Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization[1]

by

Dimitri P. Bertsekas and Asuman E. Ozdaglar[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, two new properties, pseudonormality and quasinormality, emerge as central within the taxonomy of interesting constraint characteristics. In the case where there is no abstract set constraint, these properties provide the connecting link between the classical constraint qualifications and two distinct pathways to the existence of Lagrange multipliers: one involving the notion of quasiregularity and Farkas' Lemma, and the other involving the use of exact penalty functions. The second pathway also applies in the general case where there is an abstract set constraint.

---

1. INTRODUCTION

We consider finite-dimensional optimization problems of the form

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

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 smooth (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, so $x'y$ denotes the inner product of the vectors x and y. We will use throughout the standard Euclidean norm $\|x\| = (x'x)^{1/2}$.

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 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., Bazaraa, Sherali, and Shetty [BSS93], Rockafellar and Wets [RoW98]) is that there exist 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

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

For a nonempty cone T, we will use the well-known relation $T \subset \big(T^* \big)^*$, 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 by $N_X(x)$, which is obtained from the polar cone $T_X(x)^*$ 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 $z^k \in T_X(x^k)^*$ 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 $T_X(\cdot)^*$. 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

2

and Wets [RoW98], and Borwein and Lewis [BoL00]; for the case where X is a closed subset of $\Re^n$, our definition of $N_X(x)$ coincides with the ones used by these authors). In general, we have $T_X(x)^* \subset N_X(x)$ for any $x \in X$. However, $N_X(x)$ may not be equal to $T_X(x)^*$, and in fact it may not even be a convex set. In the case where $T_X(x)^* = N_X(x)$, 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, Rockafellar and Wets [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 some $x \in X$, then $T_X(x)$ is convex (Rockafellar and Wets [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., Bazaraa, Sherali, and Shetty [BSS93], Bertsekas [Ber99], Hestenes [Hes75], Rockafellar [Roc93], Rockafellar and Wets [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 smooth 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:

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

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

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

where $A(x^*) = \{ j \mid g_j(x^*) = 0 \}$ is the index set of inequality constraints that are active at x^*. Condition (1.6) is referred to as the *complementary slackness* condition (CS for short). A pair (λ^*, μ^*) satisfying Eqs. (1.4)-(1.6) will be called a *Lagrange multiplier vector corresponding to f and x^**. When there is no danger of confusion, we refer to (λ^*, μ^*) simply as a *Lagrange multiplier vector* or a *Lagrange multiplier*. We observe that the set of Lagrange multiplier vectors corresponding to a given f and x^* is a (possibly empty) closed and convex set.

The condition (1.4) is consistent with the traditional characteristic property of Lagrange multipliers: rendering the Lagrangian function stationary at x^* [cf. Eq. (1.3)]. When X is a convex set, Eq. (1.4) is equivalent to

$$\left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) \right)' (x - x^*) \geq 0, \qquad \forall \, x \in X. \tag{1.7}$$

This is because when X is convex, $T_X(x^*)$ is equal to the closure of the set of feasible directions $F_X(x^*)$, which is in turn equal to the set of vectors of the form $\alpha(x - x^*)$, where $\alpha > 0$ and $x \in X$. If $X = \Re^n$, Eq. (1.7) becomes

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

which together with the nonnegativity condition (1.5) and the CS condition (1.6), comprise the familiar Karush-Kuhn-Tucker conditions.

In the case where $X = \Re^n$, it is well-known (see e.g., Bertsekas [Ber99], p. 332) that for a given smooth f for which x^* is a local minimum, there exist Lagrange multipliers if and only if

$$\nabla f(x^*)'y \geq 0, \qquad \forall \ y \in V(x^*),$$

where $V(x^*)$ is the *cone of first order feasible variations at* x^*, given by

$$V(x^*) = \big\{ y \mid \nabla h_i(x^*)'y = 0, \ i = 1, \ldots, m, \ \nabla g_j(x^*)'y \leq 0, \ j \in A(x^*) \big\}.$$

This result, a direct consequence of Farkas' Lemma, leads to the classical theorem that the constraint set admits Lagrange multipliers at x^* if $T_C(x^*) = V(x^*)$. In this case we say that x^* is a *quasiregular point* or that *quasiregularity holds at* x^* [other terms used are x^* "satisfies Abadie's constraint qualification" (Abadie [Aba67], Bazaraa, Sherali, and Shetty [BSS93]), or "is a regular point" (Hestenes [Hes75])].

