Let f be a convex function and $x^\star \in domf$. Assume that $D_f^<(x^\star) \neq \emptyset$ and $D_f^{\leq}(x^\star)$ is a polyhedral cone. Then f has GLFS property at $x^\star$.*

Proof: Since $D_f^{\leq}(x^\star)$ is a polyhedral cone, it is closed and convex. Therefore

$$D_f^{\leq}(x^\star) = cl(D_f^{\leq}(x^\star)).$$

On the other hand

$$D_f^=(x^\star) \subset \{d : f'(x^\star; d) = 0\} \quad \text{(see [4, Lemma 1.4(b)])}$$

and

$$D_f^<(x^\star) = \{d : f'(x^\star; d) < 0\} \neq \emptyset$$

imply that

$$\begin{aligned}cl(D_f^{\leq}(x^\star)) &= cl(D_f^<(x^\star)) \\ &= \{d : f'(x^\star; d) \leq 0\}.\end{aligned}$$

This together with

$$D_f^{\leq}(x^\star) = D_f^=(x^\star) \cup D_f^<(x^\star)$$

imply that

$$D_f^<(x^\star) \cup D_f^=(x^\star) = D_f^<(x^\star) \cup \{d : f'(x^\star; d) = 0\}.$$

Since

$$D_f^<(x^\star) \cap \{d : f'(x^\star; d) = 0\} = \emptyset$$

and

$$D_f^<(x^\star) \cap D_f^=(x^\star) = \emptyset,$$

it follows that

$$D_f^{=}(x^\star) = \{d : f'(x^\star; d) = 0\},$$

thus proving the theorem.

□

The following example shows that the converse of Theorem 1.4 is not true.

Example 1.2 Consider the convex function

$$f(x_1, x_2, x_3) = \sqrt{x_1^2 + x_2^2} - x_3$$

at $x^\star = (0,0,0)^T$.
It can easily be seen that

$$D_f^{=}(x^\star) = \{d \in R^3 : d_3 = \sqrt{d_1^2 + d_2^2}\}.$$

This is a cone yet not a convex cone (note that f is not differentiable at $x^\star$). Furthermore

$$D_f^{<}(x^\star) = \{d \in R^3 : d_3 > \sqrt{d_1^2 + d_2^2}\},$$

which is a nonempty set. Therefore

$$D_f^{\leq}(x^\star) = \{d \in R^3 : d_3 \geq \sqrt{d_1^2 + d_2^2}\},$$

which is a closed and convex cone but not polyhedral. On the other hand

$$\begin{aligned} f'(x^\star; d) &= \lim_{\lambda \to 0^+} \frac{\sqrt{\lambda^2 d_1^2 + \lambda^2 d_2^2} - \lambda d_3}{\lambda} \\ &= \sqrt{d_1^2 + d_2^2} - d_3. \end{aligned}$$

Therefore

$$\{d : f'(x^\star; d) = 0\} = D_f^{=}(x^\star),$$

i.e., f has GLFS property at $x^\star$ and $D_f^{<}(x^\star) \neq \emptyset$, yet $D_f^{\leq}(x^\star)$ is not a polyhedral cone.

Theorem 1.5 *Assume that f is differentiable at $x^\star$. Then f has LFS property at $x^\star$ if, and only if, it has GLFS property at $x^\star$.*

Proof: (Sufficiency:) Assume that f has GLFS property at $x^\star$. Then

$$D_f^{=}(x^\star) = \{d : f'(x^\star; d) = 0\}$$

and, since f is differentiable at $x^\star$,

$$f'(x^\star; d) = \nabla f(x^\star) d.$$

Therefore

$$D_f^{=}(x^\star) = N(\nabla f(x^\star)),$$

proving that f has LFS property at $x^\star$.
(Necessity:) Immediately follows from the definition of the LFS functions.

□

We now extend the definition of LFS functions from differentiable to non-smooth convex functions.

Definition 1.5 *Consider a convex function f. Assume that $x^\star \in dom f$. Then f has LFS property at $x^\star$ if it satisfies one of the following two conditions at $x^\star$:*

(i) $D_f^{<}(x^\star) \neq \emptyset$ and $D_f^{\leq}(x^\star)$ is polyhedral;

(ii) $D_f^{<}(x^\star) = \emptyset$, f has GLFS property at $x^\star$, and $D_f^{=}(x^\star)$ is polyhedral.

Note that $D_f^{<}(x^\star) \neq \emptyset$ is the same as $0 \notin \partial f(x^\star)$, and $D_f^{<}(x^\star) = \emptyset$ is the same as $0 \in \partial f(x^\star)$.

At this point we first recall some more results from [21] to help us identify some of the LFS functions. We will then give examples of LFS functions and also examples of functions that are GLFS but not LFS.

The set

$$\{(x, \mu) : x \in dom f,\ \mu \in R,\ f(x) \leq \mu\}$$

is called the *epigraph* of f and is denoted by $epif$.

Definition 1.6 *A polyhedral convex function is a convex function whose epigraph is polyhedral.*

Polyhedral convex functions enjoy the following two properties (see [21]).

Lemma 1.5 *Assume that f is a polyhedral convex function and $x^\star \in dom f$. Then $\partial f(x^\star)$ is a polyhedral convex set and $f'(x^\star; d)$ is a polyhedral convex function.*

Lemma 1.6 *If f^1 and f^2 are polyhedral convex functions, then so is $f^1 + f^2$.*

Using the above lemmas, we can now conclude the following important result.

Corollary 1.3 *If f is a polyhedral convex function, and if it has GLFS property at $x^\star$, then f has LFS property at $x^\star$.*

Proof: By Lemma 1.5, $f'(x^*; d)$ is a polyhedral convex function. Therefore, its level sets are polyhedral. It follows that $\{d : f'(x^*; d) \leq 0\}$ is a polyhedral cone. Furthermore, since f has GLFS property at x^*, the cone

$$D_f^{\leq}(x^*) = \{d : f'(x^*; d) \leq 0\}$$

is also a polyhedral cone. This implies that f always satisfies one of the two conditions of Definition 1.5 and thus has LFS property at x^*.

□

Functions

$$f(x) = \|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

and

$$f(x) = \|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$$

are polyhedral convex functions. Since linear functions are polyhedral functions, it can easily be seen from Lemma 1.6 that

$$f(x) = a^T x + \sum_{j \in B} |x_j|$$

where $B \subset \{1, 2, \ldots, n\}$ is also a polyhedral convex function.

We will show that these functions also enjoy the GLFS property at any point in their domain and thus have LFS property at any point in their domain. The next result is nontrivial, so we state it as a lemma.

Lemma 1.7 *Consider*

$$f(x) = a^T x + \sum_{j \in B} |x_j|,$$

where $B \subset \{1, 2, \ldots, n\}$. *Then* f *has GLFS property at any* $x^* \in R^n$.

Proof: Since $D_f^{=}(x^*) \subset \{d : f'(x^*; d) = 0\}$, it is enough to show that

$$\{d : f'(x^*; d) = 0\} \subset D_f^{=}(x^*).$$

Consider

$$f'(x^*; d) = \lim_{\lambda \to 0^+} \frac{[a^T(x^* + \lambda d) + \sum_{j \in B} |x_j^* + \lambda d_j|] - [a^T x^* + \sum_{j \in B} |x_j^*|]}{\lambda}$$

where $B = B_1 \cup B_2$, with

$$B_1 = \{j \in B : x_j^* = 0\}$$

and

$$B_2 = \{j \in B : x_j^* \neq 0\}.$$

For $\lambda > 0$ sufficiently close to zero,

$$sgn(x_j^\star + \lambda d_j) = sgn(x_j^\star), \quad j \in B_2.$$

Therefore

$$f'(x^\star; d) = \lim_{\lambda \to 0^+} \frac{\lambda a^T d + \sum_{j \in B_2} [sgn(x_j^\star)(x_j^\star + \lambda d_j) - sgn(x_j^\star)(x_j^\star)] + \sum_{j \in B_1} |\lambda d_j|}{\lambda}.$$

It then follows that

$$f'(x^\star; d) = a^T d + \sum_{j \in B_2} sgn(x_j^\star) d_j + \sum_{j \in B_1} |d_j|. \tag{1.5}$$

This means that

$$\{d : f'(x^\star; d) = 0\} = \{d : a^T d + \sum_{j \in B_2} sgn(x_j^\star) d_j + \sum_{j \in B_1} |d_j| = 0\}. \tag{1.6}$$

Now, for every $d \in \{d : f'(x^\star; d) = 0\}$ there exists an $\bar{\alpha} > 0$ such that

$$sgn(x_j^\star + \alpha d) = sgn(x_j^\star), \quad j \in B_2 \text{ and } \quad 0 \leq \alpha \leq \bar{\alpha}. \tag{1.7}$$

Therefore, given any d that satisfies (1.6), there exists an $\bar{\alpha} > 0$ such that for $0 \leq \alpha \leq \bar{\alpha}$ the following holds:

$$\begin{aligned}
f(x^\star + \alpha d) &= a^T(x^\star + \alpha d) + \sum_{j \in B} |x_j^\star + \alpha d_j| \\
&= a^T x^\star + \alpha d + \sum_{j \in B_2} sgn(x_j^\star)(x_j^\star + \alpha d_j) + \sum_{j \in B_1} \alpha |d_j| \\
&= [a^T x^\star + \sum_{j \in B_2} sgn(x_j^\star)(x_j^\star)] + \alpha [a^T d + \sum_{j \in B_2} sgn(x_j^\star) d_j + \sum_{j \in B_1} |d_j|] \\
&= a^T x^\star + \sum_{j \in B_2} |x_j^\star| \text{ by (1.6)} \\
&= f(x^\star).
\end{aligned}$$

This means that $d \in D_f^=(x^\star)$, proving the lemma.

□

It immediately follows from the above lemma that the function

$$f(x) = \|x\|_1$$

has GLFS property at any $x^\star \in R^n$. The next result is also nontrivial, so it is again stated as a lemma.

Lemma 1.8 *The function*

$$f(x) = \|x\|_\infty$$

has GLFS property at any $x^* \in R^n$.

Proof: It is shown in [21, Chapter 23] that

$$\partial f(0) = conv\{\pm e_1, \cdots, \pm e_n\}$$

and if $x^* \neq 0$, then

$$\partial f(x^*) = conv\{sgn(x_j^*)e_j : j \in J_{x^*}\}$$

where J_{x^*} is defined by $J_{x^*} = \{j \in \{1, \ldots, n\} : |x_j^*| = f(x^*)\}$.

Now, for $x^* = 0$,

$$\begin{aligned} f'(x^*; d) &= \lim_{\lambda \to 0+} \frac{\max\{|\lambda d_i| : 1 \leq i \leq n\}}{\lambda} \\ &= \|d\|_\infty. \end{aligned}$$

Therefore

$$\{d : f'(x^*; d) = 0\} = D_f^=(x^*).$$

If $x^* \neq 0$, then

$$\partial f(x^*) = \sum_{j \in J_{x^*}} \lambda_j sgn(x_j^*) e_j$$

where

$$\sum_{j \in J_{x^*}} \lambda_j = 1, \ \lambda_j \geq 0.$$

(Note that $x^* \neq 0$ implies that $sgn(x_j^*) \neq 0$ for all $j \in J_{x^*}$.) Therefore, by Lemma 1.3,

$$f'(x^*; d) = \max\{\sum_{j \in J_{x^*}} \lambda_j sgn(x_j^*) d_j \ : \sum_{j \in J_{x^*}} \lambda_j = 1, \ \lambda_j \geq 0\}. \quad (1.8)$$

The following four cases can occur:

(i) If $sgn(x_j^*)d_j > 0$ for every $j \in J_{x^*}$, then $f'(x^*; d)$ will be strictly positive;

(ii) If $sgn(x_j^*)d_j < 0$ for every $j \in J_{x^*}$, then $f'(x^*; d)$ will be strictly negative;

(iii) If $sgn(x_j^*)d_j > 0$ for at least one $j \in J_{x^*}$, and $sgn(x_j^*)d_j \leq 0$ for the rest of the indices $j \in J_{x^*}$, then $f'(x^*; d)$ will be strictly positive;

(iv) If $d_{\hat{j}} = 0$ for at least one $\hat{j} \in J_{x^*}$, and $sgn(x_j^*)d_j \leq 0$, $j \neq \hat{j}$, then $f'(x^*; d)$ will be zero.

Therefore

$$\{d : f'(x^\star; d) = 0\} = \left\{ (d_j) : \begin{array}{ll} d_j \in R & \text{if } j \notin J_{x^\star} \\ sgn(x_j^\star) d_j \leq 0 & \\ \text{(with at least one equality)} & \text{if } j \in J_{x^\star} \end{array} \right\}. \tag{1.9}$$

Since $|x_i^\star| < |x_j^\star|$ for $i \notin J_{x^\star}$ and $j \in J_{x^\star}$, then for any d satisfying (1.9) there exists an $\bar{\alpha} > 0$ such that for all $0 \leq \alpha \leq \bar{\alpha}$

$$\max_{i \notin J_{x^\star}} \{|x_i^\star + \alpha d_i|\} \leq |x_j^\star|.$$

Besides, $sgn(x_j^\star) d_j \leq 0$ (with at least one equality) together with the fact that $\alpha \geq 0$ imply that

$$|x_j^\star + \alpha d_j| \leq |x_j^\star|, \quad j \in J_{x^\star} \text{ (with at least one equality).}$$

It, therefore, follows that

$$\begin{aligned} f(x^\star + \alpha d) &= \max_{j \in J_{x^\star}} \{|x_j^\star + \alpha d|\} \\ &= |x_j^\star| \\ &= f(x^\star). \end{aligned}$$

Thus, $d \in D_f^=(x^\star)$ i.e., $\{d : f'(x^\star; d) = 0\} \subset D_f^=(x^\star)$ proving the lemma.

□

The following result has been borrowed from [28].

Lemma 1.9 *Consider the function*

$$\begin{aligned} f(x) &= dist(x - x^\star, K) \\ &\triangleq \inf_{z \in K} \|(x - x^\star) - z\| \end{aligned}$$

where K is a convex cone and x, $x^\star \in R^n$. Then f is a convex function on R^n and has GLFS property at $x^\star$. Furthermore

$$f'(x^\star; d) = dist(d, K) = \begin{cases} 0 & \text{if } d \in K \\ \text{positive} & \text{otherwise.} \end{cases}$$

Now using Corollary 1.3 and the last three lemmas, we conclude the following results:

(i) $f(x) = a^T x + \sum_{j \in B} |x_j|$, in particular $f(x) = \|x\|_1$, are LFS at any $x^\star \in R^n$;

(ii) $f(x) = \|x\|_\infty$ is LFS at any $x^\star \in R^n$;

(iii) $f(x) = dist(x - x^\star, K)$ is LFS at $x^\star$ where x, $x^\star$ and K are as in Lemma 1.9 and K is, in addition, a polyhedral cone.

Let us remind the reader that not every GLFS function is LFS. Consider, for example, the function

$$f(x_1, x_2, x_3) = \sqrt{x_1^2 + x_2^2} - x_3$$

at $x^\star = (0,0,0)^T$, from Example 1.2. As we showed in that example this function has GLFS property at $x^\star$ yet does not have LFS property at that point. Furthermore, the function

$$f(x) = dist(x - x^\star, K)$$

has GLFS property at $x^\star$ whenever K is a convex cone but does not have LFS property at that point if K is not a polyhedral cone.

1.3 Characterizations of Optimality

Let us split the constraints of the convex program (P) into those that have LFS property at a given feasible point $x^\star$ and those that do not. Let the indices of the former group belong to $\mathcal{Q}$ and the indices of the latter to $\mathcal{R}$, where $\mathcal{P} = \mathcal{Q} \cup \mathcal{R}$. We obtain the following program :

$$(P') \qquad \begin{array}{ll} \text{Min} & f^0(x) \\ \text{s.t.} & \\ & f^i(x) \leq 0 \quad i \in \mathcal{Q} \\ & f^i(x) \leq 0 \quad i \in \mathcal{R} = \mathcal{P} \setminus \mathcal{Q}. \end{array}$$

Using this form of a convex program, we will give necessary and sufficient conditions for optimality in primal, dual and saddle-point inequality forms.

1.3.1 Primal and Dual Characterizations

Associated with a convex program in the form (P') we define

$$\mathcal{R}^= = \{i \in \mathcal{R} : x \in F \Rightarrow f^i(x) = 0\},$$

$$\mathcal{R}(x^\star) = \{i \in \mathcal{R} : f^i(x^\star) = 0\}$$

and

$$\mathcal{Q}(x^\star) = \{i \in \mathcal{Q} : f^i(x^\star) = 0\}.$$

Note that the same algorithms for computation of $\mathcal{P}^=$ (see, e.g., [4,6,39]) can be applied to compute $\mathcal{R}^=$ by considering only the constraints in the set $\mathcal{R}$.

Lemma 1.10 *Consider the convex program in the form (P'). If $\mathcal{R}(x^\star) \setminus \mathcal{R}^= \neq \emptyset$, then there exists $\hat{x} \in F$ such that*

$$f^k(\hat{x}) < 0, \quad k \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=.$$

Proof: From the definition of $\mathcal{R}^=$, it follows that

$$\mathcal{R}(x^\star) \setminus \mathcal{R}^= = \{k \in \mathcal{R}(x^\star) : \exists x^k \in F \ni f^k(x^k) < 0\}.$$

By convexity of the f^k's, the "center of gravity" of $\{x^k : k \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=\}$,

$$\hat{x} = \frac{1}{card(\mathcal{R}(x^\star) \setminus \mathcal{R}^=)} \sum_{k \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} x^k,$$

is feasible and $f^k(\hat{x}) < 0, \;\; k \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=$.

Theorem 1.6 *A feasible point $x^\star$ is an optimal solution of the convex program (P') if, and only if,*

$$D_0^<(x^\star) \cap D^<_{\mathcal{R}(x^\star)\setminus\mathcal{R}^=}(x^\star) \cap conv\{D^=_{\mathcal{R}^=}(x^\star)\} \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) = \emptyset. \qquad (1.10)$$

Proof: (Sufficiency:) Assume that $x^\star \in F$ is not optimal. Then there exists $\hat{x} \in F$ such that

$$\begin{aligned} f^0(\hat{x}) &< f^0(x^\star) \\ f^i(\hat{x}) &\leq 0 \quad i \in \mathcal{Q}(x^\star) \\ f^i(\hat{x}) &< 0 \quad i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^= \\ f^i(\hat{x}) &= 0 \quad i \in \mathcal{R}^=. \end{aligned}$$

This and the convexity of the functions imply that

$$\hat{d} = (\hat{x} - x^\star) \in D_0^<(x^\star) \cap D^<_{\mathcal{R}(x^\star)\setminus\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap D^{\leq}_{\mathcal{R}^=}(x^\star).$$

But this also means that $\hat{d}$ is a feasible direction, and since $\mathcal{R}^= \subset \mathcal{P}^=$, it follows from Lemma 1.1(g) that

$$\hat{d} \in D^=_{\mathcal{R}^=}(x^\star).$$

Therefore

$$\hat{d} = (\hat{x} - x^\star) \in D_0^<(x^\star) \cap D^<_{\mathcal{R}(x^\star)\setminus\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap D^=_{\mathcal{R}^=}(x^\star),$$

implying that

$$\hat{d} = (\hat{x} - x^\star) \in D_0^<(x^\star) \cap D^<_{\mathcal{R}(x^\star)\setminus\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^=_{\mathcal{R}^=}(x^\star)\},$$

which is a nonempty set.
(Necessity:) Assume now that

$$D_0^{<}(x^\star) \cap D^{<}_{\mathcal{R}(x^\star)\setminus\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} \neq \emptyset.$$

This means that there exist

$$d \in D_0^{<}(x^\star) \cap D^{<}_{\mathcal{R}(x^\star)\setminus\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^{=}_{\mathcal{R}^=}(x^\star)\}$$

and $\bar{\alpha} > 0$ sufficiently small such that for any $0 < \alpha \leq \bar{\alpha}$, we have

$$\begin{array}{rcll} f^0(x^\star + \alpha d) & < & f^0(x^\star) & \\ f^i(x^\star + \alpha d) & \leq & 0, & i \in \mathcal{Q}(x^\star) \\ f^i(x^\star + \alpha d) & < & 0, & i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^= \\ f^i(x^\star + \alpha d) & \leq & 0, & i \in \mathcal{R}^=. \end{array}$$

Besides, $f^j(x^\star) < 0,\ j \in \mathcal{P} \setminus (\mathcal{R}(x^\star) \cup \mathcal{Q}(x^\star))$ which implies that

$$f^j(x^\star + \alpha d) \leq 0, \quad j \in \mathcal{P} \setminus \mathcal{P}(x^\star)$$

for all $\alpha > 0$ sufficiently small, by continuity. Hence, $\hat{x} = x^\star + \alpha d$ for all $\alpha > 0$ sufficiently small, is feasible, and $f^0(\hat{x}) < f^0(x^\star)$, contradicting optimality of $x^\star$.

Theorem 1.7 (Dual Version of Theorem 1.6) *A feasible point $x^\star$ is an optimal solution of the convex program (P') if, and only if, there exist vectors*

$$0 \neq y^0 \in \{D_0^{<}(x^\star)\}^+ \ , \ y^i \in \{D^{<}_{f^i}(x^\star)\}^+, \ i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^= \ ,$$

and

$$y \in \left\{conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\right\}^+ \tag{1.11}$$

such that

$$y^0 + \sum_{i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=} y^i + y = 0.$$

Proof: By Theorem 1.6, $x^\star \in F$ is optimal if, and only if, (1.10) holds. The cones $D_0^{<}(x^\star)$ and $D^{<}_{f^i}(x^\star)$, $i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=$, are open and convex cones. The cone $conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)$ is also convex but not necessarily open. Therefore, by the Dubovitskii-Milyutin theorem, (see, e.g., [4]), $x^\star$ is optimal if, and only if, there exist vectors

$$y^0 \in \{D_0^{<}(x^\star)\}^+ \ , \ y^i \in \{D^{<}_{f^i}(x^\star)\}^+, \ i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^= \ ,$$

and

$$y \in \left\{conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\right\}^+ \quad \text{not all zero} \tag{1.12}$$

such that

$$y^0 + \sum_{i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} y^i + y = 0.$$

Now we must show that $y^0 \neq 0$. Note that (1.11) implies (1.12) which proves sufficiency. To prove necessity, assume that $x^\star \in F$ is optimal and $y^0 = 0$ in (1.12). Then only two cases may occur. If $\mathcal{R}(x^\star) \setminus \mathcal{R}^= = \emptyset$, then in (1.12)

$$y^0 + y = 0.$$

But this and $y^0 = 0$ imply that $y = 0$. We have obtained a contradiction since not all vectors in (1.12) are zero. On the other hand, if $\mathcal{R}(x^\star) \setminus \mathcal{R}^= \neq \emptyset$ then by (1.12) there exists y^j, for some $j \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=$, such that $y^j \neq 0$ (since $y^0 = 0$). Now applying the Dubovitskii-Milyutin theorem to

$$\sum_{i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} y^i + y = 0,$$

it follows that

$$D^{<}_{\mathcal{R}(x^\star) \setminus \mathcal{R}^=}(x^\star) \cap conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) = \emptyset. \tag{1.13}$$

But since $\mathcal{R}(x^\star) \setminus \mathcal{R}^= \neq \emptyset$, by Lemma 1.10, there exists $\hat{x} \in F$ such that

$$\begin{array}{lcll} f^i(\hat{x}) & \leq & 0, & i \in \mathcal{Q}(x^\star) \\ f^i(\hat{x}) & < & 0, & i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^= \\ f^i(\hat{x}) & = & 0, & i \in \mathcal{R}^=. \end{array}$$

Hence it follows that

$$\hat{d} \triangleq \hat{x} - x^\star \in D^{<}_{\mathcal{R}(x^\star) \setminus \mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap D^{\leq}_{\mathcal{R}^=}(x^\star).$$

But this also means that $\hat{d} \in F(x^\star)$, which imlies that $\hat{d} \in D^{=}_{\mathcal{R}^=}(x^\star)$ by Lemma 1.1(g). Therefore

$$\hat{d} = (\hat{x} - x^\star) \in D^{<}_{\mathcal{R}(x^\star) \setminus \mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap D^{=}_{\mathcal{R}^=}(x^\star),$$

which is a nonempty set. Thus

$$D^{<}_{\mathcal{R}(x^\star) \setminus \mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} \neq \emptyset,$$

a contradiction to (1.13). This completes the proof.

□

We now use this result to derive a dual characterization involving subgradients.

Theorem 1.8 *Consider the convex program (P'), and assume that $x^\star$ is a feasible solution satisfying*

$$x^\star \in \bigcap_{i \in \{0\} \cup \mathcal{R}(x^\star) \setminus \mathcal{R}^=} int(dom f^i)$$

and that $conv\{D^=_{\mathcal{R}^=}(x^\star)\}$ is closed. Then $x^\star$ is optimal if, and only if, the system

$$(h^0)^T + \sum_{i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} \lambda_i (h^i)^T \in cl\left[\{D^=_{\mathcal{R}^=}(x^\star)\}^+ + \{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+\right]$$

$$\lambda_i \geq 0, \quad i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^= \tag{1.14}$$

is consistent for some $h^i \in \partial f^i(x^\star)$, $i \in \{0\} \cup \mathcal{R}(x^\star) \setminus \mathcal{R}^=$.

Proof: (Sufficiency:) Let

$$h \triangleq (h^0)^T + \sum_{i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} \lambda_i (h^i)^T \in cl\left[\{D^=_{\mathcal{R}^=}(x^\star)\}^+ + \{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+\right] \tag{1.15}$$

where $h^i \in \partial f^i(x^\star)$, $i \in \{0\} \cup \mathcal{R}(x^\star) \setminus \mathcal{R}^=$. Since by definition,

$$f^i(x) \geq f^i(x^\star) + h^i(x - x^\star), \quad i \in \{0\} \cup \mathcal{R}(x^\star) \setminus \mathcal{R}^=$$

for every $x \in R^n$, then

$$\begin{aligned} f^0(x) + \sum_{i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} \lambda_i f^i(x) &\geq f^0(x^\star) + \sum_{i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} \lambda_i f^i(x^\star) \\ &+ (h^0 + \sum_{i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} \lambda_i h^i)(x - x^\star). \end{aligned} \tag{1.16}$$

Moreover, since $f^i(x) \leq 0$, for $i \in \mathcal{P}$ and $x \in F$, and $f^i(x^\star) = 0$, $i \in \mathcal{R}(x^\star)$, it follows from (1.15) and (1.16) that

$$f^0(x) \geq f^0(x^\star) + h(x - x^\star) \quad \text{for every } x \in F.$$

By convexity of F, we have $x - x^\star \in F(x^\star)$. Thus

$$x - x^\star \in D^=_{\mathcal{R}^=}(x^\star). \tag{1.17}$$

Besides, since $F(x^\star) \subset D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)$, we also have $x - x^\star \in D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)$. Therefore

$$x - x^\star \in \left(conv\{D^=_{\mathcal{R}^=}(x^\star)\} \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\right). \tag{1.18}$$

On the other hand, since the cone $D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)$ is polyhedral and the cone $conv\{D^=_{\mathcal{R}^=}(x^\star)\}$ is closed by assumption, by Lemma 1.1(h,i) we have

$$cl\left[\{D^=_{\mathcal{R}^=}(x^\star)\}^+ + \{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+\right] = \left\{conv\{D^=_{\mathcal{R}^=}(x^\star)\} \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\right\}^+. \tag{1.19}$$

Hence (1.15), (1.18), and (1.19) imply that $h(x-x^\star) \geq 0$. Therefore

$$f^0(x^\star) \leq f^0(x) \quad \text{for every } x \in F.$$

(Necessity:) Assume that $x^\star \in F$ is optimal. If $x^\star$ is a (unconstrained) minimizer of f^0, then $D_0^{\leq}(x^\star) = \emptyset$ which also means $0 \in \partial f^0(x^\star)$ and the theorem holds with $\lambda_i = 0$, $i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=$. If $x^\star$ is not a minimizer of f^0, then $D_0^{\leq}(x^\star) \neq \emptyset$ and therefore

$$\{D_0^{\leq}(x^\star)\}^+ = \{\mu_0 \partial f^0(x^\star) \ : \ \mu_0 \leq 0\}. \tag{1.20}$$

Also, for each $i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=$, $x^\star$ is not a minimizer of f^i since by Lemma 1.10 there exists $\hat{x} \in F$ such that

$$f^i(\hat{x}) < 0, \quad i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=.$$

This and the assumption that $x^\star$ is in the interior of the domain of f^i, $i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=$, imply that, by Lemma 1.4, for each $i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=$

$$\{D_{f^i}^{<}(x^\star)\}^+ = \{\mu_i \partial f^i(x^\star) \ : \mu_i \leq 0\}. \tag{1.21}$$

Again polyhedrality of $D_{\mathcal{Q}(x^\star)}^{\leq}(x^\star)$ and closedness of $conv\{D_{\mathcal{R}^=}^{=}(x^\star)\}$ imply that (1.19) holds. It follows from Theorem 1.7 that

$$-y = y^0 + \sum_{i\in\mathcal{R}(x^\star)\setminus\mathcal{R}^=} y^i. \tag{1.22}$$

Now by (1.20) and (1.21)

$$y^0 \in \{\mu_0 \partial f^0(x^\star) \ : \mu_0 \leq 0\}$$

and

$$y^i \in \{\mu_i \partial f^i(x^\star) \ : \mu_i \leq 0\}, \ i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=.$$

Hence, it follows that there exist

$$\mu_0 < 0 \ \text{(since } y^0 \neq 0) \ , \mu_i \leq 0 \ , i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^= \text{ and}$$

$$h^i \in \partial f^i(x^\star), \quad i \in \{0\}\cup\mathcal{R}(x^\star)\setminus\mathcal{R}^=$$

such that

$$-y = \mu_0 h^0 + \sum_{i\in\mathcal{R}(x^\star)\setminus\mathcal{R}^=} \mu_i h^i.$$

Let $h = -\frac{y}{\mu_0}$ and $\lambda_i = \frac{\mu_i}{\mu_0}$ for $i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=$. Then

$$h = h^0 + \sum_{i\in\mathcal{R}(x^\star)\setminus\mathcal{R}^=} \lambda_i h^i$$

where $h \in cl\left[\{D^{=}_{\overline{\mathcal{R}}^{=}}(x^\star)\}^+ + \{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+\right]$. This completes the proof.

□

If in the above theorem we assume further that $x^\star$ also belongs to the interior of the domain of f^k, $k \in \mathcal{Q}(x^\star)$, then LFS property of f^k, $k \in \mathcal{Q}(x^\star)$, at $x^\star$, by Lemma 1.1(i), Lemma 1.4 and Corollary 1.1, implies that

$$\begin{aligned}\{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+ &= \sum_{k\in\mathcal{Q}(x^\star)} \{D^{\leq}_{f^k}(x^\star)\}^+ \\ &= \sum_{k\in\mathcal{Q}(x^\star)} \{\mu_k \partial f^k(x^\star) \ : \mu_k \leq 0\}.\end{aligned}$$

This yields the following result.

Corollary 1.4 *Consider the convex program* (P') *, and assume that* $x^\star \in F$ *is such that*

$$x^\star \in \bigcap_{i\in\{0\}\cup\mathcal{P}(x^\star)\setminus\mathcal{R}^{=}} int(dom f^i),$$

and that $conv\{D^{=}_{\overline{\mathcal{R}}^{=}}(x^\star)\}$ *is closed. Then* $x^\star$ *is optimal if, and only if, the system*

$$(h^0)^T + \sum_{i\in\mathcal{R}(x^\star)\setminus\mathcal{R}^{=}} \lambda_i (h^i)^T \in cl\left[\{D^{=}_{\overline{\mathcal{R}}^{=}}(x^\star)\}^+ + \sum_{k\in\mathcal{Q}(x^\star)} \{\mu_k \partial f^k(x^\star) \ : \mu_k \leq 0\}\right]$$

$$\lambda_i \geq 0, \quad i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^{=} \tag{1.23}$$

is consistent for some $h^i \in \partial f^i(x^\star)$, $i \in \{0\}\cup\mathcal{R}(x^\star)\setminus\mathcal{R}^{=}$.

Note that the closure can be omitted in Corollary 1.4 if the cone $D^{=}_{\overline{\mathcal{R}}^{=}}(x^\star)$ is polyhedral. This immediately yields our next result.

Corollary 1.5 *Consider the convex program* (P') *, and assume that* $x^\star \in F$ *is such that*

$$x^\star \in \bigcap_{i\in\{0\}\cup\mathcal{P}(x^\star)\setminus\mathcal{R}^{=}} int(dom f^i).$$

Furthermore, assume that the cone $D^{=}_{\overline{\mathcal{R}}^{=}}(x^\star)$ *is polyhedral. Then* $x^\star$ *is optimal if, and only if, the system*

$$(h^0)^T + \sum_{i\in\mathcal{P}(x^\star)\setminus\mathcal{R}^{=}} \lambda_i (h^i)^T \in \{D^{=}_{\overline{\mathcal{R}}^{=}}(x^\star)\}^+$$

$$\lambda_i \geq 0, \quad i \in \mathcal{P}(x^\star)\setminus\mathcal{R}^{=} \tag{1.24}$$

is consistent for some $h^i \in \partial f^i(x^\star)$, $i \in \{0\}\cup\mathcal{P}(x^\star)\setminus\mathcal{R}^{=}$.

Finally, if all the functions f^k, $k \in \mathcal{P}(x^\star)$ have LFS property at $x^\star$, then Corollary 1.5 is further simplified to our next result.

Corollary 1.6 *Consider the convex program (P'), and assume that $x^\star \in F$ is such that*

$$x^\star \in \bigcap_{i\in\{0\}\cup\mathcal{P}(x^\star)} int(dom f^i).$$

Furthermore, assume that f^k, $k \in \mathcal{P}(x^\star)$, have LFS property at $x^\star$. Then $x^\star \in F$ is optimal if, and only if,

$$0 \in \partial f^0(x^\star) + \sum_{i\in\mathcal{P}(x^\star)} \lambda_i \partial f^i(x^\star) \tag{1.25}$$

for some $\lambda_i \geq 0$, $i \in \mathcal{P}(x^\star)$.

We recall that conditions under which (1.25) holds are also constraint qualifications. Thus, using Corollaries 1.5 and 1.6, we can conclude the following important result.

Corollary 1.7 *Consider the convex program (P') at a feasible point $x^\star$. Then the following conditions are constraint qualifications:*

(i) All the functions f^i, $i \in \mathcal{P}(x^\star)$, have LFS property at $x^\star$;

(ii) There exists $\hat{x} \in F$ such that $f^j(\hat{x}) < 0$, $j \in \mathcal{R}$.

The following example shows that closedness of $conv\{D^{=}_{\mathcal{R}^=}(x^\star)\}$ is required in Theorem 1.8. The constraints of the example have been borrowed from [28].

Example 1.3 Consider the program

$$\begin{array}{ll} \text{Min} & f^0(x) = x_1 - x_2 \\ \text{s.t.} & \\ & f^1(x) \leq 0 \\ & f^2(x) \leq 0, \end{array}$$

where

$$f^1(x) = \begin{cases} (x_1^2 + x_2^2 - 1)^2 & \text{if } x_1^2 + x_2^2 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f^2(x) = dist(x - x^\star, K), \quad K = \{x \in R^2 : x_1 \geq 0, \ x_2 \geq 0\},$$

around $x^\star = (1,0)^T$. Then

$$F = \{x^\star\} \ , \mathcal{Q}(x^\star) = \{2\} \text{ and } \mathcal{R}^= = \{1\}.$$

Furthermore, the cone

$$conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} = D^{=}_{\mathcal{R}^=}(x^\star) = \{d \in R^2 : d_1 < 0\} \cup \{0\}$$

is not closed. Moreover

$$D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) = \{d : f'(x^\star; d) = 0\} = K, \quad \nabla f^0(x^\star) = (1,-1),$$

$$K^+ = K, \; \{D^{=}_{\mathcal{R}^=}(x^\star)\}^+ = \{(d_1, 0)^T \; : d_1 \leq 0\}$$

and

$$cl\left[\{D^{=}_{\mathcal{R}^=}(x^\star)\}^+ + \{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+\right] = \{d \in R^2 : d_2 \leq 0\}.$$

The optimality condition from Theorem 1.8 becomes

$$(\nabla f^0(x^\star))^T \in cl\left[\{D^{=}_{\mathcal{R}^=}(x^\star)\}^+ + \{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+\right],$$

i.e.,

$$(1,-1)^T \in \{d \in R^2 : d_2 \geq 0\}.$$

This is clearly an inconsistent system.

□

The following example shows that the polyhedrality of the cone $D^{=}_{\mathcal{R}^=}(x^\star)$ is required in Corollary 1.5.

Example 1.4 Consider the program

$$\begin{array}{ll} \text{Min} & f^0 = x_2^2 + x_3 \\ \text{s.t.} & \\ & f^1 = x_1 \leq 0 \\ & f^2 = dist(x - x^\star, C) \leq 0, \end{array}$$

where

$$C = \{x \in R^3 : \; 2x_1x_2 \geq x_3^2, \;\; x_1 \geq 0, \; x_2 \geq 0\}$$

is the self-polar ("ice-cream") cone and $x^\star = (0,0,0)^T$. (Note that f^2 is not differentiable at $x^\star$.) Then

$$F = C \cap \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} : x_1 \leq 0 \right\} = \left\{ \begin{bmatrix} 0 \\ x_2 \\ 0 \end{bmatrix} : x_2 \geq 0 \right\},$$

$$\mathcal{R}^= = \{2\}, \; \mathcal{Q}(x^\star) = \{1\},$$

$$D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) = \{x \in R^3 \; : x_1 \leq 0\} \text{ and } D^{=}_{\mathcal{R}^=}(x^\star) = C.$$

Furthermore

$$\{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+ = \left\{ \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} : x_1 \leq 0 \right\} \text{ and } \{D^{=}_{\mathcal{R}^=}(x^\star)\}^+ = C^+ = C.$$

The optimality condition from Corollary 1.5 becomes

$$(\nabla f^0(x^\star))^T + \lambda_1(\nabla f^1(x^\star))^T \in C,$$

i.e.,

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

for some $\lambda_1 \geq 0$ where $2c_1c_2 \geq c_3,\ c_1 \geq 0,\ c_2 \geq 0$. This is obviously an inconsistent system.

The asymptotic version from Theorem 1.8 becomes

$$(\nabla f^0(x^\star))^T \in cl\left[\{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+ + C\right],$$

i.e.,

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \lim_{k\to\infty}(d^k + c^k)$$

for some sequences

$$d^k = \begin{bmatrix} d_1^k \\ 0 \\ 0 \end{bmatrix},\ d_1^k \leq 0$$

and $c^k \in C,\ \ k = 1, 2, \ldots$. Indeed, the choice

$$d^k = \begin{bmatrix} -k \\ 0 \\ 0 \end{bmatrix},\quad c^k = \begin{bmatrix} k \\ \frac{1}{2k} \\ 1 \end{bmatrix}$$

confirms the consistency of the asymptotic version.

□

If all the functions in the convex program (P') are differentiable, then Theorem 1.8 and Corollary 1.5 are further simplified to our next two corollaries. These two corollaries are also directly derived in [31].

Corollary 1.8 *Consider the convex program (P') where all the functions are differentiable at $x^\star$. Furthermore, assume that $D^{\equiv}_{\mathcal{R}^=}(x^\star)$ is closed. Then $x^\star \in F$ is optimal if, and only if, the system*

$$(\nabla f^0(x^\star))^T + \sum_{k\in\mathcal{R}(x^\star)\setminus\mathcal{R}^=} \lambda_k(\nabla f^k(x^\star))^T \in cl\left[\{D^{\equiv}_{\mathcal{R}^=}(x^\star)\}^+ + \{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+\right]$$

is consistent for some

$$\lambda_k \geq 0,\quad k \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=.$$

Corollary 1.9 *Consider the convex program (P') where all the functions are differentiable at $x^\star$. Furthermore, assume that $D^{=}_{\mathcal{R}^=}(x^\star)$ is polyhedral. Then $x^\star \in F$ is optimal if, and only if, the system*

$$(\nabla f^0(x^\star))^T + \sum_{k \in \mathcal{P}(x^\star) \setminus \mathcal{R}^=} \lambda_k (\nabla f^k(x^\star))^T \in \{D^{=}_{\mathcal{R}^=}(x^\star)\}^+$$

is consistent for some

$$\lambda_k \geq 0, \quad k \in \mathcal{P}(x^\star) \setminus \mathcal{R}^=.$$

The following example shows that polyhedrality of $D^{=}_{\mathcal{R}^=}(x^\star)$ is also required in the above corollary although all the functions are differentiable. A slightly different example is given in [31].

