JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 35, No. 4, DECEMBER 1981

A Strengthened Test for Optimality[1]

H. WOLKOWICZ[2]

Communicated by A. V. Fiacco

Abstract. In this paper, we strengthen recent characterizations of optimality for the convex program

$$\text{(P)} \qquad \mu = \inf\{f(x): g^k(x) \le 0, k = 1, \ldots, m\},$$

where the functions f and g^k, $k = 1, \ldots, m$, are convex functions on R^n. The characterizations presented here are stronger, in the sense that the Lagrange multiplier relation holds over a larger set. This strengthens information about the stability of the solution with respect to perturbations in the right-hand side of the constraints. In particular, we show that, in the characterizations of optimality in Refs. 1–2, the set $D^{=}_{\mathcal{P}^{=}}(x^*)$, the intersection of the cones of directions of constancy of the *equality constraints*, can be replaced by the larger and simpler set $D^{=}_{h}(x^*)$, the cone of directions of constancy of a single function h. We also discuss how to choose h to get the *strongest* characterization.

Key Words. Characterization of optimality, Lagrange multipliers, faithfully convex functions, gradients, cones of directions of constancy, strongest optimality conditions.

1. Introduction

Consider the *convex program*

$$\text{(P)} \qquad \mu = \inf\{f(x): g^k(x) \le 0, k \in \mathcal{P} = \{1, \ldots, m\}\},$$

where $f, g^k: R^n \to R$ are differentiable convex functions. Without loss of generality, we assume that none of the functions is constant. We will require that some of the constraints g^k be *faithfully convex functions*, i.e., convex

[1] This research was partially supported by NSERC-A-3388.

[2] Assistant Professor, Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada.

0022-3239/81/1200-0497$03.00/0

functions which are not affine along any line segment, unless they are affine along the entire line extending the segment (e.g., Ref. 3). The class of faithfully convex functions is large and it includes all analytic convex functions as well as all strictly convex functions. Characterizations of optimality for Program (P) have been given in Refs. 1–2. These characterizations hold without any constraint qualification, and they implicitly include a Lagrange multiplier relation which holds on the convex set

$$x^* + D^{=}_{\mathscr{P}^=}(x^*),$$

where $D^{=}_{\mathscr{P}^=}(x^*)$ is the intersection of the cones of directions of constancy of the *equality constraints* at the optimum x^*. It is of interest to have strong optimality conditions, in the sense that the Lagrange multiplier relation holds on as large a convex set as possible (e.g., Ref. 4). For example, if we have $\lambda_k \geq 0$ satisfying

$$f(x) + \sum_{k \in \mathscr{P}} \lambda_k g^k(x) \geq \mu, \qquad \text{for all } x \in \Omega, \tag{1}$$

then, if Ω is all of R^n, our solution is stable with respect to (feasible) perturbations of the right-hand sides of the constraints (e.g., Ref. 5). Similarly, the larger the set Ω is, the more we can guarantee that the solution is stable with respect to certain perturbations (e.g., Refs. 5–6).

In this paper, we show that we can replace $D^{=}_{\mathscr{P}^=}(x^*)$ by various larger cones. In particular, if

$$\alpha_k \geq 0, \qquad \text{for all } k \in \mathscr{P}^=,$$

with $\alpha_k > 0$ if g^k is not affine, then we can use the cone $D^{=}_h(x^*)$ when

$$h = \sum_{k \in \mathscr{P}^=} \alpha_k g^k$$

is faithfully convex; i.e., we can replace the intersection of the cones of constancy $D^{=}_{\mathscr{P}^=}(x^*)$ by the larger and simpler cone of constancy of the single function h. Note that the α_k are arbitrary (nonnegative) constants. Thus we can change the values of the constants α_k and obtain a variety of different cones $D^{=}_h(x^*)$. This theory simplifies the algorithm for finding the set $\mathscr{P}^=$, given in Ref. 1 (see Remark 3.2.).

In addition, we show how to weaken the differentiability and faithfully convex assumptions; we discuss how to find the scalars α_k which give the *strongest* optimality conditions, and compare our results with those in Ref. 7, which use the *badly behaved set* of constraints.

Note that our results in Sections 2 and 3 hold when R^n is replaced by a locally convex (Hausdorff) topological vector space.

