Proceedings of Allerton Conference, September 2000

Enhanced Optimality Conditions and Exact Penalty Functions[1]

by

Dimitri P. Bertsekas and Asuman E. Koksal[2]

Abstract

We consider optimization problems with equality, inequality, and abstract set constraints, and we explore various characteristics of the constraint set that imply the existence of Lagrange multipliers. We prove a generalized version of the Fritz-John theorem, and we introduce new and general conditions that extend and unify the major constraint qualifications. Among these conditions, a new property, pseudonormality, provides the connecting link between the classical constraint qualifications and the use of exact penalty functions.

1. INTRODUCTION

We consider finite-dimensional optimization problems of the form

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C, \end{aligned} \tag{1.1}$$

where the constraint set C consists of equality and inequality constraints as well as an additional abstract set constraint $x \in X$:

$$C = X \cap \big\{x \mid h_1(x) = 0, \ldots, h_m(x) = 0\big\} \cap \big\{x \mid g_1(x) \leq 0, \ldots, g_r(x) \leq 0\big\}. \tag{1.2}$$

We assume throughout the paper that f, h_i, g_j are continuously differentiable functions from $\Re^n$ to $\Re$, and X is a nonempty closed set. In our notation, all vectors are viewed as column vectors, and a prime denotes transposition.

Necessary conditions for the above problem can be expressed in terms of tangent cones, normal cones, and their polars. In our terminology, a vector y is a *tangent* of a set $S \subset \Re^n$ at a

[1] Research supported by NSF under Grant ACI-9873339.

[2] Dept. of EE and CS, M.I.T., Cambridge, MA, 02139, dimitrib@.mit.edu, asuman@mit.edu.

vector $x \in S$ if either $y = 0$ or there exists a sequence $\{x^k\} \subset S$ such that $x^k \neq x$ for all k and

$$x^k \to x, \qquad \frac{x^k - x}{\|x^k - x\|} \to \frac{y}{\|y\|}.$$

An equivalent definition often found in the literature (e.g., [BSS93], [RoW98]) is that there exists a sequence $\{x^k\} \subset S$ with $x^k \to x$, and a positive sequence $\{\alpha^k\}$ such that $\alpha^k \to 0$ and $(x^k - x)/\alpha^k \to y$. The set of all tangents of S at x is denoted by $T_S(x)$ and is also referred to as the *tangent cone* of S at x. The *polar cone* of any cone T is defined by

$$T^\perp = \big\{z \mid z'y \leq 0,\ y \in T\big\}.$$

For a nonempty cone T, we will use the well-known relation $T \subset \big(T^\perp\big)^\perp$, which holds with equality if T is closed and convex.

For a closed set X and a point $x \in X$, we will also use the *normal cone* of X at x, denoted $N_X(x)$, which is obtained from the polar cone $T_X(x)^\perp$ by means of a closure operation. In particular, we have $z \in N_X(x)$ if there exist sequences $\{x^k\} \subset X$ and $\{z^k\}$ such that $x^k \to x$, $z^k \to z$, and $z^k \in T_X(x^k)^\perp$ for all k. Equivalently, the graph of $N_X(\cdot)$, viewed as a point-to-set mapping, $\big\{(x, z) \mid z \in N_X(x)\big\}$, is the closure of the graph of $T_X(\cdot)^\perp$. The normal cone, introduced by Mordukhovich [Mor76], has been studied by several authors, and is of central importance in nonsmooth analysis (see the books by Aubin and Frankowska [AuF90], Rockafellar and Wets [RoW98], and Borwein and Lewis [BoL00]). In general, we have $T_X(x)^\perp \subset N_X(x)$ for any $x \in X$. In the case where $T_X(x)^\perp = N_X(x)$, we will say that X is *regular* at x. The term "regular at x in the sense of Clarke" is also used in the literature (see, [RoW98], p. 199). Two properties of regularity that are important for our purposes are that (1) if X is convex, then it is regular at each $x \in X$, and (2) if X is regular at an $x \in X$, then $T_X(x)$ is convex ([RoW98], pp. 203 and 221).

A classical necessary condition for a vector $x^* \in C$ to be a local minimum of f over C is

$$\nabla f(x^*)'y \geq 0, \qquad \forall\ y \in T_C(x^*), \tag{1.3}$$

where $T_C(x^*)$ is the tangent cone of C at x^* (see e.g., [BSS93], [Ber99], p. 335, [Hes75], [Roc93], [RoW98]). Necessary conditions that involve Lagrange multipliers relate to the specific representation of the constraint set C in terms of the constraint functions h_i and g_j. In particular, we say that the constraint set C of Eq. (1.2) *admits Lagrange multipliers* at a point $x^* \in C$ if for every continuously differentiable cost function f for which x^* is a local minimum of problem (1.1) there exist vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ that satisfy the following conditions:

$$\left(\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\ y \in T_X(x^*), \tag{1.4}$$

$$\mu_j^* \geq 0, \qquad \forall\, j = 1, \dots, r, \tag{1.5}$$

$$\mu_j^* = 0, \qquad \forall\, j \notin A(x^*), \tag{1.6}$$

where $A(x^*) = \{j \mid g_j(x^*) = 0\}$ is the index set of inequality constraints that are active at x^*. We refer to such a pair (λ^*, μ^*) as a *Lagrange multiplier vector* or simply a *Lagrange multiplier*.