Example 1.5 Consider the program

$$\begin{aligned} \text{Min} \quad & f^0 = x_2^2 + x_3 \\ \text{s.t.} \quad & \\ & f^1 = x_1 \leq 0 \\ & f^2 = (dist(x, C))^2 \leq 0, \end{aligned}$$

where again

$$C = \{x \in R^3 : 2x_1x_2 \geq x_3^2, \ \ x_1 \geq 0, \ x_2 \geq 0\}$$

is the self-polar ("ice-cream") cone and $x^\star = (0,0,0)^T$. Note that f^2 is differentiable at $x^\star$ since

$$(f^2)'(x^\star; d) = \lim_{t \to 0+} \frac{\inf_{z \in C} \|td - z\|^2}{t} = 0 \quad \text{for every } d \in R^3.$$

Therefore, all the functions are differentiable. Furthermore

$$D^{=}_{f^2}(x^\star) = C,$$

$$\mathcal{R}^= = \{2\}, \ \mathcal{Q}(x^\star) = \{1\},$$

$$D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) = \{x \in R^3 \ : x_1 \leq 0\} \text{ and } D^{=}_{\mathcal{R}^=}(x^\star) = C.$$

The rest of the example is exactly the same as in Example 1.4.

1.3.2 A Saddle-Point Characterization

Before deriving a saddle-point characterization, we introduce the following Lagrangian function for a convex program in the form (P').

$$\mathcal{L}^{\leq}_{\mathcal{R}}(x, \lambda) = f^0(x) + \sum_{k \in \mathcal{P} \setminus \mathcal{R}^=} \lambda_k f^k(x).$$

Recall that

$$R^{=} = \{i \in \mathcal{R} : x \in F \Rightarrow f^i(x) = 0\}.$$

We denote

$$F_{\mathcal{R}}^{=} = \{x \in R^n : f^i(x) = 0, \, i \in \mathcal{R}^{=}\}.$$

Furthermore, let $q = card(\mathcal{P} \setminus \mathcal{R}^{=})$.

Now using Corollary 1.5, we will derive the next characterization of optimality in the saddle-point inequality form.

Theorem 1.9 *Consider the convex program* (P'). *Assume that* $x^\star \in F_{\mathcal{R}}^{=}$ *is such that*

$$x^\star \in \bigcap_{k \in \{0\} \cup \mathcal{P}(x^\star) \setminus \mathcal{R}^{=}} int(dom f^k),$$

and the cone $D_{\mathcal{R}^{=}}^{=}(x^\star)$ *is polyhedral. Then* $x^\star$ *is optimal if, and only if, there exists* $\lambda^\star \in R_+^q$ *such that*

$$\mathcal{L}_{\mathcal{R}}^{\leq}(x^\star, \lambda) \leq \mathcal{L}_{\mathcal{R}}^{\leq}(x^\star, \lambda^\star) \leq \mathcal{L}_{\mathcal{R}}^{\leq}(x, \lambda^\star) \tag{1.26}$$

for all $\lambda \in R_+^q$ *and for all* $x \in \{x^\star + D_{\mathcal{R}^{=}}^{=}(x^\star)\}$.

Proof: (Sufficiency) Assume that $x^\star \in F_{\mathcal{R}}^{=}$ satisfies (1.26). Then the left-hand inequality of (1.26) implies that for every $\lambda_k \geq 0$,

$$\sum_{\mathcal{P} \setminus \mathcal{R}^{=}} \lambda_k f^k(x^\star) \leq \sum_{\mathcal{P} \setminus \mathcal{R}^{=}} \lambda_k^\star f^k(x^\star). \tag{1.27}$$

First we show that $x^\star$ is feasible, i.e., $f^k(x^\star) \leq 0, \ k \in \mathcal{P} \setminus \mathcal{R}^{=}$. If $f^k(x^\star) > 0$ for some $k \in \mathcal{P} \setminus \mathcal{R}^{=}$, then (1.27) is violated by choosing λ_k sufficiently large. Next, since $x^\star \in F$ and $\lambda_k^\star \in R_+^q$

$$\sum_{k \in \mathcal{P} \setminus \mathcal{R}^{=}} \lambda_k^\star f^k(x^\star) \leq 0.$$

But from (1.27) with $\lambda_k = 0$

$$\sum_{k \in \mathcal{P} \setminus \mathcal{R}^{=}} \lambda_k^\star f^k(x^\star) \geq 0.$$

Therefore

$$\sum_{k \in \mathcal{P} \setminus \mathcal{R}^{=}} \lambda_k^\star f^k(x^\star) = 0. \tag{1.28}$$

Now, the right-hand inequality of (1.26) implies that

$$f^0(x^\star) + \sum_{\mathcal{P} \setminus \mathcal{R}^{=}} \lambda_k^\star f^k(x^\star) \leq f^0(x) + \sum_{\mathcal{P} \setminus \mathcal{R}^{=}} \lambda_k^\star f^k(x) \quad \text{for every } x - x^\star \in D_{\mathcal{R}^{=}}^{=}(x^\star). \tag{1.29}$$

Since

$$F(x^\star) \subset D_{\mathcal{R}^{=}}^{=}(x^\star),$$

then (1.29) holds, in particular, for all x satisfying

$$x - x^\star \in F(x^\star)$$

and hence for all $x \in F$, by convexity of F. Thus, by (1.28) and (1.29),

$$f^0(x^\star) \leq f^0(x) + \sum_{\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star f^k(x),$$

for every $x \in F$. This implies that for every $x \in F$,

$$f^0(x^\star) \leq f^0(x)$$

(since $f^k(x) \leq 0$ for every $x \in F$). This completes the sufficiency part.

(Necessity:) Let $x^\star$ be an optimal solution of (P'). Then by Corollary 1.5, there exist

$$\lambda_k^\star \geq 0, \;\; k \in \mathcal{P}(x^\star) \setminus \mathcal{R}^=, \;\; \text{and} \;\; h^k \in \partial f^k(x^\star), \;\; k \in \{0\} \cup (\mathcal{P}(x^\star) \setminus \mathcal{R}^=)$$

such that

$$h^0 + \sum_{k\in\mathcal{P}(x^\star)\setminus\mathcal{R}^=} \lambda_k^\star h^k \in \{D^=_{\mathcal{R}^=}(x^\star)\}^+.$$

Let $\lambda_k^\star = 0$ for $k \in \mathcal{P} \setminus \mathcal{P}(x^\star)$. Then the above can be augmented in the following way: There exist

$$\lambda_k^\star \geq 0, \;\; k \in \mathcal{P} \setminus \mathcal{R}^=, \;\; \text{and} \;\; h^k \in \partial f^k(x^\star), \;\; k \in \{0\} \cup (\mathcal{P} \setminus \mathcal{R}^=)$$

such that

$$h^0 + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star h^k \in \{D^=_{\mathcal{R}^=}(x^\star)\}^+. \tag{1.30}$$

From $h^k \in \partial f^k(x^\star)$, $k \in \mathcal{P} \setminus \mathcal{R}^=$, it follows that for every $x \in R^n$,

$$f^k(x) \geq f^k(x^\star) + h^k(x - x^\star).$$

Therefore

$$f^0(x) + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star f^k(x) \geq f^0(x^\star) + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star f^k(x^\star) + (h^0 + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star h^k)(x - x^\star).$$

But, (1.30) implies that for $x - x^\star \in D^=_{\mathcal{R}^=}(x^\star)$, we have

$$(h^0 + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star h^k)(x - x^\star) \geq 0.$$

Thus, for every $x - x^\star \in D^=_{\mathcal{R}^=}(x^\star)$,

$$f^0(x) + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star f^k(x) \geq f^0(x^\star) + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star f^k(x^\star),$$

proving the right-hand inequality. On the other hand, it is easily seen that

$$\sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star f^k(x^\star) = 0$$

which implies that, for every $\lambda_k \geq 0$, we have

$$\begin{aligned} f^0(x^\star) + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k f^k(x^\star) &\leq f^0(x^\star) \\ &\leq f^0(x^\star) + \sum_{k\in\mathcal{P}\setminus\mathcal{R}^=} \lambda_k^\star f^k(x^\star). \end{aligned}$$

This proves the left-hand inequality.

□

If, in Theorem 1.9, $\mathcal{R}^= = \emptyset$, then the Lagrangian function becomes the usual Lagrangian, $\mathcal{L}(x,\lambda)$, and the sets $F^=_{\mathcal{R}}$ and $x^\star + D^=_{\mathcal{R}^=}(x^\star)$ become R^n. One such case is when all the functions f^i, $i \in \mathcal{P}(x^\star)$ have LFS property at $x^\star$ yielding our next result.

Corollary 1.10 *Consider the convex program (P'). Assume that $x^\star \in R^n$ is such that*

$$x^\star \in \bigcap_{k\in\{0\}\cup\mathcal{P}(x^\star)} int(dom f^k),$$

and that all the functions f^i, $i \in \mathcal{P}(x^\star)$ have LFS property at $x^\star$. Then $x^\star$ is optimal if, and only if, there exists $\lambda^\star \in R^m_+$ such that

$$\mathcal{L}(x^\star,\lambda) \leq \mathcal{L}(x^\star,\lambda^\star) \leq \mathcal{L}(x,\lambda^\star) \tag{1.31}$$

for all $\lambda \in R^m_+$ and for all $x \in R^n$.

We will illustrate the result given in Theorem 1.9 by an example.

Example 1.6 Consider the program

$$\begin{aligned} \text{Min} \quad & f^0 = -x_1 + e^{x_2} \\ \text{s.t.} \quad & \\ & f^1 = |x_1| + |x_2| - 1 \leq 0 \\ & f^2 = |x_1^2 - x_2| - 1 \leq 0 \\ & f^3 = g(x) \leq 0 \end{aligned}$$

where

$$g(x) = \begin{cases} 0 & \text{if } x_1 \geq 0, \ x_2 \geq 0 \\ x_1^2 & \text{if } x_1 \leq 0, \ x_2 \geq 0 \\ x_2^2 & \text{if } x_1 \geq 0, \ x_2 \leq 0 \\ x_1^2 + x_2^2 & \text{if } x_1 \leq 0, \ x_2 \leq 0 \end{cases}$$

at $x^\star = (1,0)^T$. Then

$$F = \{x \in R^2: \ 0 \le x_1 \le 1, \ 0 \le x_2 \le 1\},$$

$$\mathcal{Q}(x^\star) = \{1\}, \quad \mathcal{R}(x^\star) = \{2,3\}, \quad \mathcal{R}^= = \{3\}$$

and

$$D^{\equiv}_{\mathcal{R}^=}(x^\star) = \{d \in R^2 : d_2 \ge 0\}.$$

The saddle-point optimality condition from Theorem 1.9 becomes

$$0 \le 0 \le -x_1 + e^{x_2} + \lambda_1^\star(|x_1| + |x_2| - 1) + \lambda_2^\star(|x_1^2 - x_2| - 1)$$

for some $\lambda_1^\star \ge 0$, $\lambda_2^\star \ge 0$ and for all $x \in \{(1,x_2)^T : x_2 \ge 0\}$. Indeed, the choice $\lambda_1^\star = 1$ and $\lambda_2^\star = 0$ confirms the consistency of the saddle-point version. Therefore, $x^\star$ is optimal. Furthermore, we have

$$\partial f^0(x^\star) = \{(-1,1)\}, \ \partial f^1(x^\star) = \{(1, 2\alpha - 1) : \ 0 \le \alpha \le 1\}, \ \partial f^2(x^\star) = \{(2,-1)\}$$

and

$$\{D^{\equiv}_{\mathcal{R}^=}(x^\star)\}^+ = \{(0,d_2)^T : d_2 \ge 0\}.$$

Therefore, the optimality condition from Corollary 1.5 becomes

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ 2\alpha - 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ \beta \end{bmatrix} \tag{1.32}$$

for some $0 \le \alpha \le 1$, $\beta \ge 0$, $\lambda_1 \ge 0$ and $\lambda_2 \ge 0$. Indeed, the choice

$$\lambda_1 = 1, \quad \lambda_2 = 0, \quad \alpha = 1 \text{ and } \beta = 2$$

confirms consistency of the above system and optimality of $x^\star$.

Remark: Wolkowicz [28] used the approach of Gould and Tolle to derive several optimality criteria which use the cones of directions of constancy. He defined $\mathcal{P}^b(x^\star)$, the set of "badly behaved" constraints, as follows :

$$\mathcal{P}^b(x^\star) = \{k \in \mathcal{P}^= : (D_k^{>}(x^\star) \cap C_{\mathcal{P}(x^\star)}(x^\star)) \setminus cl[D^{\equiv}_{\mathcal{P}^=}(x^\star)] \ne \emptyset\}$$

where

$$C_{\mathcal{P}(x^\star)}(x^\star) = \{d \in R^n : (f^k)'(x^\star; d) \le 0, \quad k \in \mathcal{P}(x^\star)\}.$$

These are the constraints in $\mathcal{P}^=$ whose analytic properties (given by the directional derivatives) do not fully describe the geometry of the feasible set (given by the feasible directions). The set $\mathcal{P}^b(x^\star)$ is also the set of constraints that creates problems in the Kuhn-Tucker theory. With

$$B_{\mathcal{P}(x^\star)}(x^\star) = \{h : h = \sum_{k \in \mathcal{P}(x^\star)} \lambda_k h^k \text{ for some } \lambda_k \ge 0, \ h^k \in \partial f^k(x^\star)\},$$

he derived the following result.

Theorem 1.10 *Consider the convex program (P). Suppose that $x^\star \in F$, that the set Ω satisfies*

$$\mathcal{P}^b(x^\star) \subset \Omega \subset \mathcal{P}^=$$

and that both $conv D_\Omega^=(x^\star)$ and $-B_{\mathcal{P}(x^\star)}(x^\star) + \{D_\Omega^=(x^\star)\}^+$ are closed. Then $x^\star$ is optimal if, and only if, the system

$$0 \in \partial f^0(x^\star) + \sum_{k \in \mathcal{P}(x^\star)} \lambda_k \partial f^k(x^\star) - \{D_\Omega^=(x^\star)\}^+ \tag{1.33}$$

is consistent for some $\lambda_k \geq 0$, $k \in \mathcal{P}(x^\star)$.

He further concluded that

$$\mathcal{P}^b(x^\star) = \emptyset \quad \text{and} \quad B_{\mathcal{P}(x^\star)}(x^\star) \;\; \text{closed}$$

is a necessary and sufficient condition for the Kuhn-Tucker theory to hold at $x^\star$, i.e., it is a weakest constraint qualification. he referred to this condition as "WCQ".

The interesting point is that if the convex function f has GLFS property at $x^\star$, then it is never "badly behaved" at that point. Therefore, using Wolkowicz results, we notice that if in the program (P') all the active constraints at $x^\star$ have LFS property at that point, then $B_{\mathcal{P}(x^\star)}(x^\star)$ will be polyhedral and therefore closed. Furthermore, since functions with LFS property at $x^\star$ have GLFS property at $x^\star$ by their definition, we have $\mathcal{P}^b(x^\star) = \emptyset$. Hence, the condition "WCQ" is indeed satisfied, implying that Corollary 1.6 is a special case of "WCQ".

1.4 Applications in Single- and Multi-Objective Programming

The characterizations of optimality derived in this chapter are simpler than the usual characterizations of optimality in the sense that $\mathcal{R}^=$ is a smaller set than $\mathcal{P}^=$. However, when considering the parametric programs in input optimization, we notice that the set $\mathcal{P}^=(\theta)$ plays a very important role in determining the stability of the mathematical model, whereas the set $\mathcal{R}^=(\theta)$ may give us no information about the stability of the model. For example, for any linear model the set $\mathcal{R}^=(\theta)$ is always empty, giving no information about the stability of the model. Therefore, while considering mathematical models, we always need the set $\mathcal{P}^=(\theta)$. However, the LFS properties simplify the calculation of the set $\mathcal{P}^=(\theta)$ at a fixed θ as shown below.

Consider the convex model (P, θ). If at $\theta = \theta^\star$ there exists $x^\star \in F(\theta^\star)$ at which all the functions have LFS property, then the algorithm by Zlobec and Craven [39] for calculation of $\mathcal{P}^=(\theta^\star)$ is significantly simplified as follows:

The algorithm begins with $\Omega(\theta^\star) = \mathcal{P}(x^\star, \theta^\star)$ and ends with $\Omega(\theta^\star) = \mathcal{P}^=(\theta^\star)$. At every iteration the program

$$(L, \Omega(\theta^\star)) \qquad \begin{array}{ll} \underset{(d)}{\text{Min}} & \sum_{k \in \Omega(\theta^\star)} \nabla f^k(x^\star, \theta^\star) d \\ \text{s.t.} & \\ & \nabla f^k(x^\star, \theta^\star) d \leq 0, \quad k \in \mathcal{P}(x^\star, \theta^\star) \\ & \|d\|_1 \leq 1 \end{array}$$

is solved.

Initialization: Find $x^\star \in F(\theta^\star)$ such that all the constraints have LFS property at $x^\star$, for the given $\theta^\star$. Calculate $\mathcal{P}(x^\star, \theta^\star)$ and set $\Omega(\theta^\star) = \mathcal{P}(x^\star, \theta^\star)$.
Step 1: Find an optimal solution $\tilde{d}$ of $(L, \Omega(\theta^\star))$. Calculate

$$\tilde{g}(\theta^\star) = \sum_{k \in \Omega(\theta^\star)} \nabla f^k(x^\star, \theta^\star) \tilde{d}.$$

Step 2: If $\tilde{g}(\theta^\star) < 0$, determine

$$K = \{k \in \mathcal{P}(x^\star, \theta^\star) : \nabla f^k(x^\star, \theta^\star) \tilde{d} < 0\};$$

set $\Omega(\theta^\star) = \Omega(\theta^\star) \setminus K$ and return to step 1. Otherwise, stop; $\Omega(\theta^\star) = \mathcal{P}^=(\theta^\star)$.

Recall that in the algorithm for the general case, given in [39], the program solved in each iteration is

$$(P, \Omega(\theta^\star)) \qquad \begin{array}{ll} \underset{(d)}{\text{Min}} & \sum_{k \in \Omega(\theta^\star)} \nabla f^k(x^\star, \theta^\star) d \\ \text{s.t.} & \\ & \nabla f^k(x^\star, \theta^\star) d + \Theta \|d - \delta^k\|_1 \leq 0 \\ & \delta^k \in D_{f^k}(x^\star) \\ & \|d\|_1 \leq 1, \ \|\delta^k\| \leq 1, \quad k \in \mathcal{P}(x^\star, \theta^\star). \end{array}$$

Note that the constraints represent the set of feasible directions of $(P, \theta^\star)$ at $x^\star$. A parameter $\Theta > 0$ is involved, and the iterations must be continued until the optimal value $\tilde{g}(\Theta, \theta^\star) < 0$ for all sufficiently small $\Theta > 0$. In the algorithm that we gave above for the LFS case, on the other hand, there is no Θ involved since the set of feasible directions at $x^\star$ for $\theta = \theta^\star$ is simplified to $\{d : \nabla f^k(x^\star, \theta^\star) d \leq 0\}$ due to LFS property.

We will now illustrate the above result by an example.

Example 1.7 Consider the constraints

$$f^1 \; = \; -2x_1 + \theta_1 x_1^2 \leq 0$$

$$
\begin{aligned}
f^2 &= x_1 \le 0 \\
f^3 &= -x_2 - \theta_2 x_3 \le 0 \\
f^4 &= \theta_3 e^{x_2+x_3} - 1 \le 0 \\
f^5 &= x_1 + 2x_2 \le 0
\end{aligned}
$$

at $\theta = \theta^* = (1,1,1)^T$.

Initialization: Let $x^* = (0,0,0)^T$. Then

$$
\begin{aligned}
\mathcal{P}(x^*, \theta^*) &= \{1,2,3,4,5\} \\
\nabla f^1(x^*, \theta^*) &= (-2,0,0) \\
\nabla f^2(x^*, \theta^*) &= (1,0,0) \\
\nabla f^3(x^*, \theta^*) &= (0,-1,-1) \\
\nabla f^4(x^*, \theta^*) &= (0,1,1) \\
\nabla f^5(x^*, \theta^*) &= (1,2,0).
\end{aligned}
$$

Iteration 1:

Step 1: The corresponding $(L, \mathcal{P}(x^*, \theta^*))$ is

$$
\begin{aligned}
&\min_{(d)} \quad 2d_2 \\
&\text{s.t.} \\
&\qquad -2d_1 \le 0 \\
&\qquad d_1 \le 0 \\
&\qquad -d_2 - d_3 \le 0 \\
&\qquad d_2 + d_3 \le 0 \\
&\qquad d_1 + 2d_2 \le 0 \\
&\qquad |d_i| \le 1, \quad i \in \{1,2,3\}
\end{aligned}
$$

and its optimal solution is $\tilde{d} = (0,-1,1)^T$ with $\tilde{g} = -2$.

Step 2: $K = \{5\}$ and $\Omega(\theta^*) = \mathcal{P}(x^*, \theta^*) \setminus K = \{1,2,3,4\}$.

Iteration 2:

Step 1: The linear program $(L, \Omega(x^*, \theta^*))$ is now

$$
\begin{aligned}
&\min_{(d)} \quad -d_1 \\
&\text{s.t.} \\
&\qquad -2d_1 \le 0 \\
&\qquad d_1 \le 0 \\
&\qquad -d_2 - d_3 \le 0 \\
&\qquad d_2 + d_3 \le 0 \\
&\qquad d_1 + 2d_2 \le 0 \\
&\qquad |d_i| \le 1, \quad i \in \{1,2,3\}
\end{aligned}
$$

and an optimal solution of it is $\tilde{d} = (0, -1, 1)^T$ with $\tilde{g} = 0$.
Step 2:

$$\mathcal{P}^{=}(\theta^*) = \Omega(\theta^*) = \{1, 2, 3, 4\}.$$

□

Another application of LFS functions will now be given in multi-objective programs. Let us consider the multi-objective program

$$(MP) \qquad \begin{array}{ll} \text{Min} & \{\phi^k(x) : k \in Q\} \\ \text{s.t.} & \\ & x \in F = \{x \in R^n : f^i(x) \le 0, \ i \in \mathcal{P}\} \end{array}$$

where $\mathcal{P} = \{1, \cdots, m\}$ and $Q = \{1, \cdots, q\}$. We recall that $x^* \in F$ is a Pareto minimum (or "efficient point") if there is no other $x \in F$ such that

$$\phi^k(x) \le \phi^k(x^*), \ \ k \in Q$$

with at least one strict inequality. If there is a constant $\beta > 0$ such that, whenever

$$\phi^k(x) < \phi^k(x^*), \ \ x \in F, \text{ for some } k \in Q,$$

we have

$$\frac{\phi^k(x^*) - \phi^k(x)}{\phi^{\bar{k}}(x) - \phi^{\bar{k}}(x^*)} \le \beta$$

for some $\bar{k} \in Q$ satisfying $\phi^{\bar{k}}(x) > \phi^{\bar{k}}(x^*)$, then such $x^* \in F$ is called a strong Pareto minimum (or "properly efficient point") (see, e.g., [11]). There exists a simple characterization of strong Pareto minima.

Theorem 1.11 *Consider the convex multi-objective program (MP). A point $x^* \in F$ is a strong Pareto minimum if, and only if, x^* is an optimal solution of the single-objective program*

$$\begin{array}{ll} Min & \sum_{k \in Q} \lambda_k^* \phi^k(x) \\ s.t. & \\ & x \in F \end{array}$$

for some weights $\lambda_k^* > 0, \ \ k \in Q$.

We will now prove that if all functions in (MP) have LFS property, then the strong Pareto and usual Pareto minima coincide.

Theorem 1.12 *Consider the convex multi-objective program (MP), where all the functions have LFS property at a feasible point x^* such that*

$$x^* \in \Big(\bigcap_{i \in \mathcal{P}(x^*)} int(dom f^i)\Big) \cap \Big(\bigcap_{k \in Q} int(dom \phi^k)\Big).$$

Then $x^\star$ is a Pareto optimum if, and only if, $x^\star$ solves the problem

$$(SP,\lambda^\star) \qquad \begin{array}{ll} Min & \sum_{k\in Q}(1+\lambda_k^\star)\phi^k(x) \\ s.t. & \\ & f^i(x)\le 0, \quad i\in\mathcal{P} \end{array}$$

for some weights $\lambda_k^\star \ge 0$, $k\in Q$.

Proof: Assume that $x^\star$ solves $(SP,\lambda^\star)$. Then $x^\star$ is a strong Pareto optimum and therefore a Pareto optimum. Conversely, if $x^\star$ is a Pareto optimum then $x^\star$ solves the problem

$$\begin{array}{ll} \text{Min} & \sum_{k\in Q}\phi^k(x) \\ \text{s.t.} & \\ & \phi^k(x)\le\phi^k(x^\star), \quad k\in Q \\ & f^i(x)\le 0, \quad i\in\mathcal{P}. \end{array}$$

Since all functions have LFS property at $x^\star$, we use Corollary 1.6 to conclude that the system

$$0\in\sum_{k\in Q}(1+\lambda_k^\star)\partial\phi^k(x^\star)+\sum_{i\in\mathcal{P}(x^\star)}u_i^\star\partial f^i(x^\star)$$

$$\lambda_k^\star\ge 0,\ k\in Q$$

$$u_i^\star\ge 0,\ i\in\mathcal{P}(x^\star)$$

is consistent. Hence it follows that $x^\star$ solves the program $(SP,\lambda^\star)$. This completes the proof.

Corollary 1.11 *Consider the convex multi-objective program (MP) where all the functions have LFS property at a feasible point $x^\star$ such that*

$$x^\star\in\Big(\bigcap_{i\in\mathcal{P}(x^\star)} int(dom f^i)\Big)\cap\Big(\bigcap_{k\in Q} int(dom\phi^k)\Big).$$

Then $x^\star\in F$ is a Pareto optimum if, and only if, $x^\star$ is a strong Pareto optimum.

We denote the Lagrangian of the multi-objective program (MP) by

$$\mathcal{L}_p(x,\lambda^\star,u)=\sum_{k\in Q}(1+\lambda_k^\star)\phi^k(x)+\sum_{i\in\mathcal{P}}u_if^i(x).$$

Then the following saddle-point characterization immediately follows.

Theorem 1.13 *Consider the convex multi-objective program* (MP) *where all the functions have LFS property at a feasible point* x^* *such that*

$$x^* \in \Big(\bigcap_{i \in \mathcal{P}(x^*)} int(dom f^i)\Big) \cap \Big(\bigcap_{k \in Q} int(dom \phi^k)\Big).$$

Then x^* *is a Pareto optimum if, and only if, there exists* $\lambda^* \geq 0$ *and* $u^* \geq 0$ *such that*

$$\mathcal{L}_p(x^*, \lambda^*, u) \leq \mathcal{L}_p(x^*, \lambda^*, u^*) \leq \mathcal{L}_p(x, \lambda^*, u^*)$$

for all $x \in R^n$ *and* $u \geq 0$.

Remark: Note that the condition on the interior of the domain of functions can be dropped if the functions are defined on the entire space R^n.

□

In what follows, we will give a characterization of Pareto optimality using a linearization approach. We consider the convex program (MP) where all the functions are differentiable and have LFS property at a point x in the intersection of their domains. Around x, we approximate the functions ϕ^k, $k \in Q$ and f^i, $i \in \mathcal{P}$ by their linear parts

$$\phi^k(x+d) \cong \phi^k(x) + \nabla\phi^k(x)d, \quad k \in Q$$
$$f^i(x+d) \cong f^i(x) + \nabla f^i(x)d, \quad i \in \mathcal{P}$$

thus obtaining the following approximation to the program (MP)

$$(LMP, x) \qquad \begin{array}{ll} \text{Min} & \{\phi^k(x) + \nabla\phi_k(x)d,\ k \in Q\} \\ \text{s.t.} & \\ & f^i(x) + \nabla f^i(x)d \leq 0,\ i \in \mathcal{P}. \end{array}$$

Program (LMP, x) is called *linearization of* (MP) *at* x. The following two single-objective programs, corresponding to (MP) and (LMP, x), respectively, will be used in a characterization of Pareto optima under the linearization:

$$(SP, \lambda) \qquad \begin{array}{ll} \text{Min} & \sum_{k \in Q}(1+\lambda_k)\phi^k(x) \\ \text{s.t.} & \\ & f^i(x) \leq 0, \quad i \in \mathcal{P} \end{array}$$

and

$$(LSP, \lambda, x) \qquad \begin{array}{ll} \text{Min} & \sum_{k \in Q}(1+\lambda_k)(\phi^k(x) + \nabla\phi_k(x)d,\ k \in Q) \\ \text{s.t.} & \\ & f^i(x) + \nabla f^i(x)d \leq 0,\ i \in \mathcal{P}. \end{array}$$

Theorem 1.14 *Consider the convex multi-objective program* (MP) *and a feasible point* $x^\star$ *at which all the functions have LFS property. Then* $x^\star$ *is a Pareto optimum for* (MP) *if, and only if,* $d = 0$ *is a Pareto optimum for* $(LMP, x^\star)$.

Proof: Assume that $x^\star \in F$ is a Pareto optimum for (MP). Then there exists $\lambda^\star \geq 0$ such that $x^\star$ is an optimal solution of $(SP, \lambda\ )$. This implies that $d = 0$ is an optimal solution of $(LSP, \lambda^\star, x^\star)$ (see [38, Theorem 1]). It then follows that $d = 0$ is a Pareto optimum for $(LMP, x^\star)$.

Conversely, if $d = 0$ is a Pareto optimum for $(LMP, x^\star)$, then there exists $\lambda^\star \geq 0$ such that $d = 0$ is an optimal solution of $(LSP, \lambda^\star, x^\star)$, which implies that $x^\star$ is an optimal solution of $(SP, \lambda^\star)$. Therefore, $x^\star$ is a Pareto optimum for (MP).

Chapter 2

LFS Functions and Generalized Convexity

In this chapter we extend the definition of differentiable convex LFS functions to the differentiable generalized convex functions. We will show that for such functions the Karush-Kuhn-Tucker condition is necessary for optimality. Again if some constraints do not belong to this class, we will regroup them to those that have LFS property at x^* and those that do not and give characterizations of optimality for a program rewritten in this form. We will further characterize Pareto optimality for differentiable pseudoconvex multi-objective programs (without considering LFS functions) and later consider the special case where all the functions are pseudoconvex with LFS property.

2.1 Quasi- and Pseudoconvex LFS Functions

We first recall some basic notions from generalized convexity (see, e.g., [1,17]) and then introduce quasiconvex and pseudoconvex LFS functions. We will give geometric and algebraic characterizations for special classes of differentiable quasiconvex functions with LFS property as well.

Definition 2.1 *A function $f : R^n \to R$, defined on a convex set in R^n, is called quasiconvex if all its level sets*

$$\{x : f(x) \leq \alpha\}, \quad \alpha \in R$$

are convex.

The following are equivalent (see [4]):

(a) f is quasiconvex;

(b) $f(x+d) \leq f(x),\ \ 0 \leq \lambda \leq 1 \Rightarrow f(x+\lambda d) \leq f(x)$ for every $x,\ x+d \in domf$;

(c) $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$ for any $0 \leq \lambda \leq 1$ and for every $x,\ y \in domf$.

Definition 2.2 *A function $f : R^n \to R$, defined on a convex set in R^n, is called strictly quasiconvex if, for all $x,\ x+d \in domf$,*

$$f(x+d) < f(x),\ \ 0 < \lambda \leq 1 \Rightarrow f(x+\lambda d) < f(x).$$

The following are equivalent (see [1]):

(a) f is strictly quasiconvex;

(b) $f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$ for any $0 < \lambda < 1$ and for all $x,\ y \in domf$ such that $f(x) \neq f(y)$.

Lemma 2.1 *Let f be differentiable on an open convex set $\Gamma \subset R^n$. Then f is quasiconvex if, and only if, for any $x,\ y \in \Gamma$ such that*

$$f(y) \leq f(x),$$

we have

$$\nabla f(x)(y-x) \leq 0.$$

(The proof of the above lemma can be found in [1].)

Definition 2.3 *A function $f : R^n \to R$, defined and differentiable on a convex set Γ in R^n, is called pseudoconvex on Γ if, for all $x,\ y \in \Gamma$,*

$$\nabla f(x)(y-x) \geq 0 \Rightarrow f(y) \geq f(x). \tag{2.1}$$

We can place a further restriction on a pseudoconvex function by requiring the implied inequality in (2.1) to be a strict inequality for $x \neq y$. In this case, the function is called *strictly pseudoconvex*.

Lemma 2.2 *If f is differentiable and pseudoconvex and $x^\star \in domf$, then*

(a) $D_f^{<}(x^\star) = \{d : \nabla f(x^\star)d < 0\}$;

(b) $D_f^{=}(x^\star)$ is a convex cone and $D_f^{=}(x^\star) \subset \{d : \nabla f(x^\star)d = 0\}$;

(c) $D_f^{\leq}(x^\star) = \{d : \nabla f(x^\star)d \leq 0$ with equality only if $d \in D_f^{=}(x^\star)\}$.

Recall that Lemma 2.2(a) does not hold for strictly quasiconvex functions. Furthermore, Lemma 2.2(b) may fail for differentiable quasiconvex functions. For counterexamples see [5].

Without lower semicontinuity, strict quasiconvexity does not imply quasiconvexity (as the well-known example

$$f(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$$

shows). However, since we assume that all the functions are differentiable, the following implications can be concluded.

convexity $\Rightarrow$ pseudoconvexity $\Rightarrow$ strict quasiconvexity $\Rightarrow$ quasiconvexity.

Definition 2.4 *A differentiable quasiconvex function $f : R^n \rightarrow R$ has a locally flat surface at $x^\star$ if*

$$N(\nabla f(x^\star)) = D_f^{=}(x^\star).$$

For pseudoconvex functions with LFS property at any point $x^\star \in R^n$ where $\nabla f(x^\star) \neq 0$, we have the following geometric characterization.

Theorem 2.1 *Let f be a pseudoconvex function and $x^\star \in R^n$. If $\nabla f(x^\star) \neq 0$, then f has the LFS property at $x^\star$ if, and only if, its cone of directions of nonascent is polyhedral at $x^\star$.*

Proof: (Sufficiency:) Since $\nabla f(x^\star) \neq 0$, we observe that $D_f^{<}(x^\star) \neq \emptyset$. Here pseudoconvexity of f and polyhedrality of $D_f^{\leq}(x^\star)$, by Lemma 2.2(a,c), imply that

$$\begin{aligned} D_f^{\leq}(x^\star) &= \{d : \nabla f(x^\star) d \leq 0\} \\ &= D_f^{<}(x^\star) \cup N(\nabla f(x^\star)). \end{aligned}$$

Besides (see, e.g., [5, Lemma 3.1(b)])

$$D_f^{\leq}(x^\star) = D_f^{<}(x^\star) \cup D_f^{=}(x^\star).$$

In either case $D_f^{\leq}(x^\star)$ is represented as a unoin of two disjoint sets. Hence

$$D_f^{=}(x^\star) = N(\nabla f(x^\star)).$$

(Necessity:) If f has the LFS property at $x^\star$, then, by Lemma 2.2,

$$D_f^{\leq}(x^\star) = \{d : \nabla f(x^\star) d \leq 0\},$$

which is a polyhedral cone.

□

Ferland (see, e.g., [1]) showed that, if f is a twice differentiable quasiconvex function on a solid (with a nonempty interior) convex set $\Gamma \subset R^n$, then f is pseudoconvex at any point $x^\star \in \Gamma$ where $\nabla f(x^\star) \neq 0$. Therefore, we can derive the following geometric characterization for such quasiconvex functions using Theorem 2.1.

Corollary 2.1 *Let f be a twice differentiable quasiconvex function and $\Gamma \subset R^n$ be a convex set with a nonempty interior. Furthermore, assume that $x^\star \in \Gamma$ and $\nabla f(x^\star) \neq 0$. Then f has LFS property at $x^\star$ if, and only if, its cone of directions of nonascent is polyhedral at $x^\star$.*

In the following two lemmas, we will introduce classes of quasiconvex and pseudoconvex functions for which the cones of directions of constancy can be easily calculated.

Lemma 2.3 *Let the function $f : R^n \to R$ be given as*

$$f(x) = h(Ax + b) \tag{2.2}$$

where A is an $m \times n$ matrix and $h : R^m \to R$ is a strictly pseudoconvex function. Then f is pseudoconvex and

$$D_f^=(x^\star) = N(A)$$

independently of $x^\star \in R^n$.

Proof: First we prove pseudoconvexity of f. For all $x,\ x^\star \geq 0$,

$$\begin{aligned} \nabla f(x^\star)(x - x^\star) \geq 0 \ &\Rightarrow\ (\nabla h(Ax^\star + b))A(x - x^\star) \geq 0 \\ &\Rightarrow\ (\nabla h(Ax^\star + b))(Ax - Ax^\star) \geq 0 \\ &\Rightarrow\ h(Ax + b) \geq h(Ax^\star + b) \text{ by strict pseudoconvexity of } h \\ &\Rightarrow\ f(x) \geq f(x^\star). \end{aligned}$$

Now if $d \in N(A)$, that is $Ad = 0$, then $d \in D_f^=(x^\star)$, by (2.2), regardless of the special assumptions on h. Conversely, let $d \in D_f^=(x^\star)$. Then there exists an $\bar{\alpha} > 0$ such that, for all $0 < \alpha \leq \bar{\alpha}$,

$$\begin{aligned} f(x^\star + \alpha d) - f(x^\star) &= h((Ax^\star + b) + \alpha Ad) - h(Ax^\star + b) \\ &= 0, \end{aligned}$$

which shows that h is constant on the interval

$$[Ax^\star + b, Ax^\star + b + \bar{\alpha} Ad].$$

This is a contradiction to strict pseudoconvexity of h unless $Ad = 0$. Therefore $d \in N(A)$.

□

Lemma 2.3 can be extended to a much larger class of differentiable quasiconvex functions under some additional assumptions.

Lemma 2.4 *Let the function $f : R^n \to R$ be given as*

$$f(x) = h(Ax + b) \tag{2.3}$$

where A is an $m \times n$ matrix, $b \in R^m$ and $h : R^m \to R$ is a differentiable quasiconvex function. Furthermore, assume that $x^\star \in R^n$ and that $\nabla f(x^\star) \neq 0$. Then f is quasiconvex and

$$D_f^=(x^\star) = N(A).$$

Proof: The quasiconvexity of f trivially follows from the quasiconvexity of h, since for all x, $y \in dom f$ and $0 \leq \lambda \leq 1$,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= h(A(\lambda x + (1-\lambda)y) + b) \\ &\leq h(\lambda(Ax+b) + (1-\lambda)(Ay+b)) \\ &\leq \max\{h(Ax+b), h(Ay+b)\} \\ &\leq \max\{f(x), f(y)\}. \end{aligned}$$

Now if $d \in N(A)$, then trivially $d \in D_f^=(x^\star)$ by (2.3). Conversely, let $d \in D_f^=(x^\star)$. Then there exists an $\bar{\alpha} > 0$ such that, for all $0 < \alpha \leq \bar{\alpha}$,

$$\begin{aligned} f(x^\star + \alpha d) - f(x^\star) &= h((Ax^\star + b) + \alpha Ad) - h(Ax^\star + b) \\ &= 0, \end{aligned}$$

which again shows that h is constant on the interval

$$[Ax^\star + b, Ax^\star + b + \bar{\alpha}Ad].$$

This implies that

$$\nabla h(Ax^\star + b) = 0,$$

which further implies that

$$\nabla f(x^\star) = (\nabla h(Ax^\star + b))A = 0.$$

We have reached a contradiction (since, by assumption, $\nabla f(x^\star) \neq 0$) unless $Ad = 0$. Therefore, $d \in N(A)$.

□

Lemma 2.3 and Lemma 2.4 help us find an algebraic characterization for identification of quasiconvex or pseudoconvex LFS functions as the following theorem shows.

Theorem 2.2 *Let f be a differentiable quasiconvex function of the form (2.3) and $x^* \in R^n$ such that $\nabla f(x^*) \neq 0$. Then f has the LFS property at x^*, if and only if*

$$rank[A] = 1.$$

The proof of the rank condition is exactly the same as the proof of the rank condition for the faithfully convex case, given in [30].

Example 2.1 The following functions are strictly pseudoconvex with the LFS property. The claim can be verified by Theorem 2.1. The function

$$f(x) = \frac{1}{a^T x + \beta}$$

is LFS at any $x^* \in R^n$ such that $a^T x^* + \beta \neq 0$ (here $a \in R^n$ and $\beta \in R$). Also

$$g(x) = (a^T x)^3 + a^T x$$

is LFS at any $x^* \in R^n$.

On the other hand, the following functions are quasiconvex with the LFS property. The function

$$f(x) = (a^T x + \beta)^{2m+1},$$

where m is a positive integer, is LFS at any $x^* \neq 0$. Also the function

$$g(x) = (a^T x + \beta)^5 + (a^T x + \beta)^3$$

is LFS at any $x^* \neq 0$.

□

We will now turn our attention to characterizations of optimality with LFS functions in generalized convexity.

