Mathematical Programming Study 19 (1982) 77–100.
North-Holland Publishing Company

CHARACTERIZATIONS OF OPTIMALITY WITHOUT CONSTRAINT QUALIFICATION FOR THE ABSTRACT CONVEX PROGRAM

J.M. BORWEIN* and H. WOLKOWICZ**

Department of Mathematics, Dalhousie University, Halifax, N.S. B3H 4H8, Canada

Received 21 June 1979
Revised manuscript received 15 July 1980

We consider the general abstract convex program

(P) minimize $f(x)$, subject to $g(x) \in -S$,

where f is an extended convex functional on X, $g : X \to Y$ is S-convex, S is a closed convex cone and X and Y are topological linear spaces. We present primal and dual characterizations for (P). These characterizations are derived by reducing the problem to a standard Lagrange multiplier problem. Examples given include operator constrained problems as well as semi-infinite programming problems.

Key words: Cone-convex, Locally Convex Topological Vector Space, Optimality Conditions, Subdifferential, Directional Derivative, Faithfully Convex, Lagrange Multipliers, Slater's Condition, Semi-infinite Programs.

1. Introduction

We consider the abstract convex program

$$\text{(P)} \qquad \begin{array}{ll} \text{minimize} & f(x), \\ \text{subject to} & g(x) \in -S, \end{array}$$

where $f : X \to R \cup \{\infty\}$ is convex, $g : X \to Y$, X and Y are real locally convex (Hausdorff) spaces, S is a closed convex cone and g is S-convex. Recently Ben-Israel et al. [3] have presented a characterization of optimality, without a constraint qualification, in the case that the constraint g is given by the finite number of real-valued constraints $g^k(x) \leq 0$, $k = 1, \dots, m$. They relied heavily on the convexity properties of the functions and, in particular, they used the cone of directions of constancy of the 'equality' constraints (see e.g. Abrams and Kerzner [1]).

Many people have considered optimality criteria for the abstract program (P) (see e.g. Holmes [14], Kurcyusz [15], Zowe and Kurcyusz [25], Luenberger [16] and Neustadt [17] and the references therein). Their criteria required a constraint

* Research partially supported by NRC A4493. Presently at the Department of Mathematics, Carnegie-Mellon University, Pittsburgh, PA, U.S.A.

** Research partially supported by NSERC A 3388. Presently at the Department of Mathematics, The University of Alberta, Edmonton, Alta., Canada.

qualification. Craven and Zlobec [8] have extended the work in [3] in order to get a characterization of optimality for (P) that does not require any constraint qualification. They, however, required the following assumptions: (i) the cone S has nonempty interior (this automatically guarantees that the dual cone has a compact base; thus, when they choose a compact subset of the dual cone, they might as well choose a base); (ii) the feasible set contains a relative radial (core) point; (iii) continuity and differentiability properties; and (iv) the infimum is attained. They also used the cone of constancy of the 'equality' constraints in a redundant manner.

In this paper we give a characterization which avoids the above-mentioned assumptions. Rather than attempt to extend the results in [3], our results are based on reducing (P) so as to be able to apply the 'Standard Lagrange Multiplier theorem'. We then give two classes of optimality criteria.

The organization of the paper is as follows. Section 2 presents several preliminary definitions and results. In particular, Lemma 2.2 finds a 'Slater point' for any compact subset of the 'nonequality' constraints; Lemma 2.3 and Corollary 2.1 characterize the existence of a compact base and Slater's condition; and Lemma 2.5 gives a dual relationship between the cone of subgradients and the linearizing cone.

Section 3 presents the 'Standard Lagrange Multiplier theorem' for program (P) with the added constraint $x \in \Omega$, where Ω is a convex subset of X (see Theorem 3.1). Several different types of optimality criteria are given. These criteria use directional derivatives, subgradients and the Lagrangian function.

Section 4 presents the complete characterization of optimality without any constraint qualification (see Theorem 4.1 and Corollary 4.1). These results are derived using the results in Section 3. Several corollaries are also given, including the result which leads to the BBZ conditions [3], (see Corollary 4.2 and the following remarks).

Section 5 contains several examples and applications which illustrate the theory developed in the first four sections.

2. Preliminaries

In this section we present some preliminary definitions and results needed in the sequel. We consider the *convex program*

$$\text{(P)} \qquad \begin{array}{ll} \text{minimize} & f(x), \\ \text{subject to} & g(x) \in -S, \end{array}$$

where $f: X \to R \cup \{\infty\}$ is convex, $g: X \to Y$, X and Y are real locally convex (Hausdorff) spaces, S is a closed convex cone in Y and g is S-convex, i.e. $S + S \subset S$, $\lambda S \subset S$ for all positive λ and for all x, y in X,

$$g(\lambda x+(1-\lambda)y)-\lambda g(x)-(1-\lambda)g(y)\in -S \quad \text{for all } 0<\lambda<1.$$

We let

$$F=\{x\in X: g(x)\in -S\}$$

denote the *feasible set* of (P). X^* and Y^* will denote the *topological duals* of X and Y respectively, both equipped with the w^*-topology (see e.g. [19]). If $K\subset X$ (resp. Y), then

$$K^+=\{\phi\in X^* \text{ (resp. } Y^*\text{)}: \phi k\geq 0 \text{ for all } k\in K\}$$

is the *polar* (cone) of K. When $\Lambda\subset X^*$ (resp. Y^*) then its polar is

$$\Lambda^+=\{x\in X \text{ (resp. } Y\text{)}: \phi x\geq 0 \text{ for all } \phi\in\Lambda\}.$$

The polar is always a closed convex set. Moreover,

$$K^{++}=\overline{\text{cone}}\, K \tag{2.1}$$

and

$$\Big(\bigcap_{i\in I} K_i\Big)^+=\Big(\overline{\sum_{i\in I} K_i^+}\Big), \tag{2.2}$$

where $\overline{\text{cone}}\, K$ denotes the closure of the convex cone generated by K, $K\subset X$ (or X^*), $K_i\subset X$ are closed convex cones and $\sum\cdot$ denotes finite sums (see e.g. [6, 10]). In addition, it is easy to see that $qg(\cdot)$ is a real-valued convex function for each $q\in S^+$.

We now choose $\mathscr{P}\subset S^+$ so that $\mathscr{P}$ is a *generating set* for S^+, i.e.

$$S^+=\overline{\text{cone}}\,\mathscr{P}. \tag{2.3}$$

Note that (2.3) is equivalent to $\mathscr{P}^+=S$. The following lemma is clear.

Lemma 2.1. *Let $x\in X$ and $Q\subset S^+$. Then*

$$qg(x)\leq 0 \textit{ for all } q\in Q, \quad \textit{if and only if } g(x)\in -Q^+, \tag{2.4}$$

$$qg(x)\leq 0 \textit{ for all } q\in \mathscr{P}, \quad \textit{if and only if } x\in F. \tag{2.5}$$

Thus $\mathscr{P}$ may be considered the indexing set for the (convex) constraints

$$g_q(x)=qg(x)\leq 0, \quad q\in\mathscr{P}.$$

Furthermore, we denote the *partial feasible sets* by

$$F^Q=\{x\in X: qg(x)\leq 0 \text{ for all } q\in Q\}.$$

The set of *binding constraints* at $a\in F$, with respect to a set $Q\subset S^+$, is

$$Q(a)=\{q\in Q: qg(a)=0\}.$$

An important subset of P, independent of x, is the *equality set*

$$\mathcal{P}^{=} = \{q \in \mathcal{P}: qg(x) = 0 \text{ for all } x \in F\}.$$

This is the set of indices q for which the constraint $qg(x)$ vanishes on the entire feasible set. The set $\mathcal{P}^{=}$ was used in [1, 3] to characterize optimality for the convex program (P), with a finite number of constraints. It was then used by Craven and Zlobec [8] in their optimality criteria for the program (P).

We now let

$$\mathcal{P}^{<} = \mathcal{P} \backslash \mathcal{P}^{=}.$$

We will need the following property of $\mathcal{P}^{<}$. As mentioned above, we shall assume that Y^* has the w^*-topology.