Conditions that guarantee the admittance of Lagrange multipliers are called *constraint qualifications*, and have been investigated extensively in the literature. Some of the most useful ones are the following:

CQ1: $X = \Re^n$ and x^* is a regular point in the sense that the equality constraint gradients $\nabla h_i(x^*)$, $i = 1, \dots, m$, and the active inequality constraint gradients $\nabla g_j(x^*)$, $j \in A(x^*)$, are linearly independent.

CQ2: $X = \Re^n$, the equality constraint gradients $\nabla h_i(x^*)$, $i = 1, \dots, m$, are linearly independent, and there exists a $y \in \Re^n$ such that

$$\nabla h_i(x^*)' y = 0, \quad i = 1, \dots, m, \qquad \nabla g_j(x^*)' y < 0, \quad \forall\, j \in A(x^*).$$

For the case where there are no equality constraints, this is known as the Arrow-Hurwitz-Uzawa constraint qualification, introduced in [AHU61]. In the more general case where there are equality constraints, it is known as the Mangasarian-Fromovitz constraint qualification, introduced in [MaF67].

CQ3: $X = \Re^n$, the functions h_i are linear and the functions g_j are concave.

It is well-known that all of the above constraint qualifications imply that the constraint set admits Lagrange multipliers (see e.g., [Ber99] or [BSS93]). It is also well-known that constraint qualifications and Lagrange multipliers are related to exact penalty functions. In particular, let us say that the constraint set C *admits an exact penalty* at the feasible point x^* if for every continuously differentiable function f for which x^* is a strict local minimum of f over C, there is a scalar $c > 0$ such that x^* is also a local minimum of the function

$$F_c(x) = f(x) + c\left(\sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{r} g_j^+(x)\right)$$

over $x \in X$, where we denote

$$g_j^+(x) = \max\{0, g_j(x)\}.$$

We have the following:

(a) If X is convex and the constraint set admits an exact penalty at x^* it also admits Lagrange multipliers at x^* (this follows from Prop. 3.112 of Bonnans and Shapiro [BoS00]; see also the subsequent Prop. 5).

(b) Each of the above constraint qualifications CQ1-CQ3 implies that C admits an exact penalty (the case of CQ1 was treated by Pietrzykowski [Pie69]; the case of CQ2 was treated by Zangwill [Zan67], Han and Mangasarian [HaM79], and Bazaraa and Goode [BaG82]; the case of CQ3 will be dealt with in the present paper – see the subsequent Props. 2 and 4).

In this paper, we establish the connections between constraint qualifications, Lagrange multipliers, and exact penalty functions. Much of our analysis is motivated by an enhanced set of Fritz John necessary conditions which are introduced in the next section. These conditions were proved in a somewhat weaker form for the case where $X = \Re^n$ in a largely overlooked analysis by Hestenes [Hes75] (see the discussion in Section 2). They were generalized for the case where X is a general closed convex set in the first author's recent textbook [Ber99] (Prop. 3.3.11), and they will be further generalized in Section 2 for the case where X is a closed but not necessarily convex set.

In Section 3, we introduce the new notion of *constraint pseudonormality*, and we discuss its connection with classical results relating constraint qualifications and the admittance of Lagrange multipliers. We also give a new and natural extension of the Mangasarian-Fromovitz constraint qualification that applies to the case where $X \neq \Re^n$. Finally, in Section 4, we make the connection between pseudonormality and exact penalty functions. In particular, we show that pseudonormality implies the admittance of an exact penalty, while being implied by the major constraint qualifications. In the process we prove in a unified way that the constraint set admits an exact penalty for a much larger variety of constraint qualifications than has been known hitherto.

2. ENHANCED FRITZ JOHN CONDITIONS

The Fritz John necessary optimality conditions [Joh48] are often used as the starting point for the analysis of Lagrange multipliers. Unfortunately, these conditions in their classical form are insufficient to derive the existence of Lagrange multipliers under some of the standard constraint qualifications, such as linearity of the constraint functions h and g. Recently, the classical Fritz John conditions have been strengthened through the addition of an extra necessary condition, and their effectiveness has been significantly enhanced (see [Hes75] for the case $X = \Re^n$, and [Ber99] for the case where X is a closed convex set). A further extension is given by the following proposition.

Proposition 1: Let x^* be a local minimum of problem (1.1)-(1.2). Then there exist scalars μ_0^*, $\lambda_1^*, \ldots, \lambda_m^*$, and $\mu_1^*, \ldots, \mu_r^*$, satisfying the following conditions:

(i) $-\left(\mu_0^* \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$.

(ii) $\mu_j^* \geq 0$ for all $j = 0, 1, \ldots, r$.

(iii) $\mu_0^*, \lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*$ are not all equal to 0.