Since quasiregularity is a somewhat abstract property, it is useful to have more readily verifiable conditions for the admittance of Lagrange multipliers. Such conditions 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 $\nabla h_i(x^*)$, $i = 1, \ldots, m$, 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 $\nabla h_i(x^*)$, $i = 1, \ldots, m$, are linearly independent, and there exists a $y \in \Re^n$ such that

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

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 the quasiregularity condition $T_C(x^*) = V(x^*)$, and therefore imply that the constraint set admits Lagrange multipliers (see e.g., Bertsekas [Ber99], or Bazaraa, Sherali, and Shetty [BSS93]; a survey of constraint qualifications is given by Peterson [Pet73]). These results constitute the classical pathway to Lagrange multipliers for the case where $X = \Re^n$.

However, there is another equally powerful approach to Lagrange multipliers, based on exact penalty functions, which has not received much attention thus far. In particular, let us say that the constraint set C *admits an exact penalty* at the feasible point x^* if for every smooth 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

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

Note that, like admittance of Lagrange multipliers, admittance of an exact penalty is a *property of the constraint set C*, and does not depend on the cost function f of problem (1.1).

We intend to use exact penalty functions as a vehicle towards asserting the admittance of Lagrange multipliers. For this purpose, there is no loss of generality in requiring that x^* be a strict local minimum, since we can replace a cost function $f(x)$ with the cost function $f(x) + \|x - x^*\|^2$ without affecting the problem's Lagrange multipliers. On the other hand if we allow functions f involving multiple local minima, it is hard to relate constraint qualifications such as the preceding ones, the admittance of an exact penalty, and the admittance of Lagrange multipliers, as we show in Example 11 of Section 7.

Note two important points, which illustrate the significance of exact penalty functions as a unifying vehicle towards guaranteeing the admittance of Lagrange multipliers.

(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. 8, which generalizes the Bonnans-Shapiro result by assuming that X is regular at x^* instead of being convex.)

(b) All of the above constraint qualifications CQ1-CQ3 imply 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. 3 and 9.)
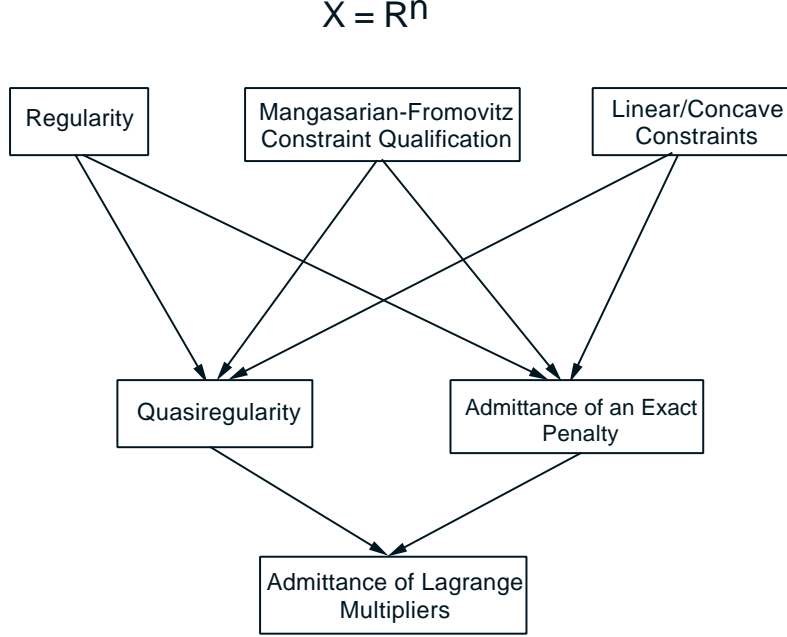
$$X = R^n$$

Figure 1. Characterizations of the constraint set C that imply admittance of Lagrange multipliers in the case where $X = \Re^n$.

Figure 1 summarizes the relationships discussed above for the case $X = \Re^n$, and highlights the two distinct pathways to the admittance of Lagrange multipliers. The two key notions, quasiregularity and admittance of an exact penalty, do not seem to be directly related (see Examples 6 and 7 in Section 7), but we will show in this paper that they are connected through the new notion of *constraint pseudonormality*, which implies both while being implied by the constraint qualifications CQ1-CQ3. Another similar connecting link is the notion of *constraint quasinormality*, which is implied by pseudonormality.