2. Preliminaries

We consider the convex program (P) presented in Section 1. We assume that the *feasible set*

$$F = \{x \in R^n : g^k(x) \le 0, \text{ for all } k \in \mathcal{P} = \{1, \ldots, m\}\}$$

is nonempty. $\mathcal{P}(x)$ will denote the *binding* (*active*) constraints at x. The *equality set* (e.g., Ref. 1) is

$$\mathcal{P}^= = \{k \in \mathcal{P} : g^k(x) = 0, \text{ for all } x \in F\}.$$

We then set

$$\mathcal{P}^<(x) = \mathcal{P}(x) \backslash \mathcal{P}^=.$$

For the relation $\mathcal{R}$, we define

$$D_{g^k}^{\mathcal{R}}(x) = \{d \in R^n : \exists \bar{\alpha} > 0, \text{ with } g^k(x + \alpha d) \mathcal{R} g^k(x), \text{ for all } 0 < \alpha \le \bar{\alpha}\}.$$

If $\mathcal{R}$ is $=$, $\le$, $<$, or $>$, we get the *cones of directions* of *constancy*, *nonincrease*, *decrease*, and *increase*, respectively (e.g., Ref. 2). We let

$$D_k^{\mathcal{R}}(x) = D_{g^k}^{\mathcal{R}}(x), \qquad D_\Omega^{\mathcal{R}}(x) = \bigcap_{k \in \Omega} D_k^{\mathcal{R}}(x).$$

When h is faithfully convex, then

$$D_h^= = D_h^=(x) \text{ is a subspace independent of } x; \tag{2}$$

and, if $x \in F$, then

$$D_{\mathcal{P}^=}^=(x) \cap D_{\mathcal{P}^<(x)}^<(x) \ne \varnothing \tag{3}$$

(e.g., Ref. 2). We let $\nabla h(x)$ denote the *gradient* of h at x. Then, the *linearizing cone* at x is

$$C(x) = \{d : \nabla g^k(x) d \le 0, \text{ for all } k \in \mathcal{P}(x)\}.$$

For a set $K \subset R^n$, the *nonnegative polar* (*apolar*) *cone* of K is

$$K^+ = \{\phi \in R^n : \phi k \ge 0, \text{ for all } k \in K\},$$

while the *annihilator* of K is

$$K^\perp = K^+ \cap -K^+.$$

Then, if K, L are closed convex cones in R^n,

$$K \subset L \Rightarrow L^+ \subset K^+, \qquad (K \cap L)^+ = \text{cl}(K^+ + L^+), \tag{4}$$

where cl denotes closure, and

$$(K \cap L)^+ = K^+ + L^+, \tag{5}$$

if K is polyhedral and L is a subspace or also polyhedral (e.g., Ref. 8). Note that $-C^+(x)$ is equal to the cone generated (finitely) by the gradients of g^k, $k \in \mathscr{P}(x)$.

The *badly behaved set* of constraints at x (Ref. 7) is

$$\mathscr{P}^b(x) = \{k \in \mathscr{P}^= : C(x) \cap D_k^>(x) \backslash \text{cl}\, D_{\mathscr{P}^=}^=(x)) \neq \varnothing\}.$$

This is the set of constraints that cause trouble in the Kuhn–Tucker theory. In Ref. 7, it was shown that, for $x \in F$,

$$\mathscr{P}^b(x) = \varnothing \tag{6}$$

is a weakest constraint qualification at x. Note that affine functions are never badly behaved. In fact, this is true of all functions g^k whose directions of constancy correspond exactly to the directions in which the directional derivatives are 0.

The next two theorems present some known characterizations of optimality.

Theorem 2.1. Suppose that $x^* \in F$ and $\mathscr{P}^b(x^*) \subset \Omega \subset \mathscr{P}^=$. Then, the following are equivalent:

(i) x^* solves Program (P);

(ii) the system

$$\nabla f(x^*) + \sum_{k \in \mathscr{P}^<(x^*)} \lambda_k \nabla g^k(x^*) \in (D_{\mathscr{P}^=}^=(x^*))^+, \qquad \lambda_k \geq 0,$$

is consistent;

(iii) the system

$$\nabla f(x^*) + \sum_{k \in \mathscr{P}(x^*)} \lambda_k \nabla g^k(x^*) \in (D_{\mathscr{P}^=}^=(x^*))^+, \qquad \lambda_k \geq 0,$$

is consistent;

(iv) if both conv $D_\Omega^=(x^*)$ and $C^+(x) + (D_\Omega^=(x))^+$ are closed, then the system

$$\nabla f(x^*) + \sum_{k \in \mathscr{P}(x^*)} \lambda_k \nabla g^k(x^*) \in (D_\Omega^=(x))^+, \qquad \lambda_k \geq 0,$$

is consistent, where conv denotes convex hull.

Proof. That (ii) characterizes optimality was shown in Refs. 1–2. That (ii) is equivalent to both (iii) and (iv) was shown in Ref. 7. □

Theorem 2.2. Consider the system

$$\nabla f(x^*) + \sum_{k\in\mathcal{P}(x^*)} \lambda_k \nabla g^k(x^*) \in G, \qquad \lambda_k \geq 0. \tag{7}$$

Then, the consistency of (7) characterizes optimality of Program (P) at x^* (feasible) iff

$$(D^{\leqq}_{\mathcal{P}(x^*)}(x^*))^+ = C^+(x^*) + G. \tag{8}$$

Proof. The case of differentiable constraints is given in Ref. 4. The convex nondifferentiable case is treated in Ref. 7. □