(iv) If the set $I \cap J$ is nonempty, where $I = \{i \mid \lambda_i^* \neq 0\}$ and $J = \{j \mid \mu_j^* > 0\}$, then given any neighborhood B of x^* and any $\epsilon > 0$, there is an $x \in B \cap X$ such that

$$f(x) < f(x^*), \qquad \lambda_i^* h_i(x) > 0, \quad \forall\, i \in I, \qquad g_j(x) > 0, \quad \forall\, j \in J, \tag{2.1}$$

$$|h_i(x)| \leq \epsilon\, m(x), \quad \forall\, i \notin I, \qquad g_j(x) \leq \epsilon\, m(x), \quad \forall\, j \notin J, \tag{2.2}$$

where

$$m(x) = \min\Big\{\min\{|h_i(x)| \mid i \in I\}, \min\{g_j(x) \mid j \in J\}\Big\}. \tag{2.3}$$

Proof: We use a quadratic penalty function approach. For each $k = 1, 2, \ldots$, consider the "penalized" problem

$$\text{minimize} \quad F^k(x) \equiv f(x) + \frac{k}{2}\sum_{i=1}^m \big(h_i(x)\big)^2 + \frac{k}{2}\sum_{j=1}^r \big(g_j^+(x)\big)^2 + \frac{1}{2}||x - x^*||^2$$
$$\text{subject to} \quad x \in X \cap S,$$

where $S = \{x \mid ||x - x^*|| \leq \epsilon\}$, and $\epsilon > 0$ is such that $f(x^*) \leq f(x)$ for all feasible x with $x \in S$. Since $X \cap S$ is compact, by Weierstrass' theorem, we can select an optimal solution x^k of the above problem. We have for all k

$$f(x^k) + \frac{k}{2}\sum_{i=1}^m \big(h_i(x^k)\big)^2 + \frac{k}{2}\sum_{j=1}^r \big(g_j^+(x^k)\big)^2 + \frac{1}{2}||x^k - x^*||^2 = F^k(x^k) \leq F^k(x^*) = f(x^*) \tag{2.4}$$

and since $f(x^k)$ is bounded over $X \cap S$, we obtain

$$\lim_{k \to \infty} ||h_i(x^k)|| = 0, \quad i = 1, \ldots, m, \qquad \lim_{k \to \infty} ||g_j^+(x^k)|| = 0, \quad j = 1, \ldots, r;$$

otherwise the left-hand side of Eq. (2.4) would become unbounded from above as $k \to \infty$. Therefore, every limit point $\overline{x}$ of $\{x^k\}$ is feasible, i.e., $\overline{x} \in C$. Furthermore, Eq. (2.4) yields $f(x^k) + (1/2)||x^k - x^*||^2 \leq f(x^*)$ for all k, so by taking the limit as $k \to \infty$, we obtain

$$f(\overline{x}) + \frac{1}{2}||\overline{x} - x^*||^2 \leq f(x^*).$$

Since $\overline{x} \in S$ and $\overline{x}$ is feasible, we have $f(x^*) \leq f(\overline{x})$, which when combined with the preceding inequality yields $||\overline{x} - x^*|| = 0$ so that $\overline{x} = x^*$. Thus the sequence $\{x^k\}$ converges to x^*, and it follows that x^k is an interior point of the closed sphere S for all k greater than some $\overline{k}$.

For $k \geq \overline{k}$, we have by the necessary condition (1.3), $\nabla F^k(x^k)'y \geq 0$ for all $y \in T_X(x^k)$, or equivalently $-\nabla F^k(x^k) \in T_X(x^k)^\perp$, which is written as

$$-\left(\nabla f(x^k) + \sum_{i=1}^m \xi_i^k \nabla h_i(x^k) + \sum_{j=1}^r \zeta_j^k \nabla g_j(x^k) + (x^k - x^*)\right) \in T_X(x^k)^\perp, \tag{2.5}$$

where

$$\xi_i^k = k h_i(x^k), \qquad \zeta_j^k = k g_j^+(x^k). \tag{2.6}$$

Denote,

$$\delta^k = \sqrt{1 + \sum_{i=1}^m (\xi_i^k)^2 + \sum_{j=1}^r (\zeta_j^k)^2}, \tag{2.7}$$

$$\mu_0^k = \frac{1}{\delta^k}, \qquad \lambda_i^k = \frac{\xi_i^k}{\delta^k}, \quad i = 1, \ldots, m, \qquad \mu_j^k = \frac{\zeta_j^k}{\delta^k}, \quad j = 1, \ldots, r. \tag{2.8}$$

Then by dividing Eq. (2.5) with δ^k, we obtain

$$-\left(\mu_0^k \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^r \mu_j^k \nabla g_j(x^k) + \frac{1}{\delta^k}(x^k - x^*)\right) \in T_X(x^k)^\perp. \tag{2.9}$$

Since by construction we have

$$(\mu_0^k)^2 + \sum_{i=1}^m (\lambda_i^k)^2 + \sum_{j=1}^r (\mu_j^k)^2 = 1, \tag{2.10}$$

the sequence $\{\mu_0^k, \lambda_1^k, \ldots, \lambda_m^k, \mu_1^k, \ldots, \mu_r^k\}$ is bounded and must contain a subsequence that converges to some limit $\{\mu_0^*, \lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*\}$.