Unfortunately, when X is a strict subset of $\Re^n$ the situation changes significantly because there does not appear to be a satisfactory extension of the notion of quasiregularity, which implies admittance of Lagrange multipliers. For example, the classical constraint qualification of Guignard [Gui69] resembles quasiregularity, but requires additional conditions that are not easily verifiable. In particular, Guignard ([Gui69], Th. 2) has shown that the constraint set admits

Lagrange multipliers at x^* if

$$V(x^*) \cap \overline{\mathrm{conv}}\big(T_X(x^*)\big) = \overline{\mathrm{conv}}\big(T_C(x^*)\big), \tag{1.8}$$

and the vector sum $V(x^*)^* + T_X(x^*)^*$ is a closed set [here $\overline{\mathrm{conv}}(S)$ denotes the closure of the convex hull of a set S]. Guignard's conditions are equivalent to

$$V(x^*)^* + T_X(x^*)^* = T_C(x^*)^*,$$

which in turn can be shown to be a necessary and sufficient condition for the admittance of Lagrange multipliers at x^* based on the classical results of Gould and Tolle [GoT71], [GoT72]. In the special case where $X = \Re^n$, we have $T_X(x^*) = \Re^n$, $T_X(x^*)^* = \{0\}$, and the condition (1.8) becomes $V(x^*) = \overline{\mathrm{conv}}\big(T_C(x^*)\big)$ [or equivalently $V(x^*)^* = T_C(x^*)^*$], which is a similar but slightly less restrictive constraint qualification than quasiregularity. However, in the more general case where $X \neq \Re^n$, condition (1.8) and the closure of the set $V(x^*)^* + T_X(x^*)^*$ seem hard to verify. (Guignard [Gui69] has only treated the cases where X is either $\Re^n$ or the nonnegative orthant.)

In this paper, we focus on 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 that are introduced in the next section. Weaker versions of these conditions were shown in a largely overlooked analysis by Hestenes [Hes75] for the case where $X = \Re^n$, and in the first author's recent textbook [Ber99] for the case where X is a closed convex set (see the discussion in Section 2). They are strengthened and further generalized in Section 2 for the case where X is a closed but not necessarily convex set. In particular, we show the existence of Fritz-John multipliers that satisfy some additional sensitivity-like conditions. These conditions motivate the introduction of two new types of Lagrange multipliers, called *informative* and *strong*. We show that informative and strong Lagrange multipliers exist when the tangent cone is convex and the set of Lagrange multipliers is nonempty.

In Section 3, we introduce the notions of pseudonormality and quasinormality, and we discuss their connection with classical results relating constraint qualifications and the admittance of Lagrange multipliers. Quasinormality serves almost the same purpose as pseudonormality when X is regular, but fails to provide the desired theoretical unification when X is not regular (compare with Fig. 6). For this reason, it appears that pseudonormality is a theoretically more interesting notion than quasinormality. In addition, in contrast with quasinormality, pseudonormality admits an insightful geometrical interpretation. In Section 3, we also introduce a new and natural extension of the Mangasarian-Fromovitz constraint qualification, which applies to the case where $X \neq \Re^n$ and implies pseudonormality.

In Section 4, we make the connection between pseudonormality, quasinormality, 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. We note that exact penalty functions have traditionally been viewed as a computational device and they have not been earlier integrated within the theory of constraint qualifications in the manner described here. Let us also note that exact penalty functions are related to the notion of *calmness*, introduced and suggested as a constraint qualification by Clarke [Cla76], [Cla83]. However, there are some important differences between the notions of calmness and admittance of an exact penalty. In particular, calmness is a property of the problem (1.1) and depends on the cost function f, while admittance of an exact penalty is a property of the constraint set and is independent of the cost function. More importantly for the purposes of this paper, calmness is not useful as a unifying theoretical vehicle because it does not relate well with other major constraint qualifications. For example CQ1, one of the most common constraint qualifications, does not imply calmness of problem (1.1), as is indicated by Example 11 of Section 7, and reversely, calmness of the problem does not imply CQ1.

In Section 5, we discuss some special results that facilitate proofs of admittance of Lagrange multipliers and of an exact penalty. In Section 6, we generalize some of our analysis to the case of a convex programming problem and we provide a geometric interpretation of pseudonormality. Finally, in Section 7 we provide examples and counterexamples that clarify the interrelations between the different characterizations that we have introduced.

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 not sufficient to derive the admittance of Lagrange multipliers under some of the standard constraint qualifications, such as when $X = \Re^n$ and the constraint functions h_i and g_j are linear (cf. CQ3). 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 Hestenes [Hes75] for the case $X = \Re^n$, and Bertsekas [Ber99], Prop. 3.3.11, for the case where X is a closed convex set). The following proposition extends these results by allowing the set X to be nonconvex, and by also showing that the Fritz John multipliers can be selected to have some special sensitivity-like properties [see condition (iv) below].