The above result follows from the so-called Pshenichnii condition (see, e.g., Ref. 8, p. 87):

x^* (feasible) solves Program (P), iff $\nabla f(x^*) \in T^+(F, x^*)$,

where $T(F, x^*)$ is the tangent cone of F at x^*. In the convex case (e.g., Ref. 7),

$$T(F, x^*) = \mathrm{cl}(\mathrm{cone}(F - x^*)) = \mathrm{cl}(D^{\leqq}_{\mathcal{P}(x^*)}(x^*)).$$

Note that the stronger optimality condition, as seen by having larger sets Ω in (1), is equivalent to having smaller sets G in (7). For, if the infimum of Program (P) is attained at x^*, then

$$\sum_{k\in\mathcal{P}} \lambda_k g^k(x) \leq 0, \qquad \text{for all } x \in F;$$

(1) is equivalent to

$$\mu = \inf\{f(x) + \sum_{k\in\mathcal{P}} \lambda_k g^k(x) : x \in \Omega\};$$

and we get the complementary slackness condition

$$\sum_{k\in\mathcal{P}} \lambda_k g^k(x^*) = 0,$$

or equivalently

$$\lambda_k = 0, \qquad \text{if } k \notin \mathcal{P}(x^*).$$

Thus, (1) is equivalent to

$$\mu = f(x^*) = f(x^*) + \sum_{k\in\mathcal{P}} \lambda_k g^k(x^*) = \inf\{f(x) + \sum \lambda_k g^k(x) : x \in \Omega\}.$$

By the Pshenichnii condition, this is equivalent to (7), with

$$G = (\Omega - x^*)^+.$$

By (4), this implies that larger sets Ω yield smaller sets G.

3. Strengthened Test for Optimality

In Refs. 7, 9, we showed that $D^{=}_{\mathscr{P}^=}(x^*)$ may be replaced by $D^{\leq}_{\mathscr{P}^=}(x^*)$, $D^{=}_{\Omega}(x^*)$, or $D^{\leq}_{\Omega}(x^*)$, where

$$\mathscr{P}^b(x^*) \subset \Omega \subset \mathscr{P}^=$$

and certain closure conditions hold. We now see that we can replace $D^{=}_{\mathscr{P}^=}(x^*)$ by the larger and simpler set $D^{=}_{h}(x^*)$, where

$$h = \sum_{k \in \mathscr{P}^=} \alpha_k g^k$$

and the α_k are *any* nonnegative scalars, with $\alpha_k > 0$, if g^k is *not* affine.

Theorem 3.1. Suppose that

$$\alpha_k \geq 0, \qquad \text{for all } k \in \mathscr{P}^=,$$

with $\alpha_k > 0$ if g^k is *not* affine, and

$$h = \sum_{k \in \mathscr{P}^=} \alpha_k g^k$$

is faithfully convex. Then x^* (feasible) is optimal for Program (P) iff the system

$$\nabla f(x^*) + \sum_{k \in \mathscr{P}(x^*)} \lambda_k \nabla g^k(x^*) \in (D^{=}_{h}(x^*))^+, \qquad \lambda_k \geq 0, \tag{9}$$

is consistent.

Proof. Since the point $x^* \in F$ is fixed throughout, we will omit it in this proof when the meaning is clear. For example, $D^{\leq}_{\mathscr{P}}$ denotes $D^{\leq}_{\mathscr{P}(x^*)}(x^*)$. Now, by Theorem 2.2, we need only show that

$$G = (D^{=}_{h})^+ \text{ satisfies (8).} \tag{10}$$

Let us first show that

$$\operatorname{cl} D^{\leq}_{\mathscr{P}} = D^{=}_{h} \cap C. \tag{11}$$

We have (e.g., Ref. 2)

$$D^{\leq}_{\mathscr{P}} = D^{=}_{\mathscr{P}^{=}} \cap D^{\leq}_{\mathscr{P}^{<}} \subset D^{=}_{h} \cap C, \tag{12}$$

by definition of h and C, and since

$$d \in D^{\leq}_{\mathscr{P}}$$

implies that the directional derivative

$$\nabla g^k d \leq 0, \qquad \text{for all } k \in \mathscr{P}.$$

The inclusion

$$\text{cl}\, D^{\leq}_{\mathscr{P}} \subset D^{=}_{h} \cap C$$

now follows, since the right-hand side is closed. Conversely, suppose that

$$d \in D^{=}_{h} \cap C, \tag{13}$$

and $\mathscr{U}$ is a neighborhood of the origin in X. To show the reverse inclusion, we need to show that

$$(\mathscr{U} + d) \cap D^{\leq}_{\mathscr{P}} \neq \varnothing. \tag{14}$$

By (3), choose

$$\hat{d} \in D^{=}_{\mathscr{P}^{=}} \cap D^{<}_{\mathscr{P}^{<}}; \tag{15}$$

and, for $0 < \lambda \leq 1$, let

$$d_\lambda = \lambda \hat{d} + (1 - \lambda) d.$$

Now,

$$d_\lambda \in D^{=}_{h} \cap C, \qquad \text{for all } 0 < \lambda \leq 1, \tag{16}$$

since $D^{=}_{h} \cap C$ is convex and [by (15), the definition of h, see (12)]

$$\hat{d} \in D^{=}_{\mathscr{P}^{=}} \cap D^{<}_{\mathscr{P}^{<}} = D^{=}_{h} \cap D^{=}_{\mathscr{P}^{=}} \cap D^{<}_{\mathscr{P}^{<}} \subset D^{=}_{h} \cap C.$$