From Eq. (2.9) and the definition of the normal cone $N_X(x^*)$, we see that μ_0^*, λ_i^*, and μ_j^* must satisfy condition (i). From Eqs. (2.6) and (2.8), μ_0^*, λ_i^*, and μ_j^* must satisfy condition (ii), and from Eq. (2.10), they must satisfy condition (iii). Finally, to show that condition (iv) is satisfied, assume that $I \cup J$ is nonempty (otherwise we are done), and note that for all sufficiently large k within the index set $\mathcal{K}$ of the convergent subsequence, we must have $\lambda_i^* \lambda_i^k > 0$ for all $i \in I$ and $\mu_j^* \mu_j^k > 0$ for all $j \in J$. Therefore, for these k, from Eqs. (2.6) and (2.8), we must have $\lambda_i^* h_i(x^k) > 0$ for all $i \in I$ and $\mu_j^* g_j(x^k) > 0$ for all $j \in J$, while from Eq. (2.4), we have $f(x^k) < f(x^*)$ for k sufficiently large (the case where $x^k = x^*$ for infinitely many k is excluded by the assumption that $I \cup J$ is nonempty). Furthermore, the conditions $|h_i(x^k)| \leq \epsilon m(x_k)$ for $i \notin I$ and $g_j(x^k) \leq \epsilon m(x_k)$ for $j \notin J$ are equivalent to

$$|\lambda_i^k| \leq \epsilon \min\big\{\min\{|\lambda_i^k| \mid i \in I\}, \min\{\mu_j^k \mid j \in J\}\big\}, \qquad \forall\, i \notin I,$$

and

$$\mu_j^k \leq \epsilon \min\big\{\min\{|\lambda_i^k| \mid i \in I\}, \min\{\mu_j^k \mid j \in J\}\big\}, \qquad \forall\, j \notin J,$$

respectively, so they evidently hold for all sufficiently large k in $\mathcal{K}$. Since every neighborhood of x^* must contain all x^k with sufficiently large k in $\mathcal{K}$, this proves condition (iv). **Q.E.D.**

Note that if X is regular at x^*, i.e., $N_X(x^*) = T_X(x^*)^\perp$, condition (i) of Prop. 1 becomes $-\Big(\mu_0^* \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\Big) \in T_X(x^*)^\perp$ or equivalently

$$\left(\mu_0^* \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\, y \in T_X(x^*).$$

If in addition, the scalar μ_0^* can be shown to be strictly positive, then by normalization we can choose $\mu_0^* = 1$, and condition (i) of Prop. 1 becomes equivalent to the Lagrangian stationarity condition (1.4). Thus, if X is regular at x^* and we can guarantee that $\mu_0^* = 1$, the vector $(\lambda^*, \mu^*) = \{\lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*\}$ is a Lagrange multiplier vector that satisfies condition (iv) of Prop. 1. A key fact is that this condition is stronger than the complementary slackness condition (1.6) [if $\mu_j^* > 0$, then according to condition (iv), the corresponding jth inequality constraint must be violated arbitrarily close to x^* [cf. Eq. (2.1)], implying that $g_j(x^*) = 0$]. The strengthening of the complementary slackness condition will turn out to be of crucial significance in the next section.

Condition (iv) of Prop. 1 also has another important implication. It asserts that the multiplier vector (λ^*, μ^*) has a special sensitivity-like property: through the signs of the λ_i^* and μ_j^*, it identifies the constraints that if violated, permit the reduction of the cost function. We call such a Lagrange multiplier *informative*, and in a separate, more extensive report on the subject of this paper, we show an interesting new result: *if $T_X(x^*)$ is convex and there exists at least one*

Lagrange multiplier vector at a local minimum x^* of problem (1.1)-(1.2), then there exists at least one Lagrange multiplier vector that is informative [convexity of $T_X(x^*)$, which is guaranteed if X is regular at x^*, is an essential property for this result to hold].

To place Prop. 1 in perspective, we note that its line of proof, based on the quadratic penalty function, is due to McShane [McS73]. Hestenes [Hes75] observed that McShane's proof can be used to strengthen the complementary slackness condition to assert the existence, within any neighborhood B of x^*, of an $x \in B \cap X$ such that

$$\lambda_i^* h_i(x) > 0, \quad \forall\, i \in I, \qquad g_j(x) > 0, \quad \forall\, j \in J, \tag{2.11}$$

which is slightly weaker than condition (iv) of Prop. 1 [there is no requirement that x simultaneously satisfies $f(x) < f(x^*)$ and Eq. (2.2)]. McShane and Hestenes considered only the case where $X = \Re^n$. The case where X is a closed convex set was considered in Bertsekas [Ber99], where a generalized version of the Mangasarian-Fromovitz constraint qualification was also proved. The extension to the case where X is a general closed set and the strengthened version of condition (iv) are presented in the present paper for the first time.

3. PSEUDONORMALITY AND CONSTRAINT QUALIFICATIONS

Proposition 1 leads to the introduction of a general constraint qualification under which the scalar μ_0^* in Prop. 1 cannot be zero. In particular, let us say that a feasible vector x^* of problem (1.1)-(1.2) is *pseudonormal* if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$, such that:

(i) $-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^r \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$.

(ii) $\mu_j \geq 0$, for all $j = 1, \ldots, r$, and $\mu_j = 0$ for all $j \notin A(x^*)$.

(iii) In every neighborhood B of x^* there is an $x \in B \cap X$ such that

$$\sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^r \mu_j g_j(x) > 0. \tag{3.1}$$

It can be seen that if x^* is a pseudonormal local minimum, the Fritz John conditions of Prop. 1 cannot be satisfied with $\mu_0^* = 0$, so that μ_0^* can be taken equal to 1. Then, if X is regular at x^*, the vector $(\lambda^*, \mu^*) = (\lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*)$ is an informative Lagrange multiplier.