Lemma 2.2. *Suppose that $K \subset \mathcal{P}^{<}$ is compact in Y^*. Then there exists $\hat{x} \in X$ such that*

$$\hat{x} \in F \quad \text{and} \quad qg(\hat{x}) < 0 \quad \text{for all } q \in K.$$

Proof. *For $x \in F$,* let

$$U(x) = \{q \in Y^*: qg(x) < 0\}.$$

By continuity of the linear functionals $g(x)$, $U(x)$ is open. Moreover, $K \subset \bigcup_{x \in F} U(x)$, since $K \subset \mathcal{P}^{<}$. Therefore, by compactness, we can find $x_1, x_2, \ldots, x_n \in F$ such that

$$K \subset \bigcup_{i=1}^{n} U(x_i).$$

Let

$$\hat{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Then $\hat{x} \in F$ and

$$\begin{aligned} qg(\hat{x}) &= qg\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\ &\leq \frac{1}{n}\sum_{i=1}^{n} qg(x_i) \quad \text{by convexity,} \\ &< 0 \quad \text{for all } q \in K, \end{aligned}$$

since x_i is feasible for each i and $qg(x_i) < 0$ for at least one i.

The next lemma is essentially as given in [24]. We have added extra details for our applications. Note that $\sigma(Y^*, Y)$ denotes the w^*-topology on Y^*; $\tau(Y, Y^*)$ is the topology on Y of uniform convergence on $\sigma(Y^*, Y)$ compact, convex, symmetric subsets of Y^*; U° is the *bipolar*

$$U^\circ = \{y: |yu| \leq 1 \text{ for all } u \in U\};$$

and we call Q a *base* for S^+ if Q is closed, convex, $0 \notin Q$ and

$$S^+ = \bigcup_{\lambda \geq 0} \lambda Q.$$

Lemma 2.3. (a) *If* $0 \neq y_0 \in \tau - \operatorname{int} S$, *then*

$$Q = S^+ \cap y_0^{-1}\{1\}$$

is a σ-compact (convex) base for S^+.

(b) *If Q is a σ-compact (convex) base for S^+ and $y_0 \in Y$ satisfies*

$$0 < \delta_1 \leq q y_0 \quad \textit{for all } q \in Q, \tag{2.6}$$

for some $\delta_1 > 0$, then

$$y_0 + \delta_1 Q^\circ = y_0 + U$$

is a τ-neighbourhood of y_0 inside S. In particular, such a y_0 always exists.

Thus S has nonempty τ-interior if and only if S^+ has a σ-compact base.

Proof. (a) Let $s^+ \neq 0$ lie in S^+. Since y_0 is nonzero and in $\tau - \operatorname{int} S$, we get $s^+ y_0 > 0$. Thus, since Q is σ-closed and convex, it is a base for S^+. Furthermore, if we set $V = (y_0 - S) \cap (S - y_0)$, then V is a τ-neighbourhood of 0 and, for $q \in Q$ and $v \in V$, we get

$$-1 = -q(y_0) \leq q(v) \leq q(y_0) = 1.$$

Thus, $Q \subset V^\circ$ and the Banach–Alaoglu theorem [19] implies that V° and therefore also Q is σ-compact.

(b) By (2.6), it follows that

$$y_0 + \delta_1 Q^\circ \subset Q^+ = S^{++} = S.$$

Moreover, Q° is a τ-neighbourhood of 0. This relies on the fact that the symmetric hull of Q is still compact and has the same polar. That such a y_0 always exists follows since Q is σ-compact, convex and does not contain 0.

We will use the above lemma in the form given in part (a) of the following corollary. The (b) part of the corollary is included to emphasize the equivalence of Slater's condition (2.8) and the semi-infinite condition (2.7). This has not always been clear in the literature [6, 24]. Note that the τ-topology is the strongest topology on Y for which Y^* is still the topological dual of Y (see [19]). Thus, for most purposes, we could have given Y the τ-topology from the start.

Corollary 2.1. *Suppose that Q is σ-compact and a generating set for S^+ i.e. $S^+ = \overline{\operatorname{cone}}\, Q$; and either Q is convex or Y^* is quasicomplete (which holds if Y is barrelled).*

(a) *If*

$$qg(\hat{x}) < 0 \quad \textit{for all } q \in Q, \tag{2.7}$$

then $\tau - \operatorname{int} S \neq \emptyset$ *and*

$$-g(\hat{x}) \in \tau - \operatorname{int} S. \tag{2.8}$$

(b) *Conversely, if* (2.8) *holds and* $g(\hat{x})$ *is nonzero, then*

$$Q = S^+ \cap \{q\colon qg(\hat{x}) = -1\}$$

is a σ*-compact* (*convex*) *base for* S^+ *for which* (2.7) *holds.*

Proof. (a) Since $S^+ = \overline{\operatorname{cone}}\, Q$ (and, if Q is not convex, Y^* is σ-quasicomplete), we get that $D = \overline{\operatorname{conv}}\, Q$ is σ-compact (see [14, p. 61]); $S^+ = \overline{\operatorname{cone}}\, D$; and, by (2.7),

$$-\delta = \sup_{d \in D} dg(\hat{x}) = \sup_{q \in Q} qg(\hat{x}) < 0.$$

In particular, $0 \notin D$ and $D \cap \{q\colon qg(\hat{x}) = -\delta\}$ is a σ-compact base for S^+. The result now follows from Lemma 2.3(b).

(b) Follows directly from Lemma 2.3(a).

Note that every Banach space and in fact every Fréchet space is barrelled. In these cases, Y^* is quasicomplete (see e.g. [19; 14, p. 135]).

Now, suppose that $h: X \to R \cup \{\infty\}$ is a convex function. Following [3], for each of the relations

'relation' is '=', '<', '≤', '<' or '≥',

we define

$$D_h^{\text{'relation'}}(x) = (d \in X\colon \text{ there exists } \bar{\alpha} > 0 \text{ with } h(x + \alpha d) \text{ 'relation' } h(x) \text{ for all } 0 < \alpha \leq \bar{\alpha}\}.$$

These are the *cones of directions* of constancy, decrease, nonincrease, increase and nondecrease respectively. For simplicity of notation, we let

$$D_q^{\text{'relation'}}(x) = D_{qg}^{\text{'relation'}}(x) \quad \text{for } q \in Y^*,$$

$$D_Q^{\text{'relation'}}(x) = \bigcap_{q \in Q} D_q^{\text{'relation'}}(x) \quad \text{for } Q \subset Y^*.$$

Furthermore, if $Q \subset S^+$ we let

$$g_Q(x) = \sup_{q \in Q} qg(x)$$

and set

$$D(x) = D_{g_{\mathscr{P}}}^{=}(x).$$

The following lemma gives several useful properties of the 'uniform' function g_Q. Note that dom(h) denotes the points at which the function h is finite.

Lemma 2.4. *If $Q \subset S^+$ is compact, then:*

(a) *g_Q is convex and* $\text{dom}(g_Q) = X$ *(in particular, g_Q is continuous on lines);*

(b) *if in addition g is 'weakly' lower semicontinuous (i.e. qg is lower semicontinuous for each q in Q) and X is barrelled, then g is continuous.*

Proof. (a) For $0 \le \lambda \le 1$,

$$\max_Q qg(\lambda x_1 + (1-\lambda)x_2) \le \max_Q(\lambda qg(x_1) + (1-\lambda)qg(x_2)) \quad \text{by convexity,}$$

$$\le \lambda \max_Q qg(x_1) + (1-\lambda) \max_Q qg(x_2).$$

This implies that g_Q is convex. That $\text{dom}(g_Q) = X$ follows from compactness of Q.

(b) Since each $qg(\cdot)$ is lower semicontinuous, it follows that $g_Q(\cdot)$ is lower semicontinuous, as it is the supremum of such functions. The result now follows since any lower semicontinuous convex function is continuous at an interior point of its domain, which in this case is the whole of X (see [9]).

Following Rockafellar [21], we say that a convex function $h: X \to R$ is *faithfully convex* if: h is affine on a line segment only if it is affine on the whole line containing that segment. If $X = R^n$, then it follows from results in [21], that the cone of directions of constancy at x, $D_h^{=}(x)$, is a subspace independent of $x \in R^n$ (see e.g. [3]). This has been shown directly for X a locally convex space in [23], if h is continuous. If qg is faithfully convex for each $q \in Q \subset S^+$, then it is clear that

$$D^{=}_{g_{Q(a)}}(a) \supset D^{=}_{Q(a)}(a).$$

(Note that if h is analytic, then h is faithfully convex. Thus, if g is 'weakly analytic' i.e. qg is analytic for all $q \in Y^*$, then qg is faithfully convex for all $q \in S^+$ (see e.g. [21]).) Various examples of these cones of constancy are given in Section 5.