Moreover, let us show that

$$d_\lambda \in D^{=}_{\mathscr{P}^{=}}, \qquad \text{for all } 0 < \lambda \leq 1. \tag{17}$$

Since

$$D^{=}_{\mathscr{P}^{=}} = D^{\leq}_{\mathscr{P}^{=}}$$

(Ref. 10), we need only show that

$$d_\lambda \in D^{\leq}_{\mathscr{P}^{=}}.$$

Suppose that this fails for a fixed $0 < \lambda \leq 1$, i.e.,

$$d_\lambda \in D^{>}_{l}, \qquad \text{for some } l \in \mathscr{P}^{=}.$$

Since $d_\lambda \in C$, we have that

$$d_\lambda \in D_k^{\leqq},$$

for all $k \in \mathscr{P}(x)$ for which g^k is affine. Therefore, $\alpha_l > 0$. Moreover, since $d_\lambda \in D_h^{=}$ and h is faithfully convex, we see that

$$-\alpha_l g^l(x^* + \alpha d_\lambda) = \sum_{k \in \mathscr{P}^= \setminus \{l\}} \alpha_k g^k(x^* + \alpha d_\lambda), \qquad \alpha \in R.$$

But, since a nonnegative linear combination of convex functions is convex, the above implies that both

$$\alpha_l g^l(x^* + \alpha d_\lambda) \quad \text{and} \quad -\alpha_l g^l(x^* + \alpha d_\lambda)$$

are convex functions of α; therefore,

$$g^l(x^* + \alpha d_\lambda)$$

is affine on the line $x^* + \alpha d_\lambda$, $\alpha \in R$. But then

$$d_\lambda \in D_l^{>}$$

implies that

$$\nabla g^l d_\lambda > 0,$$

which contradicts the fact that $d_\lambda \in C$. Thus, (17) holds.

In addition, we have

$$d_\lambda \in D_{\mathscr{P}^<}^{<}, \qquad \text{for all } 0 < \lambda \leq 1, \tag{18}$$

since (e.g., Ref. 2)

$$\begin{aligned} \hat{d} \in D_{\mathscr{P}^<}^{<} &= \{d: \nabla g^k d < 0, \text{ for all } k \in \mathscr{P}^<\} \\ &= \operatorname{int}\{d: \nabla g^k d \leq 0, \text{ for all } k \in \mathscr{P}^<\}, \end{aligned}$$

where int denotes interior and

$$d \in C \subset \{d: \nabla g^k d \leq 0, \text{ for all } k \in \mathscr{P}^<\}.$$

By (12), we see that (17) and (18) imply that

$$d_\lambda \in D_{\mathscr{P}}^{\leqq}, \qquad \text{for all } 0 < \lambda \leq 1.$$

Thus, by choosing λ sufficiently small, we get (14). This completes the proof of (11). To show (10), we now need only apply (5). □

Note that $\mathscr{P}^= = \varnothing$ iff Slater's condition is satisfied, i.e., there exists $\hat{x} \in X$ with

$$g^k(\hat{x}) < 0, \qquad \text{for all } k = 1, \ldots, m.$$

Thus,

$$D^{=}_{\mathscr{P}^{=}}(x^*) = R^n$$

iff Slater's condition holds (unless the constraints g^k, $k \in \mathscr{P}^=$, are identically 0). Therefore, the optimality criteria given in Refs. 1–2 (see our Theorem 2.1) reduce to the classical Kuhn–Tucker conditions [$G = 0$ in (7)] iff Slater's condition holds. The conditions in the above Theorem 3.1 are tighter and reduce to the Kuhn–Tucker conditions, for example whenever the *generalized Slater's condition* holds, i.e., there exists $\hat{x} \in F$ with

$$g^k(x) < 0, \text{ for all } k, \text{ except possibly those for which } g^k \text{ is affine.}$$

This holds, since we can choose $\alpha_k = 0$ when g^k is affine. Let us illustrate this with the following simple example.