Let us also mention another interesting property of the constraint set, called *quasinormality*, which is based on the condition (2.11). Quasinormality was introduced, for the case $X = \Re^n$, by Hestenes [Hes75], who showed how it can be used to unify various constraint qualifications. Pseudonormality implies quasinormality, and is thus intermediate between the constraint qualifications of this paper and quasinormality. As we will show in Section 4, pseudonormality is more naturally suited for showing results relating to the existence of exact penalties. Let us also note that in a separate report we generalize the notion of quasinormality to the case where $X \neq \Re^n$, and we discuss its relation to a slightly weaker version of pseudonormality where the vector x in Eq. (3.1) is additionally required to satisfy $\lambda_i h_i(x) \geq 0$ for all i and $\mu_j g_j(x) \geq 0$ for all j.

We now give some additional constraint qualifications, which together with CQ1-CQ3, given in Section 1, will be seen to imply pseudonormality of a feasible vector x^*.

CQ4: $X = \Re^n$ and for some integer $\overline{r} < r$, the following superset $\overline{C}$ of the constraint set C,

$$\overline{C} = \big\{x \mid h_i(x) = 0,\, i = 1, \ldots, m,\, g_j(x) \leq 0,\, j = \overline{r} + 1, \ldots, r\big\},$$

is pseudonormal at x^*. Furthermore, there exists a $y \in \Re^n$ such that

$$\nabla h_i(x^*)'y = 0, \quad i = 1, \ldots, m, \qquad \nabla g_j(x^*)'y \leq 0, \quad \forall\, j \in A(x^*),$$

$$\nabla g_j(x^*)'y < 0, \quad \forall\, j \in \{1, \ldots, \overline{r}\} \cap A(x^*).$$

Since CQ1-CQ3 imply pseudonormality, a fact to be shown in the subsequent Prop. 2, we see that CQ4 generalizes all the constraint qualifications CQ1-CQ3.

CQ5:

(a) The equality constraints with index above some $\overline{m} \leq m$:

$$h_i(x) = 0, \qquad i = \overline{m} + 1, \ldots, m,$$

are linear.

(b) There does not exist a vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*) \tag{3.2}$$

and at least one of the scalars $\lambda_1, \ldots, \lambda_{\overline{m}}$ is nonzero.

(c) The subspace

$$V_L(x^*) = \big\{y \mid \nabla h_i(x^*)'y = 0,\, i = \overline{m} + 1, \ldots, m\big\}$$

has a nonempty intersection with the interior of $N_X(x^*)^{\perp}$.

(d) There exists a $y \in N_X(x^*)^{\perp}$ such that

$$\nabla h_i(x^*)'y = 0, \quad i = 1, \ldots, m, \qquad \nabla g_j(x^*)'y < 0, \quad \forall\, j \in A(x^*).$$

We refer to CQ5 as the *generalized Mangasarian-Fromovitz constraint qualification*, since it reduces to CQ2 when $X = \Re^n$ and none of the equality constraints is assumed to be linear. This constraint qualification has several special cases, which we list below.

CQ5a:

(a) There does not exist a nonzero vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*).$$

(b) There exists a $y \in N_X(x^*)^{\perp}$ such that

$$\nabla h_i(x^*)'y = 0, \quad i = 1, \ldots, m, \qquad \nabla g_j(x^*)'y < 0, \quad \forall\, j \in A(x^*).$$

CQ5b: There are no inequality constraints, the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, are linearly independent, and the subspace

$$V(x^*) = \big\{y \mid \nabla h_i(x^*)'y = 0,\, i = 1, \ldots, m\big\}$$

contains a point in the interior of $N_X(x^*)^{\perp}$.

CQ5c: X is convex, there are no inequality constraints, the functions h_i, $i = 1, \ldots, m$, are linear, and the linear manifold

$$\big\{y \mid h_i(x) = 0,\, i = 1, \ldots, m\big\}$$

contains a point in the interior of X.

CQ5d: X is convex, the functions g_j are convex, there are no equality constraints, and there exists a feasible vector $\overline{x}$ satisfying

$$g_j(\overline{x}) < 0, \qquad \forall\, j \in A(x^*).$$

CQ5a is the special case of CQ5 where all equality constraints are assumed nonlinear. CQ5b is a special case of CQ5 (where there are no inequality constraints and no linear equality constraints) based on the fact that if $\nabla h_i(x^*)$, $i = 1, \ldots, m$, are linearly independent and the subspace $V(x^*)$ contains a point in the interior of $N_X(x^*)^\perp$, then it can be shown that assumption (b) of CQ5 is satisfied. Finally, the convexity assumptions in CQ5c and CQ5d can be used to establish the corresponding assumption (c) and (d) of CQ5, respectively. Note that CQ5d is the well-known Slater constraint qualification, introduced in [Sla50].

Let us also mention the following constraint qualification.

CQ6: There are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$ such that conditions (i) and (ii) of the definition of pseudonormality are satisfied.

CQ6 is the constraint qualification introduced by Rockafellar [Roc93], who used McShane's line of proof to derive the Fritz John conditions in the classical form where complementary slackness replaces condition (iv) in Prop. 1. Clearly CQ6 is a more restrictive condition than pseudonormality, since it does not require condition (iii) of the definition of pseudonormality. It can be shown that CQ6 together with regularity of X at x^*, is equivalent to CQ5a. This is proved by Rockafellar and Wets [RoW98] in the case where $X = \Re^n$, and can be verified in the more general case where $X \neq \Re^n$ by using their analysis given in p. 226 of [RoW98]. However, CQ3, CQ4, and CQ5 do not imply CQ6. Thus CQ6 is not as effective in unifying various existing constraint qualifications as pseudonormality, which is implied by all the constraint qualifications CQ1-CQ6, as shown in the following proposition.