2.2 Optimality Conditions

In this section, we will give a condition under which the Karush-Kuhn-Tucker condition is necessary for optimality. We will, in addition, show that the saddle-point-necessary optimality condition does not hold when all the functions are pseudoconvex even with LFS property. First we need an important property of quasiconvex LFS functions.

Lemma 2.5 *Let f be a differentiable quasiconvex function. If f has LFS property at $x^* \in R^n$, then*

$$D_f^{\leq}(x^*) = \{d : \nabla f(x^*) d \leq 0\}.$$

Proof: The LFS property at $x^\star$ implies that

$$D_f^{=}(x^\star) = \{d : \nabla f(x^\star)d = 0\}.$$

Clearly

$$\{d : \nabla f(x^\star)d < 0\} \subset D_f^{<}(x^\star).$$

Therefore

$$\begin{aligned} \{d : \nabla f(x^\star)d \leq 0\} &\subset D_f^{<}(x^\star) \cup N(\nabla f(x^\star)) \\ &= D_f^{<}(x^\star) \cup D_f^{=}(x^\star) \\ &= D_f^{\leq}(x^\star) \text{ (see, e.g., [5, Lemma 3.1(b)]).} \end{aligned} \tag{2.4}$$

On the other hand

$$d \in D_f^{\leq}(x^\star) \Rightarrow f(x^\star + \alpha d) \leq f(x^\star), \quad 0 \leq \alpha \leq \bar{\alpha}, \quad \text{for some } \bar{\alpha} > 0,$$

which, by Lemma 2.1, implies that

$$\nabla f(x^\star)d \leq 0.$$

Therefore

$$D_f^{\leq}(x^\star) \subset \{d : \nabla f(x^\star)d \leq 0\}. \tag{2.5}$$

It then follows from (2.4) and (2.5) that

$$D_f^{\leq}(x^\star) = \{d : \nabla f(x^\star)d \leq 0\}.$$

□

We will state optimality conditions for the programs of the form

$$(QP) \qquad \begin{aligned} \text{Min} \quad & f^0(x) \\ \text{s.t.} \quad & \\ & f^i(x) \leq 0, \quad i \in \mathcal{P} \end{aligned}$$

where all the functions are differentiable and quasiconvex. The definitions of the sets F, $F(x^\star)$, $\mathcal{P}$, $\mathcal{P}^{=}$ and $\mathcal{P}(x^\star)$ (given in Chapter 1), where $x^\star$ is a feasible point of (QP), remain unchanged.

Theorem 2.3 *Consider the program (QP). Assume that at an optimal solution $x^\star$, the constraints have the LFS property. Then the KKT system is consistent.*

Proof: Lemma 2.5 and the LFS property of the constraints at $x^\star$ imply that

$$F(x^\star) = \bigcap_{i \in \mathcal{P}(x^\star)} D_{f^i}^{\leq}(x^\star) = \{d : \nabla f^i(x^\star)d \leq 0, \ i \in \mathcal{P}(x^\star)\},$$

which is a constraint qualification (see [29]).

□

The following theorem is a well-known result (see, e.g., [29]). It will be introduced here for the sake of completeness.

Theorem 2.4 *Consider the differentiable and quasiconvex program (QP). Assume that f^0 is, in addition, pseudoconvex. If $x^\star$ satisfies the KKT system, then $x^\star$ is an optimal solution of (QP).*

We can now characterize optimality.

Corollary 2.2 *Consider the differentiable program (QP) where f^0 is pseudoconvex and the constraints are quasiconvex and have the LFS property at $x^\star \in R^n$. Then $x^\star$ is optimal if, and only if, the KKT system is satisfied.*

It is well known that saddle-point-necessary optimality condition does not hold for pseudoconvex programs. We will show in the following example that saddle-point-necessary optimality condition does not hold even when all the functions are pseudoconvex and enjoy the LFS property at an optimal solution $x^\star$.

Example 2.2 Consider

$$\begin{aligned} &\text{Min} \quad -x \\ &\text{s.t.} \\ &\qquad x + x^3 \leq 0. \end{aligned}$$

Here $x^\star = 0$ is the unique optimal solution. The right-hand inequality of the saddle-point condition becomes

$$0 \leq -x + \lambda^\star(x + x^3)$$

for some $\lambda^\star \geq 0$ and for all $x \in R$, clearly an inconsistent system.

2.3 Programs with Non-LFS Functions

Let us first recall some optimality conditions regarding generalized convex functions. For more details and proofs see, e.g., [5].

Lemma 2.6 *Let $x^\star$ be a feasible point of the program (QP), where f^0 is strictly quasiconvex and the constraints are quasiconvex. Then $x^\star$ is optimal if, and only if*

$$D_0^{<}(x^\star) \cap F(x^\star) = \emptyset.$$

As discussed in [5] optimality can be characterized in terms of the single subset $\mathcal{P}^=$ of $\mathcal{P}(x^\star)$, but it requires the strict quasiconvexity of all functions.

Theorem 2.5 *Let $x^\star$ be a feasible solution of (QP) where all the functions are strictly quasiconvex. Then $x^\star$ is optimal if, and only if,*

$$D_0^{<}(x^\star) \cap D_{\mathcal{P}^{<}(x^\star)}^{<}(x^\star) \cap D_{\mathcal{P}^{=}}^{=}(x^\star) = O,$$

or, dually, if, and only if, there exist vectors,

$$0 \neq y^0 \in \{D_0^{<}(x^\star)\}^+, \quad y^i \in \{D_{f^i}^{<}(x^\star)\}^+, \quad i \in \mathcal{P}^{<}(x^\star), \quad y \in \{D_{\mathcal{P}^{=}}^{=}(x^\star)\}^+$$

such that

$$y^0 + \sum_{i \in \mathcal{P}^{<}(x^\star)} y^i + y = 0.$$

We will derive an optimality condition similar to the condition in Theorem 2.5, using the index set $\mathcal{R}^{=}$, to be defined shortly, instead of $\mathcal{P}^{=}$. We will then have to assume all the functions are quasiconvex since Theorem 2.5 will be a special case of our result when $\mathcal{Q}(x^\star) = \emptyset$. We will further simplify this optimality condition when the functions are pseudoconvex using Lemma 2.2.

Now consider the program

$$(SQP) \qquad \begin{array}{ll} \text{Min} & f^0(x) \\ \text{s.t.} & \\ & f^i(x) \leq 0, \quad i \in \mathcal{P} \end{array}$$

where all the functions are strictly quasiconvex and differentiable. Let us split all the constraints of (SQP) into those that have the LFS property at a given $x^\star \in F$ and those that do not. Let the indices of the former group belong to $\mathcal{Q}$ and the indices of the latter to $\mathcal{R}$, where $\mathcal{P} = \mathcal{Q} \cup \mathcal{R}$. We can now rewrite (SQP) as

$$(SQP') \qquad \begin{array}{ll} \text{Min} & f^0(x) \\ \text{s.t.} & \\ & f^i(x) \leq 0, \quad i \in \mathcal{Q} \\ & f^j(x) \leq 0, \quad j \in \mathcal{R}. \end{array}$$

As in the convex case, we use the following index sets:

$$\mathcal{R}^{=} = \{i \in \mathcal{R} : f^i(x) = 0 \ \ \forall\, x \in F\},$$

$$\mathcal{Q}(x^\star) = \{i \in \mathcal{Q} : f^i(x^\star) = 0\}$$

and

$$\mathcal{R}(x^\star) = \{j \in \mathcal{R} : f^j(x^\star) = 0\}.$$

Lemma 2.7 *Consider the program (SQP'). If $\mathcal{R}(x^\star) \setminus \mathcal{R}^{=} \neq \emptyset$, then there exists a point $\hat{x} \in F$ such that*

$$f^i(\hat{x}) < 0, \quad i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^{=}.$$

Proof: Clearly

$$\mathcal{R}(x^\star) \setminus \mathcal{R}^= = \{k \in \mathcal{R}(x^\star) : \exists x^k \in F \ni f^k(x^k) < 0\}.$$

Let

$$\hat{x} = \frac{1}{card\{\mathcal{R}(x^\star) \setminus \mathcal{R}^=\}} \sum_{k \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=} x^k.$$

Then $\hat{x}$ is feasible since F is a convex set. Furthermore, the following claim follows from strict quasiconvexity of the constraints. Given any $i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=$, if

$$f^i(x^l) = f^i(x^i), \quad l \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=,$$

then trivially

$$f^i(\hat{x}) \leq f^i(x^i) < 0.$$

Otherwise

$$f^i(\hat{x}) < \max_{k \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=} \{f^i(x^k)\} \leq 0.$$

Therefore

$$f^i(\hat{x}) < 0, \quad i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=.$$

Theorem 2.6 *A feasible point $x^\star$ is an optimal solution of (SQP') if, and only if,*

$$D_0^<(x^\star) \cap D^{\leq}_{\mathcal{R}(x^\star)\setminus\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} = \emptyset; \tag{2.6}$$

or, dually if, and only if, there exist vectors,

$$0 \neq y^0 \in \{D_0^<(x^\star)\}^+, \;\; y^i \in \{D^<_{f^i}(x^\star)\}^+, \;\; i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=$$

and

$$y \in \{conv\{D^{=}_{\mathcal{R}^=}(x^\star)\} \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+ \tag{2.7}$$

such that

$$y^0 + \sum_{i \in \mathcal{R}(x^\star)\setminus\mathcal{R}^=} y^i + y = 0.$$

Proof: (Primal version:) (Sufficiency:) Suppose that (2.6) holds and that $x^\star$ is not optimal. Then, by Lemma 2.6,

$$D_0^<(x^\star) \cap F(x^\star) \neq \emptyset.$$

This means that there exists $d \in R^n$ such that

$$d \in D_0^<(x^\star) \cap D^{\leq}_{\mathcal{R}(x^\star)\setminus\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap D^{\leq}_{\mathcal{R}^=}(x^\star) \neq \emptyset.$$

By Lemma 2.7, there exists $\hat{x} \in F$ such that $f^i(\hat{x}) < 0$, $i \in \mathcal{R}(x^\star)\backslash\mathcal{R}^=$. Let $\hat{d} = \hat{x} - x^\star$. Then from strict quasiconvexity it follows that

$$\hat{d} \in D^<_{\mathcal{R}(x^\star)\backslash\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap D^{\leq}_{\mathcal{R}^=}(x^\star).$$

Denote $d^\lambda = \lambda\hat{d} + (1-\lambda)d$ for $0 < \lambda \leq 1$. Then, by the choice of d and $\hat{d}$, we have

$$d^\lambda \in D^<_{\mathcal{R}(x^\star)\backslash\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap D^{\leq}_{\mathcal{R}^=}(x^\star).$$

But $d^\lambda \in F(x^\star)$ and $\mathcal{R}^= \subset \mathcal{P}^=$. Therefore (see [5, Lemma 3.9(a)])

$$d^\lambda \in D^=_{\mathcal{R}^=}(x^\star).$$

It follows that

$$d^\lambda \in D^<_{\mathcal{R}(x^\star)\backslash\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap D^=_{\mathcal{R}^=}(x^\star),$$

and hence

$$d^\lambda \in D^<_{\mathcal{R}(x^\star)\backslash\mathcal{R}^=}(x^\star) \cap D^{\leqq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^=_{\mathcal{R}^=}(x^\star)\}.$$

Besides

$$\begin{aligned} f^0(x^\star + td^\lambda) &= f^0(\lambda(x^\star + t\hat{d}) + (1-\lambda)(x^\star + td)) \\ &< f^0(x^\star), \qquad \text{for } t \text{ sufficiently small,} \end{aligned}$$

which follows by continuity from $f^0(x^\star + td) < f^0(x^\star)$. Therefore

$$d^\lambda \in D^<_0(x^\star) \cap D^<_{\mathcal{R}(x^\star)\backslash\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^=_{\mathcal{R}^=}(x^\star)\},$$

which is a contradiction to (2.6).
(Necessity:) Assume that $x^\star$ is optimal and

$$D^<_0(x^\star) \cap D^<_{\mathcal{R}(x^\star)\backslash\mathcal{R}^=}(x^\star) \cap D^{\leqq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^=_{\mathcal{R}^=}(x^\star)\} \neq \emptyset.$$

Then there exist a direction

$$d \in D^<_0(x^\star) \cap D^<_{\mathcal{R}(x^\star)\backslash\mathcal{R}^=}(x^\star) \cap D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star) \cap conv\{D^=_{\mathcal{R}^=}(x^\star)\}$$

and an $\bar{\alpha} > 0$ such that, for all $0 < \alpha \leq \bar{\alpha}$,

$$\begin{aligned} f^0(x^\star + \alpha d) &< f^0(x^\star) \\ f^j(x^\star + \alpha d) &< 0, \quad j \in \mathcal{R}(x^\star) \setminus \mathcal{R}^= \\ f^j(x^\star + \alpha d) &\leq 0, \quad j \in \mathcal{Q}(x^\star) \cup \mathcal{R}^=. \end{aligned}$$

Furthermore, $f^j(x^\star) < 0$, $j \in \mathcal{P} \setminus \mathcal{P}(x^\star)$, and continuity imply that

$$f^j(x^\star + \alpha d) \leq 0, \quad j \in \mathcal{P} \setminus \mathcal{P}(x^\star),$$

for $\alpha > 0$ sufficiently small. This means that $x^\star$ is not optimal, which is again a contradiction.
(Dual version:) The dual characterization follows from the Dubovitskii-Milyutin theorem (see Theorem 1.7).

Theorem 2.7 *Consider the program (SQP') and $x^\star \in F$. Assume that all the functions*

$$f^i, \quad i \in \{0\} \cup \mathcal{R}(x^\star),$$

are pseudoconvex and $conv\{D^{\overline{\overline{\ }}}_{\mathcal{R}^=}(x^\star)\}$ is closed. Then $x^\star$ is optimal if, and only if, the system

$$(\nabla f^0(x^\star))^T + \sum_{i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=} \lambda_i (\nabla f^i(x^\star))^T \in cl\left[\{D^{\leq}_{\mathcal{Q}(x^\star)}(x^\star)\}^+ + \{D^{\overline{\overline{\ }}}_{\mathcal{R}^=}(x^\star)\}^+\right]$$

is consistent for some $\lambda_i \geq 0, \quad i \in \mathcal{R}(x^\star) \setminus \mathcal{R}^=$.

Proof: The theorem is an immediate result of Theorem 2.6, Lemma 2.2 and the basic properties of the polar set.

Corollary 2.3 *Consider the program (SQP') and $x^\star \in F$. Assume that all the functions are pseudoconvex and that the cone $D^{\overline{\overline{\ }}}_{\mathcal{R}^=}(x^\star)$ is polyhedral. Then $x^\star$ is optimal if, and only if, the system*

$$(\nabla f^0(x^\star))^T + \sum_{i \in \mathcal{P}(x^\star) \setminus \mathcal{R}^=} \lambda_i (\nabla f^i(x^\star))^T \in \{D^{\overline{\overline{\ }}}_{\mathcal{R}^=}(x^\star)\}^+$$

is consistent for some $\lambda_i \geq 0, \quad i \in \mathcal{P}(x^\star) \setminus \mathcal{R}^=$.

Note that if all the functions are convex, then the above two results recover Corollary 1.8 and Corollary 1.9.

Remark: Recently, the characterization of local minima of differentiable nonlinear programs in terms of some classical second-order conditions were investigated by Pang [20]. He obtained various necessary and sufficient conditions for a Karush-Kuhn-Tucker point to be an isolated (and/or a strict) local minimum of a differentiable nonlinear program. There were no convexity or psedoconvexity assumptions. In his paper, he also referred to the differentiable convex LFS functions. The interesting point is that, under convexity or pseudoconvexity assumptions, some of his conditions on the objective function and the inequality constraints reduce to having LFS property.

2.4 Generalized Convexity and Pareto Optimality

Recently, several characterizations of Pareto optimality for convex multi-objective programs have appeared (see, e.g., [3,30,32,35]). In this section, we will show that

Pareto optimality can be characterized for pseudoconvex multi-objective programs as well. We will further consider a special case where all the functions are quasiconvex and have the LFS property at a feasible point $x^\star$, and give a necessary condition for optimality in that case.

Consider the multi-objective program

$$(MP) \qquad \begin{array}{ll} \text{Min} & \{\phi^k(x) : k \in Q\} \\ \text{s.t.} & \\ & f^i(x) \leq 0, \quad i \in \mathcal{P}. \end{array}$$

Recall that $F = \{x \in R^n : f^i(x) \leq 0, \quad i \in \mathcal{P}\}$. Following the definitions in [30,32], for any $x^\star \in F$, let

$$\begin{aligned} F_0(x^\star) &= \{x \in R^n : \phi^k(x) \leq \phi^k(x^\star), \quad k \in Q\} \\ \mathcal{P}^=(x^\star) &= \{i \in \mathcal{P} : x \in F \cap F_0(x^\star) \Rightarrow f^i(x) = 0\} \end{aligned}$$

and, for every $r \in Q$,

$$\begin{aligned} F_r(x^\star) &= \{x \in F : \phi^k(x) \leq \phi^k(x^\star), \ k \in Q \setminus \{r\}\} \\ Q_r^=(x^\star) &= \{k \in Q \setminus \{r\} : x \in F_r(x^\star) \Rightarrow \phi^k(x) = \phi^k(x^\star)\} \\ Q^=(x^\star) &= \bigcup_{r \in Q} Q_r^=(x^\star). \end{aligned}$$

Our first result is an extension of two lemmas from [30] to generalized convexity.

Lemma 2.8 *Consider the pseudoconvex multi-objective program (MP). A feasible point $x^\star$ is not Pareto optimal if, and only if, both $Q \setminus Q^=(x^\star) \neq \emptyset$ and there exists $\hat{x}$ such that*

$$\begin{aligned} \phi^j(\hat{x}) &< \phi^j(x^\star), \quad j \in Q \setminus Q^=(x^\star) \\ \phi^i(\hat{x}) &= \phi^i(x^\star), \quad i \in Q^=(x^\star) \\ \hat{x} &\in F. \end{aligned} \tag{2.8}$$

Proof: (Necessity:) If $x^\star$ is not a Pareto optimum, then there exists $\bar{x} \in F$ and $k_0 \in Q$ such that

$$\begin{aligned} \phi^{k_0}(\bar{x}) &< \phi^{k_0}(x^\star) \\ \phi^k(\bar{x}) &\leq \phi^k(x^\star), \quad k \in Q \setminus \{k_0\}. \end{aligned} \tag{2.9}$$

It follows from the definition of $Q^=(x^\star)$ that

$$\begin{aligned} Q \setminus Q^=(x^\star) &= \bigcap_{r \in Q} Q \setminus Q_r^=(x^\star) \\ &= \{i \in Q : \forall\, r \in Q \setminus \{i\}, \exists\, \hat{x} \in F_r(x^\star) \text{ such that } \phi^i(\hat{x}) < \phi^i(x^\star)\}. \end{aligned} \tag{2.10}$$

Thus $k_0 \in Q \setminus Q^{=}(x^\star)$, which is therefore nonempty.

We will now show that for every $j \in Q \setminus Q^{=}(x^\star)$, there exists $y^j \in F$ such that

$$\begin{aligned} \phi^j(y^j) &< \phi^j(x^\star) \\ \phi^l(y^j) &\leq \phi^l(x^\star), \quad l \in Q \setminus \{j\}. \end{aligned} \tag{2.11}$$

Obviously, for $j = k_0$, (2.11) is true by (2.9). Let $y^{k_0} = \bar{x}$. We will costruct y^j for $j \neq k_0$, in the following way. Note that $j \in Q \setminus Q^{=}_{k_0}(x^\star)$. This, by the definition of $Q^{=}_{k_0}(x^\star)$, means that there exists $z^j \in F$ such that

$$\begin{aligned} \phi^j(z^j) &< \phi^j(x^\star) \\ \phi^l(z^j) &\leq \phi^l(x^\star), \quad l \in Q \setminus \{k_0\}. \end{aligned} \tag{2.12}$$

Consider

$$y^j = \lambda z^j + (1-\lambda)\bar{x}, \quad 0 < \lambda < 1.$$

It follows, from the convexity of F, that for $j \in Q \setminus Q^{=}_{k_0}$, $y^j \in F$. Furthermore (2.9) and (2.12) together with pseudoconvexity (strict quasiconvexity) imply that

$$\begin{aligned} \phi^j(y^j) &\leq \max\{\phi^j(z^j), \phi^j(\bar{x})\} < \phi^j(x^\star) \\ \phi^l(y^j) &\leq \max\{\phi^l(z^j), \phi^l(\bar{x})\} \leq \phi^l(x^\star), \quad l \in Q \setminus \{k_0, j\}. \end{aligned}$$

By choosing y^j sufficiently close to $\bar{x}$, i.e., by choosing λ sufficiently close to zero, we can conclude, by continuity, that

$$\phi^{k_0}(y^j) < \phi^{k_0}(x^\star).$$

Therefore (2.11) is proved. Now let

$$\hat{x} = \sum_{j \in Q \setminus Q^{=}(x^\star)} \lambda_j y^j,$$

where

$$\sum_{j \in Q \setminus Q^{=}(x^\star)} \lambda_j = 1, \quad \lambda_j > 0, \;\; j \in Q \setminus Q^{=}(x^\star).$$

Again, by convexity of f, $\hat{x} \in F$. Furthermore, for every $i \in Q^{=}(x^\star)$, it follows from pseudoconvexity (strict quasiconvexity) and (2.11) that

$$\phi^i(\hat{x}) \leq \max_{j \in Q \setminus Q^{=}(x^\star)} \{\phi^i(y^j)\} \leq \phi^i(x^\star). \tag{2.13}$$

In fact, equality holds in (2.13), since otherwise the definition of $Q^{=}(x^\star)$ is contradicted. On the other hand, for every $j \in Q \setminus Q^{=}(x^\star)$ we have the following situation. If $\phi^j(y^l) = \phi^j(y^j)$, $l \in Q \setminus Q^{=}(x^\star)$, then

$$\phi^j(\hat{x}) \leq \phi^j(y^j) < \phi^j(x^\star).$$

Otherwise

$$\phi^j(\hat{x}) < \max_{l \in Q \setminus Q^=(x^\star)} \{\phi^j(y^l)\} \leq \phi^j(x^\star).$$

Therefore

$$\phi^j(\hat{x}) < \phi^j(x^\star), \quad j \in Q \setminus Q^=(x^\star),$$

proving the necessity part.

(Sufficiency:) This part immediately follows from the definition of a Pareto optimum. □

Using this lemma, we can conclude the following more explicit result.

Lemma 2.9 *Consider the pseudoconvex multi-objective program (MP). Then a feasible point $x^\star$ is not Pareto optimal if, and only if, both $Q \setminus Q^=(x^\star) \neq \emptyset$ and there exists a $\hat{y} \in F$ such that*

$$\begin{aligned} \phi^k(\hat{y}) &< \phi^k(x^\star), \quad k \in Q \setminus Q^=(x^\star) \\ \phi^l(\hat{y}) &= \phi^l(x^\star), \quad l \in Q^=(x^\star) \\ f^j(\hat{y}) &< 0, \quad j \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star) \\ f^i(\hat{y}) &= 0, \quad i \in \mathcal{P}^=(x^\star). \end{aligned} \tag{2.14}$$

Proof:(Sufficiency:) This part immediately follows from the definition of a Pareto minimum.

(Necessity:) Suppose that $x^\star$ is not a Pareto minimum. Then

$$F_0(x^\star) \cap F \neq \emptyset.$$

Let $x \in F_0(x^\star) \cap F$. Now, by the definition of $\mathcal{P}^=(x^\star)$, for any feasible point x,

$$f^i(x) = 0, \quad i \in \mathcal{P}^=(x^\star).$$

Moreover, by the definition of $\mathcal{P}^=(x^\star)$, for every $j \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star)$, there exists x^j such that

$$f^j(x^j) < 0 \text{ and } x^j \in F_0(x^\star) \cap F.$$

Choose

$$\bar{y} = \sum_{j \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star)} \lambda_j x^j$$

where $\lambda_j > 0, \ \ j \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star)$ and $\sum_{j \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star)} \lambda_j = 1$. Then, by convexity of $F \cap F_0(x^\star)$,

$$\bar{y} \in F \cap F_0(x^\star).$$

Besides, by pseudoconvexity (which implies strict quasiconvexity) we have the following situation. Given any $i \in \mathcal{P}(x^\star \setminus \mathcal{P}^=(x^\star)$, if

$$f^i(x^l) = f^i(x^i), \quad l \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star),$$

then simply

$$f^i(\hat{y}) \le f^i(x^i) < 0.$$

Otherwise

$$f^i(\bar{y}) < \max_{j \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star)} \{f^i(x^j)\} \le 0.$$

Therefore

$$f^i(\bar{y}) < 0, \quad i \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star). \tag{2.15}$$

Since $x^\star$ is not a Pareto minimum, then, by Lemma 2.8, there exists $\hat{x} \in F$ such that

$$\begin{aligned} \phi^k(\hat{x}) &< \phi^k(x^\star), \quad k \in Q \setminus Q^=(x^\star) \\ \phi^l(\hat{y}) &= \phi^l(x^\star), \quad l \in Q^=(x^\star). \end{aligned}$$

Now, let $\hat{y} = \frac{\hat{x}+\bar{y}}{2}$. Then $\hat{y} \in F$ and

$$\begin{aligned} \phi^k(\hat{y}) &\le \phi^k(\hat{x}) < \phi^k(x^\star), \quad \text{if } k \in Q \setminus Q^=(x^\star), \ \phi^k(\hat{x}) = \phi^k(\bar{y}) \\ \phi^k(\hat{y}) &< \max\{\phi^k(\hat{x}), \phi^k(\bar{y})\} \le \phi^k(x^\star), \quad \text{if } k \in Q \setminus Q^=(x^\star), \ \phi^k(\hat{x}) \ne \phi^k(\bar{y}) \\ \phi^l(\hat{y}) &\le \max\{\phi^l(\hat{x}), \phi^l(\bar{y})\} \le \phi^l(x^\star), \quad \text{if } l \in Q^=(x^\star). \end{aligned}$$

Similarly

$$\begin{aligned} f^j(\hat{y}) &< 0, \quad j \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star) \\ f^i(\hat{y}) &\le 0, \quad i \in \mathcal{P}^=(x^\star). \end{aligned}$$

Hence, by the definitions of the index sets $Q^=(x^\star)$ and $\mathcal{P}^=(x^\star)$, we conclude that

$$\begin{aligned} \phi^l(\hat{y}) &= \phi^l(x^\star), \quad l \in Q^=(x^\star) \\ f^i(\hat{y}) &= 0, \quad i \in \mathcal{P}^=(x^\star). \end{aligned}$$

This completes the proof.

Remark: In fact, as the proofs show, Lemma 2.8 also holds when all the objective functions are strictly quasiconvex and the constraints are quasiconvex. Lemma 2.9 also holds when all the functions are strictly quasiconvex.

□

We will use the following abbreviations:

$$D^=_{Q^=(x^\star)}(x^\star) = \bigcap_{k \in Q^=(x^\star)} D^=_{\phi^k}(x^\star)$$

and

$$D^{=}_{\mathcal{P}^{=}(x^\star)}(x^\star) = \bigcap_{i \in \mathcal{P}^{=}(x^\star)} D^{=}_{f^i}(x^\star).$$

Then we have the following major theorem.

Theorem 2.8 *Consider the pseudoconvex multi-objective program* (MP)*. Then a point* $x^\star \in F$ *is a Pareto minimum if, and only if, either* $Q = Q^{=}(x^\star)$ *or there exist*

$$\begin{aligned} w^\star &= (w_k^\star) \geq 0, \quad k \in Q \setminus Q^{=}(x^\star), \quad w^\star \neq 0, \\ u^\star &= (u_i^\star) \geq 0, \quad i \in \mathcal{P} \setminus \mathcal{P}^{=}(x^\star) \end{aligned} \tag{2.16}$$

such that

$$\sum_{k \in Q \setminus Q^{=}(x^\star)} w_k^\star \nabla \phi^k(x^\star) + \sum_{i \in \mathcal{P}(x^\star) \setminus \mathcal{P}^{=}(x^\star)} u_i^\star \nabla f^i(x^\star) \in \{D^{=}_{Q^{=}(x^\star)}(x^\star) + D^{=}_{\mathcal{P}^{=}(x^\star)}(x^\star)\}^+.$$

Proof: By Lemma 2.9, the feasible point $x^\star$ is not Pareto minimal if, and only if, both $Q \setminus Q^{=}(x^\star) \neq \emptyset$ and there exists $\hat{y} \in F$ such that (2.14) is satisfied. This together with pseudoconvexity and the definitions of $Q^{=}(x^\star)$ and $\mathcal{P}^{=}(x^\star)$ imply that the following system

$$\begin{aligned} \nabla \phi^k(x^\star) d &< 0, \quad k \in Q \setminus Q^{=}(x^\star) \\ \nabla f^i(x^\star) d &< 0, \quad i \in \mathcal{P}(x^\star) \setminus \mathcal{P}^{=}(x^\star) \\ d \in \{D^{=}_{Q^{=}(x^\star)}(x^\star) &\cap D^{=}_{\mathcal{P}^{=}(x^\star)}(x^\star)\} \end{aligned} \tag{2.17}$$

is consistent. This means that $x^\star \in F$ is optimal if either $Q \setminus Q^{=}(x^\star) = \emptyset$ or (2.17) is inconsistent. Applying the Dubovitskii-Milyutin theorem to the inconsistent system (2.17) yields that $x^\star$ is optimal if, and only if, either $Q \setminus Q^{=}(x^\star) = \emptyset$ or there exist

$$\begin{aligned} w^\star &= (w_k^\star) \geq 0, \quad k \in Q \setminus Q^{=}(x^\star) \\ u^\star &= (u_i^\star) \geq 0, \quad i \in \mathcal{P} \setminus \mathcal{P}^{=}(x^\star), \end{aligned} \tag{2.18}$$

not all zero, such that

$$\sum_{k \in Q \setminus Q^{=}(x^\star)} w_k^\star \nabla \phi^k(x^\star) + \sum_{i \in \mathcal{P}(x^\star) \setminus \mathcal{P}^{=}(x^\star)} u_i^\star \nabla f^i(x^\star) \in \{D^{=}_{Q^{=}(x^\star)}(x^\star) + D^{=}_{\mathcal{P}^{=}(x^\star)}(x^\star)\}^+.$$

We must now show that $w^\star \neq 0$. Note that (2.18) implies (2.16) which proves the sufficiency. To prove necessity, assume that $w^\star = 0$. Then we can have two cases: If $\mathcal{P}(x^\star) \setminus \mathcal{P}^{=}(x^\star) = \emptyset$ then, (2.18) is violated; if $\mathcal{P}(x^\star) \setminus \mathcal{P}^{=}(x^\star) \neq \emptyset$ then, at least one $u_i^\star$, $i \in \mathcal{P}(x^\star) \setminus \mathcal{P}^{=}(x^\star)$ is nonzero ,which together with the Dubovitskii-Milyutin theorem applied to (2.18) imply that

$$D^{<}_{\mathcal{P}(x^\star) \setminus \mathcal{P}^{=}(x^\star)}(x^\star) \cap D^{=}_{Q^{=}(x^\star)}(x^\star) \cap D^{=}_{\mathcal{P}^{=}(x^\star)}(x^\star) = \emptyset. \tag{2.19}$$

But by (2.15) there exists $\bar{y} \in F \cap F_0(x^\star)$ such that

$$f^i(\bar{y}) < 0, \quad i \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=(x^\star),$$

which together with the definitions of $Q^=(x^\star)$ and $\mathcal{P}^=(x^\star)$ imply that

$$\bar{d} = \bar{y} - x^\star \in D^<_{\mathcal{P}(x^\star)\setminus\mathcal{P}^=(x^\star)}(x^\star) \cap D^{\overline{=}}_{Q^=(x^\star)}(x^\star) \cap D^{\overline{=}}_{\mathcal{P}^=(x^\star)}(x^\star). \tag{2.20}$$

But (2.20) contradicts (2.19). Therefore $w^\star \neq 0$ and the proof is complete.

□

In our final theorem of this section, we will illustrate the role of LFS quasiconvex functions in multi-objective programming.

Theorem 2.9 *Consider the multi-objective program (MP) where all the functions are quasiconvex, differentiable and have the LFS property at a feasible point $x^\star$. If $x^\star$ is a Pareto minimum then the system*

$$\sum_{k\in Q} w_k \nabla\phi^k(x^\star) + \sum_{i\in\mathcal{P}(x^\star)} u_i \nabla f^i(x^\star) = 0$$

$$\begin{aligned} w_k &> 0, \quad k \in Q \\ u_i &\geq 0, \quad i \in \mathcal{P}(x^\star) \end{aligned}$$

is consistent.

Proof: Assume that $x^\star$ is a Pareto minimum. Then, by the Charnes and Cooper observation (see, e.g., [8,32]), $x^\star$ solves

$$\begin{aligned} \text{Min} \quad & \sum_{k\in Q} \phi^k(x) \\ \text{s.t.} \quad & \\ & \phi^k(x) \leq \phi^k(x^\star), \quad k \in Q \\ & f^i(x) \leq 0, \quad i \in \mathcal{P}. \end{aligned}$$

Since all functions are quasiconvex with LFS property at $x^\star$, we use Theorem 2.3 to conclude that the system

$$\sum_{k\in Q} (1 + \lambda_k) \nabla\phi^k(x^\star) + \sum_{i\in\mathcal{P}(x^\star)} u_i \nabla f^i(x^\star) = 0$$

$$\begin{aligned} \lambda_k &\geq 0, \quad k \in Q \\ u_i &\geq 0, \quad i \in \mathcal{P}(x^\star) \end{aligned}$$

is consistent. Let $w_k = 1 + \lambda_k$, $k \in Q$. This completes the proof.

Chapter 3

Inverse Programming

This chapter deals mainly with *mathematical programming models.* These are mathematical programs that contain two sets of variables. One set are the parameters that can be directly influenced or controlled, and the other set are the decision variables. The level of optimization, dealing with *continuous* optimization of mathematical models is termed *input optimization.* We first recall and elaborate on some of the basic notions in input optimization such as *stability* and *optimal input.* We will then introduce and study the *inverse programming problem* for a large class of LFS functions, namely linear functions.

3.1 Modified Optimality Conditions

In the two previous chapters, we have studied characterizations of optimality that use a subset of $\mathcal{P}^=$. In this section, another index set that is larger than $\mathcal{P}^=$ will be introduced and studied in convex programming. Using this larger index set to characterize optimality is motivated by the Charnes-Cooper formulation of Pareto optimality [8]. This set is used to characterize optimality in multi-objective programming [30]. However, we study it in single-objective programming. This set will then be studied for parametric problems in the next section.

We recall the convex program (P) from Chapter 1 and, associated with a fixed feasible point $x^\star$, define the following sets:

$$\begin{aligned} \mathcal{P}^=_\star &= \{k \in \mathcal{P} : x \in F,\ \ f^0(x) \le f^0(x^\star) \Rightarrow f^k(x) = 0\}; \\ F^=(x^\star) &= \{x \in R^n : f^k(x) = 0,\ \ k \in \mathcal{P}^=_\star\}; \\ F^{\le}(x^\star) &= \{x \in R^n : f^k(x) \le 0,\ \ k \in \mathcal{P}^=_\star\}. \end{aligned}$$

Note that the set $F^=(x^\star)$ is not generally convex and $\mathcal{P}^=_\star \neq \mathcal{P}^=$, as the following example shows.

Example 3.1 Consider

$$\begin{aligned} \text{Min} \quad & f^0 = (x_1 - 1)^2 + x_2^2 \\ \text{s.t.} \quad & \\ & f^1 = (x_1+1)^2 + (x_2-1)^2 - 2 \le 0 \\ & f^2 = (x_1+1)^2 + (x_2+1)^2 - 2 \le 0 \end{aligned}$$

at $x^\star = (0,0)^T$. Here $\mathcal{P}_\star^= = \{1,2\}$ and

$$F^=(x^\star) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \end{bmatrix} \right\}.$$

Clearly, $F^=(x^\star)$ is not a convex set. Furthermore, $\mathcal{P}^= = \emptyset$ which is different from the index set $\mathcal{P}_\star^=$.

□

We begin our theory of optimality, using the index set $\mathcal{P}_\star^=$, with a simple but important lemma.

Lemma 3.1 *Consider the convex program (P). If* $\mathcal{P} \setminus \mathcal{P}_\star^= \neq \emptyset$*, then there exists* $\hat{x} \in F$ *such that*

$$f^0(\hat{x}) \le f^0(x^\star) \text{ and } f^k(\hat{x}) < 0, \quad k \in \mathcal{P} \setminus \mathcal{P}_\star^=. \tag{3.1}$$

Proof: Clearly

$$\mathcal{P} \setminus \mathcal{P}_\star^= = \{k \in \mathcal{P} : \exists x^k \in F \ni f^0(x^k) \le f^0(x^\star) \text{ and } f^k(x^k) < 0\}.$$

For x^k's from $\mathcal{P} \setminus \mathcal{P}_\star^=$, choose

$$\hat{x} = \frac{1}{card\{\mathcal{P} \setminus \mathcal{P}_\star^=\}} \sum_{k \in \mathcal{P} \setminus \mathcal{P}_\star^=} x^k.$$

Then, by the convexity of f^i, $i \in \{0\} \cup \mathcal{P}$, we conclude that $\hat{x} \in F$ and that (3.1) is satisfied.

□

Let us define

$$\mathcal{L}^\star(x,u) = f^0(x) + \sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} u_i f^i(x)$$

to be the restricted Lagrangian function associated with a fixed $x^\star$. Also let $c = card\{\mathcal{P} \setminus \mathcal{P}_\star^=\}$. Then we have the following characterization.

Theorem 3.1 *A point* $x^\star \in F$ *is an optimal solution of the convex program (P) if, and only if, there exist* $u_i^\star \ge 0$, $i \in \mathcal{P} \setminus \mathcal{P}_\star^=$, *such that*

$$\mathcal{L}^\star(x^\star, u) \le \mathcal{L}^\star(x^\star, u^\star) \le \mathcal{L}^\star(x, u^\star) \tag{3.2}$$

for all $x \in conv\{F^=(x^\star)\}$ *and for all* $u \in R_+^c$.

Proof: (Sufficiency:) Assume that (3.2) holds. Then the left-hand inequality of (3.2) implies that

$$f^0(x^\star) + \sum_{i\in\mathcal{P}\setminus\mathcal{P}^=_\star} u_i f^i(x^\star) \le f^0(x^\star) + \sum_{i\in\mathcal{P}\setminus\mathcal{P}^=_\star} u^\star_i f^i(x^\star). \tag{3.3}$$

Now, for $u_i = 0,\ i \in \mathcal{P}\setminus\mathcal{P}^=_\star$, the inequality (3.3) implies

$$\sum_{i\in\mathcal{P}\setminus\mathcal{P}^=_\star} u^\star_i f^i(x^\star) \ge 0.$$

But

$$\sum_{i\in\mathcal{P}\setminus\mathcal{P}^=_\star} u^\star_i f^i(x^\star) \le 0,$$

since $x^\star \in F$ and $u^\star_i \ge 0,\ \ i \in \mathcal{P}\setminus\mathcal{P}^=_\star$. Therefore

$$\sum_{i\in\mathcal{P}\setminus\mathcal{P}^=_\star} u^\star_i f^i(x^\star) = 0. \tag{3.4}$$

Note that

$$\{F \cap \{x : f^0(x) \le f^0(x^\star)\}\} \subset F^=(x^\star) \subset conv\{F^=(x^\star)\}.$$

Now for every $x \in conv\{F^=(x^\star)\}$, and therefore for every $x \in F \cap \{x : f^0(x) \le f^0(x^\star)\}$,

$$\begin{aligned} f^0(x^\star) &\le f^0(x^\star) + \sum_{i\in\mathcal{P}\setminus\mathcal{P}^=_\star} u^\star_i f^i(x^\star) \\ &\le f^0(x) + \sum_{i\in\mathcal{P}\setminus\mathcal{P}^=_\star} u^\star_i f^i(x) \text{ by the right-hand inequality of (3.2)} \\ &\le f^0(x). \end{aligned}$$

Therefore

$$f^0(x) \le f^0(x^\star) \quad \text{for every } x \in F \cap \{x : f^0(x) \le f^0(x^\star)\}$$

which means that $x^\star$ is optimal.