Example 3.1. Let

$$g^1(x) = x_1 - x_2 \quad \text{and} \quad g^2(x) = -g^1(x).$$

Then,

$$\mathscr{P}^= = \{1, 2\}$$

and

$$D^{=}_{\mathscr{P}^{=}}(x) = \{d \in R^2 : d_1 = d_2\}.$$

Let

$$\alpha_1 = \alpha_2 = 1 \quad \text{and} \quad h = \sum_{k=1}^{2} \alpha_k g^k.$$

Then,

$$D^{=}_{h}(x) = R^2,$$

since $h = 0$, and (9) reduces to the classical Kuhn–Tucker conditions. Note that Theorem 2.1(iv) will always reduce to the classical case if x^* is a *regular point*, since

$$\mathscr{P}^b(x^*) = \varnothing$$

is a weakest constraint qualification, see Ref. 7. If we add

$$g^3(x) = \begin{cases} 0, & \text{if } x_1, x_2 \geq 0, \\ x_1^2 + x_2^2, & \text{otherwise,} \end{cases}$$

and let the objective function

$$f(x) = x_1 + x_2,$$

then

$$\mathcal{P}^{=} = \{1, 2, 3\},$$

the Kuhn–Tucker conditions fail at the optimum $x^* = 0$, while

$$D^{=}_{\mathcal{P}^{=}}(0) = \{d \in R^2: d_1 = d_2 \geq 0\},$$

$$D^{=}_{h}(0) = \{d \in R^2: d_1, d_2 \geq 0\},$$

where we have chosen again

$$\alpha_k = 1 \quad \text{and} \quad h = \sum_{k \in \mathcal{P}^{=}} \alpha_k g^k.$$

Remark 3.1. In the above example, the constraint g^3 is not faithfully convex. In fact, the faithfully convex assumption (and the differentiability assumption) can be replaced by the following weaker assumption:

$$\operatorname{conv} D^{=}_{h}(x^*) \quad \text{and} \quad -B(x^*) + (D^{=}_{h}(x^*))^+ \tag{19}$$

are closed, where $B(x^*)$ is the *cone of subgradients* at x^*, i.e., the convex cone generated by the *subdifferentials* $\partial g^k(x^*)$, with $k \in \mathcal{P}(x^*)$.

The cone $B(x^*)$ replaces the cone $C(x^*)$ in (8). The above closure assumptions are necessary, when applying (5) in the proof, and replace the polyhedrality assumptions which hold in the faithfully convex differentiable case. The rest of the proof holds with minor modifications. It can also be shown that α_k may be equal to 0, if

$$k \notin \mathcal{P}^b(x^*).$$

Note that

$$h = \sum_{k \in \Omega} \alpha_k g^k,$$

where $\Omega \subset \mathcal{P}$, is faithfully convex if g^k, $k \in \Omega$, is faithfully convex; i.e., a nonnegative linear combination of faithfully convex functions is faithfully convex. This can be seen by applying the argument used to prove (17).

Let us now show that the closure assumptions in (19) are necessary, by violating condition (8) when

$$G = (D^{=}_{h}(x^*))^+.$$

Example 3.2. Consider Program (P) with the three constraints, defined on $x = (x_i) \in R^3$: $g^1(x) = x_1$, $g^2(x) = -x_1$, $g^3(x) = (\inf_{z \in K} \|x - z\|)^2$, where K is the self-polar *ice-cream cone*

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

Let

$$x^* = 0.$$

Then,

$$x^* \in F = \{x : x_1 = x_3 = 0, x_2 \geq 0\}$$

and

$$\mathcal{P}^= = \mathcal{P} = \{1, 2, 3\}.$$

If we choose

$$\alpha_k = 1, \qquad \text{for all } k \in \mathcal{P}^=,$$

then

$$h = g^3;\ D_h^=(x^*) = (D_h^=(x^*))^+ = K.$$

Furthermore,

$$-B(x^*) = C^+(x^*) = \operatorname{span}\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}\right\}.$$

Note that the directional derivatives of g^3 at x^* are all 0 and

$$\partial g^3(x^*) = \{0\},$$

i.e.,

$$\nabla g^3(x^*) = 0.$$

Let us show that

$$-B(x^*) + (D_h^=(x^*))^+ \text{ is not closed.}$$

Choose

$$k^i = \begin{pmatrix}i\\1/i\\1\end{pmatrix} \in K \qquad \text{and} \qquad l^i = \begin{pmatrix}-i\\0\\0\end{pmatrix} \in -B(x^*), \qquad i = 1, 2, \ldots.$$

Then,

$$k^i + l^i \to \begin{pmatrix}0\\0\\1\end{pmatrix} \notin -B(x^*) + K.$$

Note that all the constraints are differentiable at x^*.

Example 3.3. Consider Program (P) with the three constraints, defined on R^2,

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

$$g^2(x)=x_1-1,$$

$$g^3(x)=-g^2(x).$$

Then,

$$x^*=(1,0)^t$$

is the only feasible point,

$$(D_{\mathscr{P}}^{\leqq}(x^*))^+=R^2, \qquad B(x^*)=\operatorname{span}\left\{\begin{pmatrix}1\\0\end{pmatrix}\right\},$$

and

$$D_1^{=}(x^*)=\{d\in R^2: d_1<0\}\cup\{0\}.$$

Thus, $D_1^{=}(x^*)$ is convex, but *not* closed. If we let

$$\alpha_1=\alpha_2=\alpha_3=1$$

again, then

$$h=g^1$$

and

$$(D_h^{=}(x^*))^+=B(x^*)=\{x\in R^2: x_2=0\}\subsetneq(D_{\mathscr{P}}^{\leqq}(x^*))^+.$$