Proposition 2: A feasible point x^* of problem (1.1)-(1.2) is pseudonormal if any one of the constraint qualifications CQ1-CQ6 is satisfied.

Proof: We will not consider CQ2 since it is a special case of CQ5. It is also evident that CQ6 implies pseudonormality. Thus we will prove the result for the cases CQ1, CQ3, CQ4, and CQ5 in that order. In all cases, the method of proof is by contradiction, i.e., we assume that there are scalars λ_i, $i = 1, \ldots, m$, and μ_j, $j = 1, \ldots, r$, which satisfy conditions (i)-(iii) of the definition of pseudonormality. We then assume that each of the constraint qualifications CQ1, CQ3, CQ4, and CQ5 is in turn also satisfied, and we arrive at a contradiction.

CQ1: Since $X = \Re^n$, implying that $N_X(x^*) = \{0\}$, and we have $\mu_j = 0$ for all $j \notin A(x^*)$ by condition (ii), we can write condition (i) as

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0.$$

Linear independence of $\nabla h_i(x^*)$, $i = 1, \ldots, m$, and $\nabla g_j(x^*)$, $j \in A(x^*)$, implies that $\lambda_i = 0$ for all i and $\mu_j = 0$ for all $j \in A(x^*)$. This, together with the condition $\mu_j = 0$ for all $j \notin A(x^*)$, violates condition (iii).

CQ3: By the linearity of h_i and the concavity of g_j, we have for all $x \in \Re^n$,

$$h_i(x) = h_i(x^*) + \nabla h_i(x^*)'(x - x^*), \qquad i = 1, \ldots, m,$$

$$g_j(x) \le g_j(x^*) + \nabla g_j(x^*)'(x - x^*), \qquad j = 1, \ldots, r.$$

By multiplying these two relations with λ_i and μ_j, and by adding over i and j, respectively, we obtain

$$\begin{aligned} \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x) &\le \sum_{i=1}^{m} \lambda_i h_i(x^*) + \sum_{j=1}^{r} \mu_j g_j(x^*) \\ &\quad + \left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \right)' (x - x^*) \\ &= 0, \end{aligned} \tag{3.3}$$

where the last equality holds because we have $\lambda_i h_i(x^*) = 0$ for all i and $\mu_j g_j(x^*) = 0$ for all j [by condition (ii)], and

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) = 0$$

[by condition (i)]. On the other hand, by condition (iii), there is an x satisfying $\sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x) > 0$, contradicting Eq. (3.3).

CQ4: It is not possible that $\mu_j = 0$ for all $j \in \{1, \ldots, \overline{r}\}$, since if this were so, the pseudonormality assumption for $\overline{C}$ would be violated. Thus we have $\mu_j > 0$ for some $j \in \{1, \ldots, \overline{r}\} \cap A(x^*)$. It follows that for the vector y appearing in the statement of CQ4, we have $\sum_{j=1}^{\overline{r}} \mu_j \nabla g_j(x^*)'y < 0$, so that

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)'y + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)'y < 0.$$

This contradicts the equation

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) = 0,$$

[cf. condition (i)].

CQ5: We first show by contradiction that at least one of the $\lambda_1, \ldots, \lambda_{\overline{m}}$ and μ_j, $j \in A(x^*)$ must be nonzero. If this were not so, then by using a translation argument we may assume that x^* is the origin, and the linear constraints have the form $a_i'x = 0$, $i = \overline{m}+1, \ldots, m$. Using condition (i) we have

$$-\sum_{i=\overline{m}+1}^{m} \lambda_i a_i \in N_X(x^*). \tag{3.4}$$

Let $\overline{y}$ be the interior point of $N_X(x^*)^\perp$ that satisfies $a_i'\overline{y} = 0$ for all $i = \overline{m}+1, \ldots, m$, and let S be an open sphere centered at the origin such that $\overline{y} + d \in N_X(x^*)^\perp$ for all $d \in S$. We have from Eq. (3.4),

$$\sum_{i=\overline{m}+1}^{m} \lambda_i a_i' d \geq 0, \qquad \forall\ d \in S,$$

from which we obtain $\sum_{i=\overline{m}+1}^{m} \lambda_i a_i = 0$. This contradicts condition (iii), which requires that there exists some $x \in S \cap X$ such that $\sum_{i=\overline{m}+1}^{m} \lambda_i a_i' x > 0$.

Next we show by contradiction that we cannot have $\mu_j = 0$ for all j. If this were so, by condition (i) there must exist a nonzero vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

$$-\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*). \tag{3.5}$$

By what has been proved above, the multipliers $\lambda_1, \ldots, \lambda_{\overline{m}}$ of the nonlinear constraints cannot be all zero, so Eq. (3.5) contradicts assumption (b) of CQ5.