For a convex function $h: X \to R \cup \{\infty\}$, we let

$$\nabla h(x; d) = \lim_{t \downarrow 0} \frac{h(x+td) - h(x)}{t}$$

denote the *directional derivative* of h at x in the direction d, while

$$\partial h(x) = \{\phi \in X^*: h(z) \ge h(x) + \phi(z-x) \text{ for all } z \in X\}$$

denotes the *subdifferential* of h at x. If $h(x)$ is finite, then $\nabla h(x; d)$ exists, for all d, although it may be plus or minus infinity. Moreover, if h is continuous at x, then $\partial h(x)$ is a nonempty, convex compact subset of X^* and

$$\nabla h(x; d) = \max\{\phi d \colon \phi \in \partial h(x)\}. \tag{2.9}$$

For two convex functions h_1 and h_2, for which there is a point at which one is finite and the other continuous, we get that

$$\partial(\alpha_1 h_1 + \alpha_2 h_2)(x) = \alpha_1 \partial h_1(x) + \alpha_2 \partial h_2(x) \quad \text{for all } x, \tag{2.10}$$

where α_1 and α_2 are nonnegative real scalars (see e.g. [14]).

For $Q \subset S^+$, the *linearizing cone* at $x \in F$, with respect to Q, is

$$C_Q(x) = \{d \in X \colon \phi d \leq 0 \text{ for all } \phi \in \partial qg(x) \text{ and all } q \in Q\}.$$

If $qg(\cdot)$ is continuous at x for all $q \in Q$, then (2.9) implies that

$$C_Q(x) = \{d \in X \colon \nabla qg(x; d) \leq 0 \text{ for all } q \in Q\}. \tag{2.11}$$

The *cone of subgradients* at x is

$$B_Q(x) = \operatorname{cone} \bigcup_{q \in Q} \partial qg(x).$$

In the continuous case, (2.10) implies that

$$B_Q(x) = \operatorname{cone} \bigcup_{q \in \operatorname{conv} Q} \partial qg(x). \tag{2.12}$$

The linearizing cone and the cone of subgradients have the following dual property.

Lemma 2.5. *Suppose that $Q \subset S^+$. Then*

$$\overline{B_Q(x)} = -C_Q^+(x). \tag{2.13}$$

Proof. Since

$$\begin{aligned} -C_{\{q\}}^+(x) &= \partial qg(x)^{++} && \text{by definition,} \\ &= \overline{B_{\{q\}}(x)} && \text{by (2.1),} \end{aligned}$$

we conclude, as $C_{\{q\}}(x)$ is always closed, that

$$\begin{aligned} -C_Q^+(x) &= -\overline{\sum_{q \in Q} C_{\{q\}}^+(x)} && \text{by (2.2),} \\ &= \overline{B_Q(x)}. \end{aligned}$$

(Note that $\sum \cdot$ denotes finite sums. Moreover, we did not require continuity of $qg(\cdot)$.)

This lemma will provide the link between optimality criteria which use subdifferentials, and criteria which use directional derivatives.

If g has a Gateaux-derivative at x, it is easy to see that the linearizing cone

$$C_Q(x) = \{d \in X : \nabla g(x)d \in -Q^+\}. \tag{2.14}$$

This will, of course, still hold if g only has a one-sided derivative. More specifically we have the following characterization of the linearizing cone for binding constraints.

Lemma 2.6. *Suppose g is Gateaux differentiable at x and that Q generates $S^+(x)$ ($Q^{++} = S^+(x)$). Then*

$$C_Q(x) = \{d \in X : \nabla g(x)d \in -\overline{\text{cone}}(S + g(x))\}. \tag{2.15}$$

Proof. Since $Q^+ = (S^+(x))^+$ it suffices by (2.14) to observe that

$$S^+(x) = (S + g(x))^+.$$

The result then follows on taking polars as in (2.1).

Note that since S is convex,

$$\overline{\text{cone}}(S + g(x)) = T(S, -g(x)),$$

where $T(S, -g(x))$ is the cone of tangent directions to S at $-g(x)$ (see e.g. [11, 13, 14]).

3. Characterization of optimality with constraint qualification

Our results are based on the following 'Standard Lagrange Multiplier theorem'. Recall that if Slater's condition (3.2) below holds, then we can choose the 'generating set' $\mathcal{P}$ compact and have cone $\mathcal{P} = S^+$ (see Corollary 2.1). Moreover, we then have

$$\text{cone}\,\mathcal{P}(a) = S^+(a).$$

Theorem 3.1. *Consider the convex program*

$$\text{(P)} \qquad \begin{aligned} &\text{minimize} \quad f(x), \\ &\text{subject to} \quad g(x) \in -S \text{ and } x \in \Omega, \end{aligned}$$

where f, g and S are as above and Ω is a convex subset of X. Let $\boldsymbol{F} = g^{-1}(-S) \cap \Omega$ denote the feasible set of (**P**); let

$$\mu_0 = \inf\{f(x) : g(x) \in -S \text{ and } x \in \Omega\} \tag{3.1}$$

be the (finite) solution value of (**P**); *and further, suppose that there exists*

$$\hat{x} \in \text{dom}\, f \cap \Omega \cap g^{-1}(-\text{int}\, S) \quad (\textit{Slater's condition}). \tag{3.2}$$

Then, there exists $s^+ \in S^+$ *such that*

$$\mu_0 = \inf\{f(x) + s^+g(x): x \in \Omega\}. \tag{3.3}$$

If in addition $a \in \boldsymbol{F} \cap \operatorname{dom} f$, *then we get the following four statements* (i)–(iv) *of optimality and their mutual relationships.*

If

(i) *a solves* (***P***),

then

(ii) *there exists* $s^+ \in S^+$ *such that*

$$f(a) = \inf\{f(x) + s^+g(x): x \in \Omega\} \tag{3.4}$$

$$s^+g(a) = 0 \quad (\textit{complementary slackness}); \tag{3.5}$$

which implies

(iii) (a) *that there exists* $s^+ \in S^+(a)$ *such that the system*

$$\nabla f(a; d) < 0, \quad \nabla s^+g(a; d) \le 0, \quad d \in \operatorname{cone}(\Omega - a), \tag{3.6}$$

is inconsistent,

or equivalently (*with Slater's condition*),

(b) *the system*

$$\nabla f(a; d) < 0, \quad \nabla s^+g(a; d) \le 0 \quad \textit{for all } s^+ \in S^+(a), \quad d \in \operatorname{cone}(\Omega - a),$$

is inconsistent,

or equivalently (*with Slater's condition,* $\operatorname{cone} \mathcal{P}(a) = S^+(a)$, *and* $qg(\cdot)$ *continuous at a for each* $q \in \mathcal{P}(a)$),

(c)

$$D_f^{<}(a) \cap C_{\mathcal{P}(a)}(a) \cap \operatorname{cone}(\Omega - a) = \emptyset, \tag{3.7}$$

which then entails

(iv) (*with* $qg(\cdot)$ *continuous at a for each* $q \in \mathcal{P}(a)$ *and f continuous at a*)

(a) (*with* $\operatorname{cone} \mathcal{P}(a) = S^+(a)$)

$$\partial f(a) \cap (-B_{\mathcal{P}(a)}(a) + (\Omega - a)^+) \ne \emptyset,$$

or equivalently

(b)

$$0 \in \partial f(a) + \partial s^+g(a) - (\Omega - a)^+ \quad \textit{for some } s^+ \in S^+(a). \tag{3.8}$$

Conversely, without Slater's condition in the hypothesis, we have that (iv) *implies both* (iii) *and* (ii), *each of which then imply* (i). *Hence, when Slater's condition holds and both f and g are continuous at a, we get that (i) through (iv) are equivalent.*

Proof. The proof of (3.3) and of (i) implies (ii) follows from [16, p. 217], with

Table 1. Hypotheses required in Theorem 3.1

implies implied by	*a* optimal (i)	Lagrangian (ii)	Directional derivatives (iii)(a)	 (iii)(b)	 (iii)(c)	subdifferentials (iv)(a)	 (iv)(b)
(i)		Slater's	Slater's	Slater's	Slater's qg cont. $c\mathcal{P}q = S$	Slater's qg cont. f cont. $c\mathcal{P}a = S$	Slater's qg cont. f cont.
(ii)					qg cont. $c\mathcal{P}a = S$	qg cont. f cont. $c\mathcal{P}a = S$	qg cont. f cont.
(iii)(a)		Slater's			gq cont. $c\mathcal{P}a = S$	Slater's qg cont. f cont. $c\mathcal{P}a = S$	Slater's qg cont. f cont.
(iii)(b)		Slater's	Slater's gq cont.		qg cont. $c\mathcal{P}a = S$	Slater's qg cont. f cont. $c\mathcal{P}a = S$	Slater's qg cont. f cont.
(iii)(c)		Slater's	Slater's			Slater's qg cont. f cont. $c\mathcal{P}a = S$	Slater's qg cont. f cont.
(iv)(a)					qg cont. $c\mathcal{P}a = S$		
(iv)(b)					$c\mathcal{P}a = S$	qg cont. $c\mathcal{P}a = S$	

Slater's: Slater's condition
qg cont.: $qg(\cdot)$ continuous at a, for all $q \in \mathcal{P}(a)$
f cont.: f continuous at a.
$c\mathcal{P}a = S$: cone $\mathcal{P}(a) = S^+(a)$.