8

Proposition 1: Let x^* be a local minimum of problem (1.1)-(1.2). Then there exist scalars $\mu_0^*, \lambda_1^*, \ldots, \lambda_m^*$, 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 index set $I \cup J$ is nonempty where

$$I = \{i \mid \lambda_i^* \neq 0\}, \qquad J = \{j \neq 0 \mid \mu_j^* > 0\},$$

there exists a sequence $\{x^k\} \subset X$ that converges to x^* and is such that for all k,

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

$$|h_i(x^k)| = o\big(w(x^k)\big), \quad \forall\, i \notin I, \qquad g_j^+(x^k) = o\big(w(x^k)\big), \quad \forall\, j \notin J, \qquad (2.2)$$

where

$$w(x) = \min\left\{\min_{i \in I} |h_i(x)|, \min_{j \in J} g_j(x)\right\}. \qquad (2.3)$$

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

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

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 $\bar{x}$ of $\{x^k\}$ is feasible, i.e., $\bar{x} \in C$. 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

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

9

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 $-\nabla F^k(x^k) \in T_X(x^k)^*$, which is written as

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

where

$$\xi_i^k = kh_i(x^k), \qquad \zeta_j^k = kg_j^+(x^k). \qquad (2.6)$$

Denote,

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

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

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

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

Since by construction we have

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

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 defining property of the normal cone $N_X(x^*)$ $[x^k \to x^*, \{x^k\} \subset X$, $z^k \to z^*$, and $z^k \in T_X(x^k)^*$ for all k, imply that $z^* \in N_X(x^*)]$, we see that μ_0^*, λ_i^*, and μ_j^* must satisfy condition (i). From Eqs. (2.6) and (2.8), μ_0^* and μ_j^* must satisfy condition (ii), and from Eq. (2.10), μ_0^*, λ_i^*, and μ_j^* must satisfy condition (iii). Finally, to show that condition (iv) is satisfied, assume that $I \cup J$ is nonempty, 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 $|h_i(x^k)| = o(w(x^k))$ for all $i \notin I$, and $g_j^+(x^k) = o(w(x^k))$ for all $j \notin J$ are equivalent to

$$|\lambda_i^k| = o\left(\min\left\{\min_{i \in I}|\lambda_i^k|, \min_{j \in J}\mu_j^k\right\}\right), \qquad \forall\, i \notin I,$$

and

$$\mu_j^k = o\left(\min\left\{\min_{i \in I}|\lambda_i^k|, \min_{j \in J}\mu_j^k\right\}\right), \qquad \forall\, j \notin J,$$

respectively, so they hold for $k \in \mathcal{K}$. This proves condition (iv).   **Q.E.D.**

Note that if X is regular at x^*, i.e., $N_X(x^*) = T_X(x^*)^*$, condition (i) of Prop. 1 becomes

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

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 CS 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 $g_j(x^*) = 0$.] For ease of reference, we refer to condition (iv) as the *complementary violation condition* (CV for short).† This condition will turn out to be of crucial significance in the next section.

To place Prop. 1 in perspective, we note that its line of proof, based on the quadratic penalty function, originated with McShane [McS73]. Hestenes [Hes75] observed that McShane's proof can be used to strengthen the CS condition to assert the existence of a sequence $\{x^k\}$ such that

$$\lambda_i^* h_i(x^k) > 0, \quad \forall\, i \in I, \qquad \mu_j^* g_j(x^k) > 0, \quad \forall\, j \in J, \tag{2.10}$$

which is slightly weaker than CV as defined here [there is no requirement that x^k, simultaneously with violation of the constraints with nonzero multipliers, satisfies $f(x^k) < f(x^*)$ and Eq. (2.2)].

---

† This term is in analogy with "complementary slackness," which is the condition that for all j, $\mu_j^* > 0$ implies $g_j(x^*) = 0$. Thus "complementary violation" reflects the condition that for all j, $\mu_j^* > 0$ implies $g_j(x) > 0$ for some x arbitrarily close to x^* (and simultaneously for all j with $\mu_j^* > 0$).

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 given in the present paper for the first time.

To illustrate the use of the generalized Fritz John conditions of Prop. 1 and the CV condition in particular, consider the following example.