Remark 3.2. The above theorem simplifies the algorithm for finding $\mathscr{P}^{=}$, given in Ref. 1. After solving the system

$$0=\sum_{k\in\mathscr{P}(x)}\alpha_k\nabla g^k(x), \qquad \alpha_k\geq 0,$$

one need only find $D_h^{=}(x)$, where

$$h=\sum_{k\in\mathscr{P}}\alpha_k g^k,$$

rather than finding

$$\bigcap_{k\in\mathscr{P}(x)} D_{\alpha_k g^k}^{=}(x).$$

The remaining steps are similarly simplified. We show, in the next section, that the constants α_k have a special significance (see Corollary 4.1).

Remark 3.3. Characterizations of optimality for the abstract convex nondifferentiable program with cone constraints are given in Refs. 9–11. Deriving stronger and strongest optimality criteria for this problem is studied in Ref. 11.

4. Strongest Optimality Conditions

Since the cone of directions of constancy of a faithfully convex function is a subspace independent of x, we see that finding the *strongest* optimality criteria corresponds, in our case, to solving the problem:

$$\text{(M)} \quad \text{maximize (dimension } D_h^{=}),$$

$$\text{subject to } h = \sum_{k \in \mathcal{P}^=} \alpha_k g^k, \ \alpha_k \geq 0,$$

$$\text{with } \alpha_k > 0, \ \text{ if } g^k \text{ is } \textit{not} \text{ affine,}$$

where we assume that the constraints g^k, $k \in \mathcal{P}^=$, are faithfully convex functions on R^n. In Ref. 3, it was shown that every faithfully convex function g^k has the representation

$$g^k(x) = p^k(A_k x + b^k) + c^k x + \beta_k, \tag{20}$$

where A_k is a $\tau_k \times n$ matrix, b^k is a vector in R^{τ_k}, c^k is a vector in $(R^n)^*$, p^k is a strictly convex function defined on R^{τ_k}, and β_k is a real scalar. We then get that

$$D_{g^k}^{=} = \mathcal{N}\begin{bmatrix} A_k \\ c^k \end{bmatrix}, \tag{21}$$

where $\mathcal{N}(\cdot)$ denotes null space.

Theorem 4.1. Suppose that the constraints g^k, $k \in \mathcal{P}^=$, are faithfully convex. Then, finding the constants α_k in Theorem 3.1, which yield the strongest optimality conditions, is equivalent to solving the system

$$\sum_{k \in \mathcal{P}^=} \alpha_k (c^k)^t \in \sum_{k \in \mathcal{P}^=} \mathcal{R}(A_k^t), \qquad \alpha_k \geq 0,$$

$$\text{with } \alpha_k > 0, \qquad \text{if } g^k \text{ is } \textit{not} \text{ affine,} \tag{22}$$

where $\mathcal{R}(A_k^t)$ denotes the range space of the transpose of the matrix A_k, and c^k and A^k are obtained from the representation (20).

Proof. By (4) and (5) (with $\perp$ replacing $+$), we get

$$\sum_{k \in \mathcal{P}^=} \alpha_k (c^k)^t \in \sum_{k \in \mathcal{P}^=} \mathcal{R}(A_k^t),$$

iff

$$\bigcap_{k\in\mathcal{P}^=}\mathcal{N}(A_k)\subset\mathcal{N}\left(\sum_{k\in\mathcal{P}^=}\alpha_k c^k\right).$$

Since finding the strongest optimality criteria corresponds to solving Problem (M), it is now sufficient to show that, for given α_k,

$$D_h^= = W, \tag{23}$$

where

$$W=\bigcap_{k\in\mathcal{P}^=}\mathcal{N}(A_k)\cap\mathcal{N}\left(\sum_{k\in\mathcal{P}^=}\alpha_k c^k\right).$$

That $D_h^= \supset W$ is clear from (20) and (21). To prove the converse, suppose that

$$x\in F \quad \text{and} \quad d\in D_h^=.$$

If $l\in\mathcal{P}^=$ and g^l is not affine, then

$$h(x)=0$$

and

$$-\alpha_l g^l(x+\alpha d)=\sum_{k\in\mathcal{P}^=}\alpha_k g^k(x+\alpha d), \qquad \text{for all } \alpha\in R.$$

As in the proof of Theorem 3.1, this shows that g^l is affine on the line $x-\alpha d$, $\alpha\in R$. Since p is strictly convex, we see that

$$d\in\mathcal{N}(A_l).$$

Thus,

$$d\in\bigcap_{k\in\mathcal{P}^=}\mathcal{N}(A_k),$$

where we let $A_k=0$, if g^k is affine; therefore, by (20),

$$0=h(x+\alpha d)-h(x)=\sum_{k\in\mathcal{P}^=}\alpha_k c^k(\alpha d), \qquad \alpha\in R,$$

which implies that

$$d\in\mathcal{N}\left(\sum_{k\in\mathcal{P}^=}\alpha_k c^k\right).$$

This completes the proof of (23). □

Note that, if the system (22) has no solution, then the dimension of $D_h^=$ remains the same for all choices of scalars α_k in Problem (M) (see

Remark 4.1 below). When the representations (20) are not readily available, we can apply the following corollary.