Hence we must have $\mu_j > 0$ for at least one j, and since $\mu_j \geq 0$ for all j with $\mu_j = 0$ for $j \notin A(x^*)$, we obtain

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)' y + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)' y < 0,$$

for the vector y of $N_X(x^*)^\perp$ that appears in assumption (d) of CQ5. Thus,

$$-\left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \right) \notin \big(N_X(x^*)^\perp\big)^\perp.$$

Since $N_X(x^*) \subset \big(N_X(x^*)^\perp\big)^\perp$, this contradicts condition (i). **Q.E.D.**

A consequence of Prop. 2 is that each of the constraint qualifications CQ1-CQ6 implies that x^* is pseudonormal, so if X is regular at x^*, by Prop. 1, the constraint set C admits informative Lagrange multipliers at x^*. In the next two sections, we will also show similar implications regarding the admittance of an exact penalty at x^*.

4. PSEUDONORMALITY AND EXISTENCE OF EXACT PENALTY FUNCTIONS

We will show that pseudonormality implies that the constraint admits an exact penalty, which in turn, together with regularity of X at x^*, implies that the constraint set admits Lagrange multipliers. We first use the generalized Mangasarian-Fromovitz constraint qualification CQ5 to obtain a necessary condition for a local minimum of the exact penalty function.

Proposition 3: Let x^* be a local minimum of

$$F_c(x) = f(x) + c\left(\sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{r} g_j^+(x)\right)$$

over X. Then there exist $\lambda_1^*, \ldots, \lambda_m^*$ and $\mu_1^*, \ldots, \mu_r^*$ such that

$$-\left(\nabla f(x^*) + c\left(\sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right)\right) \in N_X(x^*),$$

$$\lambda_i^* = 1 \quad \text{if } h_i(x^*) > 0, \qquad \lambda_i^* = -1 \quad \text{if } h_i(x^*) < 0, \qquad \lambda_i^* \in [-1, 1] \quad \text{if } h_i(x^*) = 0,$$

$$\mu_j^* = 1 \quad \text{if } g_j(x^*) > 0, \qquad \mu_j^* = 0 \quad \text{if } g_j(x^*) < 0, \qquad \mu_j^* \in [0, 1] \quad \text{if } g_j(x^*) = 0.$$

Proof: The problem of minimizing $F_c(x)$ over $x \in X$ can be converted to the problem

$$\text{minimize} \quad f(x) + c\left(\sum_{i=1}^{m} w_i + \sum_{j=1}^{r} v_j\right)$$

$$\text{subject to} \quad x \in X, \quad h_i(x) \le w_i, \ -h_i(x) \le w_i, \ i = 1, \ldots, m, \quad g_j(x) \le v_j, \ 0 \le v_j, \ j = 1, \ldots, r,$$

which involves the auxiliary variables w_i and v_j. It can be seen that at the local minimum of this problem that corresponds to x^* the constraint qualification CQ5 is satisfied. Thus, by Prop. 2, there exist multipliers satisfying the conditions of Prop. 1, which with straightforward calculation, yield scalars $\lambda_1^*, \ldots, \lambda_m^*$ and $\mu_1^*, \ldots, \mu_r^*$, satisfying the desired conditions. **Q.E.D.**

Proposition 4: If x^* is a feasible vector of problem (1.1)-(1.2), which is pseudonormal, the constraint set admits an exact penalty at x^*.

Proof: Assume the contrary, i.e., that there exists a continuously differentiable f such that x^* is a strict local minimum of f over the constraint set C, while x^* is not a local minimum over $x \in X$ of the function

$$F_k(x) = f(x) + k\left(\sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{r} g_j^+(x)\right)$$

for all $k = 1, 2, \ldots$ Let $\epsilon > 0$ be such that

$$f(x^*) < f(x), \qquad \forall\, x \in C \text{ with } x \neq x^* \text{ and } \|x - x^*\| \leq \epsilon. \tag{4.1}$$

Suppose that x^k minimizes $F_k(x)$ over the (compact) set of all $x \in X$ satisfying $\|x - x^*\| \leq \epsilon$. Then, since x^* is not a local minimum of F_k over X, we must have that $x^k \neq x^*$, and that x^k is infeasible for problem (1.2), i.e.,

$$\sum_{i=1}^{m} |h_i(x^k)| + \sum_{j=1}^{r} g_j^+(x^k) > 0. \tag{4.2}$$

We have

$$F_k(x^k) = f(x^k) + k\left(\sum_{i=1}^{m} |h_i(x^k)| + \sum_{j=1}^{r} g_j^+(x^k)\right) \leq f(x^*), \tag{4.3}$$

so it follows that $h_i(x^k) \to 0$ for all i and $g_j^+(x^k) \to 0$ for all j. The sequence $\{x^k\}$ is bounded and if $\overline{x}$ is any of its limit points, we have that $\overline{x}$ is feasible. From Eqs. (4.1) and (4.3) it then follows that $\overline{x} = x^*$. Thus $\{x^k\}$ converges to x^* and we have $\|x^k - x^*\| < \epsilon$ for all sufficiently large k. This implies the following necessary condition for optimality of x^k (cf. Prop. 3):

$$-\left(\frac{1}{k}\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k)\right) \in N_X(x^k), \tag{4.4}$$

where

$$\lambda_i^k = 1 \quad \text{if } h_i(x^k) > 0, \qquad \lambda_i^k = -1 \quad \text{if } h_i(x^k) < 0, \qquad \lambda_i^k \in [-1, 1] \quad \text{if } h_i(x^k) = 0,$$

$$\mu_j^k = 1 \quad \text{if } g_j(x^k) > 0, \qquad \mu_j^k = 0 \quad \text{if } g_j(x^k) < 0, \qquad \mu_j^k \in [0, 1] \quad \text{if } g_j(x^k) = 0.$$