(Necessity:) Assume that $x^\star$ is an optimal solution of the convex program (P). We can assume, without loss of generality, that the first c indices constitute the set $\mathcal{P}\setminus\mathcal{P}^=_\star$. Define, in R^{c+1}, the sets

$$K_1 = \left\{ y : y \ge \begin{bmatrix} f^0(x) \\ f^1(x) \\ \vdots \\ f^c(x) \end{bmatrix} \text{ for at least one } x \in conv\{F^=(x^\star)\} \right\}$$

and

$$K_2 = \left\{ y : y < \begin{bmatrix} f^0(x^\star) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}.$$

Both sets are convex and $K_1 \cap K_2 = \emptyset$, for if the latter were not true, there would exist $\hat{x} \in conv\{F^=(x^\star)\}$ and $\hat{y} \in R^{c+1}$ such that

$$\begin{bmatrix} f^0(\hat{x}) \\ f^1(\hat{x}) \\ \vdots \\ f^c(\hat{x}) \end{bmatrix} \leq \hat{y} < \begin{bmatrix} f^0(x^\star) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This would mean that $\hat{x} \in F$ and $f^0(\hat{x}) < f^0(x^\star)$, contradicting optimality of $x^\star$. Therefore, there exists a hyperplane that separates K_1 from clK_2. That is, there exists $a \neq 0$ and α such that

$$(a, y^1) \geq \alpha \geq (a, y^2) \quad \text{for every } y^1 \in K_1 \quad \text{and for every } y^2 \in clK_2.$$

Clearly, a is nonnegative. Specify, for an arbitrary $x \in conv\{F^=(x^\star)\}$,

$$y_1 = \begin{bmatrix} f^0(x) \\ f^1(x) \\ \vdots \\ f^c(x) \end{bmatrix} \quad \text{and} \quad y_2 = \begin{bmatrix} f^0(x^\star) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Then

$$a_0 f^0(x) + \sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} a_i f^i(x) \geq a_0 f^0(x^\star).$$

We claim that $a_0 > 0$, for if this were not true, we would have

$$\sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} a_i f^i(x) \geq 0 \quad \text{for every } x \in conv\{F^=(x^\star)\}. \tag{3.5}$$

But, by Lemma 3.1, there would exist $\hat{x} \in F$ such that

$$f^0(\hat{x}) \leq f^0(x^\star) \text{ and } f^k(\hat{x}) < 0, \quad k \in \mathcal{P} \setminus \mathcal{P}_\star^=.$$

But not all a_i, $i \in \mathcal{P} \setminus \mathcal{P}_\star^=$, are equal to zero. Therefore

$$\sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} a_i f^i(\hat{x}) < 0,$$

a contradiction to (3.5). Now let

$$u_i^\star = \frac{a_i}{a_0}, \quad i \in \mathcal{P} \setminus \mathcal{P}_\star^=.$$

Then

$$\mathcal{L}^\star(x, u^\star) = f^0(x) + \sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} u_i^\star f^i(x) \geq f^0(x^\star) \quad \text{for every } x \in conv\{F^=(x^\star)\}.$$

In particular, for $x = x^\star$ this gives

$$\sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} u_i^\star f^i(x^\star) \geq 0.$$

But, since $x^\star \in F$ and $u_i^\star \geq 0, \quad i \in \mathcal{P} \setminus \mathcal{P}_\star^=$,

$$\sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} u_i^\star f^i(x^\star) \leq 0.$$

Therefore

$$\sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} u_i^\star f^i(x^\star) = 0$$

and so $\mathcal{L}^\star(x^\star, u^\star) = f^0(x^\star)$. Hence, by (3.6),

$$\mathcal{L}^\star(x, u^\star) \geq \mathcal{L}^\star(x^\star, u^\star),$$

proving the right-hand inequality. Besides

$$\mathcal{L}^\star(x^\star, u^\star) \geq f^0(x^\star) + \sum_{i \in \mathcal{P} \setminus \mathcal{P}_\star^=} u_i f^i(x^\star)$$

proving the left-hand inequality.

□

If in Theorem 3.1, we use the convex set $F^{\leq}(x^\star)$, instead of $conv\{F^=(x^\star)\}$, we obtain the following theorem, the proof of which is exactly the same as the proof of Theorem 3.1 except that $conv\{F^=(x^\star)\}$ is replaced by $F^{\leq}(x^\star)$.

Theorem 3.2 *A point $x^\star \in F$ is an optimal solution of the convex program (P) if, and only if, there exist $u_i^\star \geq 0, \ i \in \mathcal{P} \setminus \mathcal{P}_\star^=$, such that*

$$\mathcal{L}^\star(x^\star, u) \leq \mathcal{L}^\star(x^\star, u^\star) \leq \mathcal{L}^\star(x, u^\star) \qquad (3.6)$$

for all $x \in F^{\leq}(x^\star)$ and for all $u \in R_+^c$.

The following Corollary then follows.

Corollary 3.1 *A point $x^\star \in F$ is an optimal solution of the convex program (P) if, and only if, there exist $u_i^\star \geq 0$, $i \in \mathcal{P} \setminus \mathcal{P}_\star^=$, such that*

$$\mathcal{L}^\star(x^\star, u) \leq \mathcal{L}^\star(x^\star, u^\star) \leq \mathcal{L}^\star(x, u^\star) \tag{3.7}$$

for all $x \in F^=(x^\star)$ and for all $u \in R_+^c$.

Proof: (Sufficiency:) The proof is exactly the same as the proof given in the sufficiency part of Theorem 3.1 since

$$\{F \cap \{x : f^0(x) \leq f^0(x^\star)\}\} \subset F^=(x^\star) \subset conv\{F^=(x^\star)\}.$$

(Necessity:) Assume that $x^\star$ is optimal. Then by Theorem 3.1 there exist $u_i^\star \geq 0$, $i \in \mathcal{P} \setminus \mathcal{P}_\star^=$, such that (3.2) holds, for all $x \in conv\{F^=(x^\star)\}$ and all $u \in R_+^c$. Note that

$$F^=(x^\star) \subset conv\{F^=(x^\star)\}.$$

By choosing the same $u^\star$ as in the proof of Theorem 3.1, (3.7) holds for all $x \in F^=(x^\star)$ and all $u \in R_+^c$.

□

We recall from [32] the set

$$F^= = \{x \in R^n : f^i(x) = 0, \ i \in \mathcal{P}^=\}$$

and the usual Lagrangian for the convex program (P)

$$\mathcal{L}(x, u) = f^0(x) + \sum_{i \in \mathcal{P} \setminus \mathcal{P}^=} u_i f^i(x).$$

Then it is well known (see, e.g., [32,36]) that $x^\star \in F^=$ is an optimal solution of the convex program (P) if, and only if, there exists $u = (u_i) \geq 0$, $i \in \mathcal{P} \setminus \mathcal{P}^=$, such that

$$\mathcal{L}(x^\star, u) \leq \mathcal{L}(x^\star, u^\star) \leq \mathcal{L}(x, u^\star).$$

Unlike the minimal index set and the Lagrangian function of the above well-known result, the index set and the Lagrangian function of the saddle-point results derived in this section depend on $x^\star$, the candidate for optimality. Furthermore, it is well known that $F^=$ is a convex set and obviously independent of the feasible point $x^\star$ while we showed that $F^=(x^\star)$ is not generally convex and depends on $x^\star$.

Our objective is now to give a subdifferentiable and a differentiable version of Corollary 3.1. To this end, we need several preliminary results. The following result is very trivial, so we give it without proof.

$$conv\{F^=(x^\star)\} - x^\star = conv\{F^=(x^\star) - x^\star\}.$$

Hence

$$\{conv\{F^{=}(x^\star)\} - x^\star\}^+ = \{conv\{F^{=}(x^\star) - x^\star\}\}^+ = \{F^{=}(x^\star) - x^\star\}^+.$$

The following important lemma can therefore be derived.

Lemma 3.2 *Consider the convex program (P) and a point $x^\star \in F$. Then*

$$\{F^{=}(x^\star) - x^\star\}^+ \subset \{D^{=}_{\mathcal{P}^{=}_\star}(x^\star)\}^+. \tag{3.8}$$

Proof: Let $y \in \{F^{=}(x^\star) - x^\star\}^+$. Then

$$yz \geq 0 \quad \text{for every } z \in \{F^{=}(x^\star) - x^\star\}.$$

Given any $d \in D^{=}_{\mathcal{P}^{=}_\star}(x^\star)$, there exists an $\bar{\alpha} > 0$ sufficiently small such that, for all $0 < \alpha \leq \bar{\alpha}$, we have

$$f^k(x^\star + \alpha d) = 0, \quad k \in \mathcal{P}^{=}_\star.$$

So $x^\star + \alpha d \in F^{=}(x^\star)$ for all $0 < \alpha \leq \bar{\alpha}$, or

$$\alpha d \in \{F^{=}(x^\star) - x^\star\}.$$

Therefore $y(\alpha d) \geq 0$, which implies that $yd \geq 0$. This means that

$$y \in \{D^{=}_{\mathcal{P}^{=}_\star}(x^\star)\}^+.$$

$\square$

Note that the reverse inclusion $\{D^{=}_{\mathcal{P}^{=}_\star}(x^\star)\}^+ \subset \{F^{=}(x^\star) - x^\star\}^+$ is not generally true as the next example shows.

Example 3.2 Consider again the problem

$$\begin{aligned} \text{Min} \quad & f^0 = (x_1 - 1)^2 + x_2^2 \\ \text{s.t.} \quad & \\ & f^1 = (x_1 + 1)^2 + (x_2 - 1)^2 - 2 \leq 0 \\ & f^2 = (x_1 + 1)^2 + (x_2 + 1)^2 - 2 \leq 0 \end{aligned}$$

at $x^\star = (0,0)^T$. Then

$$\mathcal{P}^{=}_\star = \{1,2\},\ F^{=}(x^\star) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \end{bmatrix} \right\} \text{ and } D^{=}_{\mathcal{P}^{=}_\star}(x^\star) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}.$$

Therefore

$$\{D^{=}_{\mathcal{P}^{=}_\star}(x^\star)\}^+ = R^2,$$

while

$$\{(F^{=}(x^\star) - x^\star\}^+ = \{d \in R^2 : d_1 \leq 0\}.$$

Let us introduce the Lagrangian function

$$L^\star(x,u^\star) = f^0(x) + \sum_{k\in\mathcal{P}(x^\star)\setminus\mathcal{P}^=_\star} u_k^\star f^k(x).$$

Also, let $\hat{c} = card\{\mathcal{P}(x^\star) \setminus \mathcal{P}^=_\star\}$. Then Theorem 3.1 can be rephrased as follows:

Lemma 3.3 *Consider a convex program (P). A point $x^\star \in F$ is optimal if, and only if, $x^\star$ minimizes $L^\star(x,u^\star)$ on $conv\{F^=(x^\star)\}$ (or on $F^{\leq}(x^\star)$) for some $u^\star \in R^{\hat{c}}_+$.*

Proof: Since we already have the feasibility of $x^\star$, to prove the lemma, it suffices to show that we can replace $\mathcal{L}^\star(x,u^\star)$ with $L^\star(x,u^\star)$ in (3.2). We have, by equation (3.4),

$$\sum_{i\in\mathcal{P}\setminus\mathcal{P}^=_\star} u_i^\star f^i(x^\star) = 0.$$

But $u_i^\star f^i(x^\star) \leq 0, \;\; i \in \mathcal{P}$. Hence, we have

$$u_i^\star f^i(x^\star) = 0, \; i \in \mathcal{P}.$$

This implies that

$$u_i^\star = 0, \;\; i \in \mathcal{P} \setminus \mathcal{P}(x^\star).$$

So the Lagrangian in (3.2) is indeed equal to $L^\star(x,u^\star)$.

□

We now derive a characterization for the subdifferentiable case.

Theorem 3.3 *Consider a convex program (P). A point $x^\star \in F$ is optimal if, and only if, there exist*

$$h^i \in \partial f^i(x^\star), \;\; i \in \{0\} \cup (\mathcal{P}(x^\star) \setminus \mathcal{P}^=_\star)$$

and

$$u_k^\star \geq 0, \;\; k \in \mathcal{P}(x^\star) \setminus \mathcal{P}^=_\star$$

such that

$$(h^0)^T + \sum_{k\in\mathcal{P}(x^\star)\setminus\mathcal{P}^=_\star} u_k^\star (h^k)^T \in \{D^=_{\mathcal{P}^=_\star}(x^\star)\}^+. \tag{3.9}$$

Proof: (Sufficiency:) Since $h^i \in \partial f^i(x^\star)$, $i \in \{0\} \cup (\mathcal{P}(x^\star) \setminus \mathcal{P}^=_\star)$, for every $x \in R^n$, we have

$$f^i(x) \geq f^i(x^\star) + h^i(x - x^\star), \;\; i \in \{0\} \cup (\mathcal{P}(x^\star) \setminus \mathcal{P}^=_\star). \tag{3.10}$$

Therefore

$$\begin{aligned} f^0(x) + \sum_{k\in\mathcal{P}(x^\star)\setminus\mathcal{P}^=_\star} u_k^\star f^k(x) \;\; &\geq \;\; f^0(x^\star) + \sum_{k\in\mathcal{P}(x^\star)\setminus\mathcal{P}^=_\star} u_k^\star f^k(x^\star) \\ &\quad + (h^0 + \sum_{k\in\mathcal{P}(x^\star)\setminus\mathcal{P}^=_\star} u_k^\star h^k)(x - x^\star). \end{aligned}$$

Now, for $x \in F \cap \{x : f^0(x) \le f^0(x^\star)\}$, it follows that

$$f^0(x) \ge f^0(x^\star) + (h^0 + \sum_{k \in \mathcal{P}(x^\star) \setminus \mathcal{P}_\star^=} u_k^\star h^k)(x - x^\star).$$

Besides, for such x, the convexity of the set $F \cap \{x : f^0(x) \le f^0(x^\star)\}$ implies that

$$x - x^\star \in \{F(x^\star) \cap D_{f^0}^{\le}(x^\star)\}.$$

(Note that $F(x^\star)$ is the set of feasible directions at $x^\star$ defined in Chapter 1.) By the definition of the set $\mathcal{P}_\star^=$, this implies that

$$x - x^\star \in D_{\mathcal{P}_\star^=}^=(x^\star).$$

Therefore

$$(h^0 + \sum_{k \in \mathcal{P}(x^\star) \setminus \mathcal{P}_\star^=} u_k^\star h^k)(x - x^\star) \ge 0 \quad \text{for every } x \in F \cap \{x : f^0(x) \le f^0(x^\star)\}.$$

This implies that

$$f^0(x) \ge f^0(x^\star) \quad \text{for every } x \in F \cap \{x : f^0(x) \le f^0(x^\star)\},$$

proving optimality of $x^\star$.

(Necessity:) Assume that $x^\star$ is optimal. Then, by Theorem 3.3, $x^\star$ minimizes $L^\star(x, u^\star)$ for some $u^\star \in R_+^{\hat{c}}$ and for all $x \in conv\{F^=(x^\star)\}$, where $\hat{c} = card\{\mathcal{P}(x^\star) \setminus \mathcal{P}_\star^=\}$. This means that there exists

$$h = (h^0 + \sum_{k \in \mathcal{P}(x^\star) \setminus \mathcal{P}_\star^=} u_k^\star h^k) \in \partial L^\star(x^\star, u^\star)$$

such that

$$h(x - x^\star) \ge 0 \quad \text{for every } x \in conv\{F^=(x^\star)\}.$$

Therefore

$$h^T \in \{conv\{F^=(x^\star)\} - x^\star\}^+ = \{F^=(x^\star) - x^\star\}^+.$$

It follows, by Lemma 3.2, that

$$h^T \in \{D_{\mathcal{P}_\star^=}^=(x^\star)\}^+.$$

□

A realization of the above result to differentiable functions is obvious.

Corollary 3.2 *Consider a convex and differentiable program (P). Then $x^\star \in F$ is optimal if, and only if, there exist*

$$u_k^\star \ge 0, \quad k \in \mathcal{P}(x^\star) \setminus \mathcal{P}_\star^=,$$

such that

$$(\nabla f^0(x^\star))^T + \sum_{k \in \mathcal{P}(x^\star) \setminus \mathcal{P}_\star^=} u_k^\star (\nabla f^k(x^\star))^T \in \{D_{\mathcal{P}_\star^=}^=(x^\star)\}^+. \tag{3.11}$$

Note that the difference between the characterizations in Theorem 3.3 and Corollary 3.2 and the usual characterizations of optimality in the subdifferentiable and differentiable forms (see, e.g., [4]) is that the index set $\mathcal{P}^=_\star$ is used instead of the usual index set $\mathcal{P}^=$. As we stated earlier the former set depends on $x^\star$, the candidate for optimality, while the latter does not.

One interesting point about the modified optimality conditions derived in this section is that although they require the calculation of $\mathcal{P}^=_\star$ each time a point $x^\star$ is tested for optimality, they usually provide more trivial sufficient optimality conditions. In other words, since $\mathcal{P}^= \subset \mathcal{P}^=_\star$, these optimality conditions usually involve fewer Lagrange multipliers, and the set $F^=(x^\star)$ is usually a smaller set than $F^=$. We will now illustrate some of these optimality conditions by the following example.

Example 3.3 Consider, once more, the problem

$$\begin{aligned} \text{Min} \quad & f^0 = (x_1 - 1)^2 + x_2^2 \\ \text{s.t.} \quad & \\ & f^1 = (x_1 + 1)^2 + (x_2 - 1)^2 - 2 \leq 0 \\ & f^2 = (x_1 + 1)^2 + (x_2 + 1)^2 - 2 \leq 0 \end{aligned}$$

at $x^\star = (0,0)^T$, where

$$\mathcal{P}^=_\star = \{1,2\},\ F^=(x^\star) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \begin{bmatrix} -2 \\ 0 \end{bmatrix} \right\} \text{ and } D^=_{\mathcal{P}^=_\star}(x^\star) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}.$$

Furthermore

$$\{D^=_{\mathcal{P}^=_\star}(x^\star)\}^+ = R^2.$$

In this case $\mathcal{P} \setminus \mathcal{P}^=_\star = \emptyset$, and therefore

$$\mathcal{L}^\star(x,u) = f^0(x).$$

The optimality condition from Corollary 3.1 becomes: (Note that there are no Lagrange multipliers in this case.)

$$f^0(x^\star) \leq f^0(x)$$

for every $x \in F^=(x^\star)$. Clearly, this is a consistent system, verifying optimality of $x^\star$.

On the other hand

$$\nabla f^0(x^\star) = (-2,0).$$

Therefore, the optimality condition from Corollary 3.2 becomes

$$\begin{bmatrix} -2 \\ 0 \end{bmatrix} \in R^2,$$

which is clearly a consistent system.

□

We now show how to calculate the set $\mathcal{P}^{=}_{\star}$. Any method for calculating the index set $\mathcal{P}^{=}$ should work for $\mathcal{P}^{=}_{\star}$ as well, provided that the constraint $f^0(x) - f^0(x^\star) \leq 0$ is added to the constraints of the program (P). Here, we give an algorithm which is a slight extension of the algorithm for calculating $\mathcal{P}^{=}$, given in [39], except that an extra constraint due to the objective function has been added to the problem. The algorithm starts with $\Omega = \mathcal{P}(x^\star)$ and ends with $\Omega = \mathcal{P}^{=}_{\star}$. At every iteration the program

$$(P^\star, \theta, \Omega) \qquad \begin{aligned} \text{Min} \quad & \sum_{k \in \Omega} \nabla f^k(x^\star) d \\ \text{s.t.} \quad & \\ & \nabla f^k(x^\star) d + \theta \| d - \delta^k \|_1 \leq 0 \\ & \delta^k \in D_{f^k}(x^\star) \\ & |d_i| \leq 1, \quad |\delta^k_i| \leq 1 \quad k \in \mathcal{P}(x^\star) \cup \{0\}, \; i = 1, \ldots, n \end{aligned}$$

is solved. Let $K = \{k \in \mathcal{P}(x^\star) : \nabla f^k(x^\star) \tilde{d} < 0\}$. Then the steps of the algorithm are as follows:

Initialization: Calculate $\mathcal{P}(x^\star)$ and $D_{f^k}(x^\star)$, $k \in \mathcal{P}(x^\star) \cup \{0\}$. Specify a tolerance $\epsilon > 0$, and set $\theta > \epsilon$ and $\Omega = \mathcal{P}(x^\star)$.

Step 1: Find an optimal solution of $(P^\star, \theta, \Omega)$ $\tilde{d}$, $\tilde{\delta}$. Calculate its optimal value

$$\tilde{g}(\theta) = \sum_{k \in \Omega} \nabla f^k(x^\star) \tilde{d}.$$

Step 2: If $\tilde{g}(\theta) < 0$, determine K, set $\Omega = \Omega \setminus K$ and return to Step 1. Otherwise continue.

Step 3: If $\tilde{g}(\theta) = 0$, set $\theta = \frac{1}{2}$. If $\theta > \epsilon$, return to Step 1. If $\theta \leq \epsilon$ stop; $\Omega = \mathcal{P}^{=}_{\star}$.

□

If all the functions, including the objective function are faithfully convex, i.e, if $f^k(x) = \phi^k(A_k x + b_k) + a_k^T x + \alpha$, $k \in \{0\} \cup \mathcal{P}$, where ϕ^k, $k \in \{0\} \cup \mathcal{P}$, are strictly convex, then the program $(P^\star, \theta, \Omega)$ is significantly simplified to the following program:

$$(L^\star, \theta, \Omega) \qquad \begin{aligned} \text{Min} \quad & \sum_{k \in \Omega} \nabla f^k(x^\star) d \\ \text{s.t.} \quad & \\ & \nabla f^k(x^\star) d + \theta \left(|a_k^T d| + \sum_{i=1}^{m_k} |A^i_k d| \right) \\ & |d_i| \leq 1, \quad k \in \mathcal{P}(x^\star) \cup \{0\}, \; i = 1, \ldots, n \end{aligned}$$

where A^i_k is the i_{th} row of A_k (see [39]). The above algorithm for calculating $\mathcal{P}^{=}_{\star}$ then applies with minor changes.

Initialization: Calculate $\mathcal{P}(x^\star)$ and identify A_k, b_k, a_k, $k \in \mathcal{P}(x^\star) \cup \{0\}$. Specify a tolerance $\epsilon > 0$; set $\theta > \epsilon$ and $\Omega = \mathcal{P}(x^\star)$.

Step 1: Find an optimal solution of $(L^\star, \theta, \Omega)$. Calculate its optimal value

$$\tilde{g}(\theta) = \sum_{k \in \Omega} \nabla f^k(x^\star)\tilde{d}.$$

Steps 2 and 3 remain unchanged.

In order to demonstrate how the algorithm works, we borrow the next example from [4] and modify it by adding an objective function and fixing a feasible point $x^\star$.

Example 3.4 Consider

$$\begin{array}{ll}
\text{Min} & f^0 = x_1^2 + e^{x_3} \\
\text{s.t.} & \\
& f^1 = e^{x_1} + x_2^2 - 1 \leq 0 \\
& f^2 = x_1^2 + x_2^2 + e^{-x_3} - 1 \leq 0 \\
& f^3 = x_1 + x_4^2 + x_5^2 - 1 \leq 0 \\
& f^4 = e^{-x_2} - 1 \leq 0 \\
& f^5 = (x_1 - 1)^2 + x_2^2 - 1 \leq 0 \\
& f^6 = x_1 + e^{-x_4} - 1 \leq 0 \\
& f^7 = x_2 + e^{-x_5} - 1 \leq 0
\end{array}$$

and $x^\star = (0,0,0,0,0)^T$.

Initialization: Here

$$\begin{aligned}
\mathcal{P}(x^\star) &= \{1,2,4,5,6,7\} \\
\nabla f^0(x^\star) &= (0,0,1,0,0) \\
\nabla f^1(x^\star) &= (1,0,0,0,0) \\
\nabla f^2(x^\star) &= (0,0,-1,0,0) \\
\nabla f^4(x^\star) &= (0,-1,0,0,0) \\
\nabla f^5(x^\star) &= (-2,0,0,0,0) \\
\nabla f^6(x^\star) &= (1,0,0,-1,0) \\
\nabla f^7(x^\star) &= (0,1,0,0,-1).
\end{aligned}$$

Furthermore, we identify

$$\begin{aligned}
a_0 &= a_1 = a_2 = a_4 = a_5 = 0 \\
a_6 &= (1,0,0,0,0)^T \\
a_7 &= (0,1,0,0,0)^T
\end{aligned}$$

$$
\begin{aligned}
A_0 &= \begin{bmatrix} 1\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0 \end{bmatrix} \\
A_1 &= A_5 = \begin{bmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \end{bmatrix} \\
A_2 &= \begin{bmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \end{bmatrix} \\
A_4 &= (0,1,0,0,0)^T \\
A_6 &= (0,0,0,1,0)^T \\
A_7 &= (0,0,0,0,1)^T.
\end{aligned}
$$

Iteration 1: For $\theta = 0.1$ the optimal solution of the corresponding $(L^\star, \theta, \mathcal{P}(x^\star))$, i.e., of

$$
\begin{aligned}
\text{Min} \quad & g^0 = -d_3 - d_4 - d_5 \\
\text{s.t.} \quad & \\
& g^1 = d_3 + \theta(|d_1| + |d_3|) \leq 0 \\
& g^2 = d_1 + \theta(|d_1| + |d_2|) \leq 0 \\
& g^3 = -d_3 + \theta(|d_1| + |d_2| + |d_3|) \leq 0 \\
& g^4 = -d_2 + \theta(|d_2|) \leq 0 \\
& g^5 = -2d_1 + \theta(|d_1| + |d_2|) \leq 0 \\
& g^6 = d_1 - d_4 + \theta(|d_1| + |d_4|) \leq 0 \\
& g^7 = d_2 - d_5 + \theta(|d_2| + |d_5|) \leq 0 \\
& |d_i| \leq 1 \quad i = 1, \ldots, 5
\end{aligned}
$$

is $\tilde{d} = (0,0,0,1,1)^T$, which gives $\tilde{g}(\theta) = -2$. It then follows that

$$K = \mathcal{P}(x^\star) \setminus \{6,7\} = \{1,2,4,5\}.$$

Iteration 2: Using again $\theta = 0.1$, we solve the program $(L^\star, \theta, \{1,2,4,5\})$ which is Min $-d_3$, subject to the same constraints as in the above $(L^\star, \theta, \mathcal{P}(x^\star))$. It turns out that $\tilde{g}(\theta) = 0$, regardless of $\theta > 0$. Therefore

$$\mathcal{P}^{=}_{\star} = \{1,2,4,5\}.$$

□

The set $\mathcal{P}^{=}_{\star}$ in the above example has also been calculated by a different method in [30].

3.2 Basic Point-to-Set Topology

In this section we will extend the definition of $\mathcal{P}_\star^=$ to parametric problems. First let us recall some basic notions from point-to-set topology that are used in input optimization. We study the mathematical models of the form

$$(P,\theta) \qquad \begin{array}{ll} \underset{(x)}{\text{Min}} & f^0(x,\theta) \\ s.t. & \\ & f^k(x,\theta) \leq 0, \quad k \in \mathcal{P} = \{1,\ldots,m\} \end{array}$$

where $x \in R^n$ is a decision variable, $\theta \in I \subset R^p$ is a parameter and $f^i : R^n \times R^p \to R$, $i \in \{0\} \cup \mathcal{P}$, are continuous functions. We assume that all functions $f^i(.,\theta) : R^n \to R$, $i \in \{0\} \cup \mathcal{P}$, are convex for every θ. Such (P,θ) is referred to as a *convex model*. With every "input"(parameter) θ, we associate the "output" triple, that is the *feasible set*

$$F(\theta) = \{x : f^i(x,\theta) \leq 0, \quad i \in \mathcal{P}\};$$

the *set of optimal solutions* $\tilde{x}(\theta)$

$$\tilde{F}(\theta) = \{\tilde{x}(\theta)\};$$

and the *optimal value*

$$\tilde{f}(\theta) = f^0(\tilde{x}(\theta),\theta).$$

One of the basic criteria for using the model (P,θ) in practice is its "stability", i.e., continuous dependence of the output on the input. Since the mapping F is closed, it is considered to be *continuous* at θ if it is *lower semicontinuous* (or, equivalently, open) at θ. We recall that a point-to-set mapping $\Gamma : Z \to X$, between two topological vector spaces Z and X, is lower semicontinuous at some $\theta^\star$ if, for every open set $\mathcal{A}$, such that $\mathcal{A} \cap \Gamma(\theta^\star) \neq \emptyset$, there is a neighbourhood $N(\theta^\star)$ of $\theta^\star$ such that $\mathcal{A} \cap \Gamma(\theta) \neq \emptyset$ for every $\theta \in N(\theta^\star)$ (see, e.g. [2]).

We will always assume that the set of optimal solutions at $\theta^\star$ is nonempty and bounded. The objective functions with this property are called *realistic* at $\theta^\star$ and for these there is an important characterization of continuity given in [32].

Theorem 3.4 *Consider the convex model (P,θ) at $\theta^\star$. Then the following statements are equivalent:*

(i) The point-to-set mapping F is continuous at $\theta^\star$;

(ii) For every realistic objective function f^0, there is a neighbourhood $N(\theta^\star)$ of $\theta^\star$ such that $\tilde{F}(\theta) \neq \emptyset$ for every $\theta \in N(\theta^\star)$, and $\theta \in N(\theta^\star)$, $\theta \to \theta^\star$, implies that the sequence $\tilde{x}(\theta)$ is bounded and all its limit points lie in $\tilde{F}(\theta^\star)$;

(iii) For every realistic objective function f^0 there exists a neighbourhood $N(\theta^*)$ of θ^* such that $\tilde{F}(\theta) \neq \emptyset$ for every $\theta \in N(\theta^*)$, and $\theta \in N(\theta^*)$, $\theta \to \theta^*$, implies that $\tilde{f}(\theta) \to \tilde{f}(\theta^*)$.

We also recall the following definition from [32].

Definition 3.1 *Consider the convex model (P, θ). A set $S \subset R^p$ is a region of stability at $\theta^* \in S$ if for each open set $\mathcal{A} \subset R^n$ satisfying $\mathcal{A} \cap F(\theta^*) \neq \emptyset$, there exists a neighbourhood $N(\theta^*)$ of θ^* such that $\mathcal{A} \cap F(\theta) \neq \emptyset$ for each $\theta \in N(\theta^*) \cap S$.*

Therefore, S is a region attached to θ^* where the mapping F is locally lower semicontinuous. If S can be identified as a neighbourhood, i.e., if $S = N(\theta^*)$, then the model is said to be *stable* at θ^*. For example, if Slater's condition holds for the program (P, θ^*), i.e., if there exists $\hat{x} \in R^n$ such that $f^i(\hat{x}, \theta^*) < 0$, $i \in \mathcal{P}$, then the model is stable at θ^*. There have been about 20 different regions of stability reported in the literature (see, e.g., [32]). Two of such regions are

$$M(\theta) = \{\theta : F(\theta^*) \subset F(\theta)\}$$

and

$$H(\theta) = \{\theta : F(\theta^*) \subset F_*^=(\theta)\},$$

where

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

We denote the limit inferior of a set by

$$\underline{\lim}_{\theta \to \theta^*} \Gamma(\theta) \overset{\text{def}}{=} \{x \in R^n : \text{There exists a sequence } x(\theta) \in \Gamma(\theta) \ni x = \lim_{\theta \to \theta^*} x(\theta).\}.$$

The following characterization for a region of stability was recently given in [26].

Lemma 3.4 *Consider a convex model (P, θ) around some θ^* with a realistic objective function. A set $S \subset R^p$, containing θ^*, is a region of stability if, and only if,*

$$F(\theta^*) \subset \underline{\lim}_{\substack{\theta \in S \\ \theta \to \theta^*}} F_*^=(\theta).$$

We will now introduce a slightly different notion of stability. This notion is motivated by the modified optimality conditions of the previous section.

Definition 3.2 *Consider the convex model (P, θ) with a realistic objective function at $\theta^* \in I$. Then $\tilde{S}(\theta^*) \subset I$ is a $\tilde{F}$-stability region at θ^* if, for each open set $\mathcal{A} \subset R^n$ satisfying*

$$\mathcal{A} \cap \tilde{F}(\theta^*) \neq \emptyset,$$

there exists a neighbourhood $N(\theta^*)$ *of* θ^* *such that*

$$\mathcal{A} \cap \tilde{F}(\theta) \neq \emptyset \text{ for any } \theta \in N(\theta^*) \cap \tilde{S}(\theta^*).$$

Note that a $\tilde{F}$-stability region $\tilde{S}(\theta^*)$ is a region attached to θ^* that guarantees lower semicontinuity of the point-to-set mapping $\tilde{F}$ (rather than F, as is usually the case). Let us denote some properties of $\tilde{F}$-stability.

Theorem 3.5 *Consider the convex model* (P, θ) *with a realistic objective function. If* $\tilde{F}$ *is open (lower semicontinuous) at* θ^*, *then* $\tilde{f}$ *is continuous at* θ^*.

Proof: Since f^0 is assumed to be realistic at θ^*, $\tilde{F}(\theta^*) \neq \emptyset$. Now consider $x^* \in \tilde{F}(\theta^*)$ and $\theta^k \in I$ such that $\theta^k \to \theta^*$. Then, by lower semicontinuity of $\tilde{F}$ at θ^*, there exists $\tilde{x}(\theta^k) \in \tilde{F}(\theta^k)$ such that

$$\tilde{x}(\theta^k) \to x^* \quad \text{as } \theta^k \to \theta^*.$$

But f^0 is continuous in both x and θ. Therefore

$$\lim_{\theta^k \to \theta^*} f^0(\tilde{x}(\theta^k), \theta^k) = f^0(x^*, \theta^*)$$

or

$$\lim_{\theta^k \to \theta^*} \tilde{f}(\theta^k) = \tilde{f}(\theta^*).$$

Corollary 3.3 *Consider the convex model* (P, θ) *with a realistic objective function. If* $\tilde{F}$ *is open at* θ^*, *then it is also closed at* θ^*, *and therefore it is continuous at* θ^*.

Proof: By definition,

$$\tilde{F}(\theta) = F(\theta) \cap \{x : f^0(x, \theta) = \tilde{f}(\theta)\}.$$

Now since the objective function, the constraints and the optimal value function (by Theorem 3.4) are all continuous at θ^*, the closedness of $\tilde{F}$ at θ^* easily follows.

□

Therefore, $\tilde{F}$-stability at θ^* implies continuity (both openness and closedness) of the point-to-set mapping $\tilde{F}$ at θ^* and the continuity of the real valued function $\tilde{f}$ at θ^*.

At this point, we extend the equality set $\mathcal{P}^=_\star$ to one for the model (P, θ). We define

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

Associated with this equality set, we define the following sets:

$$\begin{aligned}
\tilde{\mathcal{P}}^<(\theta) &= \mathcal{P} \setminus \tilde{\mathcal{P}}^=(\theta); \\
\mathcal{F}^=(\theta) &= \{x \in R^n : f^i(x, \theta) = 0, \; i \in \tilde{\mathcal{P}}^=(\theta)\}; \\
\mathcal{F}^=_\star(\theta) &= \{x \in R^n : f^i(x, \theta) \leq 0, \; i \in \tilde{\mathcal{P}}^=(\theta^*)\}; \\
\tilde{\mathcal{F}}^=_\star(\theta) &= \mathcal{F}^=_\star(\theta) \cap \{x \in R^n : f^0(x, \theta) = \tilde{f}(\theta)\}.
\end{aligned}$$

Similar to the characterization for a region of stability in Lemma 3.4, we have the following characterization for a $\tilde{F}$-stability region.

Theorem 3.6 *Consider the convex model (P, θ) with a realistic objective function at $\theta^\star$. A set $\tilde{S}(\theta^\star) \subset R^p$ is a $\tilde{F}$-stability region at $\theta^\star$ if, and only if,*

$$\tilde{F}(\theta^\star) \quad \subset \quad \underline{\lim}_{\substack{\theta \in \tilde{S}(\theta^\star) \\ \theta \to \theta^\star}} \tilde{\mathcal{F}}_\star^=(\theta). \tag{3.12}$$

Proof: Assume that $\tilde{S}(\theta^\star)$ is a $\tilde{F}$-stability region at $\theta^\star$. Then $\tilde{F}$ is lower semicontinuous at $\theta^\star$. This means that given any sequence $\theta \in \tilde{S}(\theta^\star)$, $\theta \to \theta^\star$, and any $\tilde{x}(\theta^\star) \in \tilde{F}(\theta^\star)$, there exists $\tilde{x}(\theta) \in \tilde{F}(\theta)$ such that

$$\tilde{x}(\theta^\star) = \lim_{\theta \to \theta^\star} \tilde{x}(\theta).$$

Since $\tilde{F}(\theta) \subset \tilde{\mathcal{F}}_\star^=(\theta)$, the inclusion (3.12) holds.

Conversely, assume that the inclusion (3.12) holds for some set $\tilde{S}(\theta^\star)$. Take any $x^\star \in \tilde{F}(\theta^\star)$. If $\tilde{\mathcal{P}}^<(\theta^\star) = \emptyset$, then $\tilde{\mathcal{F}}_\star^=(\theta) = \tilde{F}(\theta)$, which trivially implies that $\tilde{S}(\theta^\star)$ is a $\tilde{F}$-stability region at $\theta^\star$. If $\tilde{\mathcal{P}}^<(\theta^\star) \neq \emptyset$, then by Lemma 3.1, arbitrarily close (for some $\epsilon > 0$) to $x^\star$ there exist points $\tilde{y}_\epsilon \in \tilde{F}(\theta^\star)$ satisfying

$$\begin{aligned} f^0(\tilde{y}_\epsilon) &= \tilde{f}(\theta^\star) \\ f^k(\tilde{y}_\epsilon, \theta^\star) &< 0 \quad k \in \tilde{\mathcal{P}}^<(\theta^\star) \\ \|\tilde{y}_\epsilon - x^\star\| &< \epsilon. \end{aligned} \tag{3.13}$$

Now, since

$$\tilde{y}_\epsilon \in \tilde{F}(\theta^\star) \quad \subset \quad \underline{\lim}_{\substack{\theta \in \tilde{S}(\theta^\star) \\ \theta \to \theta^\star}} \tilde{\mathcal{F}}_\star^=(\theta),$$

there exists a sequence $\tilde{y}(\theta) \in \tilde{\mathcal{F}}_\star^=(\theta)$ such that

$$\tilde{y}_\epsilon = \lim_{\substack{\theta \in \tilde{S}(\theta^\star) \\ \theta \to \theta^\star}} \tilde{y}(\theta).$$

But, using (3.13), for all these $\theta's$, sufficiently close to $\theta^\star$, we have

$$f^k(\tilde{y}(\theta), \theta) < 0 \quad k \in \tilde{\mathcal{P}}^<(\theta^\star). \tag{3.14}$$

Now $\tilde{y}(\theta) \in \tilde{\mathcal{F}}_\star^=(\theta)$ and (3.14) imply that $\tilde{y}(\theta) \in \tilde{F}(\theta)$. This completes the proof since $\tilde{y}(\theta) \to \tilde{y}_\epsilon$ and $\tilde{y}_\epsilon$ is arbitrarily close to $x^\star$. Hence $\tilde{F}$ is open at $\theta^\star$ with respect to $\tilde{S}(\theta^\star)$.

Remark: The inclusion (3.12), like the inclusion in Lemma 3.4, is of theoretical rather than practical importance. $\tilde{F}$-stability regions can also be defined in a similar way that the twenty or so regions of stability have been defined. However, the obstacle in calculating $\tilde{F}$-stability regions is the calculation of the set of optimal solutions (and thus the optimal value) as a function of θ explicitly. As an illustration, consider

$$\tilde{M}(\theta^\star) = \{\theta : \tilde{F}(\theta^\star) \subset \tilde{F}(\theta)\},$$

$$\tilde{H}(\theta^\star) = \{\theta : \tilde{F}(\theta^\star) \subset \tilde{\mathcal{F}}_\star^=(\theta)\}.$$

Then it is easily seen that $\tilde{M}(\theta^\star)$ and $\tilde{H}(\theta^\star)$ imply the inclusion (3.12) and are, therefore, $\tilde{F}$-stability regions. On the other hand, for calculating them we need to know $\tilde{F}(\theta)$ and $\tilde{f}(\theta)$ explicitly. This problem makes the applications of the results related to $\tilde{F}$-stability rather limited. However, there is a condition under which the two different notions of stability coincide.

Lemma 3.5 *Consider the convex model (P,θ) around $\theta^\star$ with a realistic objective function at $\theta^\star$. Assume that $S(\theta^\star)$ is a region of stability at $\theta^\star$. Furthermore, assume that the point-to-set mapping $\tilde{F}$ is open at $\theta^\star$ with respect to $S(\theta^\star)$. Then $S(\theta^\star)$ is a $\tilde{F}$-stability region at $\theta^\star$.*

The proof of the lemma is obvious.

□

This property will make the results related to $\tilde{F}$-stability more applicable as we will show later.

Let us illustrate, with two examples, that $S(\theta^\star)$ and $\tilde{S}(\theta^\star)$ are generally two different regions.