$\Omega \cap \operatorname{dom} f$ replacing Ω. Note that the proof is valid if we consider Y with the τ-topology, since its topological dual remains unchanged.

Table 1 gives the hypotheses required in the proof of Theorem 3.1.

(ii) $\Rightarrow$ (iii)(a): Suppose that $d \in \operatorname{cone}(\Omega - a)$. Then there exists $\bar{t} > 0$ such that $a + td \in \Omega$, for all $0 < t \le \bar{t}$. Therefore, (3.4) implies that

$$t^{-1}[f(a+td) - f(a)] + t^{-1}[s^+g(a+td) - s^+g(a)] \ge 0,$$

for all $0 < t \le \bar{t}$. Now

$$\lim_{t \downarrow 0} t^{-1}[f(a+td) - f(a)] = \nabla f(a; d)$$

and

$$\lim_{t \downarrow 0} t^{-1}[s^+g(a+td) - s^+g(a)] = \nabla s^+g(a; d)$$

both exist since $\mu = f(a)$ and $s^+g(a)$ are finite. Thus

$$\nabla f(a;d) + \nabla s^+g(a;d) \geq 0$$

(with the proviso that $-\infty + \infty \geq 0$). We now conclude that (3.6) is inconsistent. (Note that $s^+ \in S^+(a)$ by (3.5).)

(iii)(a) $\Rightarrow$ (iii)(b): Obvious.

(iii)(b) $\Rightarrow$ (iii)(c): Let us first show that

$$C_{\mathcal{P}(a)}(a) = C_{S^+(a)}(a). \tag{3.12}$$

Recall that $S^+(a) = \operatorname{cone} \mathcal{P}(a)$. That $C_{S^+(a)}(a) \subset C_{\mathcal{P}(a)}(a)$ is clear. To prove the converse, suppose that $d \in C_{\mathcal{P}(a)}(a)$ and $s^+ \in S^+(a) = \operatorname{cone} \mathcal{P}(a)$. Then $s^+ = \sum_{i=1}^n \alpha_i q_i$ for some $\alpha_i \geq 0$ and $q_i \in \mathcal{P}(a)$ and

$$\begin{aligned}
\nabla s^+g(a;d) &= \sup\{\phi d: \phi \in \partial s^+g(a)\} \quad \text{by (2.9)},\\
&= \sup\left\{\phi d: \phi \in \partial \sum \alpha_i q_i g(a)\right\}\\
&= \sup\left\{\phi d: \phi \in \sum \alpha_i \partial q_i g(a)\right\} \quad \text{by (2.10)},\\
&\leq \sum \alpha_i \sup\{\phi_i d: \phi_i \in \partial q_i g(a)\}\\
&= \sum \alpha_i \nabla q_i g(a;d) \leq 0.
\end{aligned}$$

This implies that $C_{\operatorname{cone}\mathcal{P}(a)}(a) \supset C_{\mathcal{P}(a)}(a)$. The result now follows since

$$D_f^{<}(a) = \{d: \nabla f(a;d) < 0\} \quad \text{(see e.g. [5])}.$$

(iii)(c) $\Rightarrow$ (iii)(a): Suppose that (3.7) holds but a is not optimal. Then there exists a feasible point x such that $f(x) < f(a)$. Thus

$$d = x - a \in D_f^{<}(a) \cap C_{\mathcal{P}(a)}(a) \cap \operatorname{cone}(\Omega - a), \tag{3.13}$$

which is a contradiction. Therefore a solves (**P**). We have seen that this, with Slater's condition, implies part (a).

(iii) $\Rightarrow$ (iv)(b): From the above proof, we have that (iii) $\Rightarrow$ (i) $\Rightarrow$ (ii). Therefore, there exists $s^+ \in S^+(a)$ such that

$$f(a) = \mu_0 = \inf\{f(x) + s^+g(x): x \in \Omega\}$$

and Pshenichnii's condition [14, p. 87] implies that

$$0 \in \partial(f + s^+g)(a) - (\Omega - a)^+,$$

which by (2.10) yields part (b). Note that if $\Omega = X$, then f need not be assumed continuous.

(iv)(b) $\Rightarrow$ (iv)(a): It is sufficient to show that

$$\partial s^+g(a) \subset B_{\mathcal{P}(a)}(a). \tag{3.14}$$

This is clear since $s^+ \in S^+(a) = \text{cone}\,\mathcal{P}(a)$ and $qg(\cdot)$ is continuous at a for all $q \in \mathcal{P}(a)$.

(iv)(a) $\Rightarrow$ (iv)(b): Let $\phi \in B_{\mathcal{P}(a)}(a)$. It is sufficient to show that

$$\phi \in \partial s^+ g(a) \quad \text{for some } s^+ \in S^+(a).$$

Now $\phi = \sum_{i=1}^k \alpha_i \phi_i$ for some $\alpha_i \geq 0$, $\phi_i \in \partial p_i g(a)$ and $p_i \in \mathcal{P}(a)$. Therefore

$$\phi_i(x-a) \leq p_i g(x) - p_i g(a)$$

for all $x \in X$ and all $i = 1, \ldots, k$. This implies that

$$\sum \alpha_i \phi_i(x-a) \leq \sum \alpha_i p_i(g(x) - g(a)),$$

i.e. $\sum \alpha_i \phi_i \in \partial(\sum \alpha_i p_i) g(a)$. We can now choose $s^+ = \sum \alpha_i p_i$.

(iv)(b) $\Rightarrow$ (ii): This follows from the Pshenichnii condition.

(iii) $\Rightarrow$ (i): See the proof of (iii)(c) $\Rightarrow$ (iii)(a).

4. Characterization of optimality without constraint qualification

Recall that $\mathcal{P}$ is a generating set for S^+ and $\mathcal{P} = \mathcal{P}^= \cup \mathcal{P}^<$. We now divide up the set $\mathcal{P}^<$ so that we can apply the 'Standard Lagrange Multiplier theorem'. Choose the set

$$Q \subset \mathcal{P}^< \tag{4.1}$$

compact in Y^* and let $\mathcal{R}$ be a remainder, i.e.

$$S^+ \supset \mathcal{R} \supset \mathcal{P}^< \backslash Q. \tag{4.2}$$

In order to apply Corollary 2.1, we will assume throughout that either Q is convex or Y is barrelled and $(\overline{\text{cone}}\, Q)(a) = \overline{\text{cone}}(Q(a))$. The notion of dividing up the indexing set of constraints was previously used in [7, 8]. Recall that, for $\Omega \subset \mathcal{P}$, we denote the '*partial feasible sets*'

$$F^\Omega = \{x \in X : qg(x) \leq 0 \text{ for all } q \in \Omega\}$$

and the '*uniform*' *function*

$$g_\Omega(x) = \sup_{q \in \Omega} qg(x).$$

Note that if $g_\Omega(x) = 0$, then

$$D^{\leq}_{g_\Omega}(a) = \text{cone}(F^\Omega - a). \tag{4.3}$$

The following theorem and corollary characterize optimality for (P), without any constraint qualification. One may of course choose $\mathcal{P}$ convex and closed. Under our assumptions $\overline{\text{conv}\, Q}$ is still compact and contained in $\mathcal{P}^<$. Thus, to simplify reading the theorem and its proof, the reader may assume that both $\mathcal{P}$ and Q are convex and thus delete $\overline{\text{conv}}$ and replace $\overline{\text{cone}}$ by cone. If Q is finite dimensional, the situation is even simpler since conv Q is always compact.