Corollary 4.1. Suppose that $\bar{x} \in F$ and that the constraints g^k, $k \in \mathscr{P}^=$, are faithfully convex. If the scalars $\alpha_k \geq 0$, with $\alpha_k > 0$ if g^k is *not* affine, satisfy

$$\sum_{k \in \mathscr{P}^=} \alpha_k \nabla g^k(\bar{x}) = 0, \tag{24}$$

then they yield the strongest optimality conditions in Theorem 3.1.

Proof. From the representation (20), we see that

$$\sum_{k \in \mathscr{P}^=} \alpha_k \nabla g^k(\bar{x}) = 0$$

iff

$$\sum_{k \in \mathscr{P}^=} \alpha_k c^k = -\sum_{k \in \mathscr{P}^=} \alpha_k \nabla p^k(A_k \bar{x} + b^k) A_k,$$

which implies that

$$\sum_{k \in \mathscr{P}^=} \alpha_k (c^k)^t \in \sum_{k \in \mathscr{P}^=} \mathscr{R}(A_k^t). \qquad \square$$

Example 4.1. Consider Program (P) with the three constraints

$$g^1(x) = x_1^2 + 2x_1x_3 + x_3^2 + x_1 + 2x_2 + x_3,$$
$$g^2(x) = x_1^2 - 2x_1x_3 + x_3^2 + 3x_1 + x_2 + 2x_3,$$
$$g^3(x) = 4x_1^2 - 4x_1x_3 + x_3^2 - 5x_1 - 5x_2 - 4x_3.$$

Then,

$$\bar{x} = (0, 0, 0)^t$$

is in F, and it is easy to check that

$$2\nabla g^1(\bar{x}) + \nabla g^2(\bar{x}) + \nabla g^3(\bar{x}) = 0,$$

that is,

$$\sum_{k=1}^{3} \alpha_k \nabla g^k(\bar{x}) = 0, \qquad \text{with } \alpha_1 = 2, \alpha_2 = \alpha_3 = 1.$$

By the Dubovitskii–Milyutin theorem of the alternative (see, e.g., Ref. 12), this implies that

$$\mathscr{P}^= = \{1, 2, 3\}.$$

In fact, the constants α_k are the constants found in the first step of the algorithm which finds $\mathscr{P}^=$ (see Ref. 1). By the above theorem, these constants have a special significance, i.e., we now define

$$h = \sum_{k \in \mathscr{P}^=} \alpha_k g^k = 2g^1 + g^2 + g^3.$$

It is easy to check that the strongest optimality conditions is given by

$$D_h^= = \{d \in R^3 \colon d_1 = d_3 = 0\}, \tag{25}$$

while

$$D_{\mathscr{P}^=}^= = \{0\}.$$

Example 4.2. In the above example, we get that

$$D_h^= = D_{\mathscr{P}^=}^= = \{0\},$$

unless

$$\sum_{k \in \mathscr{P}^=} \alpha_k \nabla g^k(\bar{x}) = 0,$$

in which case (25) holds. This gives us just two possible choices for our optimality conditions. The following example shows that we may have a large variety of choices. Consider Program (P) with the constraints

$$g^1(x) = x_1^2 + x_4^2 + 2x_1x_4 + 4x_1 + x_2 + 3x_3 + 3x_4,$$
$$g^2(x) = x_1^2 + x_4^2 - 2x_1x_4 - x_1 + x_2 + x_3 - 7x_4,$$
$$g^3(x) = -7x_1 - 3x_2 - 7x_3 + x_4.$$

Then,

$$\bar{x} = (0, 0, 0, 0)^t$$

is in F and

$$\nabla g^1(\bar{x}) = \begin{bmatrix} 4 \\ 1 \\ 3 \\ 3 \end{bmatrix}, \qquad \nabla g^2(\bar{x}) = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -7 \end{bmatrix}, \qquad \nabla g^3(\bar{x}) = \begin{bmatrix} -7 \\ -3 \\ -7 \\ 1 \end{bmatrix}.$$

Since

$$2\nabla g^1(\bar{x}) + \nabla g^2(\bar{x}) + \nabla g^3(\bar{x}) = 0,$$

we conclude that

$$\mathscr{P}^= = \{1, 2, 3\}.$$

The strongest optimality condition is given by

$$h = 2g^1 + g^2 + g^3,$$

which yields

$$D_h^{=} = \{d \in R^4 : d_1 = d_4 = 0\},$$

while

$$D_{\mathscr{P}^=}^{=} = \{0\}.$$

This can be seen by examining the representations

$$g^1(x) = p^1([1, 0, 0, 1]x + 1) + c^1x - 1,$$
$$g^2(x) = p^2([1, 0, 0, -1]x + 4) + c^2x - 4,$$

where

$$p^1(z) = p^2(z) = z^2, \qquad c^1 = (2, 1, 3, 1), \qquad c^2 = (-5, 1, 1, -3).$$