Example 3.5 Consider

$$\begin{aligned}
\underset{(x)}{\text{Min}} \quad & f^0 = -x_1 \\
\text{s.t.} \quad & \\
& f^1 = x_1 + x_2 - 1 \leq 0 \\
& f^2 = -x_1 - \theta x_2 + 1 \leq 0 \\
& f^3 = -x_1 \leq 0 \\
& f^4 = -x_2 \leq 0
\end{aligned}$$

around $\theta^\star = 1$. Here

$$\begin{aligned}
\tilde{F}(\theta) &= \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\} \text{ for any } \theta \in R \\
\tilde{f}(\theta) &= -1 \text{ for any } \theta \in R
\end{aligned}$$

$$
\begin{aligned}
\mathcal{P}^{=}(\theta) &= \begin{cases} \{1,2\} & \text{if } \theta = 1 \\ \emptyset & \text{if } \theta > 1 \\ \{1,2,4\} & \text{if } \theta < 1 \end{cases} \\
\tilde{\mathcal{P}}^{=}(\theta) &= \{1,2,4\} \text{ for any } \theta \in R \\
S(\theta^\star) &= M(\theta^\star) = \{\theta : \theta \geq 1\} \\
\tilde{S}(\theta^\star) &= \tilde{M}(\theta^\star) = \{\theta : \theta \in R\}.
\end{aligned}
$$

□

Note that in the above example $S(\theta^\star) \subset \tilde{S}(\theta^\star)$. This is not, however, the case in the following example.

Example 3.6 Consider

$$
\begin{aligned}
&\underset{(x)}{\text{Min}} \quad f^0 = -x_1 - \theta x_2 \\
&\text{s.t.} \\
&\qquad f^1 = x_1 + x_2 - 1 \leq 0 \\
&\qquad f^2 = -x_1 \leq 0 \\
&\qquad f^3 = -x_2 \leq 0
\end{aligned}
$$

around $\theta^\star = 1$. Here

$$
\begin{aligned}
\tilde{F}(\theta) &= \begin{cases} \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{if } \theta < 1 \\ \lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (1-\lambda) \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \text{if } \theta = 1 \text{ and } 0 \leq \lambda \leq 1 \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \text{if } \theta > 1 \end{cases} \\
\tilde{f}(\theta) &= \begin{cases} -1 & \text{if } \theta \leq 1 \\ -\theta & \text{if } \theta > 1 \end{cases} \\
\tilde{\mathcal{P}}^{=}(\theta) &= \begin{cases} \{1,3\} & \text{if } \theta < 1 \\ \{1\} & \text{if } \theta = 1 \\ \{1,2\} & \text{if } \theta > 1 \end{cases} \\
\mathcal{P}^{=}(\theta) &= \emptyset \text{ for any } \theta \in R \\
S(\theta^\star) &= \{\theta : \theta \in R\} \\
\tilde{S}(\theta^\star) &= \{\theta^\star\}.
\end{aligned}
$$

In this example, unlike the previous example, $\tilde{S}(\theta^\star) \subset S(\theta^\star)$.

□

For the sake of completeness, let us also recall another notion of stability, defined in [24], known as *weak-stability*. Recall that the limit superior of a set is

$$
\limsup_{\theta \to \theta^\star} G(\theta) = \{x : x(\theta') \to x,\ \{\theta'\} \subset \{\theta\},\ x(\theta') \in G(\theta')\}.
$$

Definition 3.3 *A region S is called a weak-stability region, for the parametric program (P, θ) at $\theta^* \in S$, if for every sequence $\{\theta\} \subset S$ converging to θ^* it follows that*

$$\lim_{\theta \to \theta^*} \tilde{f}(\theta) = \tilde{f}(\theta^*)$$

and that

$$\emptyset \neq \limsup_{\theta \to \theta^*} \tilde{F}(\theta) \subset \tilde{F}(\theta^*).$$

For a convex model with a realistic objective function at θ^*, a stability region at θ^* is also a weak-stability region at θ^*. We will show that for a convex model with a realistic objective function, a $\tilde{F}$-stability region at θ^* is also a weak-stability region at θ^*.

Lemma 3.6 *Consider the convex model (P, θ) with a realistic objective function at θ^*. Assume that $\tilde{S}(\theta^*)$ is a $\tilde{F}$-stability region at θ^*. Then $\tilde{S}(\theta^*)$ is also a weak stabilty region at θ^*.*

Proof: By Theorem 3.5, lower semicontinuity of $\tilde{F}$ at θ^* with respect to $\tilde{S}(\theta^*)$ implies continuity of $\tilde{f}$ at θ^*. Furthermore, closedness of $\tilde{F}$ follows from Corollary 3.3 . This completes the proof.

□

Hence, for a convex model with a realistic objective function at θ^* we have the following implications:

Stability at θ^* $\Rightarrow$ weak-stability at θ^*;
$\tilde{F}$-stability at θ^* $\Rightarrow$ weak-stability at θ^*;
Weak-stability at θ^* and lower semicontinuity of $\tilde{F}$ at θ^* $\Rightarrow$ $\tilde{F}$-stability at θ^*.

Note that in the rest of the thesis, whenever we refer to stability, we mean stability in the usual sense (i.e., continuity of the point-to-set mapping F), while $\tilde{F}$-stability and weak-stability will be explicitly mentioned.

3.3 Optimal Input

We first recall some well-known results from input optimization and then give a characterization of optimal inputs with respect to $\tilde{F}$-stability regions. Denote the *minimal index set of active constraints* by

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

its complementary set by

$$\mathcal{P}^{<}(\theta) = \mathcal{P} \setminus \mathcal{P}^{=}(\theta)$$

and the corresponding set by

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

We recall (see [32]) the following important necessary condition for stability.

Theorem 3.7 *Consider the convex model (P, θ) at some $\theta = \theta^\star \in I$. Let S be a region of stability at $\theta^\star$. Then there exists a neighbourhood $N(\theta^\star)$ such that*

$$\mathcal{P}^{=}(\theta) \subset \mathcal{P}^{=}(\theta^\star) \tag{3.15}$$

for all $\theta \in N(\theta^\star) \cap S$.

This is, in particular, important for the characterization of an *optimal input*. We recall the following definition from [32].

Definition 3.4 *Consider the convex model (P, θ) at some $\theta^\star \in I$. Let S be a region of stability at $\theta^\star$. If $\tilde{f}(\theta^\star) \leq \tilde{f}(\theta)$ for every $\theta \in N(\theta^\star) \cap S$, where $N(\theta^\star)$ is a neighbourhood of $\theta^\star$, then $\theta^\star$ is a locally optimal input with respect to S and $(P, \theta^\star)$ is the corresponding locally optimal realization of the mathematical model.*

To simplify the notation, let $q(\theta) = card\mathcal{P}^{<}(\theta)$. The characterization of an optimal input is stated in terms of the "restricted" Lagrangian defined for θ, around a fixed candidate of optimality $\theta^\star \in I$, by

$$\mathcal{L}_\star^{<}(x, \lambda; \theta) = f^0(x, \theta) + \sum_{i \in \mathcal{P}^{<}(\theta^\star)} \lambda_i f^i(x, \theta).$$

It has been formulated in [25].

Theorem 3.8 *Consider the convex model (P, θ) with a realistic objective function at some $\theta^\star \in I$. Let $\tilde{x}(\theta^\star)$ be a corresponding optimal solution and let S be an arbitrary region of stability at $\theta^\star$. Then $\theta^\star$ is a locally optimal input with respect to S if, and only if, there exists a neighbourhood $N(\theta^\star)$ of $\theta^\star$ and a nonnegative vector function*

$$\Lambda : N(\theta^\star) \cap S \longrightarrow R_+^{q(\theta^\star)}$$

such that, whenever $\theta \in N(\theta^\star) \cap S$,

$$\mathcal{L}_\star^{<}(\tilde{x}(\theta^\star), \lambda; \theta^\star) \leq \mathcal{L}_\star^{<}(\tilde{x}(\theta^\star), \Lambda(\theta^\star); \theta^\star) \leq \mathcal{L}_\star^{<}(x, \Lambda(\theta); \theta) \tag{3.16}$$

for every $\lambda \in R_+^{q(\theta^\star)}$ and every $x \in F_\star^{=}(\theta)$.

We will now show that the same characterization for an optimal input holds with respect to a $\tilde{F}$-stability region. First we show that the necessary condition for stability extends to $\tilde{F}$-stability.

Theorem 3.9 *Consider the convex model (P, θ) at θ^* with a realistic objective function. Let $\tilde{x}(\theta^*)$ be a corresponding optimal solution at θ^* and $\tilde{S}(\theta^*)$ a $\tilde{F}$-stability region at θ^*. Then there exists a neighbourhood $N(\theta^*)$ of θ^* such that*

$$\tilde{\mathcal{P}}^{=}(\theta) \subset \tilde{\mathcal{P}}^{=}(\theta^*) \tag{3.17}$$

for every $\theta \in N(\theta^*) \cap \tilde{S}(\theta^*)$.

Proof: Assume that such $N(\theta^*)$ does not exist. Then there exists a sequence $\theta^k \in \tilde{S}(\theta^*)$, $\theta^k \to \theta^*$, and an index $j_0 \in \tilde{\mathcal{P}}^{<}(\theta^*)$ such that

$$j_0 \in \tilde{\mathcal{P}}^{=}(\theta^k) \cap \tilde{\mathcal{P}}^{<}(\theta^*) \neq \emptyset$$

for infinitely many k's. Now, by definition of $\tilde{\mathcal{P}}^{<}(\theta^*)$, there exists $x^* \in \tilde{F}(\theta^*)$ such that

$$f^j(x^*, \theta^*) < 0, \;\; j \in \tilde{\mathcal{P}}^{<}(\theta^*).$$

Since $j_0 \in \tilde{\mathcal{P}}^{<}(\theta^*)$, it follows that

$$f^{j_0}(x^*, \theta^*) < 0. \tag{3.18}$$

Besides, since $j_0 \in \tilde{\mathcal{P}}^{=}(\theta^k)$ for the above sequence $\{\theta^k\}$,

$$f^{j_0}(x, \theta^k) = 0 \quad \text{for every } x \in \mathcal{F}^{=}(\theta^k).$$

But $\tilde{F}(\theta^k) \subset \mathcal{F}^{=}(\theta^k)$. Therefore

$$f^{j_0}(x, \theta^k) = 0 \quad \text{for every } x \in \tilde{F}(\theta^k). \tag{3.19}$$

On the other hand, since $\theta^k \in \tilde{S}(\theta^*)$, there exists a sequence $\tilde{x}^k \in \tilde{F}(\theta^k)$ such that $\tilde{x}^k \to x^*$ as $\theta^k \to \theta^*$. This means that

$$f^{j_0}(\tilde{x}^k, \theta^k) < 0$$

for all k's sufficiently large, by (3.18) and continuity of f^{j_0}. This contradicts (3.19). □

The inclusion (3.17) may not hold on a region of stability or even on a weak-stability region, as the following example shows.

Example 3.7 Consider again the model

$$\begin{aligned}
\underset{(x)}{\text{Min}} \quad & f^0 = -x_1 - \theta x_2 \\
\text{s.t.} \quad & \\
& f^1 = x_1 + x_2 - 1 \leq 0 \\
& f^2 = -x_1 \leq 0 \\
& f^3 = -x_2 \leq 0
\end{aligned}$$

around $\theta^\star = 1$, with

$$\tilde{\mathcal{P}}^{=}(\theta) = \begin{cases} \{1,3\} & \text{if } \theta < 1 \\ \{1\} & \text{if } \theta = 1 \\ \{1,2\} & \text{if } \theta > 1. \end{cases}$$

The model is stable for all $\theta \in R$. It is also weakly stable at $\theta^\star$, yet the inclusion (3.17) is not satisfied.

□

However, we have the following result.

Corollary 3.4 *Consider the convex model (P,θ) at $\theta^\star$ with a realistic objective function. Let $\tilde{x}(\theta^\star)$ be a corresponding optimal solution at $\theta^\star$ and $S(\theta^\star)$ a region of stability at $\theta^\star$. Assume that $\tilde{F}$ is open at $\theta^\star$ with respect to $S(\theta^\star)$. Then there exists a neighbourhood $N(\theta^\star)$ of $\theta^\star$ such that*

$$\tilde{\mathcal{P}}^{=}(\theta) \subset \tilde{\mathcal{P}}^{=}(\theta^\star) \tag{3.20}$$

for every $\theta \in N(\theta^\star) \cap S(\theta^\star)$.

The index set $\mathcal{P}^{=}$ is thought of (see [32]) as a measure of how strongly the constraints are tied up in the model. Theorem 3.7 has a simple economic interpretation: *Stable economic systems necessarily unfold towards less restricted states*, i.e., towards more "freedom". Similarly, we can think of the index set $\tilde{\mathcal{P}}^{=}(\theta)$ as a measure of how strongly the constraints and the optimal value function are tied up in the model. The inclusion in Theorem 3.9 thus means that the set of optimal solutions tends towards more "interiority" (more "freedom").

At this stage, we are ready to give a characterization that is both necessary and sufficient for an input $\theta^\star$ to locally minimize $\tilde{f}(\theta)$ with respect to a $\tilde{F}$-stability region $\tilde{S}(\theta^\star)$. We will refer to such $\theta^\star$ as a locally optimal input with respect to $\tilde{S}(\theta^\star)$. The characterization will be stated in terms of the restricted Lagrangian for θ, around a fixed candidate for optimality $\theta^\star$, by

$$\tilde{\mathcal{L}}^{<}_{\star}(x,\lambda;\theta) = f^0(x,\theta) + \sum_{i \in \tilde{\mathcal{P}}^{<}(\theta^\star)} \lambda_i f^i(x,\theta).$$

Also, let $\tilde{q}(\theta^\star) = card\tilde{\mathcal{P}}^{<}(\theta^\star)$. Then we have the following modification of Theorem 3.8.

Theorem 3.10 *Consider the convex model (P,θ) with a realistic objective function at some $\theta^\star \in I$. Let $\tilde{x}(\theta^\star)$ be any corresponding optimal solution, and let $\tilde{S}(\theta^\star)$ be a $\tilde{F}$-stability region at $\theta^\star$. Then $\theta^\star$ is a locally optimal input with respect to $\tilde{S}(\theta^\star)$ if, and only if, there exists a neighbourhood $N(\theta^\star)$ of $\theta^\star$ and a nonnegative vector function*

$$\tilde{\lambda} : N(\theta^\star) \cap \tilde{S}(\theta^\star) \to R_+^{\tilde{q}(\theta^\star)}$$

such that, whenever $\theta \in N(\theta^*) \cap \tilde{S}(\theta^*)$,

$$\tilde{\mathcal{L}}_*^<(\tilde{x}(\theta^*), \lambda; \theta^*) \le \tilde{\mathcal{L}}_*^<(\tilde{x}(\theta^*), \tilde{\lambda}(\theta^*); \theta^*) \le \tilde{\mathcal{L}}_*^<(x, \tilde{\lambda}; \theta) \tag{3.21}$$

for every $\lambda \in R_+^{\tilde{q}(\theta^*)}$ and every $x \in \mathcal{F}_*^=(\theta)$.

Proof: (Necessity:) Without loss of generality, we assume that the first $\tilde{q}(\theta^*)$ indices of $\mathcal{P}$ are precisely the set $\tilde{\mathcal{P}}^<(\theta^*)$. Then for every $\theta \in N(\theta^*) \cap \tilde{S}(\theta^*)$, where θ^* is a locally optimal input with respect to $\tilde{S}(\theta^*)$, we construct the following two sets in $R^{\tilde{q}(\theta^*)+1}$:

$$K_1(\theta) = \left\{ y : y \ge \begin{bmatrix} f^0(x,\theta) \\ f^1(x,\theta) \\ \vdots \\ f^{\tilde{q}(\theta^*)}(x,\theta) \end{bmatrix} \text{ for at least one } x \in \mathcal{F}_*^= \right\}$$

and

$$K_2 = \left\{ y : y < \begin{bmatrix} \tilde{f}(\theta^*) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}.$$

(The ordering of the vectors is given componentwise.) Since the set $\mathcal{F}_*^=(\theta)$ is convex in R^n so is $K_1(\theta)$. Convexity of K_2 is obvious. Moreover,

$$K_1(\theta) \cap K_2 = \emptyset.$$

Otherwise, there would exist sequences $\theta^k \in \tilde{S}(\theta^*)$, $\theta^k \to \theta^*$ and $x^k \in \mathcal{F}_*^=(\theta^k)$ such that

$$f^0(x^k, \theta^k) < \tilde{f}(\theta^*), \quad f^i(x^k, \theta^k) < 0, \quad i \in \tilde{\mathcal{P}}^<(\theta^*),$$

violating optimality of θ^*. Therefore, the two sets can be separated, i.e., there exists a nonzero vector $a = a(\theta)$ and also a scalar $\alpha = \alpha(\theta)$ such that

$$a^T y^1 \ge \alpha \ge a^T y^2$$

for all $y^1 \in K_1$ and all $y^2 \in clK_2$. Clearly $a \ge 0$. Specifying, for each $x \in \mathcal{F}_*^=(\theta)$,

$$y^1 = \begin{bmatrix} f^0(x,\theta) \\ f^1(x,\theta) \\ \vdots \\ f^{\tilde{q}(\theta^*)}(x,\theta) \end{bmatrix} \quad \text{and} \quad y^2 = \begin{bmatrix} \tilde{f}(\theta^*) \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

we get

$$a_0 \tilde{f}(\theta^*) \le a_0 f^0(x,\theta) + \sum_{i \in \tilde{\mathcal{P}}^<(\theta^*)} a_i f^i(x,\theta). \tag{3.22}$$

The leading coefficient a_0 must be strictly positive. Otherwise, we would have

$$\sum_{i\in\tilde{\mathcal{P}}^{<}(\theta^*)} a_i f^i(x,\theta) \geq 0 \tag{3.23}$$

for every $x \in \mathcal{F}_\star^{=}(\theta)$. But then, by Lemma 3.1, there would exist

$$\hat{x} \in \hat{F}(\theta) \subset \mathcal{F}_\star^{=}(\theta)$$

such that

$$f^i(\hat{x}) < 0, \quad i \in \hat{\mathcal{P}}^{<}(\theta).$$

This would further imply that

$$f^i(\hat{x}) < 0, \quad i \in \hat{\mathcal{P}}^{<}(\theta^*),$$

since $\tilde{\mathcal{P}}^{<}(\theta^*) \subset \tilde{\mathcal{P}}^{<}(\theta)$ for every $\theta \in \tilde{S}(\theta^*)$ close enough to θ^*, by Theorem 3.9. On the other hand, $a_i \geq 0$, $i \in \hat{\mathcal{P}}^{<}(\theta^*)$, and not all zero. Therefore, we would have

$$\sum_{i\in\tilde{\mathcal{P}}^{<}(\theta^*)} a_i f^i(x,\theta) < 0,$$

a contradiction to (3.23). Dividing (3.22) by a_0, and introducing the notation

$$\tilde{\lambda}_i, \quad i \in \tilde{\mathcal{P}}^{<}(\theta^*),$$

we obtain

$$\tilde{f}(\theta^*) \leq \tilde{\mathcal{L}}_\star^{<}(x,\tilde{\lambda};\theta). \tag{3.24}$$

But, as is expected,

$$\sum_{i\in\tilde{\mathcal{P}}^{<}(\theta^*)} \tilde{\lambda}_i(\theta^*) f^i(\tilde{x}(\theta^*),\theta^*) = 0. \tag{3.25}$$

establishing the right-hand side inequality in (3.21). The left-hand side inequality is an immediate consequence of the fact that $\tilde{x}(\theta^*) \in F(\theta^*)$ and of the nonnegativeness of $\tilde{\lambda}_i$, $i \in \tilde{\mathcal{P}}^{<}(\theta^*)$.

(Sufficiency:) Assume that the saddle-point inequalities hold as required. After setting $\lambda = 0$, we obtain

$$\sum_{i\in\tilde{\mathcal{P}}^{<}(\theta^*)} \tilde{\lambda}_i(\theta^*) f^i(\tilde{x}(\theta^*),\theta^*) \geq 0.$$

But the reverse sign is also true, since $\tilde{x}(\theta^*) \in F(\theta^*)$ and $\tilde{(\lambda)}_i(\theta^*) \geq 0$, $i \in \tilde{\mathcal{P}}^{<}(\theta^*)$. This gives (3.25) and further (3.24) for every $x \in \mathcal{F}_\star^{=}(\theta)$. Hence (3.24) holds for every $x \in \tilde{F}(\theta) \subset \mathcal{F}_\star^{=}(\theta)$. In particular, for $\tilde{x}(\theta) \in \tilde{F}(\theta)$, (3.24) gives

$$\begin{aligned}\tilde{f}(\theta^*) &\leq \tilde{f}(\theta) + \sum_{i\in\tilde{\mathcal{P}}^{<}(\theta^*)} \tilde{\lambda}_i(\theta) f^i(\tilde{x}(\theta),\theta) \\ &\leq \tilde{f}(\theta) \quad \text{for all } \theta \in N(\theta^*) \cap \tilde{S}(\theta^*).\end{aligned}$$

This completes the proof.

□

In the claim of Theorem 3.10 one can replace $\mathcal{F}_\star^=$ by $\tilde{\mathcal{F}}_\star^=$. The result is then in the following form.

Corollary 3.5 *Consider the convex model (P,θ) with a realistic objective function at some $\theta^\star \in I$. Let $\tilde{x}(\theta^\star)$ be any corresponding optimal solution and let $\tilde{S}(\theta^\star)$ be a $\tilde{F}$-stability region at $\theta^\star$. Then $\theta^\star$ is a locally optimal input with respect to $\tilde{S}(\theta^\star)$ if, and only if, there exists a neighbourhood $(N(\theta^\star)$ of $\theta^\star$ and a nonnegative vector function*

$$\tilde{\lambda} : N(\theta^\star) \cap \tilde{S}(\theta^\star) \rightarrow R_+^{\tilde{q}(\theta^\star)}$$

such that, whenever $\theta \in N(\theta^\star) \cap \tilde{S}(\theta^\star)$,

$$\tilde{\mathcal{L}}_\star^<(\tilde{x}(\theta^\star), \lambda; \theta^\star) \leq \tilde{\mathcal{L}}_\star^<(\tilde{x}(\theta^\star), \tilde{\lambda}(\theta^\star); \theta^\star) \leq \tilde{\mathcal{L}}_\star^<(x, \tilde{\lambda}; \theta) \tag{3.26}$$

for every $\lambda \in R_+^{\tilde{q}(\theta^\star)}$ and every $x \in \tilde{\mathcal{F}}_\star^=(\theta)$.

Proof: (Necessity:) Let $\theta^\star \in I$ be a locally optimal input with respect to $\tilde{S}(\theta^\star)$. Then, by Theorem 3.10, there exist $N(\theta^\star)$ and $\tilde{\lambda} \geq 0$ such that (3.21) holds on the set $\mathcal{F}_\star^=$, for every $\lambda \in R_+^{\tilde{q}(\theta^\star)}$. Note that

$$\tilde{\mathcal{F}}_\star^= \subset \mathcal{F}_\star^=.$$

Therefore, by choosing the same $\tilde{\lambda}$, the inequality (3.26) holds for every $\lambda \in R_+^{\tilde{q}(\theta^\star)}$. (Sufficiency:) Since $\tilde{F}(\theta) \subset \tilde{\mathcal{F}}_\star^=(\theta)$, this part of the proof is exactly the same as the sufficiency part of Theorem 3.10, with $\mathcal{F}_\star^=$ replaced by $\tilde{\mathcal{F}}_\star^=$.

□

We now illustrate the above two characterizations of optimal inputs with respect to $\tilde{F}$-stability regions by an example.

Example 3.8 Consider the model

$$\begin{aligned} \underset{(x)}{\text{Min}} \quad & f^0 = \frac{-x_1}{(\theta-1)^2+1} \\ \text{s.t.} \quad & \\ & f^1 = x_1 + x_2 - 1 \leq 0 \\ & f^2 = -x_1 - \theta x_2 + 1 \leq 0 \\ & f^3 = -x_1 \leq 0 \\ & f^4 = -x_2 \leq 0 \end{aligned}$$

around $\theta^\star = 1$. Then

$$\tilde{F}(\theta) = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} \text{ for any } \theta \in R$$

$$\begin{aligned}
\dot{f}(\theta) &= \frac{-1}{(\theta-1)^2+1} \quad \text{for any } \theta \in R \\
\mathcal{P}^{=}(\theta) &= \begin{cases} \{1,2\} & \text{if } \theta = 1 \\ \emptyset & \text{if } \theta > 1 \\ \{1,2,4\} & \text{if } \theta < 1 \end{cases} \\
\tilde{\mathcal{P}}^{=}(\theta) &= \{1,2,4\} \text{ for any } \theta \in R \\
S(\theta^\star) &= M(\theta^\star) = \{\theta : \theta \geq 1\} \\
\tilde{S}(\theta^\star) &= \tilde{M}(\theta^\star) = \{\theta : \theta \in R\}.
\end{aligned}$$

Furthermore

$$\mathcal{F}_\star^{=}(\theta) = \left\{ x : \begin{array}{rl} x_1 + x_2 - 1 & \leq 0 \\ -x_1 - \theta x_2 + 1 & \leq 0 \\ -x_2 & \leq 0 \end{array} \right\}.$$

The saddle-point inequality from Theorem 3.10 reduces to

$$-1 - \lambda_3 \leq -1 \leq \frac{-x_1}{(\theta-1)^2+1} + \tilde{\lambda}_3(\theta)(-x_1) \quad \text{for every } x \in \mathcal{F}_\star^{=}(\theta),$$

for some $\tilde{\lambda}_3(\theta) \geq 0$ and for all $\lambda_3 \geq 0$. The left-hand inequality is easily satisfied. With the choice $\tilde{\lambda}_3 = 0$, the right-hand inequality reduces to

$$x_1 \leq (\theta - 1)^2 + 1 \text{ for all } x_1 \in \mathcal{F}_\star^{=}(\theta),$$

which holds for all $\theta \in R$. Therefore, $\theta^\star = 1$ is a locally optimal input with respect to the $\tilde{F}$-stability region $\tilde{S}(\theta^\star) = R$. Besides, $\theta^\star = 1$ is a locally optimal input with respect to the region of stability $S(\theta^\star) = \{\theta : \theta \geq 1\}$. We can also use Corollary 3.5, in which case we have $\tilde{\mathcal{F}}_\star^{=}(\theta) = \{(1,0)^T\}$ for all $\theta \in R$. This implies that the inequality (3.26) trivially holds.

□

If the input θ does not vary, but is held fixed, then the convex model (P, θ) becomes the mathematical program (P) and Theorem 3.10 recovers the characterization in Theorem 3.2 .

As mentioned earlier, the difficulty in using the $\tilde{F}$-stability regions and the related results such as Theorem 3.10 or Corollary 3.5 is the calculation of the set of optimal solutions explicitly as a function of θ. In what follows we will study a special case for which we can use these results without worrying about the calculation of the set of optimal solutions as a function of θ. This is the case where the stability region at $\theta^\star$ and the $\tilde{F}$-stability region at $\theta^\star$ coincide, i.e., we have usual stability and in addition lower semicontinuity of the point-to-set mapping $\tilde{F}$.

Corollary 3.6 *Consider the convex model (P, θ) with a realistic objective function at some $\theta^\star \in I$. Let $\tilde{x}(\theta^\star)$ be any corresponding optimal solution, and let $S(\theta^\star)$ be* *a region of stability at θ^*. Assume that the mapping $\tilde{F}$ is open at θ^* with respect to $S(\theta^*)$. Then θ^* is a locally optimal input with respect to $S(\theta^*)$ if, and only if, there exists a neighbourhood $N(\theta^*)$ of θ^* and a nonnegative vector function*

$$\tilde{\lambda} : N(\theta^*) \cap S(\theta^*) \rightarrow R_+^{\bar{q}(\theta^*)}$$

such that, whenever $\theta \in N(\theta^*) \cap S(\theta^*)$,

$$\tilde{\mathcal{L}}_*^<(\tilde{x}(\theta^*), \lambda; \theta^*) \leq \tilde{\mathcal{L}}_*^<(\tilde{x}(\theta^*), \tilde{\lambda}(\theta^*); \theta^*) \leq \tilde{\mathcal{L}}_*^<(x, \tilde{\lambda}; \theta) \tag{3.27}$$

for every $\lambda \in R_+^{\bar{q}(\theta^*)}$ and every $x \in \mathcal{F}_*^=(\theta)$.

Note that the above characterization is different from the ones given in Theorems 3.8 and 3.10. In order to make Corollary 3.6 more applicable, it will be useful to find some sufficient conditions for lower semicontinuity of the map $\tilde{F}$ at θ^* for a stable model at θ^*. We give one such condition in the following theorem.

Theorem 3.11 *Consider a convex model (P, θ) with a realistic objective function at θ^*. Assume that the mapping F is lower semicontinuous at θ^* and that the optimal solution $\tilde{x}(\theta^*)$ at θ^* is unique. Then the mapping $\tilde{F}$ is lower semicontinuous at θ^*.*

Proof: Since the mapping F is lower semicontinuous at θ^* and f^0 is realistic at θ^*, by Theorem 3.4, there exists a neighbourhood $N(\theta^*)$ of θ^* such that $\tilde{F}(\theta) \neq \emptyset$ for every $\theta \in N(\theta^*)$. Furthermore, $\theta \in N(\theta^*)$, $\theta \rightarrow \theta^*$, implies that the sequence $\tilde{x}(\theta)$ is bounded and all its limit points lie in $\tilde{F}(\theta^*)$.

Assume now that the mapping $\tilde{F}$ is not lower semicontinuous at θ^*. Then there exist a sequence $\theta^k \longrightarrow \theta^*$ and $\delta > 0$ such that

$$\|\tilde{x}(\theta^k) - \tilde{x}(\theta^*)\| > \delta \quad \text{for every } \tilde{x}(\theta^k) \in \tilde{F}(\theta^k). \tag{3.28}$$

Now for such $\{\theta^k\}$, consider the set

$$K \triangleq cl(\bigcup_{\theta^k \in N(\theta^*)} \tilde{F}(\theta^k)).$$

Then K is nonempty and bounded, since each $\tilde{F}(\theta^k)$ is, and is closed by the way it is defined. Therefore, K is compact and so every sequence in K has a convergent subsequence. Consider the sequence

$$\tilde{x}(\theta^k) \in \tilde{F}(\theta^k), \quad \theta^k \in N(\theta^*), \; \theta^k \longrightarrow \theta^*.$$

Then, by the compactness of K, the sequence $\tilde{x}(\theta^k)$ has a convergent subsequence, i.e., there exists a subsequence of θ^k, say $\theta^{k(l)}$, such that $\theta^{k(l)} \rightarrow \theta^*$ implies that

$\tilde{x}(\theta^{k(l)}) \rightarrow x^*$. But we know from stability of the model at θ^* that $x^* \in \tilde{F}(\theta^*)$. Since $\tilde{F}(\theta^*)$ is a singleton, we must have $x^* = \tilde{x}(\theta^*)$. This means that there exists a subsequence of the sequence $\tilde{x}(\theta^k)$ such that

$$\tilde{x}(\theta^{k(l)}) \rightarrow \tilde{x}(\theta^*) \quad \text{as } l \rightarrow \infty.$$

This contradicts (3.28).

□

The following result trivially follows.

Corollary 3.7 *Consider the convex model* (P,θ) *with a realistic objective function at* θ^*. *Assume that* f^0 *is strictly convex in* x *at* θ^* *and that the mapping* F *is lower semicontinuous at* θ^*. *Then the mapping* $\tilde{F}$ *is lower semicontinuous at* θ^*.

We will now apply some of the above results to the following control problem adapted from [27] known as Zermelo's problem.

Example 3.9 Determine the steering angle θ that will minimize the time x required to go from the origin to the target

$$T = \{(y_1, y_2) : (y_1 - 5)^2 + (y_2 - 1)^2 \leq 1\},$$

for Zermelo's problem with stream speed $V = 2$. The system dynamics in this case are given by

$$\begin{aligned} \dot{y}_1 &= 2 + \cos\theta \\ \dot{y}_2 &= \sin\theta. \end{aligned}$$

Under a constant control θ, these differential equations are easily integrated and evaluated at time x and the result may be used to define the target in terms of x:

$$(2x + x\cos\theta - 5)^2 + (x\sin\theta - 1)^2 \leq 1. \tag{3.29}$$

The objective is to find a control θ that will minimize x subject to (3.29). We will solve this problem using input optimization. The mathematical model associated with the problem is

$$(P,\theta) \qquad \begin{aligned} &\underset{(x)}{\text{Min}} \quad f^0 = x \\ &\text{s.t.} \\ &\qquad f^1 = (2x + x\cos\theta - 5)^2 + (x\sin\theta - 1)^2 - 1 \leq 0. \end{aligned}$$

The graph of $f^1(x, \theta)$ is shown in Figure 3.1.

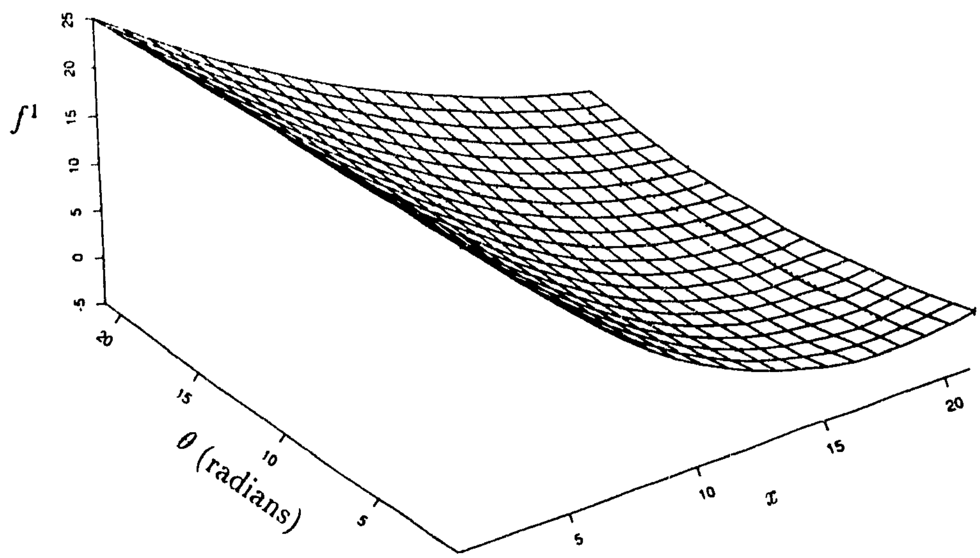

Figure 3.1: Graph of $f^1(x, \theta)$.

The set $F(\theta)$ after some algebraic manipulations reduces to the set of x that satisfy

$$x^2(5 + 4\cos\theta) + x(-20 - 10\cos\theta - 2\sin\theta) + 25 \leq 0. \tag{3.30}$$

This further implies that

$$F(\theta) = \{x \in R : a_1(\theta) \leq x \leq a_2(\theta)\},$$

where

$$a_1(\theta) = \frac{20 + 10\cos\theta + 2\sin\theta - \sqrt{\Delta(\theta)}}{10 + 8\cos\theta}$$

and

$$a_2(\theta) = \frac{20 + 10\cos\theta + 2\sin\theta + \sqrt{\Delta(\theta)}}{10 + 8\cos\theta}$$

are the roots of the quadratic in (3.30) and

$$\Delta(\theta) = 8\sin\theta(-12\sin\theta + 5\cos\theta + 10)$$

is its discriminant. It can be seen that for all $0 \leq \theta \leq 72.905°$, $\Delta(\theta)$ is nonnegative.

A graphic illustration of the set $F(\theta)$ is given in Figure 3.2

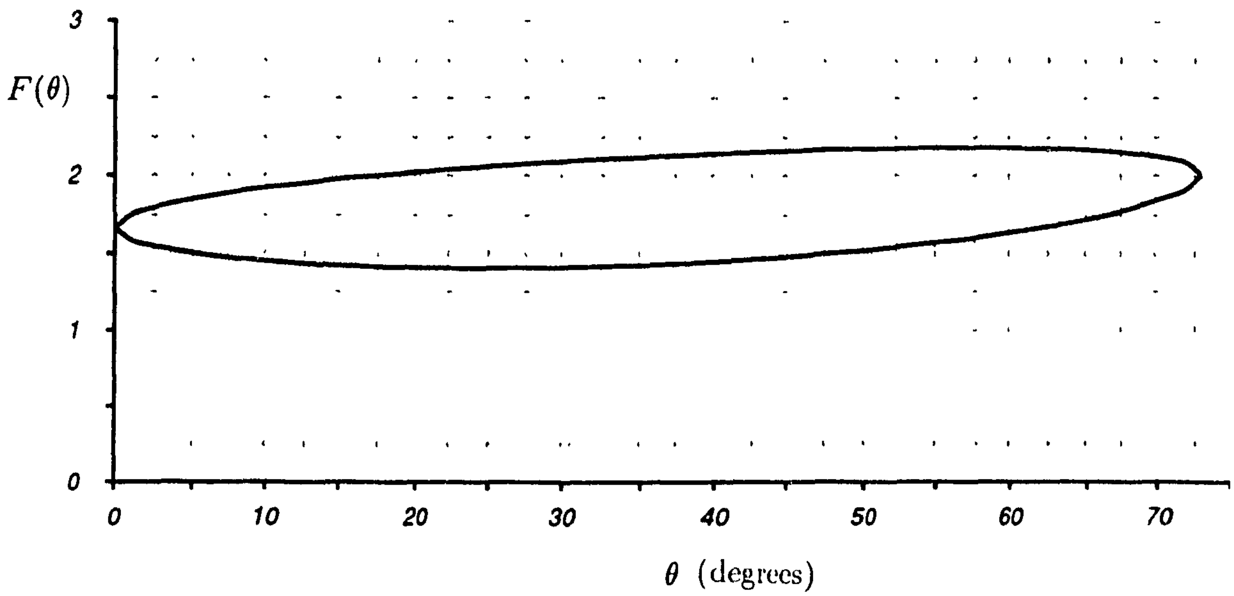

Figure 3.2: Graph of $F(\theta)$.

Let us denote by $\mathcal{F}$ the set of θ's for which the set $F(\theta)$ is nonempty. Then

$$\mathcal{F} = \{\theta : 0 \le \theta \le 72.905^\circ\}.$$

Thus, it follows that

$$\tilde{F}(\theta) = \{a_1(\theta)\}$$

and

$$\tilde{f}(\theta) = a_1(\theta) \text{ for all } \theta \in \mathcal{F}.$$

Since as a function of θ, this is a one dimensional problem we can directly use the Method of Golden Rule to find a candidate for an optimal input for this model. (The Method of Golden Rule is described in Section 3.5). Using this Method, after about 20 iterations, we get $\theta^\star \approx 24.56^\circ$ and $x^\star = \tilde{x}(\theta^\star) \approx 1.406$. This $\theta^\star$ is a candidate for an optimal input for the model (P,θ). Note that for all $\theta \in \mathcal{F} \setminus \{0, 72.905^\circ\}$, Slater's condition is satisfied in (P,θ). Therefore, (P,θ) is stable for all those perturbations of θ. To verify that this $\theta^\star$ is a locally optimal input, applying Theorem 3.8, we must find $\Lambda(\theta)$ such that the saddle-point inequality (3.16) holds in some neighbourhood of $\theta^\star$. Since Slater's condition is satisfied at $\theta^\star$,

$$\mathcal{L}(x,\lambda,\theta) = x + \lambda[(2x + x\cos\theta - 5)^2 + (x\sin\theta - 1)^2 - 1]$$

Furthermore
$$\mathcal{L}(\tilde{x}(\theta^\star), \Lambda(\theta^\star), \theta^\star) = x^\star \approx 1.406.$$

Proving the left-hand inequality is trivial. To prove the right-hand inequality, we must find $\lambda(\theta) \geq 0$ and $N(\theta^\star)$, a neighbourhood of $\theta^\star$, such that the inequality

$$\begin{aligned} 0 \leq x^2[\lambda(\theta)(5 + 4\cos\theta)] \;&+\; x[1 + \lambda(\theta)(-20 - 10\cos\theta - 2\sin\theta)] \\ &+\; [25\lambda(\theta) - x^\star] \end{aligned} \tag{3.31}$$

holds for every $\theta \in N(\theta^\star)$ and for every $x \in R$. We denote the discriminant of this quadratic by $D(\theta)$. Then we must have

$$\begin{aligned} D(\theta) \;=\; & \lambda^2(\theta)[(20 + 10\cos\theta + 2\sin\theta)^2 - 500 - 400\cos\theta] \\ & + 2\lambda(\theta)[-20 - 10\cos\theta - 2\sin\theta + 10x^\star + 8x^\star\cos\theta] \\ & + 1 \leq 0 \text{ for any } \theta \in N(\theta^\star). \end{aligned} \tag{3.32}$$

Solving this inequality, we conclude that any $\lambda(\theta)$ that satisfies

$$h_1 \leq \lambda(\theta) \leq h_2 \tag{3.33}$$

is a solution of (3.31), where

$$h_1 = \frac{20 + 10\cos\theta + 2\sin\theta - 10x^\star - 8x^\star\cos\theta - \sqrt{g(\theta)}}{(20 + 10\cos\theta + 2\sin\theta)^2 - 500 - 400\cos\theta}$$

and

$$h_2 = \frac{20 + 10\cos\theta + 2\sin\theta - 10x^\star - 8x^\star\cos\theta + \sqrt{g(\theta)}}{(20 + 10\cos\theta + 2\sin\theta)^2 - 500 - 400\cos\theta}$$

are the roots of the quadratic in (3.32), and

$$g(\theta) = 4(5 + 4\cos\theta)[(4x^{\star 2} - 10x^\star)\cos\theta - 2x^\star\sin\theta + 5x^{\star 2} - 20x^\star + 25]$$

is its discriminant. Here $g(\theta) \geq 0$ for all $\theta \in \mathcal{F}$. Hence, we can choose

$$\lambda(\theta) = \frac{20 + 10\cos\theta + 2\sin\theta - 10x^\star - 8x^\star\cos\theta}{(20 + 10\cos\theta + 2\sin\theta)^2 - 500 - 400\cos\theta}$$

which simplifies to

$$\lambda(\theta) = \frac{2.97 - 0.624\cos\theta + \sin\theta}{40\sin\theta + 10\sin 2\theta - 48\sin^2\theta}.$$

It can be seen that this $\lambda(\theta)$ is nonnegative for all $\theta \in \mathcal{F} \setminus \{0, 72.905°\}$. Therefore, $\theta^\star = 24.56°$ is indeed an optimal input for (P, θ).