Theorem 4.1. *Consider the original convex program* ($\mathcal{P}$). *Let the sets Q and $\mathcal{R}$ be as above. Let*

$$\mu_0 = \inf\{f(x)\colon g(x) \in -S\} < \infty \tag{4.4}$$

be the (finite) solution value of (P) *and assume that* $\operatorname{dom} F \supset F$. *Then there exists $q \in \overline{\text{cone}}\, Q$ such that*

$$\mu_0 = \inf\{f(x) + qg(x)\colon x \in F^{\mathcal{R}} \cap F^{\mathcal{P}^=}\}.$$

If in addition $a \in F \cap \operatorname{dom} f$ (and for simplicity $qg(\cdot)$ is continuous at a for each $q \in Q(a)$; and $f(\cdot)$ is continuous at a), then the following four statements (i)–(iv) *are equivalent*:

(iii)

(i) *a solves* (P);

(ii) *there exists $q \in \overline{\text{cone}}\, Q$ such that*

$$f(a) = \inf\{f(x) + qg(x)\colon x \in F^{\mathcal{R}} \cap F^{\mathcal{P}^=}\},$$
$$qg(a) = 0;$$

(iii) (a) *there exists $q \in \overline{\text{cone}}\, Q(a)$ such that the system*

$$\nabla f(a; d) < 0, \quad \nabla qg(a; d) \le 0, \quad d \in D(a) \cap \operatorname{cone}(F^{\mathcal{R}} - a)$$

is inconsistent,

or equivalently

(b) *the system*

$$\nabla f(a; d) < 0,\ \nabla qg(a; d) \le 0 \text{ for all } q \in \overline{\text{cone}}(Q(a)),$$
$$d \in D(a) \cap \operatorname{cone}(F^{\mathcal{R}} - a)$$

is inconsistent,

or equivalently

(c)

$$D_f^{<}(a) \cap C_{\operatorname{conv}(Q(a))}(a) \cap D(a) \cap \operatorname{cone}(F^{\mathcal{R}} - a) = \emptyset;$$

(iv) (a)

$$\partial f(a) \cap (-B_{\operatorname{conv}(Q(a))}(a) + (D(a) \cap \operatorname{cone}(F^{\mathcal{R}} - a))^+) \ne 0,$$

or equivalently

(b)

$$0 \in \partial f(a) + \partial qg(a) - (D(a) \cap \operatorname{cone}(F^{\mathcal{R}} - a))^+ \text{ for some } q \in \overline{\text{cone}}\, Q(a).$$

Proof. By Lemma 2.1, we can rewrite (P) as:

$$\begin{aligned} &\text{minimize} \quad f(x), && qg(x)\le 0 \quad \text{for all } q\in Q,\\ &\text{subject to} && qg(x)\le 0 \quad \text{for all } q\in \mathcal{R},\\ & && qg(x)\le 0 \quad \text{for all } q\in \mathcal{P}^{=}. \end{aligned} \tag{4.5}$$

This is equivalent to

$$\begin{aligned} &\text{minimize} \quad f(x),\\ &\text{subject to} \quad g(x)\in -Q^{+} \text{ and } x\in F^{\mathcal{P}^{=}}\cap F^{\mathcal{R}}. \end{aligned}$$

This problem now satisfies all the conditions of the 'Standard Lagrange Multiplier theorem', i.e. g is Q^+-convex since $S=S^{++}\subset Q^+$, $F^{\mathcal{P}^=}\cap F^{\mathcal{R}}$ is convex since $S\subset \mathcal{R}^+$ and $S\subset \mathcal{P}^{=+}$; and Lemma 2.2 and Corollary 2.1 imply that there exists $\hat{x}\in F\cap \operatorname{dom} f$ with $0\neq g(\hat{x})\in -\operatorname{int}(Q^+)$. We can now apply Theorem 3.1 with $\Omega = F^{\mathcal{P}^=}\cap F^{\mathcal{R}}$; S replaced by Q^+; and S^+ replaced by $\overline{\operatorname{cone}}\, Q$. (Note that $0\notin \overline{\operatorname{conv}}\, Q$ by Lemma 2.2; $\operatorname{cone}\overline{\operatorname{conv}}\, Q$ is closed; and $\overline{\operatorname{cone}}(Q(a)) = (\overline{\operatorname{cone}}\, Q)(a)$.) It only remains to show that

$$(D(a)\cap \operatorname{cone}(F^{\mathcal{R}}-a))^{+} = ((F^{\mathcal{P}^=}\cap F^{\mathcal{R}})-a)^{+}. \tag{4.6}$$

In fact, let us show the stronger statement

$$\begin{aligned} D(a)\cap \operatorname{cone}(F^{\mathcal{R}}-a) &= D^{\le}_{g_{\mathcal{P}^=}}(a)\cap \operatorname{cone}(F^{\mathcal{R}}-a)\\ &= \operatorname{cone}((F^{\mathcal{P}^=}\cap F^{\mathcal{R}})-a). \end{aligned} \tag{4.7}$$

That the inclusion $\subset$ holds above is clear since $D(a)\subset \operatorname{cone}(F^{\mathcal{P}^=}-a)$, (4.3) holds and since for convex sets A_1, A_2 containing a, $\operatorname{cone}(A_1-a)\cap\operatorname{cone}(A_2-a)$ agrees with $\operatorname{cone}(A_1\cap A_2-a)$. To prove the converse, suppose that $d\in \operatorname{cone}((F^{\mathcal{P}^=}\cap F^{\mathcal{R}})-a)$. Then $d\in\operatorname{cone}(F^{\mathcal{R}}-a) = D^{\le}_{g_{\mathcal{P}^=}}(a)$ and we need only show that

$$d\in D(a). \tag{4.8}$$

Let $x=a+\alpha d\in F^{\mathcal{P}^=}\cap F^{\mathcal{R}}$, $\alpha>0$, and choose $\hat{x}$ to satisfy

$$\hat{x}\in F \quad\text{and}\quad g(\hat{x})\in -\operatorname{int} Q^{+}. \tag{4.9}$$

This $\hat{x}$ always exists by Lemma 2.2 and Corollary 2.1 and moreover satisfies

$$g_Q(\hat{x})<0. \tag{4.10}$$

Now let

$$x_\lambda = \lambda x + (1-\lambda)\hat{x} \quad \text{for } 0\le\lambda\le 1.$$

Then, by convexity, $x_\lambda\in F^{\mathcal{P}^=}\cap F^{\mathcal{R}}$ and

$$g_Q(x_\lambda)<0 \quad \text{for small } \lambda>0$$

(see Lemma 2.4(a)). Thus $x_\lambda\in F$ for small $\lambda>0$, which implies that $g_{\mathcal{P}^=}(x_\lambda)=0$

for small $\lambda > 0$. Now, since $g_{\mathscr{P}^=}(x_\lambda) \le 0$ for all $0 \le \lambda \le 1$, and $g_{\mathscr{P}^=}$ is convex, we must have $g_{\mathscr{P}^=}(x_\lambda) = 0$ for all $0 \le \lambda \le 1$. This means that $g_{\mathscr{P}^=}(x) = 0$ and $d \in D(a)$.

Corollary 4.1. *The above theorem holds if we relax the restrictions on the multipliers, i.e. if, in the statements* (ii), (iii), (iv), $\overline{\text{cone}}\, Q$ *is replaced by* S^+ *and both* $\overline{\text{cone}}\, Q(a)$ *and* $\overline{\text{conv}}\, Q(a)$ *are replaced by* $S^+(a)$.

Proof. Existence of $q \in \overline{\text{cone}}\, Q$ implies existence of $q \in S^+$ since $Q \subset \mathscr{P}$. The converse follows since the statements (ii), (iii), (iv), after the replacements, still imply (i).

The above theorem and corollary characterize optimality for the convex program (P) without using any constraint qualification. Craven and Zlobec [8] have presented similar optimality conditions for (P). They did not, however, use a reduction to Slater's condition and required that the infimum be attained, $\text{int}\, S \ne \phi$ and F posses a 'radial point', i.e. that the intrinsic core of F be nonempty. Moreover, the set $D(a)$ is redundant in their optimality conditions.

In the following corollary we see that the set $\mathscr{R}$ may become redundant. We shall see that this leads to the so-called BBZ conditions in the case that S is polyhedral.

Corollary 4.2. *In the above theorem (and corollary), suppose that*

$$a \in \text{int dom}\, g_{\mathscr{R}} \quad \textit{and} \quad g_{\mathscr{R}}(a) < 0, \tag{4.11}$$

or

$$\mathscr{R} \subset \overline{\text{cone}}(\mathscr{P}^= \cup Q). \tag{4.12}$$

Then $\mathscr{R}$ *(and also* $F^{\mathscr{R}}$ *and* $\text{cone}(F^{\mathscr{R}} - a)$*) can be omitted, thus simplifying the optimality criteria. (Note that any subset of* $\mathscr{R}$ *that satisfies* (4.11) *or* (4.12) *may similarly be omitted in the previous result.)*

Proof. Only (4.12) requires proof. Since

$$\begin{aligned} x \in F \quad &\text{if and only if} \quad g(x) \in -\mathscr{P}^+ && \text{by Lemma 2.1,} \\ &\text{if and only if} \quad g(x) \in -(\mathscr{P}^= \cup Q)^+ && \text{by hypothesis,} \end{aligned}$$

we see that $\mathscr{R}$ is redundant in (4.5) and thus redundant in the rest of the proof of Theorem 4.1.