Note that, if we choose α_k arbitrarily in Theorem 3.1, we still obtain stronger optimality conditions than when we use $D_{\mathscr{P}^=}^{=}$, since

$$D_h^{=} = \mathcal{N}\left(\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{bmatrix}\right) \cap \mathcal{N}\left(\sum_{k \in \mathscr{P}^=} \alpha_k c^k\right) \subsetneq D_{\mathscr{P}^=}^{=} = \{0\}.$$

It is also interesting to note that choosing

$$\alpha_3 = 0$$

does not yield the strongest conditions. In addition, applying Theorem 2.1(iv) does not yield the strongest conditions.

Remark 4.1. In conclusion, we see that, when the constraints g^k, $k \in \mathscr{P}^=$, are faithfully convex, then we have many choices for the cone G in (7) other than $D_{\mathscr{P}^=}^{=}$. We can apply Theorem 2.1(iv) and use $D_{\Omega}^{=}$, for $\mathscr{P}^b(x^*) \subset \Omega \subset \mathscr{P}^=$. In particular, we may choose

$$\Omega = \mathscr{P}^= \backslash \{k : g^k \text{ is affine}\}.$$

Or we may apply Theorem 3.1 and use $D_h^{=}$, where

$$h = \sum_{k \in \mathscr{P}^=} \alpha_k g^k$$

and

$$\alpha_k \geq 0, \qquad \text{with } \alpha_k > 0 \text{ if } g^k \text{ is not affine}.$$

Now, using (20), we have

$$D^{=}_{\mathscr{P}^=} = \bigcap_{k\in\mathscr{P}^=} \mathcal{N}(A_k) \bigcap_{k\in\mathscr{P}^=} \mathcal{N}(c^k),$$

while the larger subspace is

$$D^{=}_{h} = \bigcap_{k\in\mathscr{P}^=} \mathcal{N}(A_k) \cap \mathcal{N}\Big(\sum_{k\in\mathscr{P}^=} \alpha_k c^k\Big).$$

If we are able to choose the α_k appropriately, i.e., so that the vector

$$\sum_{k\in\mathscr{P}^=} \alpha_k c^k$$

is linearly dependent on the rows of the matrices A_k [see (22)], then we can increase the dimension of $D^{=}_{h}$ (by at most one) and get

$$D^{=}_{h} = \bigcap_{k\in\mathscr{P}^=} \mathcal{N}(A_k).$$

References

1. ABRAMS, R. A., and KERZNER, L., *A Simplified Test for Optimality*, Journal of Optimization Theory and Applications, Vol. 25, pp. 161–170, 1978.
2. BEN-ISRAEL, A., BEN TAL, A., and ZLOBEC, S., *Optimality Conditions in Convex Programming*, IX International Symposium on Mathematical Programming, Budapest, Hungary, 1976.
3. ROCKAFELLAR, R. T., *Some Convex Programs Whose Duals Are Linearly Constrained*, Nonlinear Programming, Edited by J. B. Rosen, O. L. Mangasarian, and K. Ritter, Academic Press, New York, New York, pp. 293–322, 1970.
4. GOULD, F. J., and TOLLE, J. W., *Geometry of Optimality Conditions and Constraint Qualifications*, Mathematical Programming, Vol. 2, pp. 1–18, 1972.
5. GEOFFRION, A. M., *Duality in Nonlinear Programming: A Simplified Application-Oriented Development*, SIAM Review, Vol. 13, pp. 1–37, 1971.
6. WOLKOWICZ, H. *Shadow Prices for an Unstable Convex Program*, Utilitas Mathematica (to appear).
7. WOLKOWICZ, H., *Geometry of Optimality Conditions and Constraint Qualifications: The Convex Case*, Mathematical Programming, Vol. 19, pp. 32–60, 1980.
8. HOLMES, R. B., *Geometric Functional Analysis and Its Applications*, Springer-Verlag, Berlin, Germany, 1975.
9. BORWEIN, J., and WOLKOWICZ, H., *Characterizations of Optimality without Constraint Qualification for the Abstract Convex Program*, Dalhousie University, Halifax, Canada, Research Report No. 14, 1979.

10. BORWEIN, J., and WOLKOWICZ, H., *Characterizations of Optimality for the Abstract Convex Program with Finite-Dimensional Range,* Journal of the Australian Mathematical Society (to appear).
11. BORWEIN, J., and WOLKOWICZ, H., *Regularizing the Abstract Convex Program,* Journal of Mathematical Analysis and Applications (to appear).
12. DUBOVITSKII, A. JA., and MILYUTIN, A. A., *Extremum Problems in the Presence of Restrictions,* Vol. 5, pp. 1–80, 1965.