On the other hand, $\hat{x}(\theta^\star)$ is unique, and hence $\hat{F}$ is open at $\theta^\star$. Therefore, we can use Theorem 3.10 to verify the optimality of $\theta^\star$ as well. Note that $\hat{\mathcal{P}}^{=}(\theta^\star) = \{1\}$, so

$$\tilde{\mathcal{L}}(x, \lambda; \theta) = x,$$

(i.e., there are no multipliers) and

$$\mathcal{F}^{=}_{\star}(\theta) = \{x \in R : f^1(x, \theta) \leq 0\} = F(\theta).$$

Therefore, to prove the saddle-point inequality (3.21), it is enough to show that

$$x^\star \leq x \quad \text{for every } x \in F(\theta) \text{ and every } \theta \in N(\theta^\star),$$

where $N(\theta^\star)$ is some neighbourhood of $\theta^\star$. Because of the structure of the feasible set, it suffices to show that

$$1.406 \leq a_1(\theta) \text{ for all } \theta \in N(\theta^\star).$$

This reduces to

$$r(\theta) = -98.442\cos^2\theta - 44.992\sin\theta\cos\theta - 56.24\sin\theta - 14.826\cos\theta + 135.284 \geq 0$$

for $\theta \in N(\theta^\star)$. But $r(\theta) \geq 0$ for all $0 \leq \theta \leq 72.905°$, as illustrated in Figure 3.3.

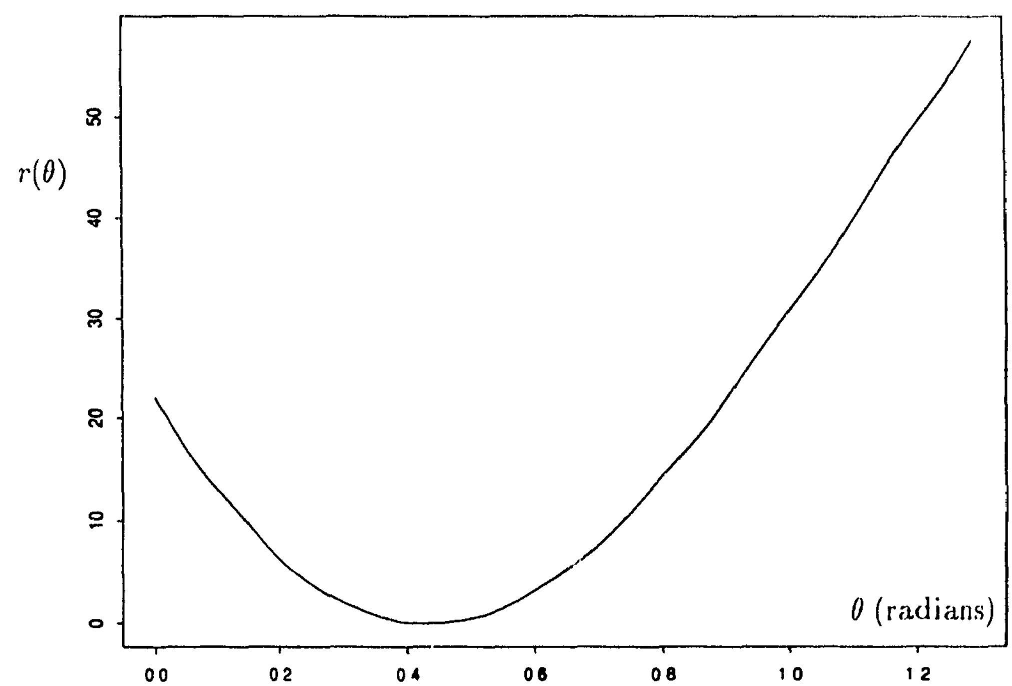

Figure 3.3: Graph of $r(\theta)$.

This confirms that $\theta^\star$ is an optimal input for (P, θ).

Remark: The notion of stability can be extended to multi-objective convex models (see, e.g., [30]) using the Charnes-Cooper [8] characterization of Pareto optimality. In this approach, the model is converted to single-objective, thus allowing the use of available input optimization results. The method can further be used to characterize Pareto optima in nonconvex, yet *convexifiable-by-a-splitting*, programs. However, this should be done with some care, as explained below.

Consider an arbitrary (generally nonconvex) multi-objective program

$$(MP) \qquad \begin{array}{l} \text{Min} \quad \{\phi^k(z) : k \in Q\} \\ \text{s.t.} \\ \qquad f^i(z) \leq 0, \quad i \in \mathcal{P} \end{array}$$

where all the functions are assumed continuous. The program is said to be convexifiable by-a-splitting if for some splitting of the variable z into $z = (x, \theta)$, all the functions are convex in x for every fixed θ. Such programs can be written as the convex model

$$(MCP,\theta) \qquad \begin{array}{l} \underset{(x)}{\text{Min}} \quad \{\phi^k(x,\theta) : k \in Q\} \\ \text{s.t.} \\ \qquad f^i(x,\theta) \leq 0, \quad i \in \mathcal{P}. \end{array}$$

(Note that here all the functions are continuous in x and θ and convex in x for each fixed θ.)

For every fixed θ, and every $x^\star$, a Pareto optimum of the convex program $(MCP, \theta^\star)$, the mapping $\mathcal{F}_\star : R^p \to R^n$ is defined by

$$\mathcal{F}_\star(\theta) = \left\{ x : \begin{array}{l} f^i(x,\theta) \leq 0, \quad i \in \mathcal{P} \\ \phi^k(x,\theta) \leq \phi^k(x^\star,\theta^\star), \quad k \in Q \end{array} \right\}.$$

Furthermore, $z^\star = (x^\star, \theta^\star)$ is a feasible point of (MP) to be tested for optimality. We recall that (see, e.g., [30]) a feasible point $z^\star$ is a strict Pareto optimum for (MP) if there is no other point $z \in F$ such that

$$\phi^k(z) \leq \phi^k(z^\star), \quad k \in Q.$$

Certainly, we have to distinguish between local and global strict Pareto optima whenever the program (MP) is not convex. So, in general, if there exists a neighbourhood $N(z^\star)$ such that there is no other point $z \in F \cap N(z^\star)$ satisfying

$$\phi^k(z) \leq \phi^k(z^\star), \quad k \in Q.$$

then $z^\star$ is a locally strict Pareto solution of (MP).

It is interesting that the point $z^\star$ is a globally strict Pareto solution of (MP) if, and only if,

$$\mathcal{F}_\star(\theta) = \emptyset \quad \text{for every } \theta \neq \theta^\star.$$

(Note that this means that $\mathcal{F}_\star$ is never lower semicontinuous at $\theta^\star$, or in other words, optimality and continuity do not go together in this case.) For, assume that $z^\star = (x^\star, \theta^\star)$ is a globally strict Pareto solution of (MP). Then by definiton of the strict Pareto minimum

$$F_0(x^\star,\theta^\star) = \left\{(x,\theta) : \begin{array}{l} f^i(x,\theta) \leq 0, \;\; i \in \mathcal{P} \\ \phi^k(x,\theta) \leq \phi^k(x^\star,\theta^\star), \;\; k \in Q \end{array} \right\} = \{(x^\star,\theta^\star)\}.$$

This implies that

$$\mathcal{F}_\star(\theta) = \emptyset \quad \text{for every } \theta \neq \theta^\star.$$

Coversely, emptiness of the set $\mathcal{F}_\star$ for $\theta \neq \theta^\star$ implies that the set $F_0(x^\star, \theta^\star)$ is a singleton and that $z^\star = (x^\star, \theta^\star)$ is a globally strict Pareto solution of (MP). For locally strict Pareto solutions of (MP), on the other hand, if $\mathcal{F}_\star = \emptyset$ for every $\theta \in N(\theta^\star)$ such that $\theta \neq \theta^\star$, where $N(\theta^\star)$ is some neighbourhood of $\theta^\star$, then $F_0(x^\star, \theta^\star) = \{(x^\star, \theta^\star)\}$. This means that $z^\star = (x^\star, \theta^\star)$ is a locally strict Pareto solution of (MP).

As an illustration, consider Zermelo's problem from Example 3.9 again. Assume that the objectives are now to minimize the time required to go from the origin to the target while staying as close to the y_1 axis as possible. More specifically, assume that the costs are

$$\phi^1 = x, \text{ and } \phi^2 = \int_0^x |y_2(x)| dx$$

with $y_2(x) = x|\sin\theta|$ and $y_2(0) = 0$. The corresponding convex multi-objective model is

$$(MCP,\theta) \qquad \begin{array}{l} \underset{(x)}{\text{Min}} \;\; \{\phi^1 = x, \phi^2 = \frac{x^2}{2}|\sin\theta|\} \\ \text{s.t.} \\ \qquad f^1 = (2x + x\cos\theta - 5)^2 + (x\sin\theta - 1)^2 - 1 \leq 0. \end{array}$$

Consider $z^\star = (x^\star, \theta^\star) = (\frac{5}{3}, 0)$. Then $z^\star$ is a strict Pareto solution of the corresponding program (MP). Therefore

$$\mathcal{F}_\star(\theta) = \emptyset \quad \text{for any } \theta \neq \theta^\star.$$

Similarly, $z^\star = (1.406, 24.56°)$ is a strict Pareto solution of (MP) implying the above set is empty for any $\theta \neq \theta^\star$.

3.4 The Inverse Problem

We recall the mathematical models of the form

$$(P,\theta) \qquad \begin{array}{l} \underset{(x)}{\text{Min}} \;\; f^0(x,\theta) \\ \text{s.t.} \\ \qquad f^i(x,\theta) \leq 0, \;\; i \in \mathcal{P} \\ \qquad \theta \in I. \end{array}$$

where the functions $f^i : R^n \times R^p \to R, \ \ i \in \{0\} \cup \mathcal{P}$, are assumed continuous in both vector variables and convex in x for every fixed θ, and $\theta \in I \subset R^p$ is the "input" (parameter). Suppose that the model is initially running with some input $\theta = \theta^0$ at which a desirable point x^* is feasible yet nonoptimal. The objective of the inverse programming is to find θ^* from θ^0 by stable perturbations of the input θ such that x^* becomes optimal, i.e., $x^* \in \tilde{F}(\theta^*)$. (Note that by stability we mean the usual stability which is continuity of the point-to-set mapping F.)

This objective may not be achievable, in which case an alternative is to find θ^* from θ^0, in a stable way, such that θ^* is a solution of the following problem:

$$\begin{aligned} &\text{Min} \quad |f^0(x^*,\theta) - \tilde{f}(\theta)| \\ &\text{s.t.} \\ &\qquad\qquad \theta \in S(x^*) \cap S, \end{aligned}$$

where

$$S(x^*) = \{\theta \in I : x^* \in F(\theta)\}$$

is the "region of feasibility of x^*" and S is the set of stable paths in I emanating from θ^0. We will use input optimization to find a solution to the inverse programming problem for linear models. As we mentioned earlier, the class of linear functions is a large well-known subclass of LFS functions.

We recall the following lemma from [9]. The proof will also be recalled for the sake of completeness. The idea is to formulate optimality conditions in terms of the optimal value.

Lemma 3.7 *Consider the program*

$$(CP) \qquad\qquad \begin{aligned} &Min \quad f^0(x) \\ &s.t. \\ &\qquad\quad x \in F, \end{aligned}$$

where f^0 is a convex and differentiable function and F is a convex set. Then a feasible point x^* is optimal if, and only if,

$$\nabla f^0(x^*)(x - x^*) \geq 0 \quad \textit{for every } x \in F.$$

Proof:(sufficiency:) Assume that

$$\nabla f^0(x^*)(x - x^*) \geq 0 \quad \text{for every } x \in F.$$

Then, by convexity of f^0,

$$f^0(x) - f^0(x^*) \geq \nabla f^0(x^*)(x - x^*) \geq 0 \quad \text{for every } x \in F.$$

Hence

$$f^0(x) \geq f^0(x^*) \quad \text{for every } x \in F,$$

which implies that x^* is optimal.
(Necessity:) Assume that $x^* \in F$ is optimal. Then

$$f^0(x^*) \leq f^0(x) \quad \text{for every } x \in F.$$

The minimum point x^* is either an interior point of F or a boundary point of F. If x^* is an interior point of F, then we must have

$$\nabla f^0(x^*) = 0 \quad \text{and hence} \quad (x - x^*)\nabla f^0(x^*) = 0.$$

For any minimum point, we have by convexity of f^0 and the set F,

$$f^0(x^*) \leq f^0(\lambda x + (1-\lambda)x^*)$$

for all $x \in F$ and $0 \leq \lambda \leq 1$. Then, for $\lambda > 0$,

$$\frac{f^0(x^* + \lambda(x - x^*)) - f^0(x^*)}{\lambda} \geq 0.$$

Taking the limit as λ approaches zero, we have

$$\nabla f^0(x^*)(x - x^*) \geq 0 \quad \text{for every } x \in F.$$

□

The following corollary immediately follows.

Corollary 3.8 *A feasible point x^* is optimal for the convex program (CP) if, and only if, the optimal value of the following program is zero:*

$$\begin{aligned} &Min \quad \nabla f^0(x^*)(x - x^*) \\ &s.t. \\ &\qquad x \in F. \end{aligned}$$

We can extend this result to mathematical models (having θ as a parameter).

Corollary 3.9 *Consider the convex model (P, θ) at some $\theta^* \in I$. Then $x^* \in F(\theta^*)$ is an optimal solution of (P, θ^*) if, and only if, the optimal value of the following model at $\theta = \theta^*$ is zero:*

$$(P1, x^*, \theta) \qquad \begin{aligned} &\underset{(x)}{Min} \quad \nabla f^0(x^*, \theta)(x - x^*) \\ &s.t. \\ &\qquad x \in F(\theta). \end{aligned}$$

We will use this result to solve the inverse problem for linear models. We first consider linear models in the standard form

$$(L,\theta) \qquad \begin{aligned} &\underset{(x)}{\text{Max}} \quad c^T(\theta)x \\ &\text{s.t.} \\ &\qquad A(\theta)x \le b(\theta) \\ &\qquad x \ge 0 \\ &\qquad \theta \in I, \end{aligned}$$

where $A(\theta) \in R^{m\times n}$ and $b(\theta) \in R^m$ are continuous in θ. We will later extend the results to the linear models in a more general form. Although the standard form is a special case of the general form, we follow this order for pedagogical reasons. For a linear model (L,θ), Corollary 3.9 yields the following important result.

Corollary 3.10 *Consider the linear model (L,θ) at some $\theta^\star \in I$. Then $x^\star \in F(\theta^\star)$ is an optimal solution of $(L,\theta^\star)$ if, and only if, the optimal value of the following linear model at $\theta = \theta^\star$ is zero:*

$$(LI,x^\star,\theta) \qquad \begin{aligned} &\underset{(x)}{Min} \quad c^T(\theta)(x^\star - x) \\ &s.t. \\ &\qquad A(\theta)x \le b(\theta) \\ &\qquad x \ge 0 \\ &\qquad \theta \in I. \end{aligned}$$

Corollary 3.10 gives the primal formulation. However, later on we will use the dual of the above model to characterize optimal parameters for the inverse problem. To this end, we need to study the dual model as well. The next result involves the dual model.

Corollary 3.11 *Consider the linear model (L,θ) at some $\theta^\star \in I$. Then $x^\star \in F(\theta^\star)$ is an optimal solution of $(L,\theta^\star)$ if, and only if, the optimal value of the following linear model at $\theta = \theta^\star$ is zero:*

$$(DI,x^\star,\theta) \qquad \begin{aligned} &\underset{(v)}{Min} \quad b^T(\theta)v - c^T(\theta)x^\star \\ &s.t. \\ &\qquad A^T(\theta)v \ge c(\theta) \\ &\qquad v \ge 0 \\ &\qquad \theta \in I. \end{aligned}$$

Proof: By Corollary 3.10, $x^\star \in F(\theta^\star)$ is an optimal solution of $(L,\theta^\star)$ if, and only if, the optimal value of $(LI,x^\star,\theta^\star)$ is zero. But the optimal value of $(LI,x^\star,\theta^\star)$ is zero if, and only if, the optimal value of the following model at $\theta = \theta^\star$ is zero:

$$\begin{aligned} \underset{(x)}{\text{Max}} \quad & c^T(\theta)(x - x^\star) \\ \text{s.t.} \quad & \\ & A(\theta)x \leq b(\theta) \\ & x \geq 0 \\ & \theta \in I, \end{aligned}$$

or, equivalently, if, and only if, the optimal value of $(DI, x^\star, \theta)$ at $\theta = \theta^\star$ is zero.

□

Note that (L, θ) and $(LI, x^\star, \theta)$ have the same constraints (and hence the same feasible set) and the same set of optimal solutions at a fixed θ. Besides, the absolute values of their optimal value functions differ by the constant $c^T(\theta)x^\star$ at a fixed θ. Therefore, studying stability of one is exactly the same as studying stability of the other. The same is true for their duals (D, θ) and $(DI, x^\star, \theta)$. We will refer to the feasible sets of (L, θ) and $(LI, x^\star, \theta)$ as $F(\theta)$ and to those of (D, θ) and $(DI, x^\star, \theta)$ as $F_D(\theta)$. Furthermore, let us denote the optimal value of $(LI, x^\star, \theta)$, that is

$$\min_{x \in F(\theta)} c^T(\theta)(x^\star - x),$$

by $\tilde{g}(\theta)$, and the optimal value of $(DI, x^\star, \theta)$, that is

$$\min_{v \in F_D(\theta)} b^T(\theta)v - c^T(\theta)x^\star,$$

by $\tilde{h}(\theta)$. Note that $\tilde{g}(\theta) = -\tilde{h}(\theta)$. We will refer to the model $(LI, x^\star, \theta)$ as the inverse model of (L, θ) and to $(DI, x^\star, \theta)$ as the dual inverse model of (L, θ). We now define the optimal input for the inverse model of (L, θ).

Definition 3.5 *Consider (L, θ) starting at θ^0. Assume that $x^\star$ is a feasible, yet nonoptimal solution of (L, θ^0). Furthermore, assume that $\theta^\star \in S(x^\star)$ and $S(\theta^\star)$ is a region of stability at $\theta^\star$. If $\tilde{g}(\theta^\star) \geq \tilde{g}(\theta)$ for every $\theta \in N(\theta^\star) \cap S(\theta^\star) \cap S(x^\star)$, where $N(\theta^\star)$ is a neighbourhood of $\theta^\star$, then we say that $\theta^\star$ is a locally "Approximate Optimal Input" for the inverse model $(LI, x^\star, \theta)$ of (L, θ), with respect to $S(\theta^\star) \cap S(x^\star)$.*

We will use the abbreviation "AOI" for "Approximate Optimal Input".

Definition 3.6 *Under the same assumptions as in Definition 3.5 if $\tilde{g}(\theta^\star) = 0$, then we say that $\theta^\star$ is an "Ideal Optimal Input" for the inverse model $(LI, x^\star, \theta)$ with respect to $S(\theta^\star) \cap S(x^\star)$.*

We will use the abbreviation "IOI" for "Ideal Optimal Input".

Remark: Note that the fact that $\tilde{g}(\theta) \leq 0$, for any $\theta \in S(x^*)$, implies that an IOI for (LI, x^*, θ) is an AOI for (LI, x^*, θ), for which $\tilde{g}(\theta^*) = 0$.

□

Obviously, at an IOI for (LI, x^*, θ), x^* is an optimal solution of (L, θ). If an IOI for this inverse model can not be found, then an AOI is searched for. A locally AOI for (LI, x^*, θ) has the following interesting property.

Lemma 3.8 *Assume that θ^* is a locally AOI for (LI, x^*, θ) with respect to $S(\theta^*) \cap S(x^*)$, $\tilde{x}(\theta^*)$ is any optimal solution of (L, θ^*) and $\tilde{x}(\theta)$ is any optimal solution of (L, θ). Then*

$$0 \leq c^T(\theta^*)(\tilde{x}(\theta^*) - x^*) \leq c^T(\theta)(\tilde{x}(\theta) - x^*) \tag{3.34}$$

for all $\theta \in N(\theta^*) \cap S(\theta^*) \cap S(x^*)$, where $N(\theta^*)$ is some neighbourhood of θ^*.

Proof: Since θ^* is a locally AOI for (LI, x^*, θ) with respect to $S(\theta^*) \cap S(x^*)$,

$$\tilde{g}(\theta^*) \geq \tilde{g}(\theta) \quad \text{for every } \theta \in N(\theta^*) \cap S(\theta^*) \cap S(x^*).$$

Therefore

$$\min_{x \in F(\theta^*)} c^T(\theta^*)(x^* - x) \geq \min_{x \in F(\theta)} c^T(\theta)(x^* - x) \quad \text{for every } \theta \in N(\theta^*) \cap S(\theta^*) \cap S(x^*),$$

or

$$c^T(\theta^*)(x^* - \tilde{x}(\theta^*)) \geq c^T(\theta)(x^* - \tilde{x}(\theta)) \quad \text{for every } \theta \in N(\theta^*) \cap S(\theta^*) \cap S(x^*).$$

But, $\tilde{g}(\theta) \leq 0$ for every $\theta \in S(x^*)$. Therefore, (3.34) holds for all $\theta \in N(\theta^*) \cap S(\theta^*) \cap S(x^*)$.

□

We will now present a condition that is both necessary and sufficient for an input $\theta^* \in I$ to be an AOI for (LI, x^*, θ) over the region $S(x^*) \cap S(\theta^*)$. As we recalled in Theorem 3.8, locally optimal inputs for the convex model (P, θ) have been characterized using the hyperplane separation theorem (see, e.g., [32]). The following property of a locally optimal input θ^* makes it possible to use that separation theorem.

$$\tilde{f}(\theta^*) = \min_{\theta \in N(\theta^*) \cap S(\theta^*)} \min_{x \in F(\theta)} f^0(x, \theta).$$

Whereas a locally AOI for (LI, x^*, θ), which we also denote by θ^*, is a solution of the *maxmin* problem

$$\tilde{g}(\theta^*) = \max_{\theta \in N(\theta^*) \cap S(\theta^*) \cap S(x^*)} \min_{x \in F(\theta)} c^T(\theta)(x^* - x).$$

This makes it impossible to follow the same proof, for it is impossible to use the hyperplane separation theorem. However, using the duality properties of linear programming, taking into account stability of the primal model, we can change the *maxmin* problem into a *minmin* problem and characterize a locally AOI for the inverse model.

Theorem 3.12 *Consider the linear model (L,θ) starting at θ^0. Assume that $x^\star \in F(\theta^0)$, yet $x^\star \notin \tilde{F}(\theta^0)$. Furthermore, assume that $\theta^\star \in S(x^\star)$, $S(\theta^\star)$ is a region of stability at $\theta^\star$ for both (L,θ) and (D,θ), and (L,θ) has a realistic objective function at $\theta^\star$. Then $\theta^\star$ is a locally AOI for $(LI, x^\star, \theta)$, with respect to $S(\theta^\star) \cap S(x^\star)$, if, and only if, it is a locally optimal input for $(DI, x^\star, \theta)$ with respect to $S(\theta^\star) \cap S(x^\star)$.*

Proof: Parameter $\theta^\star$ is an AOI for $(LI, x^\star, \theta)$, with respect to $S(\theta^\star) \cap S(x^\star)$, if, and only if, there exists a neighbourhood $N(\theta^\star)$ of $\theta^\star$ such that

$$\tilde{g}(\theta^\star) = \max_{\theta \in N(\theta^\star) \cap S(\theta^\star) \cap S(x^\star)} \tilde{g}(\theta),$$

or, if, and only if,

$$-\tilde{g}(\theta^\star) = \min_{\theta \in N(\theta^\star) \cap S(\theta^\star) \cap S(x^\star)} (-\tilde{g}(\theta)).$$

Since (L,θ) has a realistic objective function at $\theta^\star$, it follows from Theorem 3.4 that there exists a neighbourhood $N_1(\theta^\star)$ such that $\tilde{F}(\theta) \neq \emptyset$ and bounded for all $\theta \in N_1(\theta^\star) \cap S(\theta^\star)$. But for all such θ we have

$$\begin{aligned} -\tilde{g}(\theta) &= -\min_{x \in F(\theta)} c^T(\theta)(x^\star - x) \\ &= -\min_{x \in F(\theta)} (-c^T(\theta))(x - x^\star) \\ &= \max_{x \in F(\theta)} c^T(\theta)(x - x^\star) \\ &= \min_{v \in F_D(\theta)} b^T(\theta)v - c^T(\theta)x^\star \\ &= \tilde{h}(\theta). \end{aligned}$$

Therefore

$$\tilde{g}(\theta^\star) = \max_{\theta \in N(\theta^\star) \cap N_1(\theta^\star) \cap S(\theta^\star) \cap S(x^\star)} \tilde{g}(\theta),$$

if, and only if,

$$\tilde{h}(\theta^\star) = \min_{\theta \in N(\theta^\star) \cap N_1(\theta^\star) \cap S(x^\star) \cap S(\theta^\star)} \tilde{h}(\theta).$$

This completes the proof since $S(\theta^\star)$ is assumed to be also a region of stability for (D,θ) and therefore for $(DI, x^\star, \theta)$.

□

Let $(\mathcal{L}_\star^<)_{DI}(v,\lambda;\theta)$ be the restricted Lagrangian, $(F_\star^=)_D$ be the map $F_\star^=$, $q_D(\theta^\star) = card P_D^<(\theta^\star)$ and $\hat{v}(\theta^\star)$ be an optimal solution at $\theta^\star$ of the dual inverse problem $(DI, x^\star, \theta)$. Then the following theorem is an immediate result of Theorem 3.12 and Theorem 3.8

Theorem 3.13 *Consider (L,θ) starting at θ^0. Assume that $x^\star \in F(\theta^0)$, yet $x^\star \notin \tilde{F}(\theta^0)$. Furthermore, assume that $\theta^\star \in S(x^\star)$, $S(\theta^\star)$ is a region of stability at $\theta^\star$ for* *both (L,θ) and (D,θ) and (L,θ) has a realistic objective function at θ^*. Then θ^* is a locally AOI for (LI, x^*, θ) with respect to $S(\theta^*) \cap S(x^*)$ if, and only if, there exists a neighbourhood $N(\theta^*)$ of θ^* and a nonnegative vector function*

$$\Lambda : N(\theta^*) \cap S(\theta^*) \cap S(x^*) \to R_+^{q_D(\theta^*)}$$

such that, whenever $\theta \in N(\theta^*) \cap S(\theta^*) \cap S(x^*)$,

$$(\mathcal{L}_*^{\leq})_{DI}(\tilde{v}(\theta^*), \lambda; \theta^*) \leq (\mathcal{L}_*^{\leq})_{DI}(\tilde{v}(\theta^*), \Lambda(\theta^*); \theta^*) \leq (\mathcal{L}_*^{\leq})_{DI}(v, \Lambda(\theta); \theta) \qquad (3.35)$$

for all $\lambda \in R_+^{q_D(\theta^*)}$ and every $v \in (F_*^{=})_D(\theta)$.

□

If we rewrite (DI, x^*, θ) in the following standard form for a convex model, that is

$$\begin{aligned} &\underset{(v)}{\text{Min}} \quad b^T(\theta)v - c^T(\theta)x^* \\ &\text{s.t.} \\ &\qquad h^i = -a_i^T(\theta)v + c_i(\theta) \leq 0, \quad i = 1, \ldots, n \\ &\qquad h^j = -v_j \leq 0, \qquad\qquad\qquad j = 1, \ldots, m, \end{aligned}$$

where a_i^T is the i_{th} row of A^T, then

$$(\mathcal{L}_*^{\leq})_{DI}(v, \lambda; \theta) = b^T(\theta)v - c^T(\theta)x^* + \sum_{k \in \mathcal{P}_D \setminus \mathcal{P}_D^{=}(\theta^*)} \lambda_k h^k(v, \theta),$$

which implies in particular that

$$(\mathcal{L}_*^{\leq})_{DI}(\tilde{v}(\theta^*), \Lambda(\theta^*); \theta^*) = \tilde{h}(\theta^*) = -\tilde{g}(\theta^*).$$

This means that if (3.35) is satisfied, and in addition

$$(\mathcal{L}_*^{\leq})_{DI}(\tilde{v}(\theta^*), \Lambda(\theta^*); \theta^*) = 0,$$

then θ^* is an IOI for (LI, x^*, θ). However, since $\tilde{g}(\theta) \leq 0$, for any $\theta \in S(x^*)$, then $\tilde{g}(\theta^*) = 0$ implies that $\tilde{g}(\theta^*) \geq \tilde{g}(\theta)$ for every $\theta \in S(x^*)$. Thus, to verify optimality of an IOI for the inverse model, it is enough to have $\tilde{g}(\theta^*) = 0$.

Remark: Consider (L,θ) starting at θ^0. Assume that $x^* \in F(\theta^0)$, yet $x^* \notin \tilde{F}(\theta^0)$, and that S is the set of stable paths in I emanating from θ^0. (Note that for a stable model the mapping F is continuous and f^0 is realistic.) If the model (D,θ) is also stable for every perturbation in S, then the problem of finding a locally AOI for (LI, x^*, θ) in $S \cap S(x^*)$ is the same as the problem of finding a locally optimal input for (DI, x^*, θ) in $S \cap S(x^*)$.

□

In order to use Theorems 3.12 and 3.13, we must find conditions under which the continuity of the point-to-set mapping F implies continuity of F_D. In particular, we would like to find out when a region of stability for (L, θ) is also a region of stability for (D, θ). We need to refer to some important results in linear programming first. Consider the linear program

$$(L) \qquad \begin{array}{ll} \text{Max} & c^T x \\ \text{s.t.} & \\ & Ax \leq b \\ & x \geq 0 \end{array}$$

and its dual program

$$(D) \qquad \begin{array}{ll} \text{Min} & b^T v \\ \text{s.t.} & \\ & A^T v \geq c \\ & v \geq 0. \end{array}$$

We rewrite (L) in the following standard form for a convex program:

$$\begin{array}{ll} \text{Min} & -c^T x \\ \text{s.t.} & \\ & f^i = a_i x - b_i \leq 0, \quad i = 1, \ldots, m \\ & f^j = -x_j \leq 0, \qquad j = 1, \ldots, n. \end{array}$$

(Note that a_i is the i_{th} row of A.) Assume that the set of optimal solutions of (L) is nonempty and bounded. Then associated with x^*, a (finite) optimal solution of (L), are the nonnegative vectors $\lambda^* = (\lambda_i^*)$, $i \in \{1, \cdots, m\}$ and $u^* = (u_j^*)$, $j \in \{1, \cdots, n\}$, the components of which are referred to as Kuhn-Tucker multipliers (or "shadow prices"), such that

$$-c + A^T \lambda^* - u^* = 0 \tag{3.36}$$

$$\lambda_i^*(a_i x^* - b_i) = 0, \quad i = 1, \ldots, m \tag{3.37}$$

$$-u_j^* x_j^* = 0, \quad j = 1, \ldots, n \tag{3.38}$$

$$\lambda^* \geq 0 \tag{3.39}$$

$$u^* \geq 0. \tag{3.40}$$

The above conditions are referred to as Kuhn-Tucker conditions. Furthermore, the vectors λ^* and u^* are referred to as Kuhn-Tucker vectors.

Besides satisfying the Kuhn-Tucker conditions, the Kuhn-Tucker vectors have another well-known property that will be recalled in the following lemma, borrowed from [22]. The proof of the lemma will also be recalled for the sake of completeness.

Lemma 3.9 *Assume that x^* is a finite optimal solution of (L) and consider the Kuhn-Tucker vectors associated with x^*, as defined above. Then the Kuhn-Tucker vector λ^* is an optimal solution to the dual problem (D).*

Proof: Since $x^\star$ is a finite optimal solution of (L), the Kuhn-Tucker conditions at $x^\star$ are satisfied. Equation (3.38) may be written as

$$(a_{1j}\lambda_1^\star + \cdots + a_{mj}\lambda_m^\star - c_j)x_j^\star = 0 \quad j = 1, \ldots, n. \tag{3.41}$$

On the other hand, (3.36) and (3.40) imply that $u^\star = A^T\lambda^\star - c \geq 0$, or

$$A^T\lambda^\star \geq c. \tag{3.42}$$

Now (3.39) and (3.42) guarantee the existence of a solution $v = \lambda^\star$ to the system

$$\begin{aligned} A^T v &\geq c \\ v &\geq 0. \end{aligned}$$

These are the constraints of the dual problem (D). Hence, $\lambda^\star$ is a feasible solution of (D). For any feasible solutions x and v to the primal and the dual problems, we may write

$$b^T v \geq c^T x,$$

and, since $x^\star$ is a feasible solution to the primal problem,

$$b^T v \geq c^T x^\star. \tag{3.43}$$

The complementary slackness condition (3.37) implies that, for any $i = 1, \ldots, m$, either $\lambda_i^\star = 0$ or else $b_i = a_i x^\star$. It follows that

$$b^T\lambda^\star = (Ax^\star)^T\lambda^\star = x^{\star T} A^T\lambda^\star. \tag{3.44}$$

Similarly, (3.41) implies that

$$c^T x^\star = x^{\star T} c = x^{\star T} A^T\lambda^\star. \tag{3.45}$$

Here (3.44) and (3.45) yield

$$b^T\lambda^\star = c^T x^\star.$$

Therefore, $\lambda^\star$ is an optimal solution to the dual problem (D).

□

We now recall some well-known results from [10]. Consider the nonlinear program with inequality and equality constraints

$$(NP) \qquad \begin{aligned} &\text{Max } f(x) \\ &\text{s.t.} \\ &\quad g_i(x) \leq 0, \quad i = 1, \ldots, m \\ &\quad h_j(x) = 0, \quad j = 1, \ldots, q, \end{aligned}$$

where all the functions are real valued and continuously differentiable on R^n, and, for a feasible point x^*, consider the Mangasarian-Fromovitz constraint qualification

$$
(MFCQ) \qquad
\begin{array}{ll}
(i) & \text{There exists a } \hat{d} \in R^n \text{ such that} \\
 & \nabla g_i(x^*)\hat{d} < 0 \quad i \in \{i : g_i(x^*) = 0\} \\
 & \nabla h_j(x^*)\hat{d} = 0 \quad j = 1, \ldots, q; \\
(ii) & \text{The gradients } \{\nabla h_j(x^*)\},\ i = 1, \ldots, q \text{ are linearly independent.}
\end{array}
$$

At a local maximum x^*, let $K(x^*)$ denote the set of Kuhn-Tucker vectors corresponding to x^*; that is, the set of $(u, w) \in R^m \times R^q$ such that

$$
\begin{aligned}
(\nabla f(x^*))^T &= \sum_{i=1}^{m} u_i(\nabla g_i(x^*))^T + \sum_{j=1}^{q} w_j(\nabla h_j(x^*))^T \\
u_i &\geq 0 \\
u_i g_i(x^*) &= 0, \quad i = 1, \ldots, m.
\end{aligned}
$$

Then the following important result holds.

Lemma 3.10 (Gauvin [10]) *Let x^* be a local maximum for (NP). Then $K(x^*)$ is a nonempty and bounded set if, and only if, MFCQ is satisfied at x^*.*

In the case of linear programs of the form (L), the MFCQ is the well-known Slater condition. That is, we have the following immediate result.

Corollary 3.12 *Let x^* be an optimal solution of (L). Then $K(x^*)$ is a nonempty and bounded set if, and only if, Slater's condition is satisfied.*

We will use Corollary 3.12 and Lemma 3.9 to prove the following important result. The idea is to find conditions under which a region of stability for (L, θ) is also a region of stability for (D, θ).

Theorem 3.14 *Consider the linear model (L,θ) and its dual (D,θ). If (L,θ) has a realistic objective function at θ^*, then the constraints of (D,θ) satisfy Slater's condition at θ^*. Furthermore, if (D,θ) has a realistic objective function at θ^*, then it is stable at θ^*.*

Proof: Since (L,θ) has a realistic objective function at θ^*, it has a finite optimal solution at θ^*, say $\tilde{x}(\theta^*)$. This, by the duality theorem of linear programming, implies that (D,θ) has a finite optimal solution at θ^*, say $\tilde{v}(\theta^*)$. Now consider

$$
(D,\theta) \qquad
\begin{array}{ll}
\underset{(v)}{\text{Min}} & b^T(\theta)v \\
\text{s.t.} & \\
 & h^i = -a_i^T(\theta)v + c_i(\theta) \leq 0, \quad i = 1, \ldots, n, \\
 & h^j = -v_j \leq 0, \qquad\qquad\qquad j = 1, \ldots, m
\end{array}
$$

at $\theta^\star$. Let $\lambda^\star(\theta^\star) = (\lambda_i^\star(\theta^\star))$, $i = 1,\ldots,n$ and $u^\star(\theta^\star) = (u_j^\star(\theta^\star))$, $j = 1,\ldots,m$ be the Kuhn-Tucker vectors associated with $\tilde{v}(\theta^\star)$ at $\theta^\star$. Then, by Lemma 3.9, $\lambda^\star(\theta^\star)$ is an optimal solution of the dual of $(D,\theta^\star)$, which is $(L,\theta^\star)$. Since (L,θ) has a realistic objective function at $\theta^\star$, $\tilde{F}(\theta^\star)$ is a nonempty and bounded set. Therefore $\lambda^\star(\theta^\star)$ is bounded. Besides, by the Kuhn-Tucker conditions, we have

$$\begin{aligned}
b(\theta^\star) - A(\theta^\star)\lambda^\star(\theta^\star) - u^\star(\theta^\star) &= 0 && (3.46)\\
\lambda_i^\star(\theta^\star)(-a_i^T(\theta^\star)\tilde{v}(\theta^\star) + c_i(\theta^\star)) &= 0, \quad i = 1,\ldots,n && (3.47)\\
u_j^\star(\theta^\star)\tilde{v}_j(\theta^\star) &= 0, \quad j = 1,\ldots,m && (3.48)\\
\lambda^\star &\geq 0 && (3.49)\\
u^\star &\geq 0. && (3.50)
\end{aligned}$$

But (3.46) implies that

$$u^\star(\theta^\star) = -A(\theta^\star)\lambda^\star(\theta^\star) + b(\theta^\star).$$

Hence boundedness of $\lambda^\star(\theta^\star)$ implies boundedness of $u^\star(\theta^\star)$. Thus $K(\tilde{v}(\theta^\star))$, the set of Kuhn-Tucker vectors corresponding to $\tilde{v}(\theta^\star)$, is nonempty and bounded. This, by Corollary 3.12, means that the constraints of (D,θ) satisfy Slater's condition at $\theta^\star$. If (D,θ) has a realistic objective function at $\theta^\star$, then it is also stable at $\theta^\star$.

□

The following corollary immediately follows.

Corollary 3.13 *If (L,θ) has a realistic objective function at $\theta^\star$, then any region of stability for (L,θ) at $\theta^\star$ is also a region of stability for (D,θ) at $\theta^\star$.*

Let us illustrate these results by the following example.

Example 3.10 Consider

$$(L,\theta) \qquad \begin{aligned}
&\max_{(x)} \; x_2\\
&\text{s.t.}\\
&\quad x_1 + x_2 \leq 1\\
&\quad -x_1 - \theta x_2 \leq -1\\
&\quad x_1 \geq 0\\
&\quad x_2 \geq 0
\end{aligned}$$

around $\theta^\star = 1$. Then

$$\tilde{F}(\theta) = \begin{cases} \{(0,1)^T\} & \text{if } \theta \geq 1\\ \{(1,0)^T\} & \text{if } \theta < 1. \end{cases}$$

Note that (L,θ) is not stable at $\theta^\star$ However, it is stable for $S(\theta^\star) = \{\theta : \theta \geq 1\}$. Besides, (L,θ) has a realistic objective function at $\theta^\star$. Now consider its dual

$$(D,\theta) \qquad \begin{array}{ll} \underset{(v)}{\text{Min}} & v_1 - v_2 \\ \text{s.t.} & \\ & v_1 - v_2 \geq 0 \\ & v_1 - \theta v_2 \geq 1 \\ & v_1 \geq 0 \\ & v_2 \geq 0. \end{array}$$

We notice that Slater's condition is satisfied at $\theta = \theta^\star$ and

$$\tilde{F}_D(\theta^\star) = \{v^\star \geq 0 : v_1^\star - v_2^\star = 1\},$$

which means that (D,θ) does not have a realistic objective function at $\theta^\star = 1$. Note that (L,θ) does not satisfy Slater's condition at $\theta^\star$.