Corollary 4.3.

$$\begin{aligned} D(a) \cap \text{cone}(F^{\mathscr{R}} - a) = \{d\colon\ &\textit{there exists } \bar{\alpha} > 0 \textit{ with } qg(a + \alpha d) = (\textit{or} \le) \\ &qg(a) \textit{ for all } q \in \mathscr{P}^= \textit{ and all } 0 < \alpha \le \bar{\alpha}\} \cap \\ &\text{cone}(F^{\mathscr{R}} - a) \end{aligned} \tag{4.13}$$

and is a convex set. Moreover,

$$D(a) \text{ is convex if } \mathcal{R} \text{ can be omitted} \tag{4.14}$$

Proof. Let us call the first cone on the right-hand side of (4.13) $D^u(a)$. It is immediate that

$$D^u(a) \subset D^{\leq}_{g_{\mathcal{P}^=}}(a).$$

Conversely, if d lies in $D^{\leq}_{g_{\mathcal{P}^=}}(a) \cap \operatorname{cone}(F^{\mathcal{R}} - a)$, an examination of the proof of Theorem 4.1 shows that the point x_λ constructed below (4.10) is feasible and hence satisfies

$$qg(x_\lambda) = 0, \quad x_\lambda = \lambda(a + \alpha d) + (1 - \lambda)\hat{x}$$

for all q in $\mathcal{P}^=$. Exactly as before it follows that $qg(a + \alpha d) = 0$ for all a in $\mathcal{P}^=$ and d lies in $D^u(a)$. Since d lies in $\operatorname{cone}(F^{\mathcal{R}} - a)$ by (4.7) we have the desired equality. Moreover, (4.14) follows from (4.7) since $D^{\leq}_{g_{\mathcal{P}^=}}(a)$ is convex.

Thus we can modify the uniformity of the directions of constancy (or nonincrease). In the case that $qg(\cdot)$ is faithfully convex, for all but a finite number of $q \in \mathcal{P}^=$, we see that (4.13) implies that

$$D(a) \cap \operatorname{cone}(F^{\mathcal{R}} - a) = D^{=}_{\mathcal{P}^=}(a) \cap \operatorname{cone}(F^{\mathcal{R}} - a), \tag{4.15}$$

i.e. we can replace the uniform directions of constancy over $\mathcal{P}^=$ by the intersection of the directions of constancy.

Now, in the case of a finite number of constraints $g^k(x) \leq 0$, $k = 1, \ldots, m$, we may choose $S = R^m_+$, the nonnegative orthant in R^m, and $g(x) = (g^k(x))$. We may now set the generating set $\mathcal{P} = \{e_i\}_{i=1}^m$, the set of unit vectors in R^m. The sets $\mathcal{P}^=$ and $\mathcal{P}^<$ are therefore finite and so compact. Moreover, by Corollary 4.2, we may choose $Q = \mathcal{P}^<$ and $\mathcal{R} = \emptyset$, since $\mathcal{P} = \mathcal{P}^= \cup \mathcal{P}^<$. First, we note that by (4.14) and (4.15),

$$D^{=}_{\mathcal{P}^=}(a) = D^{\leq}_{\mathcal{P}^=}(a) \quad \text{is convex.}$$

(The set $\operatorname{conv} D^{=}_{\mathcal{P}^=}(a)$ was needlessly considered in [3, 4].) Next, our Theorem 4.1 reduces to the so-called BBZ conditions [1, 3, 5]:

$$a \text{ (feasible) is optimal if and only if } \partial f(a) \cap (-B_{\mathcal{P}^<(a)}(a) + D(a)^+) \neq \emptyset.$$

In addition, in the case that the infimum may not be attained, we get that

$$\mu_0 = \inf\{f(x) + \lambda g(x) \colon x \in F^{\mathcal{P}^=}\} \quad \text{for some } \lambda \in R^m_+.$$

Another situation which is easily handled is the case of two cone constraints, in two spaces,

$$g_1(x) \in -S_1 \subset Y_1, \qquad g_2(x) \in -S_2 \subset Y_2,$$

where there exists a (feasible) Slater point $\hat{x}$ for the first constraint:

$$g_1(\hat{x}) \in -\operatorname{int} S_1, \qquad g_2(\hat{x}) \in -S_2,$$

while the second cone S_2 is polyhedral, i.e. is the intersection of a finite number of half-spaces

$$S_2 = \{d \in Y_2\colon h_i d \geq 0, i = 1, 2, \dots, m\}.$$

Then the cone S_1^+ has a compact base say $\mathcal{P}_1$ while $\mathcal{P}_2 = \{h_i\}_{i=1}^m$ is a compact base for S_2^+. Furthermore, $\mathcal{P}_1^= = \emptyset$ and $\mathcal{P}_2^= = \{h_i\colon h_i g_2(x) = 0$ for all feasible $x\}$. Thus the abstract program (P) has

$$S = S_1 \times S_2 \subset Y = Y_1 \times Y_2$$

and the generating set for $S^+ = S_1^+ \times S_2^+$ is

$$\mathcal{P} = (\mathcal{P}_1 \cup \{0\}) \times (\mathcal{P}_2 \cup \{0\}).$$

(Note that the set $\mathcal{P}_1 \times \mathcal{P}_2$ is not a generating set for S^+ since it does not contain $\{0\} \times S_2$ nor $S_1 \times \{0\}$.) Corollary 4.1 now yields the characterization:

$$a \text{ (feasible) is optimal if and only if } \partial f(a) \cap (-B_{\mathcal{P}(a)} + D(a)^+) \neq \emptyset,$$

if and only if

$0 \in \partial f(a) + \partial s_1^+ g_1(a) + \partial s_2^+ g_2(a) - D(a)^+$ for some $s_1^+ \in S_1^+(a)$ and $s_2^+ \in S_2^+(a)$,

where

$$D(a) = D_{\mathcal{P}^=}^=(a) = \bigcap_{i \in \mathcal{P}_2^=} D_{h_i g_2}^=(a).$$

Remark 4.1. To derive our characterizations of optimality in the above, we divided up the constraint set into three distinct parts, i.e.

$$\mathcal{P} \subset \mathcal{P}^= \cup Q \cup \mathcal{R},$$

where $Q \subset \mathcal{P}^<$, $\mathcal{P}^< \backslash Q \subset \mathcal{R}$ and Q is chosen compact in Y^*. Corollary 4.2 further showed that we could throw away part (if not all) of $\mathcal{R}$ which satisfied (4.11) or (4.12). This led to the BBZ conditions in the case S was polyhedral and, in addition, gave a Lagrange multiplier relation in the case that the infimum may not be attained. If is also possible to throw away part (if not all) of $\mathcal{P}^=$. This leads to 'stronger' optimality criteria as well as 'weakest constraint qualifications'. For instance, this covers situations where a Lagrange multiplier exists but $\mathcal{P}^= \neq \emptyset$, i.e. Slater's condition fails. The details are given in [7]. The special case $S = R_+^m$ is treated in [23]. Moreover, the case that S is a convex cone (not necessarily closed) in R^m leads to very elegent characterizations of optimality without constraint qualification. These characterizations use the 'faces' of the cone S. These results are presented in a forthcoming paper.

5. Examples

In this section we present several examples to illustrate the theory presented above.

Example 5.1. This example first illustrates the fact that $D^{=}_{\mathscr{P}^=}(a) \neq D(a)$ in general. (Recall that $D^{=}_{\mathscr{P}^=}(a) = \bigcap_{q \in \mathscr{P}^=} D^{=}_{qg}(a)$ while $D(a) = D^{=}_{g_{\mathscr{P}^=}}(a)$.) We then see that choosing the generating set $\mathscr{P}$ in different ways can yield simpler results. In addition, the Kuhn–Tucker conditions fail here, but we characterize optimality using Theorem 4.1 (see Fig. 1).