In view of Eq. (4.2), we can find a subsequence $\{\lambda^k, \mu^k\}_{k \in \mathcal{K}}$ such that for some equality constraint index i we have $|\lambda_i^k| = 1$ and $h_i(x^k) \neq 0$ for all $k \in \mathcal{K}$ or for some inequality constraint index j we have $\mu_j^k = 1$ and $g_j(x^k) > 0$ for all $k \in \mathcal{K}$. Let (λ, μ) be a limit point of this subsequence. We then have $(\lambda, \mu) \neq (0, 0)$, $\mu \geq 0$. Using the closure of the mapping $x \mapsto N_X(x)$, Eq. (4.4) yields

$$-\left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right) \in N_X(x^*). \tag{4.5}$$

Finally, for all $k \in \mathcal{K}$, we have $\lambda_i^k h_i(x^k) \geq 0$ for all i, $\mu_j^k g_j(x^k) \geq 0$ for all j, so that, for all $k \in \mathcal{K}$, $\lambda_i h_i(x^k) \geq 0$ for all i, $\mu_j g_j(x^k) \geq 0$ for all j. Since by construction of the subsequence $\{\lambda^k, \mu^k\}_{k \in \mathcal{K}}$, we have for some i and all $k \in \mathcal{K}$, $|\lambda_i^k| = 1$ and $h_i(x^k) \neq 0$, or for some j and all $k \in \mathcal{K}$, $\mu_j^k = 1$ and $g_j(x^k) > 0$, it follows that for all $k \in \mathcal{K}$,

$$\sum_{i=1}^{m} \lambda_i h_i(x^k) + \sum_{j=1}^{r} \mu_j g_j(x^k) > 0. \tag{4.6}$$

Thus, Eqs. (4.5) and (4.6) violate the hypothesis that x^* is pseudonormal. **Q.E.D.**

Proposition 5: Let x^* be a feasible vector of problem (1.1)-(1.2), and let X be regular at x^*. If the constraint set admits an exact penalty at x^*, it admits Lagrange multipliers at x^*.

Proof: Suppose that a given continuously differentiable function $f(x)$ has a local minimum at x^*. Then the function $f(x) + \|x - x^*\|^2$ has a strict local minimum at x^*. Since C admits an exact penalty at x^*, there exist λ_i^* and μ_j^* satisfying the conditions of Prop. 3 (the term $\|x - x^*\|^2$ in the cost function is inconsequential, since its gradient at x^* is 0). In view of the regularity of X at x^*, the λ_i^* and μ_j^* are Lagrange multipliers. **Q.E.D.**

REFERENCES

[AuF90] Aubin, J.-P., and Frankowska, H., Set-Valued Analysis, Birkhauser, Boston.

[BSS93] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M., 1993. Nonlinear Programming Theory and Algorithms, (2nd Ed.), Wiley, N. Y.

[BaG82] Bazaraa, M. S., and Goode, J. J., 1982. "Sufficient Conditions for a Globally Exact Penalty Function without Convexity," Math. Programming Stud., Vol. 19, pp. 1-15.

[Ber99] Bertsekas, D. P., "Nonlinear Programming: 2nd Edition," Athena Scientific, Belmont, MA, 1999.

[BoL00] Borwein, J. M., and Lewis, A. S., Convex Analysis and Nonlinear Optimization, Springer-Verlag, N. Y.

[BoS00] Bonnans, J. F., and Shapiro, A., 2000. Perturbation Analysis of Optimization Problems, Springer-Verlag, N. Y.

[HaM79] Han, S. P., and Mangasarian, O. L., 1979. "Exact Penalty Functions in Nonlinear Programming," Math. Programming, Vol. 17, pp. 251-269.

[Hes75] Hestenes, M. R., 1975. Optimization Theory: The Finite Dimensional Case, Wiley, N. Y.

[Joh48] John, F., 1948. "Extremum Problems with Inequalities as Subsidiary Conditions," in Studies and Essays: Courant Anniversary Volume, K. O. Friedrichs, Neugebauer, O. E., and Stoker, J. J., (Eds.), Wiley-Interscience, N. Y., pp. 187-204.

[McS73] McShane, E. J., 1973. "The Lagrange Multiplier Rule," Amer. Mathematical Monthly, Vol. 80, pp. 922-925.

[Mor76] Mordukhovich, B. S., 1976. "Maximum Principle in the Problem of Time Optimal Response with Nonsmooth Constraints," J. of Applied Mathematics and Mechanics, Vol. 40, pp. 960-969.

[Pie69] Pietrzykowski, T., 1969. "An Exact Potential Method for Constrained Maxima," SIAM J. Numer. Anal., Vol. 6, pp. 294-304.

[RoW98] Rockafellar, R. T., and Wets, R. J.-B., 1998. Variational Analysis, Springer-Verlag, Berlin.

[Roc93] Rockafellar, R. T., 1993. "Lagrange Multipliers and Optimality," SIAM Review, Vol. 35, pp. 183-238.

[Zan67] Zangwill, W. I., 1967. "Nonlinear Programming via Penalty Functions," Management Sci., Vol. 13, pp. 344-358.