□

Theorem 3.13 can now be restated in the following form.

Theorem 3.15 *Consider (L,θ) starting at θ^0. Assume that $x^\star \in F(\theta^0)$ yet, $x^\star \notin \tilde{F}(\theta^0)$, $\theta^\star \in S(x^\star)$ and $S(\theta^\star)$ is a region of stability at $\theta^\star$ for (L,θ). Furthermore, assume that (L,θ) has a realistic objective function at $\theta^\star$, and let $\tilde{v}(\theta^\star)$ be a finite optimal solution of (D,θ) at $\theta^\star$. Then $\theta^\star$ is a locally AOI for $(LI,x^\star,\theta)$ with respect to $S(\theta^\star) \cap S(x^\star)$ if, and only if, there exists a neighbourhood $N(\theta^\star)$ of $\theta^\star$ and a nonnegative vector function*

$$\Lambda : N(\theta^\star) \cap S(\theta^\star) \cap S(x^\star) \to R_+^{q_D(\theta^\star)}$$

such that, whenever $\theta \in N(\theta^\star) \cap S(\theta^\star) \cap S(x^\star)$,

$$\mathcal{L}_{DI}(\tilde{v}(\theta^\star), \lambda; \theta^\star) \leq \mathcal{L}_{DI}(\tilde{v}(\theta^\star), \Lambda(\theta^\star); \theta^\star) \leq \mathcal{L}_{DI}(v, \Lambda(\theta); \theta) \qquad (3.51)$$

for all $\lambda \in R_+^{\mathrm{card}P_D}$ and every $v \in R^m$.

Proof: Since (L,θ) has a realistic objective function at $\theta^\star$, by Theorem 3.14, the constraints of $(D,\theta^\star)$ satisfy Slater's condition, under which $(\mathcal{L}_\star^{\leq})_{DI}$ becomes the usual Lagrangian $\mathcal{L}_{DI}$ for the dual inverse model. Applying Theorem 3.13 now yields the result.

Remark: Consider (L,θ) starting at θ^0. Assume that $x^\star \in F(\theta^0)$, yet $x^\star \notin \tilde{F}(\theta^0)$, and S is a region in which, for every perturbation, the objective functions of (L,θ) and (D,θ) are realistic. Then the problem of finding a locally AOI for $(LI,x^\star,\theta)$ in $S \cap S(x^\star)$ is the same as the problem of finding a locally optimal input for $(DI,x^\star,\theta)$ in $S \cap S(x^\star)$.

□

From Theorem 3.14 the following important result follows as well.

Corollary 3.14 *If (L,θ) and (D,θ) both have realistic objective functions at θ^*, then they both satisfy Slater's condition at θ^* and are therefore both stable at θ^*.*

Remark: It was shown in [19] that the set of optimal solutions of (L,θ) and (D,θ), being nonempty and bounded at θ^*, is a sufficient condition for the continuity of the optimal value function at θ^*. By Corollary 3.14, this condition is not only sufficient for continuity of the optimal value function but also sufficient for the continuity of the point-to-set mappings F and F_D at θ^* and thus sufficient for stability of both the primal and dual linear models.

Note that the objective function of (L,θ), being realistic at θ^*, does not imply that the objective function of (D,θ) is realistic at θ^* as Example 3.10 shows. The following result, however, gives a sufficient condition for the dual objective function to be realistic at θ^* whenever the objective function of the primal is realistic at θ^*. Although the result is for programs (with no parameters), it can be extended to models (since here we are interested in the behaviour of the primal and dual models at a fixed parameter θ^*).

Theorem 3.16 *Assume that the set of optimal solutions of (L) is nonempty and bounded and that all the components of the vector b are strictly positive. Then the set of optimal solutions of (D) is also nonempty and bounded.*

Proof: Since the set of optimal solutions of (L) is nonempty and bounded, for any (finite) optimal solution $\tilde{x}$ of (L), there exists an optimal solution $\tilde{v}$ of (D) such that

$$c^T\tilde{x} = b^T\tilde{v}. \tag{3.52}$$

If $\tilde{v}_{j_0} \to +\infty$ for some $j_0 \in \{1,\dots,m\}$, then

$$b^T\tilde{v} \to +\infty$$

since by assumption $b_j > 0,\ \ j = 1,\dots,m$. This is, however, a contradiction to (3.52) since $c^T\tilde{x}$ is finite.

Linear models with mixed constraints

We now consider the more general form of linear mathematical models, i.e., linear models of the form

$$(Lm,\theta) \qquad \begin{aligned} \operatorname*{Max}_{(X)}\ \ & C_1^T(\theta)X_1 + C_2^T(\theta)X_2 \\ \text{s.t.}\ \ & \\ & A_{11}(\theta)X_1 + A_{12}(\theta)X_2 \le B_1(\theta) \\ & A_{21}(\theta)X_1 + A_{22}(\theta)X_2 = B_2(\theta) \\ & X_1 \ge 0 \quad (X_2 \text{ unrestricted }), \end{aligned}$$

and their duals

$$
(Dm,\theta) \qquad \begin{array}{ll} \underset{(V)}{\text{Min}} & B_1^T(\theta)V_1 + B_2^T(\theta)V_2 \\ \text{s.t.} & \\ & A_{11}^T(\theta)V_1 + A_{21}^T(\theta)V_2 \geq C_1(\theta) \\ & A_{12}^T(\theta)V_1 + A_{22}^T(\theta)V_2 = C_2(\theta) \\ & V_1 \geq 0 \quad (V_2 \text{ unrestricted }), \end{array}
$$

where the vector $X = (x_1, \ldots, x_n)^T$ is decomposed into two blocks:

$$X_1 = (X_1^i) = (x_i), \ \ x_i \geq 0, \ \ i \in N_1$$

and

$$X_2 = (X_2^j) = (x_j), \ \ x_j \text{ (unrestricted)}, \ j \in N_2,$$

and the vector $V = (v_1, \ldots, v_m)^T$ is decomposed into the following two blocks:

$$V_1 = (V_1^i) = (v_i), \ \ v_i \geq 0, \ \ i \in M_1$$

and

$$V_2 = (V_2^j) = (v_j), \ \ v_j \text{ (unrestricted)}, \ j \in M_2.$$

Here N is the set of indices $\{1, 2, \ldots, n\}$, while M is the set $\{1, 2, \ldots, m\}$. Furthermore, N_1 and N_2 are complementary subsets of N with n_1 and n_2 elements respectively, while M_1 and M_2 are complementary subsets of M with m_1 and m_2 elements respectively.

The matrix A and the vectors C and B are decomposed into blocks corresponding to the decompositions of M and N into $M_1 + M_2$ and $N_1 + N_2$ respectively. These blocks are $A_{11}, A_{12}, A_{21}, A_{22}, C_1, C_2, B_1$ and B_2. Besides, as a function of θ, all the coefficient matrices are continuous. Any linear model of the form (Lm, θ) can, by means of well-known transformations of variables and constraints, be modified so as to involve only nonnegative variables and only inequality constraints. Because of this equivalence, the theorems and definitions used for linear models (L, θ) and (D, θ) also hold for (Lm, θ) and (Dm, θ). On the other hand, there are advantages to using the latter forms. For example, the standard device for eliminating a variable of unrestricted sign is to replace it by the difference of two new nonnegative variables. This, however, always results in unbounded sets of optimal solutions in the primal models. A similar problem also happens on the occurence of equality constraints, for to each equality constraint in the primal model, there corresponds an unrestricted variable in the dual model.

We will show that the results of inverse programming formulations for standard linear models can also be extended to linear models with mixed constraints. Theorems 3.12, 3.13, Corallaries 3.10, 3.11, Definitions 3.5, 3.6 and Lemma 3.8 trivially hold for (Lm, θ) and (Dm, θ). We will refer to the inverse model of (Lm, θ) as (LIm, X^*, θ) and to its dual inverse model as (DIm, X^*, θ). Now consider the linear program

$$(Lm)\qquad \begin{array}{ll} \text{Max} & C_1^T X_1 + C_2^T X_2 \\ \text{s.t.} & \\ & A_{11}X_1 + A_{12}X_2 \leq B_1 \\ & A_{21}X_1 + A_{22}X_2 = B_2 \\ & X_1 \geq 0 \quad (X_2 \text{ unrestricted }) \end{array}$$

and its dual

$$(Dm)\qquad \begin{array}{ll} \text{Min} & B_1^T V_1 + B_2^T V_2 \\ \text{s.t.} & \\ & A_{11}^T V_1 + A_{21}^T V_2 \geq C_1 \\ & A_{12}^T V_1 + A_{22}^T V_2 = C_2 \\ & V_1 \geq 0 \quad (V_2 \text{ unrestricted }). \end{array}$$

Then associated with an optimal solution X^* of (Lm) are the Kuhn-Tucker vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_{m_1}^*)^T \geq 0$, $w^* = (w_1^*, \ldots, w_{m_2}^*)^T$ (unrestricted) and $u^* = (u_1^*, \ldots, u_{n_1}^*)^T \geq 0$. These vectors have the following interesting property.

Theorem 3.17 *Assume that X^* is an optimal solution of (Lm) and consider the Kuhn-Tucker vectors associated with X^*, as defined above. Then the vector $(\lambda^{*T}, w^{*T})^T$ is an optimal solution to the dual problem (Dm).*

Proof: We rewrite (Lm) and (Dm) in the following standard forms:

$$(Lm')\qquad \begin{array}{ll} \text{Max} & C_1^T X_1 + C_2^T X_2' - C_2^T X_2'' \\ \text{s.t.} & \\ & A_{11}X_1 + A_{12}X_2' - A_{12}X_2'' \leq B_1 \\ & A_{21}X_1 + A_{22}X_2' - A_{22}X_2'' \leq B_2 \\ & -A_{21}X_1 - A_{22}X_2' + A_{22}X_2'' \leq -B_2 \\ & X_1 \geq 0, \quad X_2' \geq 0, \quad X_2'' \geq 0; \end{array}$$

and

$$(Dm')\qquad \begin{array}{ll} \text{Min} & B_1^T V_1 + B_2^T V_2' - B_2^T V_2'' \\ \text{s.t.} & \\ & A_{11}^T V_1 + A_{21}^T V_2' - A_{21}^T V_2'' \geq C_1 \\ & A_{12}^T V_1 + A_{22}^T V_2' - A_{22}^T V_2'' \geq C_2 \\ & -A_{12}^T V_1 - A_{22}^T V_2' + A_{22}^T V_2'' \geq -C_2 \\ & V_1 \geq 0, \quad V_2' \geq 0, \quad V_2'' \geq 0. \end{array}$$

Then the Kuhn-Tucker vectors (considered as columns) corresponding to X^* will be

$$\lambda^* = (\lambda_i^*) \geq 0, \;\; i = 1, \ldots, m_1, \quad w'^* = (w_i'^*) \geq 0, \;\; i = 1, \ldots, m_2,$$

$$w''^{\star} = (w_i''^{\star}) \geq 0, \quad i = 1, \ldots, m_2, \quad u^{\star} = (u_i^{\star}) \geq 0, \quad i = 1, \ldots, n_1,$$

$$u'^{\star} = (u_i'^{\star}) \geq 0, \quad i = 1, \ldots, n_2, \text{ and } u''^{\star} = (u_i''^{\star}) \geq 0, \quad i = 1, \ldots, n_2.$$

Now by Lemma 3.9, $(\lambda^{\star T}, w'^{\star T}, w''^{\star T})^T$ is an optimal solution of (Dm') which implies that $(\lambda^{\star T}, w'^{\star T} - w''^{\star T})^T$ is an optimal solution of (Dm). But $w^{\star} = w'^{\star} - w''^{\star}$, and the proof is complete.

□

Theorem 3.18 *Consider the linear program (Lm). Then MFCQ reduces to the following conditions:*

(a) There exists a feasible point, $\hat{X}$, such that $\hat{X}$ is a Slater point for the inequality constraints of (Lm);

(b) The coefficient matrix corresponding to the equality constraints has full row rank.

Proof: We rewrite (Lm) in the form

$$\begin{array}{ll} \text{Max} & C_1^T X_1 + C_2^T X_2 \\ \text{s.t.} & \\ & f^i = A_{11}^i X_1 + A_{12}^i X_2 \leq B_1^i \quad i \in M_1 \\ & f^j = A_{21}^j X_1 + A_{22}^j X_2 = B_2^j \quad j \in M_2 \\ & f^k = -X_1^k \leq 0 \quad k \in N_1. \end{array}$$

Then at a feasible point $X^{\star}$, MFCQ is as follows:

There exists $\hat{d} \in R^n$ such that

$$\begin{array}{lll} (i) & \nabla f^i(X^{\star})\hat{d} < 0, & i \in M_1(X^{\star}) \\ & \nabla f^j(X^{\star})\hat{d} = 0, & j \in M_2 \\ & \nabla f^k(X^{\star})\hat{d} < 0, & k \in N_1(X^{\star}) \\ (ii) & \nabla f^j(X^{\star}), & j \in M_2 \text{ are linearly independent.} \end{array}$$

Since all the functions are linear, condition (i) is equivalent to the existence of the point $\hat{X} = X^{\star} + \alpha\hat{d}$, $\alpha > 0$ sufficiently small, such that

$$\begin{array}{rcll} f^i(\hat{X}) & < & 0, & i \in M_1 \\ f^j(\hat{X}) & = & 0, & j \in M_2 \\ f^k(\hat{X}) & < & 0, & k \in N_1. \end{array}$$

This means that $\hat{X}$ is feasible and is a Slater point for the inequality constraints of (Lm). On the other hand, condition (ii) is equivalent to the rows of the coefficient matrix corresponding to the equality constraints $[A_{21}\ A_{22}]$ being linearly independent. This also means that this matrix has full row rank.

□

Theorem 3.19 *Consider the linear model (Lm, θ) around $\theta^\star$. Assume that MFCQ is satisfied at $\theta^\star$ (i.e., conditions (a) and (b) of Theorem 3.18 hold). Then the point-to-set mapping F is lower semicontinuous at $\theta^\star$. If (Lm, θ) has, in addition, a realistic objective function at $\theta^\star$, then it is a stable model at $\theta^\star$.*

Proof: Assume that MFCQ is satisfied at $\theta^\star$. Then there exists $\hat{X} \in F(\theta^\star)$ such that

$$\begin{aligned} A_{11}^i(\theta^\star)\hat{X}_1 + A_{12}^i(\theta^\star)\hat{X}_2 &< B_1^i(\theta^\star), \quad i \in M_1 \\ A_{21}^j(\theta^\star)\hat{X}_1 + A_{22}^j(\theta^\star)\hat{X}_2 &= B_2^j(\theta^\star), \quad j \in M_2 \\ -\hat{X}_1^k &< 0, \quad k \in N_1, \end{aligned}$$

where the matrix $[A_{21}(\theta^\star)\ A_{22}(\theta^\star)]$ has full row rank. This means that there exists $N(\theta^\star)$, a neighbourhood of $\theta^\star$, such that

$$\mathcal{P}^=(\theta) = \mathcal{P}^=(\theta^\star) \text{ for every } \theta \in N(\theta^\star). \tag{3.53}$$

It is well known that (see, [32]) $R_1(\theta^\star) = \{\theta : \mathcal{P}^=(\theta) = \mathcal{P}^=(\theta^\star)\}$ is a region of stability at $\theta^\star$, provided that $F^=$ is lower semicontinuous at $\theta^\star$. Here (3.53) implies that

$$R_1(\theta^\star) = N(\theta^\star). \tag{3.54}$$

On the other hand, since $[A_{21}(\theta)\ A_{22}(\theta)]$ has full row rank at $\theta^\star$, it has full row rank in a sufficiently small neighbourhood of $\theta^\star$. It follows that the mapping

$$F^= : \theta \to \{X : A_{21}(\theta)X_1 + A_{22}(\theta)X_2 = B_2(\theta)\}$$

is lower semicontinuous at $\theta^\star$. This together with (3.54) imply that the point-to-set mapping F is lower semicontinuous at $\theta^\star$. If (Lm, θ) has, in addition, a realistic objective function at $\theta^\star$, then it is stable at $\theta^\star$.

□

Finally, we need to prove the following result for the dual linear model (Dm, θ) in order to be able to use Theorem 3.13 to characterize an AOI for the inverse model $(LIm, X^\star, \theta)$.

Theorem 3.20 *If (Lm, θ) has a realistic objective function at $\theta^\star$, then the constraints of (Dm, θ) satisfy MFCQ at $\theta^\star$. If, in addition to MFCQ property, (Dm, θ) has a realistic objective function at $\theta^\star$, then it is a stable model at $\theta^\star$.*

Proof: Since (Lm, θ) has a realistic objective function at $\theta^\star$, it has a finite optimal solution at $\theta^\star$, say $\tilde{X}(\theta^\star)$. By the duality theory of linear programming, this implies that the dual problem (Dm, θ) has a finite optimal solution at θ^*, say $\tilde{V}(\theta^*)$. Consider

$$
(Dm,\theta) \qquad \begin{array}{ll} \underset{(v)}{\text{Min}} & B_1^T(\theta)V_1 + B_2^T(\theta)V_2 \\ \text{s.t.} & \\ & -(A_{11}^T)^i(\theta)V_1 - (A_{21}^T)^i(\theta)V_2 + (C_1)^i(\theta) \leq 0,\ i \in N_1 \\ & (A_{12}^T)^j(\theta)V_1 + (A_{22}^T)^j(\theta)V_2 + C_2(\theta) = 0,\ j \in N_2 \\ & -V_1^k \leq 0,\ k \in M_1. \end{array}
$$

Then associated with $\tilde{V}(\theta^*)$ are the Kuhn-Tucker vectors

$$\lambda^*(\theta^*) = (\lambda_1^*(\theta^*), \ldots, \lambda_{m_1}^*(\theta^*))^T \geq 0,$$

$$w^*(\theta^*) = (w_1^*(\theta^*), \ldots, w_{m_2}^*(\theta^*))^T \text{ (unrestricted) and}$$

$$u^*(\theta^*) = (u_1^*(\theta^*), \ldots, u_{n_1}^*(\theta^*))^T \geq 0.$$

By Theorem 3.17, $(\lambda^{*T}(\theta^*), w^{*T}(\theta^*))^T$ is an optimal solution of the dual of (Dm, θ^*), which is (Lm, θ^*). Since (Lm, θ) has a realistic objective function at θ^*, $\tilde{F}(\theta^*)$ is a nonempty and bounded set. Therefore, $\lambda^*(\theta^*)$ and $w^*(\theta^*)$ are bounded. Besides, by the Kuhn-Tucker conditions,

$$
\begin{aligned}
\begin{bmatrix} B_1(\theta^*) \\ B_2(\theta^*) \end{bmatrix} + \begin{bmatrix} -A_{11}(\theta^*) & A_{12}(\theta^*) \\ -A_{21}(\theta^*) & A_{22}(\theta^*) \end{bmatrix} \begin{bmatrix} \lambda^*(\theta^*) \\ w^*(\theta^*) \end{bmatrix} - u^*(\theta^*) &= 0 \\
\lambda_i^*((A_{11}^T)^i(\theta^*)\tilde{V}_1(\theta^*) + (A_{21}^T)^i(\theta^*)\tilde{V}_2(\theta^*) + (C_1)^i(\theta^*)) &= 0, \quad i \in N_1 \\
-u_j^*(\theta^*)(\tilde{V}_1(\theta^*))^j &= 0, \quad j \in M_1.
\end{aligned}
$$

Therefore

$$u^*(\theta^*) = \begin{bmatrix} B_1(\theta^*) \\ B_2(\theta^*) \end{bmatrix} + \begin{bmatrix} -A_{11}(\theta^*) & A_{12}(\theta^*) \\ -A_{21}(\theta^*) & A_{22}(\theta^*) \end{bmatrix} \begin{bmatrix} \lambda^*(\theta^*) \\ w^*(\theta^*) \end{bmatrix},$$

which implies that $u^*(\theta^*)$ is bounded. Hence, $K(\tilde{V}(\theta^*))$, the set of Kuhn-Tucker vectors corresponding to $\tilde{V}(\theta^*)$, is nonempty and bounded. Thus, by Lemma 3.10, the constraints of (Dm, θ) satisfy MFCQ at θ^*.

□

An important consequence immediately follows.

Corollary 3.15 *If (Lm, θ) has a realistic objective function at θ^*, then any region of stability for (Lm, θ) at θ^* is also a region of stability for (Dm, θ) at θ^*.*

We now illustrate some of these results by an example.

Example 3.11 Consider the linear model

$$\begin{array}{ll} & \underset{(X)}{\text{Max}} \ x_2 \\ & \text{s.t.} \\ & x_1 + x_2 \leq 1 \\ (Lm, \theta) & (\theta_1^2 - \theta_2^2)^2 x_1 = 0 \\ & x_1 \geq 0, \ \ x_2 \geq 0 \end{array}$$

around $\theta^* = (0,0)^T$. Note that, regardless of θ,

$$\tilde{F}(\theta) = \{(0,1)^T\}.$$

Although (Lm, θ) is not stable at θ^*, it has a realistic objective function at θ^*. Furthermore, the region

$$S(\theta^*) = \{\theta \in R^2 : \theta_1^2 = \theta_2^2\}$$

is a region of stability for (Lm, θ) at θ^*. The dual problem is

$$\begin{array}{ll} & \underset{(V)}{\text{Min}} \ v_1 \\ & \text{s.t.} \\ & v_1 + (\theta_1^2 - \theta_2^2)^2 v_2 \geq 0 \\ (Dm, \theta) & v_1 \geq 1 \\ & v_1 \geq 0, \ \ v_2 \text{ (unrestricted)}. \end{array}$$

The constraints of (Dm, θ) satisfy MFCQ (in fact Slater's) condition at θ^*. So the point-to-set mapping F_D is lower semicontinuous at θ^*. But (Dm, θ) does not have a realistic objective function at θ^* since

$$\tilde{F}_D(\theta^*) = \{(1, v_2)^T : v_2 \in R\}.$$

□

The following result thus follows from Theorems 3.13 and 3.20.

Corollary 3.16 *Consider (Lm, θ) starting at θ^0. Assume that $X^* \in F(\theta^0)$, yet $X^* \notin \tilde{F}(\theta^0)$, and at $\theta^* \in S(X^*)$, (Lm, θ) has a realistic objective function. Furthermore, assume that $S(\theta^*)$ is a region of stability for (Lm, θ) at θ^* and $\tilde{V}(\theta^*)$ is a finite optimal solution of (Dm, θ). Then θ^* is a locally AOI for (LIm, X^*, θ) with respect to $S(\theta^*) \cap S(X^*)$ if, and only if, there exist a neighbourhood $N(\theta^*)$ of θ^* and a nonnegative vector function*

$$\Lambda : N(\theta^*) \cap S(\theta^*) \cap S(X^*) \to R_+^{q_D(\theta^*)}$$

such that, whenever $\theta \in N(\theta^*) \cap S(\theta^*) \cap S(X^*)$,

$$(\mathcal{L}_*^{\leq})_{DI}(\tilde{V}(\theta^*), \lambda; \theta^*) \leq (\mathcal{L}_*^{\leq})_{DI}(\tilde{V}(\theta^*), \Lambda(\theta^*); \theta^*) \leq (\mathcal{L}_*^{\leq})_{DI}(V, \Lambda(\theta); \theta) \tag{3.55}$$

for every $\lambda \in R_+^{q_D(\theta^*)}$ and every $V \in (F_*^{=})_D(\theta)$.

Remark: Consider (Lm, θ) starting at θ^0. Assume that $X^\star \in F(\theta^0)$, yet $X^\star \notin \tilde{F}(\theta^0)$, and S is a region in which, for every perturbation, the objective functions of (Lm, θ) and (Dm, θ) are realistic. Then the problem of finding a locally AOI for $(LIm, X^\star, \theta)$ in $S \cap S(X^\star)$ is the same as the problem of finding a locally optimal input for $(DIm, X^\star, \theta)$ in $S \cap S(X^\star)$.

□

From Theorem 3.20 the following important result follows as well.

Corollary 3.17 *If (Lm, θ) and (Dm, θ) both have realistic objective functions at $\theta^\star$, then they both satisfy MFCQ at $\theta^\star$ and are therefore both stable at $\theta^\star$.*

Remark: This corollary shows that the set of optimal solutions of (Lm, θ) and (Dm, θ), being nonempty and bounded at $\theta^\star$, is not only a sufficient condition for the continuity of the optimal value function at $\theta^\star$, but also a sufficient condition for stability of both models at $\theta^\star$.

3.5 A Numerical Method

Like most input optimization problems, solving the inverse problems numerically involves the constrained optimization of an optimal value function (such as $\tilde{g}(\theta)$) which is explicitly unknown. Therefore, the same difficulties arise when solving the inverse problems numerically.

It is well known that if all the functions in (P, θ) are jointly convex in (x, θ), then the optimal value function $\tilde{f}(\theta)$ is convex (see, e.g., [18]). In this case it is possible to prove the convergence of some of the numerical algorithms for solving the problem of constrained optimization of the optimal value function. One such case was discussed in [14].

Unfortunately, jointly convexity of functions is not the case in most of the mathematical models that we study. Therefore, finding a numerical algorithm, and then proving its convergence, is a formidable task. In spite of all these difficulties, there do exist tools (though rather primitive) for solving input optimization problems in the linear and convex case. Several case studies have actually been solved using these tools (see, e.g., [7,23]).

There are presently two basic approaches to solving the input optimization problems. Both are essentially different from the usual numerical methods because they optimize the explicitly-unknown function $\tilde{f}(\theta)$ and the optimization is performed only along the paths of stability. These two approaches are called the "M-method" and the "MV-method". The M-method considers the perturbations of θ along which $F(\theta)$ monotonously increases, while the MV-method uses the *marginal value formula* and then searches $\tilde{f}(\theta)$ along a stable path. More details on these are given in [34].

In this section, we will use a method based on the MV-method to solve the inverse problem for linear models numerically. For simplicity, we will use the notation (L,θ) to refer to both linear models in the standard form and linear models with mixed constraints. The corresponding inverse linear model will also be denoted by $(LI,x^\star,\theta)$.

In order to find an AOI for $(LI,x^\star,\theta)$, starting from θ^0, with $x^\star$ a feasible but nonoptimal point, we must solve the optimization problem

$$\max_{\theta\in S\cap S(x^\star)} \tilde{g}(\theta), \tag{3.56}$$

where, as denoted earlier,

$$\tilde{g} = \min_{x\in F(\theta)} c^T(\theta)(x^\star - x),$$

S is the set of stable paths in I and $S(x^\star)$ is the region of feasibility of $x^\star$. Ideally, we would like to find a globally optimal solution of (3.56). But this is a very difficult task since $\tilde{g}(\theta)$ is generally not concave. We will develop an algorithm for finding a locally AOI for $(LI,x^\star,\theta)$ under Slater's condition or under a special case of MFCQ. The algorithm is the one given in [7] for finding an optimal input, with slight modifications. The main differences are that instead of looking for a direction of descent we look for a direction of ascent and that the algorithm is applied to the inverse model rather than the model itself. Although rather primitive, the algorithm help us show how to solve some of the inverse programming problems numerically. We will use the following marginal value theorem proved by van Rooyen and Zlobec in [26] stated for convex models.

Theorem 3.21 *Consider a convex model (P,θ) with a realistic objective function at some $\theta^\star$. Let $S(\theta^\star)$ be an arbitrary region of stability at $\theta^\star$. Assume that the mapping $F_\star^=$ is lower semicontinuous at $\theta^\star$ relative to $S(\theta^\star)$, and that the saddle-point $(\tilde{x}(\theta^\star),\tilde{u}(\theta^\star))$ is unique. Also suppose that the gradients $\nabla f^k(x,\theta)$, $k\in\{0\}\cup\mathcal{P}^<(\theta^\star)$ are continuous at $(\tilde{x}(\theta^\star),\theta^\star)$. Then for every sequence $\theta\in S(\theta^\star)$, $\theta\to\theta^\star$, and $\tilde{x}(\theta)\to\tilde{x}(\theta^\star)$, for which the limits*

$$l = \lim_{\substack{\theta\in S(\theta^\star)\\ \theta\to\theta^\star}} \frac{\theta-\theta^\star}{\|\theta-\theta^\star\|} \quad \text{and} \quad z = \lim_{\substack{\theta\in S(\theta^\star)\\ \theta\to\theta^\star}} \frac{\tilde{x}(\theta)-\tilde{x}(\theta^\star)}{\|\theta-\theta^\star\|}$$

exist, we have

$$\lim_{\substack{\theta\in S(\theta^\star)\\ \theta\to\theta^\star}} \frac{\tilde{f}(\theta)-\tilde{f}(\theta^\star)}{\|\theta-\theta^\star\|} = \nabla_x\mathcal{L}_\star^<(\tilde{x}(\theta^\star),\tilde{u}(\theta^\star);\theta^\star)z + \nabla_\theta\mathcal{L}_\star^<(\tilde{x}(\theta^\star),\tilde{u}(\theta^\star);\theta^\star)l. \tag{3.57}$$

For the existence of limits and nonuniqueness of the saddle point, see [24] or [26].

To simplify the problem, we will consider only the conditions under which, for every $\theta \in S \cap S(x^\star)$,

$$\nabla_x \mathcal{L}_\star^{\leq}(\tilde{x}(\theta), \tilde{u}(\theta); \theta) z = 0. \tag{3.58}$$

The marginal value formula is significantly simplified under such conditions. Slater's condition is one such condition (see, e.g., [34]). Another condition is when the perturbations of θ are restricted to the region

$$\mathcal{S} = \{\theta \in I : \mathcal{P}^{=}(\theta) = \mathcal{P}^{=}(\theta^0) \text{ and } F^{=}(\theta) = F^{=}(\theta^0)\}.$$

(For more details on these see [34].)

Using the primal formulation in Corollary 3.10, we first determine the regions S and $S(x^\star)$ and so $S \cap S(x^\star)$. Then We will apply the marginal value formula to the inverse model $(LI, x^\star, \theta)$, the optimal value of which is $\tilde{g}(\theta)$, to find a direction along which $\tilde{g}(\theta)$ increases. The numerical method is iterative, and each iteration yields a vector θ^k such that $\tilde{g}(\theta^{k-1}) < \tilde{g}(\theta^k) \leq 0$.

It starts at $\theta^k = \theta^0$ and at each iteration the linear program $(LI, x^\star, \theta^k)$ is solved. Since (L, θ^k) and $(LI, x^\star, \theta^k)$ both have the same optimal solutions, to solve $(LI, x^\star, \theta^k)$ it is enough to solve (L, θ^k) and then determine $\tilde{g}(\theta^k)$ from

$$\tilde{g}(\theta^k) = c^T(\theta^k) x^\star - \tilde{f}(\theta^k),$$

where $\tilde{f}(\theta^k)$ is the optimal value of (L, θ^k). Besides, $\tilde{u}(\theta^k)$ can also be determined by solving (L, θ^k). Then we calculate

$$\nabla_\theta (\mathcal{L}_\star^{\leq})_I(\tilde{x}(\theta^k), \tilde{u}(\theta^k); \theta^k),$$

where $(\mathcal{L}_\star^{\leq})_I(x, u; \theta)$ is the Lagrangian function for the inverse model $(LI, x^\star, \theta)$. Since, at any $\theta^k \in S \cap S(x^\star)$ for which $x^\star$ is not optimal, $\tilde{g}(\theta^k)$ is strictly negative, we want the product $\nabla_\theta (\mathcal{L}_\star^{\leq})_I(\tilde{x}(\theta^k), \tilde{u}(\theta^k); \theta^k) l$ to be positive. We choose the path emanating from θ^k to θ^{k+1} to be linear, i.e.,

$$\theta^{k+1} = \theta^k + \alpha d, \quad \alpha \geq 0, \quad d \in R^p.$$

The limit l defined and used in the marginal value formula can therefore be expressed as

$$l = \lim_{\theta \to \theta^\star} \frac{\theta - \theta^\star}{\|\theta - \theta^\star\|} = \lim_{\alpha \to 0^+} \frac{\alpha d}{\alpha \|d\|} = \frac{d}{\|d\|}.$$

Hence two major decisions are now required for each iteration:

(i) Selection of a direction of ascent d, emanating from θ^k;

(ii) Selection of a stepsize $\alpha > 0$ along d such that $\theta^k + \alpha d \in S \cap S(x^*)$ and $\tilde{g}(\theta^k + \alpha d) > \tilde{g}(\theta^k)$.

First let us determine the direction of ascent d. We will use a simple rule to determine the direction d. This choice of d as well as a more efficient one were studied in [7].

Definition 3.7 *The vector d is said to be a direction of steepest ascent at θ^* if $\|d\| = 1$ and d maximizes*

$$\lim_{\alpha \to 0^+} \frac{\tilde{f}(\theta^* + \alpha d) - \tilde{f}(\theta^*)}{\alpha}. \tag{3.59}$$

Theorem 3.22 *Consider the convex model (P, θ) with a realistic objective function at θ^*, where all the functions are differentiable at θ^*. Assume that $S(\theta^*)$ is a region of stability for (P, θ) at θ^*. Furthermore, assume that the saddle-point $(\tilde{x}(\theta^*), \tilde{u}(\theta^*))$ is unique and*

$$\nabla_x \mathcal{L}_*^{<}(\tilde{x}(\theta^*), \tilde{u}(\theta^*); \theta^*) z = 0.$$

Then for any direction $d \neq 0$ emanating from θ^*, that satisfies

$$\theta^* + \alpha d \in S(\theta^*), \quad 0 < \alpha \leq \bar{\alpha}, \ \textit{for some } \bar{\alpha} > 0,$$

we have

$$\lim_{\alpha \to 0^+} \frac{\tilde{f}(\theta^* + \alpha d) - \tilde{f}(\theta^*)}{\alpha} = \nabla_\theta \mathcal{L}_*^{<}(\tilde{x}(\theta^*), \tilde{u}(\theta^*); \theta^*) d. \tag{3.60}$$

Proof: Let $\theta = \theta^* + \alpha d, \ \ 0 < \alpha \leq \bar{\alpha}$. Then

$$\lim_{\theta \to \theta^*} \frac{\tilde{f}(\theta) - \tilde{f}(\theta^*)}{\|\theta - \theta^*\|} = \lim_{\alpha \to 0^+} \frac{\tilde{f}(\theta^* + \alpha d) - \tilde{f}(\theta^*)}{\alpha \|d\|}.$$

On the other hand, by (3.57),

$$\begin{aligned} \lim_{\theta \to \theta^*} \frac{\tilde{f}(\theta) - \tilde{f}(\theta^*)}{\|\theta - \theta^*\|} &= \nabla_\theta \mathcal{L}_*^{<}(\tilde{x}(\theta^*), \tilde{u}(\theta^*); \theta^*) l \\ &= \nabla_\theta \mathcal{L}_*^{<}(\tilde{x}(\theta^*), \tilde{u}(\theta^*); \theta^*) \frac{d}{\|d\|}. \end{aligned}$$

Therefore,

$$\lim_{\alpha \to 0^+} \frac{\tilde{f}(\theta^* + \alpha d) - \tilde{f}(\theta^*)}{\alpha \|d\|} = \nabla_\theta \mathcal{L}_*^{<}(\tilde{x}(\theta^*), \tilde{u}(\theta^*); \theta^*) \frac{d}{\|d\|}.$$

Since $d \neq 0$, (3.60) follows.

$\square$

Remark: It follows, from Theorem 3.22 and under the same assumptions, that if, in addition, the model is stable at $\theta^\star$, then there exists a direction $d \in B = \{d \in R^p : \|d\| = 1\}$ such that d maximizes the limit (3.59). This d is an optimal solution of the program

$$(G) \qquad \begin{array}{ll} \underset{(d)}{\text{Max}} & \nabla_\theta \mathcal{L}_\star^<(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star)d \\ \text{s.t.} & \\ & \|d\| = 1. \end{array}$$

□

Any norm can be used in (G). We will use the l_∞ norm. Thus, d is an optimal solution of the following program:

$$\begin{array}{ll} \underset{(d)}{\text{Max}} & \nabla_\theta \mathcal{L}_\star^<(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star)d \\ \text{s.t.} & \\ & \|d\|_\infty = 1. \end{array}$$

Clearly, if $(\nabla_\theta \mathcal{L}_\star^<(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star))_i = 0,\ i \in \mathcal{P}^<(\theta^\star)$, then any d of unit uniform norm is an optimal solution. Otherwise, an optimal solution is

$$d_i = sgn(\nabla_\theta \mathcal{L}_\star^<(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star))_i, \quad i \in \mathcal{P}^<(\theta^\star).$$

Therefore, considering $(LI, x^\star, \theta)$ and using l_∞ norm, at each iteration we choose d as follows: If

$$(\nabla_\theta (\mathcal{L}_\star^<)_I(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star))_i = 0, \quad i \in \mathcal{P}^<(\theta^\star),$$

then we simply set $d = 0$ and the method terminates. Otherwise, we set

$$d_i = sgn(\nabla_\theta (\mathcal{L}_\star^<)_I(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star))_i, \quad i \in \mathcal{P}^<(\theta^\star). \tag{3.61}$$

The resulting direction d is such that the product $\nabla_\theta (\mathcal{L}_\star^<)_I(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star)l > 0$ (note that here $\|d\| = 1$). If

$$|\nabla_\theta (\mathcal{L}_\star^<)_I(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star))_i| < \epsilon$$

for some $i \in \mathcal{P}^<(\theta^\star)$ and for some predetermined $\epsilon > 0$, then we set $d_i = 0$. If θ^k is located on a boundary of $S \cap S(x^\star)$ and some $d_i e$(here $e = (e_j)$, where $e_j = 1$ if $j = i$ and $e_j = 0$ otherwise,) points out of $S \cap S(x^\star)$, then that d_i is simply reset to zero. If, at this point, the resulting direction d becomes zero, only then do we solve

$$\begin{array}{ll} \underset{(d)}{\text{Max}} & \nabla_\theta (\mathcal{L}_\star^<)_I(\tilde{x}(\theta^\star), \tilde{u}(\theta^\star); \theta^\star)d \\ \text{s.t.} & \\ & \theta^k + d \in S \cap S(x^\star) \\ & \|d\|_\infty \leq 1. \end{array}$$

If we still get $d = 0$, then the method terminates. Otherwise, we normalize d and continue iterating.

After determining the direction of ascent d, we must determine the stepsize along d, i.e., α, where $\theta^{k+1} = \theta^k + \alpha d$. To do this, we may first determine $\bar{\alpha} > 0$ which is the greatest distance along d for which $\theta^k + \bar{\alpha} d \in S \cap S(x^*)$. Once $\bar{\alpha}$ is known, α could be determined by the Method of Golden Rule (see, e.g., [1,29]). This is a one-dimensional search to determine the point that maximizes a real valued function over a closed interval. This method, however, is for a unimodal function (see [1]), whereas the function $\tilde{g}(\alpha) = \tilde{g}(\theta^k + \alpha d)$ may not be unimodal on the interval $[0, \alpha]$. Nevertheless, using the method of Golden Rule may still yield an $\alpha \in [0, \alpha]$ such that $\tilde{g}(\alpha) > \tilde{g}(0)$.

We recall the following description of the Method of Golden Rule from [29]. The Method of Golden Rule is iterative and requires the use of the Fibonacci fractions

$$F_1 = \frac{3 - \sqrt{5}}{2} \approx 0.381966 \quad \text{and} \quad F_2 = \frac{\sqrt{5} - 1}{2} \approx 0.618034.$$

Given an interval $[a, b]$ of lenght $l = b - a$, the next interval is selected as follows. Let

$$w_1 = a + F_1 l \quad \text{and} \quad w_2 = a + F_2 l = b - F_1 l$$

be points on the interval. If $\tilde{g}(w_1) > \tilde{g}(w_2)$, then the new interval is $[a, w_2]$. If $\tilde{g}(w_1) < \tilde{g}(w_2)$, then the new interval is $[w_1, b]$. If $\tilde{g}(w_1) = \tilde{g}(w_2)$, then the new interval is either $[a, w_2]$ or $[w_1, b]$. The iterations are continued until the interval being considered has a length less than some predetermined $\epsilon > 0$.