Consider the semi-infinite program

$$\begin{aligned} &\text{minimize} \quad f(x) = -x, \\ &\text{subject to} \quad h(x,t) = \max{}^2(0, tx - t^2) \le 0 \quad \text{for all } t \in [0,1]. \end{aligned} \tag{5.1}$$

For each $x \in R$, $g(x) = h(x, \cdot\,)$ is a convex continuous (differentiable) function on $[0, 1]$, i.e. $g(x) \in C[0, 1]$. Now, let $Y = C[0, 1]$ and set

$$S = \{\text{nonnegative functions in } Y\}.$$

Then we can rewrite the above semi-infinite program as the abstract convex program

$$\text{(P)} \qquad \begin{aligned} &\text{minimize} \quad -x, \\ &\text{subject to} \quad g(x) \in -S. \end{aligned}$$

Let us choose the generating set

$$\mathscr{P} = \{\delta_t : 0 \le t \le 1\}, \tag{5.2}$$

the set of point measures in Y^*, where Y^* is the space of all Borel measures, i.e.

$$\delta_t g(x) = h(x, t).$$

Then $\overline{\text{cone}\, \mathscr{P}} = S^+$ and moreover, Slater's condition fails. In fact

$$\mathscr{P} = \mathscr{P}^=,$$

since $s^+ g(x) = 0$, for all $x \in F$ and all (nonnegative) measures $s^+ \in S^+$. We now

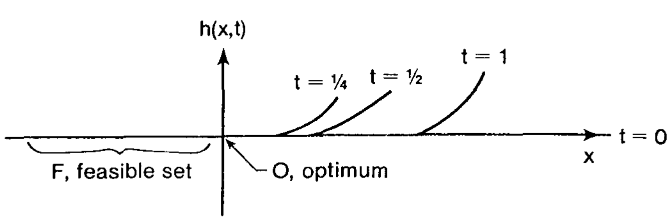

Fig. 1.

see that $D(0) = D(0)^+ = R_-$, the negative real half-line, while $D_{\mathcal{P}^=}^=(0) = R$, the whole real line. Furthermore, our optimality conditions (Theorem 4.1) hold at the optimal point 0:

$$0 \in \partial f(0) + \partial s^+ g(0) - (D(0))^+ = -1 - R_-, \tag{5.3}$$

where we have chosen the multiplier $s^+ = 0$.

Let us now change our choice of the generating set $\mathcal{P}$. The functional $y_0 \equiv 1$ is in int S. Therefore, Lemma 2.3 implies that

$$\mathcal{P} = S^+ \cap y_0^{-1}\{1\} = \{\text{probability measures}\}$$

is a compact (convex) base for S^+. (Note that $\mathcal{P} = \overline{\text{conv}}\{\delta_t : 0 \le t \le 1\}$, by the Krein–Milman theorem [14].) Then we actually have cone $\mathcal{P} = S^+$ and, as before, $\mathcal{P}^= = \mathcal{P}$. But now

$$D(0) = D_{\mathcal{P}^=}^=(0) = R_-,$$

since Lebesgue measure $\mu(t)$ is in $\mathcal{P}^=$ and $D_{\mu(t)}^=(0) = R_-$. Note that the optimality condition

$$0 \in \partial f(0) - D(0)^+$$

still holds although the Kuhn–Tucker conditions fail, i.e.

$$0 \notin \partial f(0) + \partial s^+ g(0) \quad \text{for any } s^+ \in S^+(0), \tag{5.4}$$

since $\partial f(0) = \{-1\}$ while $\partial s^+ g(0) = \{0\}$ for all $s^+ \in S^+$.

Example 5.2. This example uses a linear operator constraint to illustrate that the Kuhn–Tucker conditions may still hold at the optimum even though Slater's condition fails.

Now suppose that T is a bounded linear operator between the Banach spaces X and Y. Let $y \neq 0$ be in $\overline{\mathcal{R}(T^*)}$, the closure of the range of the adjoint of T in the w^*-topology, and consider the program

$$\begin{aligned} &\text{minimize} \quad f(x) = yx, \\ &\text{subject to} \quad Tx = b, \quad x \in X, \end{aligned} \tag{5.5}$$

where $b \in \mathcal{R}(T)$. If we let $g(x) = Tx - b$ and $S = \{0\}$, then the above program is equivalent to the abstract convex program

$$\text{(P)} \qquad \begin{aligned} &\text{minimize} \quad yx, \\ &\text{subject to} \quad g(x) \in -S. \end{aligned}$$

Since int $S = \emptyset$, we see that Slater's condition fails here. Moreover, $S^+ = Y^*$. We now choose the generating set

$$\mathcal{P} = B \quad \text{the unit ball in } Y^*.$$

Then cone $\mathcal{P} = S^+$ and $\mathcal{P}^= = \mathcal{P}$, since the feasible set $F = \bar{x} + \mathcal{N}(T)$, where $\bar{x}$ is any particular solution of $Tx = b$ and $\mathcal{N}(T)$ denotes the nullspace of T. Moreover, $D(\bar{x}) = \mathcal{N}(T)$ and $B_{\mathcal{P}(\bar{x})}(\bar{x}) = \{y^*T \mid y^* \in Y^*\} = \mathcal{R}(T^*)$ as T is its own Fréchet derivative. Therefore, Corollary 4.1 implies that $\bar{x}$ is optimal if and only if

$$\begin{aligned}0 \in \partial f(\bar{x}) + B_{\mathcal{P}(\bar{x})}(\bar{x}) - D(\bar{x})^+ &= y + \mathcal{R}(T^*) - \mathcal{N}(T)^+ \\ &= y + \overline{\mathcal{R}(T^*)},\end{aligned}$$

by the Fredholm alternative. This is a special case of (2.13). Thus every feasible point $\bar{x}$ is optimal, since we have assumed $y \in \overline{\mathcal{R}(T^*)}$.

If $y \notin \overline{\mathcal{R}(T^*)}$, then the problem is unbounded. If $y \in \mathcal{R}(T^*)$, then the Kuhn–Tucker conditions hold at the optimum:

$$0 \in \partial f(\bar{x}) + B_{\mathcal{P}(\bar{x})}(\bar{x}) = y + \mathcal{R}(T^*),$$

while if $y \in \overline{\mathcal{R}(T^*)} \backslash \mathcal{R}(T^*)$, then the Kuhn–Tucker conditions fail.

Note that even the Fritz John conditions:

$$0 \in \partial \lambda f(\bar{x}) + \partial s^+ g(\bar{x}) \tag{5.6}$$

for some $\lambda \geq 0$ and $s^+ \in S^+(a)$ not both zero, may fail at the optimum. For suppose $y \in \overline{\mathcal{R}(T^*)} \backslash \mathcal{R}(T^*)$ and $\overline{\mathcal{R}(T)} = Y$. (In the Banach space setting $\mathcal{R}(T)$ is closed exactly when $\mathcal{R}(T^*)$ is (weak*) closed and so such a y exists exactly when T is 1–1 with dense nonsurjective range. Consider $(Tx)_n = x_n/n$ in l_2 and take $y = (1/n)$ as a specific example.) Then the Fritz John conditions fail if $\lambda > 0$, since the Kuhn–Tucker conditions fail. This implies that necessarily $\lambda = 0$. But

$$\partial s^+ g(\bar{x}) = \{s^+ T\} \neq \{0\} \quad \text{for all } s^+ \neq 0 \text{ in } Y^*, \tag{5.7}$$

since $\overline{\mathcal{R}(T)} = X$, i.e. since T^* is 1–1. Note that the Fritz John conditions *do* hold when T^* is not 1–1, since we can then choose $s^+ \neq 0$ in $\mathcal{N}(T^*)$.