Applying the above Method, a suitable α is chosen. Recall that the function $\tilde{g}(\alpha)$ is not explicitly known, yet can be evaluated at any α, $\alpha \in [0, \bar{\alpha}]$. Since the Method of Golden Rule is an expensive method, there are various stopping rules for it. For example, if

$$\left| \frac{\tilde{g}(\theta^{k-2}) - \tilde{g}(\theta^{k-1})}{\tilde{g}(\theta^{k-2})} \right| \geq \delta,$$

for some predetermined $\delta > 0$, then only three iterations of Golden Rule could be performed to get an α on the k_{th} iteration. If the change is less than δ, then the Golden Rule Method is carried out until the interval being considered has a length less than some $\epsilon > 0$. It is possible that after three iterations $\tilde{g}(\alpha) < \tilde{g}(0)$. This may happen in case $\tilde{g}(\alpha)$ increases close to $\alpha = 0$ and quickly drops from then on. In this case the iterations are continued until an α is found with $\tilde{g}(\alpha) > \tilde{g}(0)$ or until there have been a certain number of iterations to guard against the infinite loops. We can therefore give the following primitive algorithm.

Algorithm

1. Determine S and $S(x^*)$ and hence $S \cap S(x^*)$.

2. Set $\theta^k = \theta^0$, and specify $\epsilon > 0$ and $\delta > 0$.

3. Update the matrix $A = A(\theta^k)$ and the vectors $b = b(\theta^k)$ and $c = c(\theta^k)$.

4. Solve (LI, x^*, θ^k) to get $\tilde{x}(\theta^k)$ and $\tilde{u}(\theta^k)$ and $\tilde{g}(\theta^k)$. If $\tilde{g}(\theta^k) = 0$ or $|\tilde{g}(\theta^k)| < \epsilon$ for some predetermined $\epsilon > 0$, stop; in the first case θ^k is an IOI and in the second case θ^k is an ϵ-approximation of an IOI for (LI, x^*, θ); if not continue.

5. Calculate $\nabla_\theta(\mathcal{L}_*^{\leq})_I(\tilde{x}(\theta^k), \tilde{u}(\theta^k); \theta^k)$, and determine d as previously described. If $d = 0$, stop; θ^k is a candidate for an AOI for (LI, x^*, θ); if not, continue.

6. Calculate $\bar{\alpha}$ and then α, by the Method of Golden Rule as described above. If

$$\left| \frac{\tilde{g}(\theta^{k-2}) - \tilde{g}(\theta^{k-1})}{\tilde{g}(\theta^{k-2})} \right| \geq \delta,$$

then perform only three Golden Rule iterations. Otherwise, iterate until the interval $[a, b]$ has length $l = b - a < \epsilon$.

7. Set $\theta^{k+1} = \theta^k + \alpha d$, and repeat from step 3.

□

Clearly, a development of numerically efficient methods for solving the inverse problem is a matter left for future research. More sophisticated methods could utilize some ideas from the predictor-corrector path-following methods, recently published in [13].

3.6 A Case Study

We will show how to apply the theoretical results of inverse programming to solve a simple case study. This case study has been selected from [15], with extra assumptions and modifications added to turn it into a suitable linear model for studying the inverse problem. This problem dates to before 1973, so some of the data may no longer be realistic.

Giant Transistor (GT), Inc., manufactures two products: regular transistors and the newer giant transistors. The plant operates at capacity (three shifts, five days a week). The Vice President in Charge of Operations conducted a study that resulted in setting up an optimum operating point for the whole plant. Following are the highlights of that study.

GT employs 625 workers per shift in the capacity of direct labour. The available labour per week was calculated as follows:

(625 workers per shift)×(three shifts)=1,875 workers

(1,875 workers)×(40 hours per week per worker)=75,000 hours per week

(75,000 hours per week)×(60 minutes per hour)=4,500,000 minutes per week.

The average wage for the entire plant is \$2.10 an hour. Studies have shown that ten minutes of labour are required for each regular transistor produced, and twenty minutes of labour are required for each giant transistor produced. For the regular transistors, the estimated labour cost per unit is

(1/6 hour per unit)×(\$2.10 an hour)=\$0.35 per unit.

For the giant transistors, the estimated labour cost per unit is

(1/3 hour per unit)×(\$2.10 an hour)=\$0.70 per unit.

For regular transistors the average material cost per unit is \$0.15, and for giant transistors it is \$0.3. The total decision cost per unit of regular transistors is

(\$0.35 for labour)+(\$0.15 for material)=\$0.50 .

The total cost per unit of giant transistors is

(\$0.7 for labour)+(\$0.30 for material)=\$1.00.

Since GT sells regular transistors for \$0.82 per unit, the profit is

\$0.82-\$0.50=\$0.32 per unit.

Giant transistors are sold for \$1.63 per unit; hence, the profit is

\$1.63-\$1.00=\$0.63 per unit.

Other than the labour constraint of 4,500,000 minutes per week for the total three shifts, the only other critical areas are those of soldering and final assembly. GT has 84 electronic soldering machines that are capable of running 24 hours per day. The available soldering time was calculated as:

(60 minutes/hour)×(8 hours/shift)×(3 shifts/day)× (5 days/week)

×(84 machines) =604,800 minutes each week.

GT has 65 final assembly machines that are capable of running 24 hours per day. The available final assembly time was calculated as:

(60 minutes/hour)×(8 hours/shift)×(3 shifts/day)× (5 days/week)

×(65 machines)=468,000 minutes each week.

Regular transistors require one minute of soldering time per unit and 1.1 minute of final assembly time per unit. Giant transistors require 3 minutes of soldering time per unit and 1.7 minutes of final assembly time per unit. The variables were defined as follows :

x_1 =number of regular transistors produced each week;

x_2 =number of giant transistors produced each week.

The objective is to maximize profit. These lead to the following program:

$$(L) \quad \begin{array}{ll} \text{Max} & 0.32x_1 + 0.63x_2 \\ \text{s.t.} & \\ & 10x_1 + 20x_2 \leq 4,500,000 \\ & x_1 + 3x_2 \leq 604,800 \\ & 1.1x_1 + 1.7x_2 \leq 468,000 \\ & x_i \geq 0, \quad i = 1, 2. \end{array}$$

The optimal solution of this program is

$$\tilde{x} = (342000,\ 54000)^T,$$

and its optimal value is

$$\tilde{f} = 143,460.$$

GT has been following this plan and has been producing and selling 342,000 regular transistors per week and 54,000 giant transistors per week.

In what follows we will add extra assumptions to the above problem to turn it into a mathematical model and then apply some of the inverse programming results to it. Assume that, while reviewing the new contracts offered, the operating officers of GT have noticed an increased demand for giant transistors while the demand for regular transistors has decreased. In particular, assume that one of the interesting contracts offered to them requires production of 140,400 regular transistors and 150,000 giant transistors per week. The operating officers of GT have to decide whether or not they are able to make their optimal production plan to be

$$x^\star = (140400,\ 150000)^T$$

by first introducing some parameters in the program and then by perturbing them in a stable way. Note that currently $x^\star$ is an interior point of the feasible set of the program and is not optimal.

Assume that by using improved equipment the labour time for production of a unit of giant transistor can be reduced from 20 minutes by anywhere up to 2 minutes (i.e., to a minimum of 18 minutes). In addition, assume that GT can reassign some of the workers to another plant, so that the total labour time in this plant can be reduced from 4,500,000 minutes/week to as low as 3,937,500 minutes/week. Finally, assume that the available soldering time can be reduced from 604,800 minutes/week to as low as 590,400 minutes/week (reducing the available soldering time means reducing the number of hours the soldering machines are running per day thus saving energy). Therefore the following parameters are introduced :

$$a_{12} = 20(1 + \theta_1) \text{ where } -0.1 \leq \theta_1 \leq 0$$

$$b_1 = 4,500,000(1+\theta_2) \text{ where } -0.125 \le \theta_2 \le 0$$

$$b_2 = 604,800(1+\theta_3) \text{ where } -0.02380952 \le \theta_3 \le 0.$$

(Note that here eight decimal digits are used for the lower bound on θ_3 to avoid roundoff errors.)

On the other hand, changing the coefficient a_{12} affects the coefficient c_2, the decision profit for a giant transistor, through

$$c_2 = 1.63 - \left(\frac{a_{12}}{60} \times 2.10 + 0.3\right).$$

Hence

$$c_2 = 1.63 - \left(\frac{20(1+\theta_1)}{60} \times 2.10 + 0.3\right)$$

or

$$c_2 = 0.63 - 0.7\theta_1.$$

Therefore, the following model is set up:

$$(L,\theta) \qquad \begin{array}{ll} \underset{(x)}{\text{Max}} & 0.32x_1 + (0.63 - 0.7\theta_1)x_2 \\ \text{s.t.} & \\ & 10x_1 + 20(1+\theta_1)x_2 \le 4,500,000(1+\theta_2) \\ & x_1 + 3x_2 \le 604,800(1+\theta_3) \\ & 1.1x_1 + 1.7x_2 \le 468,000 \\ & x_1 \ge 0 \quad x_2 \ge 0 \end{array},$$

where

$$\theta \in I = \left\{ \theta \in R^3 : \begin{array}{r} -0.1 \le \theta_1 \le 0 \\ -0.125 \le \theta_2 \le 0 \\ -0.02380952 \le \theta_3 \le 0 \end{array} \right\}.$$

Note that here the input (parameter) θ appears in the matrix A and in the vectors b and c. If a choice is possible, one must choose carefully which elements of the above coefficient matrix and vectors should depend on θ and which should remain constant. Perturbing a particular element may have no effect on the optimal value function, whereas a small perturbation in another element may well affect it. To establish which elements are most sensitive to small perturbations, one can use the following well-known formulae (see, e.g., [34]):

$$\frac{\partial \tilde{f}}{\partial c_l} = x_l^*; \quad \frac{\partial \tilde{f}}{\partial b_k} = -u_k^*; \quad \text{and} \quad \frac{\partial \tilde{f}}{\partial a_{kl}} = u_k^* x_l^*, \tag{3.62}$$

where $l = 1,\ldots,n$ and $k = 1,\ldots,m$. Note that in the above formulae

$$\tilde{f}(\theta) = -c^T(\theta)\tilde{x}(\theta).$$

Furthermore, the second formula yields the economic interpretation of Lagrange multipliers as "shadow prices".

Recall that, given the feasible point

$$x^\star = (140400,\ 150000)^T \text{ at } \theta^0 = (0,0,0)^T,$$

we want to see if it is possible to make $x^\star$ optimal by changing the introduced parameters in a stable way and, if so, find the corresponding $\theta^\star$ and a stable path from θ^0 to $\theta^\star$. First, we have to determine the region S, that is the set of θ's for which (L,θ) is stable (i.e., it has a realistic objective function and the mapping F is lower semicontinuous). The point $\hat{x} = (1,1)^T$ is Slater's point for every $\theta \in I$. Besides

$$A(\theta) > 0, \quad b(\theta) > 0 \text{ and } x \geq 0.$$

These imply that $F(\theta)$ is bounded. Therefore, $\tilde{F}(\theta)$ is a nonempty and bounded set and the mapping F is lower semicontinuous for all $\theta \in I$; hence $S = I$. Then we must determine $S(x^\star)$. Note that $x^\star$ satisfies the second and third constraints of (L,θ) for every

$$\theta \in \{\theta \in I : -0.02380952 \leq \theta_3 \leq 0\}.$$

For $x^\star$ to satisfy the first constraint of (L,θ), it must satisfy

$$10(140,400) + 20(1+\theta_1)(150,000) \leq 4,500,000(1+\theta_2),$$

or

$$2\theta_1 - 3\theta_2 \leq 0.064.$$

Therefore

$$S \cap S(x^\star) = \left\{ \theta \in R^3 : \begin{array}{r} 2\theta_1 - 3\theta_2 \leq 0.064 \\ -0.1 \leq \theta_1 \leq 0 \\ -0.125 \leq \theta_2 \leq 0 \\ -0.02380952 \leq \theta_3 \leq 0 \end{array} \right\}.$$

We now set the following model (the inverse model) suitable for solving the corresponding inverse problem.

$$\begin{array}{ll} & \underset{(x)}{\text{Min}} \ f^0 = -0.32x_1 - (0.63 - 0.7\theta_1)x_2 - 105,000\theta_1 + 139,428 \\ & \text{s.t.} \\ & \quad f^1 = 10x_1 + 20(1+\theta_1)x_2 - 4,500,000(1+\theta_2) \leq 0 \\ & \quad f^2 = x_1 + 3x_2 - 604,800(1+\theta_3) \leq 0 \\ (LI, x^\star, \theta) & \quad f^3 = 1.1x_1 + 1.7x_2 - 468,000 \leq 0 \\ & \quad f^4 = -x_1 \leq 0 \\ & \quad f^5 = -x_2 \leq 0, \end{array}$$

where $\theta \in S \cap S(x^\star)$ and $\theta^0 = (0,0,0)^T$.

Numerical solution of the inverse problem

The algorithm given in the previous section will now be used to solve the problem. Recall that $\tilde{g}(\theta)$ is the optimal value of $(LI, x^\star, \theta)$ and $\tilde{g}(\alpha) = \tilde{g}(\theta^k + \alpha d)$ at the k_{th} iteration. We will look for $\theta^\star \in S \cap S(x^\star)$ that maximizes $\tilde{g}(\theta)$ with respect to $S \cap S(x^\star)$. If $\tilde{g}(\theta^\star) = 0$, then $\theta^\star$ will be an IOI for $(LI, x^\star, \theta)$. Since Slater's condition is satisfied at any $\theta \in S \cap S(x^\star)$, the Lagrangian for $(LI, x^\star, \theta)$ will be the usual Lagrangian at any $\theta \in S$. That is

$$\begin{aligned}\mathcal{L}_I(x, u; \theta) \;=\; & -0.32x_1 - (0.63 - 0.7\theta_1)x_2 - 105{,}000\theta_1 + 139{,}428 \\ & + u_1(10x_1 + 20(1+\theta_1)x_2 - 4{,}500{,}000(1+\theta_2)) \\ & + u_2(x_1 + 3x_2 - 604{,}800(1+\theta_3)) \\ & + u_3(1.1x_1 + 1.7x_2 - 468{,}000) \\ & + u_4(-x_1) + u_5(-x_2).\end{aligned}$$

First iteration: At $\theta = \theta^0$ we have the following situation:

$$\tilde{x}(\theta^0) = \begin{bmatrix} 342{,}000 \\ 54{,}000 \end{bmatrix}; \quad \tilde{u}(\theta^0) = \begin{bmatrix} 0.0298 \\ 0 \\ 0.0199 \end{bmatrix}; \quad \tilde{g}(\theta^0) = -4{,}032.$$

This gives

$$(\nabla_\theta \mathcal{L}_I(\tilde{x}(\theta^0), \tilde{u}(\theta^0); \theta^0))^T = \begin{bmatrix} -35{,}016 \\ -134{,}100 \\ 0 \end{bmatrix}.$$

Note that $u_2(\theta^0) = 0$ and hence, by (3.62),

$$\frac{\partial \tilde{g}}{\partial \theta_3^0} = -604{,}800\tilde{u}_2(\theta^0) = 0\ .$$

Therefore, small perturbations of θ_3^0 do not affect $\tilde{g}(\theta)$. (However, as we will see later on, at other points on the path from θ^0 to $\theta^\star$, small perturbations of θ_3 will well affect $\tilde{g}(\theta)$. The direction d, using (3.61), can be determined as

$$d = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}.$$

Then we determine $\bar{\alpha}$, the biggest step that can be taken along d and still remain in the set $S \cap S(x^\star)$. Substituting

$$\theta_1 = 0 - \beta,\ \ \theta_2 = 0 - \beta \text{ and } \theta_3 = 0,$$

for $\theta \in S \cap S(x^*)$, then $\bar{\alpha} > 0$ must be the largest $\beta > 0$ that satisfies

$$-2\beta + 3\beta \leq 0.064, \quad 0 \leq \beta \leq 0.1 .$$

This gives $\bar{\alpha} = 0.064$. At this point, we will determine $\alpha > 0$ by applying the Golden Rule Method to the function $\tilde{g}(\theta)$ on the interval $[0, \bar{\alpha}]$. After three Golden Rule iterations, we obtain

$$\alpha = \bar{\alpha} = 0.064$$

and

$$\theta^1 = \theta^0 + \alpha d = \begin{bmatrix} -0.064 \\ -0.064 \\ 0 \end{bmatrix} .$$

Second iteration: At $\theta = \theta^1$ we have the following situation:

$$\tilde{x}(\theta^1) = \begin{bmatrix} 116,502.128 \\ 162,765.957 \end{bmatrix}, \quad \tilde{u}(\theta^1) = \begin{bmatrix} 0.0253 \\ 0.0672 \\ 0 \end{bmatrix} \text{ and } \tilde{g}(\theta^1) = -967.149,$$

which gives

$$(\nabla_\theta \mathcal{L}_I(\tilde{x}(\theta^1), \tilde{u}(\theta^1); \theta^1))^T = \begin{bmatrix} 91,242.65 \\ -113,776.6 \\ -40,620.25 \end{bmatrix} .$$

Note that here $\tilde{u}_2(\theta^1) \neq 0$, so, by (3.62), we have

$$\frac{\partial \tilde{g}}{\partial \theta_3^1} = -604,800\tilde{u}_2(\theta^1) = -40,642 .$$

Therefore, changes in θ_3^1 do affect $\tilde{g}(\theta)$. This implies that d can be chosen to be

$$d = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} .$$

But θ^1 is on the boundary of $S \cap S(x^*)$ and any perturbation along d_1 or along d_2 points out of the set $S \cap S(x^*)$. Thus, as discussed earlier, we simply reset $d_1 = d_2 = 0$ (to remain in that set). Therefore

$$d = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} .$$

It then follows that $\bar{\alpha} = 0.02380952$. But $\tilde{g}(\bar{\alpha}) = 0$, so $\alpha = \bar{\alpha}$ and

$$\theta^* = \theta^2 = \theta^1 + \bar{\alpha} d = \begin{bmatrix} -0.064 \\ -0.064 \\ -0.02380952 \end{bmatrix} .$$

This θ is an IOI for $(LI, x^\star, \theta)$ since $\tilde{g}(\theta^\star) = 0$. Furthermore

$$a_{12}(\theta^\star) = 18.72, \;\; b_1(\theta^\star) = 4,212,000. \;\; b_2(\theta^\star) = 590\ 100, \;\; c_2(\theta^\star) = 0.6748$$

and $\tilde{f}(\theta^\star) = 146,148$. Comparing $\tilde{f}(\theta^\star)$ with $\tilde{f}(\theta^0) = 143,160$, we notice an increase of 1.874% in the profit. Note that here any path in the set $S \cap S(x^\star)$ is stable. Thus, any path in this set from θ^0 to $\theta^\star$ is a solution to the inverse problem.

Here we happened to obtain an IOI for which the profit was more than the initial profit. However, this may not always be the case. In general, if we want the profit to increase or remain constant along a stable path, then besides maximizing $g(\theta)$ we must also maximize the difference between the profit at any θ on the chosen stable path and the initial profit. This shows that the inverse problem, applied in real-life situations, is typically a multi-objective problem.

The dual numerical approach

The dual model for studying the inverse problem is

$$(DI, x^\star, \theta) \qquad \begin{array}{ll} \underset{(v)}{\text{Min}} & 4,500,000(1+\theta_2)v_1 + 604,800(1+\theta_3)v_2 + 468,000v_3 \\ & \qquad + 105,000\theta_1 - 139,428 \\ \text{s.t.} & \\ & -10v_1 - v_2 - 1.1v_3 + 0.32 \le 0 \\ & -20(1+\theta_1)v_1 - 3v_2 - 1.7v_3 + 0.63 - 0.7\theta_1 \le 0 \\ & -v_i \le 0, \;\; i = 1,2,3. \end{array}$$

Since (L, θ) has a realistic objective function at any $\theta \in S \cap S(x^\star)$, by Theorem 3.14, the constraints of (D, θ) and therefore the constraints of $(DI, x^\star, \theta)$ satisfy Slater's condition for every perturbation in $S \cap S(x^\star)$. Furthermore, by Theorem 3.16, the dual inverse model has a realistic objective function for all these perturbations. Therefore we can look for a locally optimal input for $(DI, x^\star, \theta)$, which will be an AOI for $(LI, x^\star, \theta)$ (as discussed earlier) as well. Applying the same method to the model $(DI, x^\star, \theta)$ to find a locally optimal input with the choice

$$d_i = -sgn(\nabla_\theta \mathcal{L}_{DI}(\tilde{v}(\theta^k), \tilde{u}(\theta^k); \theta^k)),$$

for the direction d (see, e.g, [7,34]) at each iteration, we obtain the same solution as we obtained using the primal inverse model. This is because, due to the duality properties, the role of the elements of optimal solution and the corresponding Lagrange multipliers are exchanged in the dual model. Besides, we use the same saddle points (with roles of optimal solution and Lagrange multipliers exchanged) as in the primal approach because of the uniqueness of those points. This results in the same elements (but of opposite sign) in the gradient of the Lagrangian function for $(DI, x^\star, \theta)$, used to determine the direction d at each iteration. More precisely, for example, in the first iteration we have

$$(\nabla_\theta \mathcal{L}_{DI}(\tilde{v}(\theta^0), \tilde{u}(\theta^0); \theta^0))^T = \begin{bmatrix} 35,016 \\ 134,100 \\ 0 \end{bmatrix}.$$

This implies that $d = (-1, -1, 0)^T$, as obtained in the first iteration in the primal approach. However, the primal and dual numerical approches do not always yield the same solution when the saddle points along the stable paths are not unique.

A "direct" solution of the problem

Let us note that, in general, finding an IOI for $(LI, x^\star,$ directly (e.g., using the Karush-Kuhn-Tucker conditions) is a difficult task, specially in large linear models with many parameters. Besides, the direct solution does not yield a stable path from θ^0 to $\theta^\star$.

In the model for the above case study $((L, \theta))$ only the constraints f^1 and f^2 depend on θ and the rest of the constraints are not active at $x^\star$. Therefore, to find values of $\theta \in S \cap S(x^\star)$ for which $x^\star$ is optimal, we must solve the following KKT system:

$$\begin{aligned} 10u_1^\star + u_2^\star &= 0.32 \\ (20 + 200_1)u_1^\star + 3u_2^\star &= 0.63 - 0.7\theta_1 \\ u_1^\star f^1(x^\star, \theta) &= 0 \\ u_2^\star f^2(x^\star, \theta) &= 0 \\ \theta &\in S \cap S(x^\star). \end{aligned}$$

This gives, after some manipulation, the following set of all $\theta \in R^3$ at which $x^\star$ becomes an optimal solution of (L, θ):

$$S_I = \left\{ \theta \in R^3 : \begin{array}{r} 2\theta_1 - 3\theta_2 = 0.064 \\ -0.1 \le \theta_1 \le -0.0074627 \\ \theta_3 = -0.02380952 \end{array} \right\}.$$

Again, to avoid roundoff error, seven decimal digits are used for the upper bound on θ_1. It can be easily seen that the IOI that we obtained earlier numerically, i.e.,

$$\theta^\star = \begin{bmatrix} -0.064 \\ -0.064 \\ -0.02380952 \end{bmatrix}$$

also belongs to the set S_I.

An advantage of this direct approach is that we can find a $\theta \in S_I$ closest to θ^0. To find such closest θ to θ^0, we must solve the program

$$\begin{aligned} \text{Min} \quad & \|\theta^0 - \theta\| \\ \text{s.t.} \quad & \\ & \theta \in S_I \end{aligned}$$

that is

$$\begin{aligned} \text{Min} \quad & |\theta_1| + \tfrac{1}{3}|2\theta_1 - 0.064| \\ \text{s.t.} \quad & \\ & -0.1 \le \theta_1 \le -0.0074627 \; . \end{aligned}$$

This implies that the closest θ, in S_I, to θ^0 is

$$\hat{\theta} = \begin{bmatrix} -0.0074627 \\ -0.0263085 \\ -0.02380952 \end{bmatrix}.$$

For the sake of curiosity, the furthest θ, in S_I, to θ^0 is

$$\bar{\theta} = \begin{bmatrix} -0.1 \\ -0.088 \\ -0.02380952 \end{bmatrix}.$$

Remarks

(i) It is interesting that the optimal solution of $(L, \hat{\theta})$ is not unique. In fact

$$\tilde{F}(\hat{\theta}) = \left\{ \lambda \begin{bmatrix} 140,400 \\ 150,000 \end{bmatrix} + (1-\lambda) \begin{bmatrix} 380,785.23 \\ 28,903.672 \end{bmatrix}, \; 0 \le \lambda \le 1 \right\}.$$

Besides $\tilde{f}(\hat{\theta}) = 140,211.582$, implying that although $x^\star$ is optimal, the profit is smaller than the initial profit at θ^0. In fact, among the elements of S_I, $\hat{\theta}$ yields the smallest profit. On the contrary, the optimal solution of $(L, \bar{\theta})$ is unique and $\tilde{f}(\bar{\theta}) = 149,928$, which gives a larger profit than the initial profit. Also, among the elements of S_I, $\bar{\theta}$ yields the largest profit.

(ii) Considering the inverse model $(LI, x^\star, \theta)$, then we notice that $\tilde{g}(\theta) = 0$ for every θ in S_I (since $x^\star$ is optimal for any such θ). Furthermore, the set $S \cap S(x^\star)$ is a convex set, $S_I \subset S \cap S(x^\star)$ and we always have $\tilde{g}(\theta) \le 0$ for every $\theta \in S \cap S(x^\star)$. Therefore, any θ in S_I is also an IOI for the model $(LI, x^\star, \theta)$. Besides, it is possible to go from θ^0 to any of the elements of S_I through a straight line. This means that the shortest stable path in $S \cap S(x^\star)$ from θ^0 to an IOI for $(LI, x^\star, \theta)$ is the straight line from θ^0 to $\hat{\theta}$. However, this shortest path would result in a drop in the profit.

(iii) In this particular case study any convex combination of $\hat{\theta}$ and $\bar{\theta}$ is an IOI for $(LI, x^\star, \theta)$. This is, however, not true in general. Consider, for example, the model

$$\begin{aligned} \underset{(x)}{\text{Max}} \quad & x_1 + x_2 \\ \text{s.t.} \quad & \\ & 4x_1 + 2x_2 \leq 14 \\ & \theta_1 x_1 + x_2 \leq 5 \\ & x_1 + \theta_2 x_2 \leq 5 \\ & x_i \geq 0, \ \ i = 1, 2 \end{aligned}$$

starting at $\theta^0 = (0,0)^T$, with $x^\star = (2,3)^T$. Here

$$S \cap S(x^\star) = \{\theta : 0 \leq \theta_1 \leq 1, \ \ 0 \leq \theta_2 \leq 1\},$$

and $\tilde{\theta} = (1,0)^T$ and $\theta^\star = (0,1)^T$ are both IOI's for the model $(LI, x^\star, \theta)$. Now consider $\hat{\theta} = (\frac{1}{2}, \frac{1}{2})^T$. Then $\tilde{x}(\hat{\theta}) = (\frac{4}{3}, \frac{13}{3})$ and $\tilde{f}(\hat{\theta}) = \frac{17}{3}$, implying that $\hat{\theta}$ is not an IOI for the inverse model since $\tilde{g}(\hat{\theta}) = \frac{-2}{3}$.

(iv) Another choice of parameters θ_i, $i = 1, 2, 3$, in our case study would be

$$\begin{aligned} a_{12} &= 18\theta_1 + 20(1 - \theta_1) \\ b_1 &= 3{,}937{,}500\theta_2 + 4{,}500{,}000(1 - \theta_2) \\ b_2 &= 590{,}400\theta_3 + 604{,}800(1 - \theta_3) \end{aligned}$$

which would result in simpler bounds for θ_i, $i = 1, 2, 3$, i.e.,

$$0 \leq \theta_i \leq 1, \quad i = 1, 2, 3.$$

However, this choice would yield a different solution for the problem, i.e., at the corresponding $\theta^\star$, we would have

$$a_{12}(\theta^\star) = 19.2686, \ \ b_1(\theta^\star) = 4{,}294{,}293.8, \ \ b_2(\theta^\star) = 590{,}400, \ \ c_2(\theta^\star) = 0.6556$$

and $\tilde{f}(\theta^\star) = 143{,}268.108$. Note that the profit in this case would be smaller than the initial profit.

A situation where an IOI does not exist

Consider again the program (L), and assume that the only coefficient that can be changed is the assembly time per unit of a regular transistor. That is, assume that the decision makers have decided to slow down the assembly process for the regular transistors since they do not need as many regular transistors as they were initially producing. This causes the assembly time per unit of a regular transistor to become larger. Assume that the following model has therefore been set up:

$$(L,\theta)\qquad \begin{array}{ll} \text{Max} & 0.32x_1 + 0.63x_2 \\ \text{s.t.} & \\ & 10x_1 + 20x_2 \le 4,500,000 \\ & x_1 + 3x_2 \le 604,800 \\ & (1.1+\theta)x_1 + 1.7x_2 \le 468,000 \\ & x_i \ge 0, \quad i = 1,2 \end{array}$$

starting at $\theta = 0$ with $x^\star = (140400,\ 150000)^T$. It can be seen that

$$S \cap S(x^\star) = \{\theta \in R : 0 \le \theta \le 0.417\}.$$

Here, the inverse problem uses the following model:

$$(LI,x^\star,\theta)\qquad \begin{array}{ll} \underset{(x)}{\text{Min}} & -0.32x_1 - 0.63x_2 + 139,428 \\ \text{s.t.} & \\ & 10x_1 + 20x_2 \le 4,500,000 \\ & x_1 + 3x_2 \le 604,800 \\ & (1.1+\theta)x_1 + 1.7x_2 \le 468,000 \\ & x_i \ge 0, \quad i = 1,2. \end{array}$$

Note that $\tilde{g}(\theta^0) = -4,032$. Using the algorithm given at the end of the previous section, we obtain

$$\theta^\star = 0.417,\ \ \tilde{x}(\theta^\star) = (131827.429,\ 157657.524)^T \text{ and } \tilde{g}(\theta^\star) \approx -2,081\ .$$

Obviously, $\theta^\star$ can not be an IOI for $(LI,x^\star,\theta)$. Let us verify that the above $\theta^\star$ is an AOI for $(LI,x^\star,\theta)$ using the the characterization in Theorem 3.15 . Here, the dual inverse model is

$$(DI,x^\star,\theta)\qquad \begin{array}{ll} \underset{(x)}{\text{Min}} & 4,500,000v_1 + 604,800v_2 + 468,000v_3 - 139,428 \\ \text{s.t.} & \\ & -10v_1 - v_2 - (1.1+\theta)v_3 + 0.32 \le 0 \\ & -20v_1 - 3v_2 - 1.7v_3 + 0.63 \le 0 \\ & -v_i \le 0, \quad i = 1,2,3. \end{array}$$

Solving this program at $\theta = \theta^\star$ yields the solution

$$\tilde{v}(\theta^\star) = \begin{bmatrix} 0 \\ 0.144 \\ 0.116 \end{bmatrix}$$

and the shadow prices

$$\tilde{u}(\theta^\star) = \begin{bmatrix} 131,827.429 \\ 157,657.524 \\ 28,575.237 \\ 0 \\ 0 \end{bmatrix}.$$

Recall that $\tilde{h}(\theta^\star) = -\tilde{g}(\theta^\star) \approx 2,081$. We will look for a vector function $\Lambda(\theta) \geq 0$ such that the saddle-point inequality (3.51), in Theorem 3.15, holds in some neighbourhood of $\theta^\star$. In order to find such a function, we first fix $\theta = \theta^\star$ to get $\Lambda(\theta^\star)$, and then use $\Lambda(\theta)$ (by changing θ to $\theta^\star$) as the vector function that we need. The KKT system at $\theta^\star$ results in the following system of equations:

$$\begin{aligned} 10\lambda_1 + 20\lambda_2 + \lambda_3 &= 4,500,000 \\ \lambda_1 + 3\lambda_2 &= 604,800 \\ (1.1 + \theta^\star)\lambda_1 + 1.7\lambda_2 &= 468,000 \\ \lambda_i &= 0 \quad i = 4,5. \end{aligned}$$

Solving this system to get $\Lambda(\theta^\star)$ and then changing $\theta^\star$ to θ, we obtain

$$\begin{aligned} \lambda_1(\theta) &= \frac{375,840}{1.6 + 3\theta} \\ \lambda_2(\theta) &= \frac{197,280 + 604,800\theta}{1.6 + 3\theta} \\ \lambda_3(\theta) &= \frac{-504,000 + 140,4000\theta}{1.6 + 3\theta} \\ \lambda_i(\theta) &= 0 \quad i = 4,5. \end{aligned}$$

The left-hand of the saddle-point inequality in Theorem 3.15 is easily satisfied with this Λ. The right-hand inequality reduces to

$$0 \leq \frac{18,140.8 - 43,503\theta}{1.6 + 3\theta},$$

which holds for all $0 \leq \theta \leq 0.417$. Hence, $\theta^\star = 0.417$ is indeed a globally AOI for $(LI, x^\star, \theta)$ with respect to $S \cap S(x^\star)$. This means that, with the above choice of $\theta^\star$, the point $x^\star$ is the closest possible to being optimal with respect to the perturbations in $S \cap S(x^\star)$. Any path in $S \cap S(x^\star)$ from θ^0 to $\theta^\star$ is a solution to the inverse problem.

Note that here $\hat{f}(\theta^\star) = 141,509$, which is smaller than the initial profit. In this case there are no parameters in the objective function, so it is impossible to increase the profit or even keep it constant. Therefore, to control the profit in this way, we must always have parameters in the objective function.

Conclusions

In the first part of the research we have extended the concept of differentiable LFS functions to the non-smooth case and to the differentiable generalized convex case. Also, new characterizations of optimality are given with respect to a smaller minimal index set rather than the usual minimal set $\mathcal{P}^{=}$. It appears that the programs with LFS functions are the largest identifiable class of programs for which the KKT conditions are necessary for optimality.

Some of the results of the second part of the work (Chapter 3), specially the notion of $\tilde{F}$-stability and characterization of optimal inputs with respect to such regions of stability, are mostly of theoretical rather than practical importance. While the results on the inverse problem (Section 3.4) appear to be new, the numerical algorithm (Section 3.5) is basically a modification of existing numerical methods of input optimization. Therefore, the algorithm is still rather primitive and presents an open problem with many directions for possible improvements. One such improvement, for example, would be to choose a parabolic path from one θ to the next in each iteration, or a combination of straight line segments and parabolae (whichever is better, depending on the current θ), or to find a different method to choose the direction d at each iteration. In addition, a method that employs second-order input optimization results may be faster.

A detailed study of numerical methods in input optimization, and in particular for solving the inverse problem, would relate our stability research to, e.g., the path-following methods of the predictor-corrector type recently developed by Guddat, Guerra Vasguez and Jongen [13] and other researchers in the field of non-smooth optimization [12]. This is a direction of general research which the author may pursue in the near future.

The inverse problem in this research has been studied only for linear models using the elegant duality results in linear programming. Some of the results could possibly be extended to the general convex or even the nonconvex case. In addition, only single-objective models have been studied, leaving the multi-objective case open for further research.

References

[1] Avriel, M., *Nonlinear Programming: Analysis and Methods*, Prentice Hall, Inc. Englewood Cliffs, New Jersey, 1976.

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

[3] Ben-Israel, A., A. Ben-Tal, and A. Charnes, "Necessary and Sufficient Conditions for a Pareto Optimum in Convex Programming," *Econometrica,* 45 (1977) 811-820.

[4] Ben-Israel, A., A. Ben-Tal, and S. Zlobec, *Optimality in Nonlinear Programming: A Feasible Directions Approach*, Wiley-Interscience, New York, 1981.

[5] Ben-Israel, A., and B. Mond, "First-Order Optimality Conditions for Generalized Convex Functions: A Feasible Directions Approach," *Utilitas Mathematica*, Vol. 25 (1984) 249-262.

[6] Ben Tal, A., and A. Ben-Israel, "Characterizations of Optimality in Convex Programming: The Nondifferentiable Case," *Applicable Analysis* (1979) Vol. 9, 137-156.

[7] Brunet, M. P., "Numerical Experimentation with Input Optimization," M.Sc. Thesis, McGill University, 1989.

[8] Charnes, A. and W. Cooper, *Management Models and Industrial Applications to Linear Programming*, Wiley, New York, 1961.

[9] Gass, S. I., *Linear Programming: Methods and Applications*, Fifth Edition, McGraw-Hill Book Company, 1985.

[10] Gauvin, J., "A Necessary and Sufficient Regularity Condition to Have Bounded Multipliers in Nonconvex Programming," *Mathematical Programming* 12 (1977) 136-138.

[11] Geoffrion, A. M., "Proper Efficiency and the Theory of Vector Maximization," *J. Math. Anal. Appli.* 22 (1968) 618-630.

[12] Goffin, J. L., A. Haurie and J. P. Vial, "Decomposition and Nondifferentiable Optimization with the Projective Algorithm," *Management Science*, Vol. 38, No. 2 (1992) 284-302.

[13] Guddat, J., F. Guerra Vasguez and H. Th. Jongen, *Parametric Optimization: Singularities, Pathfollowing and Jumps*, Teubner and Wiley, Stuttgart, 1990.

[14] Hogan, W., "Directional Derivatives for Extremal-Value Functions with Applications to the Completely Convex Case," *Operations Research*, V.21, no.1 (1973) 188-209.

[15] Hughes, A. J., and D. E. Grawiog, *Linear Programming: An Emphasis on Decision Making*, Addison-Wesley, 1973.

[16] Kuhn, H. W. and A. W. Tucker, "Nonlinear Programming," in J. Neymann (ed.) *Proc. Second Berkley Symposium on Mathematical Statistics and Probabilty*, University of California Press (1951) 481-492.

[17] Mangasarian, O., *Nonlinear Programming*, McGraw-Hill, New York, 1969.

[18] Mangasarian, O.L., and J.B. Rosen, "Inequalities for Stochastic Nonlinear Programming Problems," *Operations Research* 12 (1964) 143-154.

[19] Martin, D. H., "On the Continuity of the Maximum in Parametric Linear Programming," *Journal of Optimization Theory and Applications*, Vol. 17, Nos. 3/4 (1975) 205-210.

[20] Pang, J.S., "On Local Minima of Nonlinear Programs," Johns Hopkins University, Maryland, U.S.A., June 1992.

[21] Rockafellar, T., *Convex Analysis*, Princeton University Press, 1970.

[22] Simmons, D. M., *Nonlinear Programming for Operations Research*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[23] Valace, S., "The M-Method in Input Optimization," McGill University, Department of Electrical Engineering, 1990.

[24] Van Rooyen, M., "Stable Parametric Optimization," Ph.D. Thesis, McGill University, 1992.

[25] Van Rooyen, M., and S. Zlobec, "A Complete Characterization of Optimal Economic Systems with Respect to Stable Perturbations," *Glasnik Matematicki* V. 25, no. 45 (1990) 235-253.

[26] Van Rooyen, M., and S. Zlobec, "The Marginal Value Formula on Regions of Stability," *Optimization* (forthcoming).

[27] Vincent, T. L., and W. J. Grantham, *Optimality in Parametric Systems*, John Wiley & Sons, Inc. , 1981.

[28] Wolkowicz, H., "Geometry of Optimality Conditions and Constraint Qualifications: The Convex Case," *Mathematical Programming* 19 (1980) 32-60.

[29] Zangwill, W. I., *Nonlinear Programming: A Unified Approach*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1969.

[30] Zhou, X., "Characterizations of Optimality in Multi-objective Programming," M.Sc. Thesis, McGill University, 1992.

[31] Zhou, X., F. Sharifi Mokhtarian, and S. Zlobec, "A Simple Constraint Qualification in Convex Programming," McGill University, Dept. of Mathematics and Statistics, 1991, submitted for publication.

[32] Zlobec, S., "Characterizing Optimality in Mathematical Programming Models," *Acta Applicandae Mathematicae* 12 (1988) 113-180.

[33] Zlobec, S., "Characterizing Optimality in Nonconvex Optimization," *Yugoslav Journal of Operations Research* 1 (1991) 3-14.

[34] Zlobec, S., "The Marginal Value Formula in Input Optimization," *Optimization* 22 (1991) 341-386.

[35] Zlobec, S., "Two Charcterizations of Pareto Minima in Convex Multicriteria Optimization," *Aplikace Matematiky* 29 (1984) 342-349.

[36] Zlobec, S., *Mathematical Programming*, Lecture Notes (McGill University 1977).

[37] Zlobec, S., "Survey of Input Optimization," *Optimization* 18 (1987) 309-348.

[38] Zlobec, S., and A. Ben-Israel, "Duality in Convex Programming: A Linearization Approach," *Math. Operationsforsch. Statist., Ser. Optimization*, Vol. 10 (1979) No. 2, 171-178.

[39] Zlobec, S., and B. D. Craven, "Stabilization and Determination of the Set of Minimal Binding Constraints in Convex Programming," *Math. Operationsforsch. Statist., Ser. Optimization*, Vol. 12 (1981) 203-220.