Example 5.3. This example shows that one should be careful when choosing the generating set $\mathcal{P}$. More precisely, we see that the assumption cone $\mathcal{P}(a) = S^+(a)$ is needed in Theorem 3.1. We again consider a semi-infinite program

$$\begin{aligned}&\text{minimize} \quad f(x) = x, \\ &\text{subject to} \quad h(x,t) = tx^2 - t - x \leq 0 \quad \text{for all } t \in [0,1],\end{aligned}$$

and let $g(x) = h(x, \cdot)$ so as to formulate an abstract convex program (P). (See Example 5.1 for the definitions of X, Y, S, etc.) Note that Slater's condition is satisfied:

$$g(1) = -1 \in \operatorname{int} -S$$

and 0 is the optimal point. We first choose the generating set

$$\mathcal{P} = \{\delta_t : 0 < t \leq 1\}.$$

Then $\mathscr{P}(0)=\emptyset$ and the optimality conditions in Theorem 3.1(iv) fail at the optimal point 0, even though Slater's condition is satisfied. Note that $\overline{\text{cone}}\, P = S^+$ but $\emptyset = \overline{\text{cone}}\, \mathscr{P}(0) \neq S^+(0)$. However, if we choose

$$\mathscr{P} = \{\delta_t : 0 \leq t \leq 1\},$$

then $\mathscr{P}(0)=\{0\}$, $B_{\mathscr{P}(0)}(0) = R_-$ and the Kuhn–Tucker conditions hold at 0:

$$0 \in \partial f(0) + B_{\mathscr{P}(0)}(0) = \{1\} + R_-.$$

Example 5.4. This example shows that part (iii) of Theorem 3.1 may characterize optimality when parts (ii) and (iv) fail to, i.e. the conditions involving directional derivatives are weaker than those involving subdifferentials. The directional derivatives here coincide with the homogeneous constraint and objective functions. Let

$$K = \{\phi = (\phi_i) \in R^2 : \phi_1^2 + (\phi_2 - 1)^2 \leq 1\},$$

$$g(x) = \sup\{\phi x : \phi \in K\} = \sqrt{x_1^2 + x_2^2} + x_2,$$

$$S = R_+.$$

Consider the program

$$\text{(P)} \qquad \begin{array}{ll} \text{minimize} & f(x) = x_1, \\ \text{subject to} & g(x) \in -S. \end{array}$$

Then the point 0 solves (P) and Slater's condition fails since $g(x) \geq 0$, for all $x \in R^2$. Now, Theorem 4.1 yields the optimality condition:

$$D_f^{<}(0) \cap C_{\mathscr{P}(0)}(0) = \emptyset.$$

However,

$$0 \notin \partial f(0) + \partial s^+ g(0) \quad \text{for any } s^+ \in S^+(0),$$

since $\partial f(0) = (1,0)$ while $\partial s^+ g(0) = s^+ K$. This shows that part (iii) of Theorem 3.1 characterizes optimality while parts (iii) and (iv) fail to. The reason for this is that the cone of subgradients $B_{\mathscr{P}(0)}(0)$ is not closed and thus not equal to $-(C_{\mathscr{P}(0)}(0))^+$.

Example 5.5 This example illustrates Corollary 4.1. Consider the semi-infinite program (see Fig. 2)

$$\begin{array}{ll} \text{minimize} & f(x) = -x, \\ \text{subject to} & h(x,t) \leq 0 \quad \text{for all } t \in \{1, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{4}, \ldots\}, \end{array}$$

where

$$h(x,t) = \begin{cases} -tx & \text{if } 0 \leq x \leq 1+t, \\ tx - 2t^2 - 2t & \text{if } 1+t \leq x. \end{cases}$$

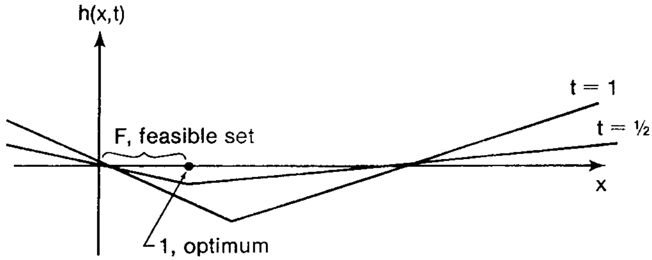

Fig. 2.

For each $x \in R$, $g(x) = h(x, \cdot)$ is a sequence in l_2. So, let $Y = l_2$ and set

$$S = \{\text{nonnegative sequences in } Y\}.$$

We can now rewrite the above program in the abstract formulation

$$\text{(P)} \qquad \begin{array}{ll} \text{minimize} & f(x), \\ \text{subject to} & g(x) \in -S. \end{array}$$

Let us choose the generating set

$$\mathcal{P} = \{e_i \colon i = 1, 2, 3, \ldots\},$$

the set of coordinate functions in l_2. Then $\overline{\operatorname{cone} \mathcal{P}} = S^+ = S$. Note that Slater's condition fails, since $\operatorname{int} S = \emptyset$, but $\mathcal{P}^= = \emptyset$. To apply Corollary 4.1, we must choose the compact set

$$Q = \{e_i \colon i = 1, 2, \ldots, k\} \subset \mathcal{P}^< = \mathcal{P},$$

while the remainder

$$\mathcal{R} = \{e_i \colon i = k + 1, 1 + 2, \ldots\}.$$

Thus

$$(F^{\mathcal{R}} - 1)^+ = \operatorname{cone}(F^{\mathcal{R}} - 1) = R_-$$

and our optimality conditions yield:

$$0 \in \partial f(1) + \partial s^+ g(1) - (F^{\mathcal{R}} - 1)^+ = \{-1\} - R_-,$$

since $S^+(1) = \{0\}$, as there are no binding constraints.

References

[1] R.A. Abrams and L. Kerzner, "A simplified test for optimality" *Journal of Optimization Theory and Applications* 25 (1978) 161–170.

[2] M.S. Bazaraa, C.M. Shetty, J.J. Goode and M.Z. Nashed, "Nonlinear programming without differentiability in Banach spaces: Necessary and sufficient constraint qualification", *Applicable Analysis* 5 (1976) 165–173.

[3] A. Ben-Israel, A. Ben-Tal and S. Zlobec, "Optimality conditions in convex programming", in: A. Prékopa, ed., *Survey of mathematical programming*, Proceedings of the 9th International Mathematical Programming Symposium (The Hungarian Academy of Sciences, Budapest and North-Holland, Amsterdam, 1979) pp. 137–156.
[4] A. Ben-Tal, A. Ben-Israel and S. Zlobec, "Characterization of optimality in convex programming without a constraint qualification", *Journal of Optimization Theory and Applications* 20 (1976) 417–437.
[5] A. Ben-Tal and A. Ben-Israel, "Characterizations of optimality in convex programming: The nondifferentiable case", *Applicable Analysis* 9 (1979) 137–156.
[6] J. Borwein, "Proper efficient points for maximizations with respect to cones", *SIAM Journal on Control and Optimization* 15 (1977) 57–63.
[7] J. Borwein and H. Wolkowicz, "Characterizations of optimality without constraint qualification for the abstract convex program", Research Report 14, Dalhousie University, Halifax, N.S.
[8] B.D. Craven and S. Zlobec, "Complete characterization of optimality of convex programming in Banach spaces", *Applicable Analysis* 11 (1980) 61–78.
[9] M. Day, *Normed linear spaces*, 3rd ed. (Springer, Berlin, 1973).
[10] I.V. Girsanov, *Lectures on mathematical theory of extremum problems*, Lecture notes in economics and mathematical systems 67 (Springer, Berlin, 1972).
[11] F.J. Gould and J.W. Tolle, "Geometry of optimality conditions and constraint qualifications", *Mathematical Programming* 2 (1972) 1–18.
[12] F.J. Gould and J.W. Tolle, "A necessary and sufficient qualification for constrained optimization", *SIAM Journal on Applied Mathematics* 20 (1971) 164–171.
[13] M. Guignard, "Generalized Kuhn–Tucker conditions for mathematical programming problems in a Banach space", *SIAM Journal on Control* 7 (1969) 232–241.
[14] R.B. Holmes, *A course on optimization and best approximation*, Lecture notes in mathematics 257 (Springer, Berlin, 1972).
[15] S. Kurcyusz, "On the existence and nonexistence of Lagrange multipliers in Banach spaces", *Journal of Optimization Theory and Applications* 20 (1976) 81–110.
[16] D.G. Luenberger, *Optimization by vector space methods* (J. Wiley, New York, 1969).
[17] L.W. Neustadt, *Optimization—A theory of necessary conditions* (Princeton University Press, Princeton, NJ, 1976).
[18] B.N. Pshenichnyi, *Necessary conditions for an extremum* (Marcel Dekker, New York, 1971).
[19] A.P. Robertson and W.J. Robertson, *Topological vector spaces* (Cambridge University Press, Cambridge, 1964).
[20] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, NJ, 1972).
[21] R.T. Rockafellar, "Some convex programs whose duals are linearly constrained", in: J.B. Rosen, O.L. Mangasarian and K. Ritter, eds., *Nonlinear programming* (Academic Press, New York, 1970) pp. 393–322.
[22] H. Wolkowicz, "Calculating the cone of directions of constancy", *Journal of Optimization Theory and Applications* 25 (1978) 451–457.
[23] H. Wolkowicz, "Geometry of optimality conditions and constraint qualifications: The convex case", *Mathematical Programming* 19 (1980) 32–60.
[24] J. Zowe, "Linear maps majorized by a sublinear map", *Archiv der Mathematik* 26 (1975) 637–645.
[25] J. Zowe and S. Kurcyusz, "Regularity and stability for the mathematical programming problem in Banach spaces", Preprint No. 37 Würzburg (1978).