Example 1

Suppose that we convert a problem with a single equality constraint, $\min_{h(x)=0} f(x)$, to the inequality constrained problem

$$\text{minimize} \ \ f(x)$$

$$\text{subject to} \ \ h(x) \leq 0, \quad -h(x) \leq 0.$$

The Fritz John conditions assert the existence of nonnegative $\mu_0^*, \lambda^+, \lambda^-$, not all zero, such that

$$\mu_0^* \nabla f(x^*) + \lambda^+ \nabla h(x^*) - \lambda^- \nabla h(x^*) = 0. \tag{2.11}$$

The candidate multipliers that satisfy the above condition as well as the CS condition $\lambda^+ h(x^*) = \lambda^- h(x^*) = 0$, include those of the form $\mu_0^* = 0$ and $\lambda^+ = \lambda^- > 0$, which provide no relevant information about the problem. However, these multipliers fail the stronger CV condition of Prop. 1, showing that if $\mu_0^* = 0$, we must have either $\lambda^+ \neq 0$ and $\lambda^- = 0$ or $\lambda^+ = 0$ and $\lambda^- \neq 0$. Assuming $\nabla h(x^*) \neq 0$, this violates Eq. (2.11), so it follows that $\mu_0^* > 0$. Thus, by dividing Eq. (2.11) by μ_0^*, we recover the familiar first order condition $\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$ with $\lambda^* = (\lambda^+ - \lambda^-)/\mu_0^*$, under the regularity assumption $\nabla h(x^*) \neq 0$. Note that this deduction would not have been possible without the CV condition.

If we can take $\mu_0^* = 1$ in Prop. 1 for all smooth f for which x^* is a local minimum, and X is regular at x^*, then the constraint set C admits Lagrange multipliers of a special type, which satisfy the stronger CV condition in place of the CS condition. The salient feature of such multipliers is the information they embody regarding constraint violation with corresponding cost reduction. This is consistent with the classical sensitivity interpretation of a Lagrange multiplier as the rate of reduction in cost as the corresponding constraint is violated. Here we are not making enough assumptions for this stronger type of sensitivity interpretation to be valid. Yet it is remarkable that with hardly any assumptions (other than their existence), Lagrange multipliers of the type obtained through Prop. 1, provide a significant amount of sensitivity information: they indicate the index set $I \cup J$ of constraints that if violated, a cost reduction can be effected [the remaining constraints, whose indices do not belong to $I \cup J$, may also be violated, but the degree of their

violation is arbitrarily small relative to the other constraints as per Eqs. (2.2) and (2.3)]. In view of this interpretation, we refer to a Lagrange multiplier vector (λ^*, μ^*) that satisfies, in addition to Eqs. (1.4)-(1.6), the CV condition [condition (iv) of Prop. 1] as being *informative*.

An informative Lagrange multiplier vector is useful, among other things, if one is interested in identifying redundant constraints. Given such a vector, one may simply discard the constraints whose multipliers are 0 and check to see whether x^* is still a local minimum. While there is no general guarantee that this will be true, in many cases it will be; for example, in the special case where f and X are convex, the g_j are convex, and the h_i are linear, x^* is guaranteed to be a global minimum, even after the constraints whose multipliers are 0 are discarded.

Now if we are interested in discarding constraints whose multipliers are 0, we are also motivated to find Lagrange multiplier vectors that have a minimal number of nonzero components (a minimal support). We call such Lagrange multiplier vectors *minimal*, and we define them as having support $I \cup J$ that does not strictly contain the support of any other Lagrange multiplier vector. Minimal Lagrange multipliers are not necessarily informative. For example, think of the case where some of the constraints are duplicates of others. Then in a minimal Lagrange multiplier vector, at most one of each set of duplicate constraints can have a nonzero multiplier, while in an informative Lagrange multiplier vector, either all or none of these duplicate constraints will have a nonzero multiplier. Nonetheless, minimal Lagrange multipliers turn out to be informative after the constraints corresponding to zero multipliers are neglected, as can be inferred by the subsequent Prop. 2. In particular, let us say that a Lagrange multiplier (λ^*, μ^*) is *strong* if in addition to Eqs. (1.4)-(1.6), it satisfies the condition

(iv′) If the set $I \cup J$ is nonempty, where $I = \{i \mid \lambda_i^* \neq 0\}$ and $J = \{j \neq 0 \mid \mu_j^* > 0\}$, then given any neighborhood B of x^*, there exists a sequence $\{x^k\} \subset X$ that converges to x^* and is such that for all k,

$$f(x^k) < f(x^*), \qquad \lambda_i^* h_i(x^k) > 0, \ \ \forall \ i \in I, \qquad g_j(x^k) > 0, \ \ \forall \ j \in J. \qquad (2.12)$$

This condition resembles the CV condition, but is weaker in that it makes no provision for negligibly small violation of the constraints corresponding to zero multipliers, as per Eqs. (2.2) and (2.3). As a result, informative Lagrange multipliers are also strong, but not reversely.

The following proposition, illustrated in Fig. 2, clarifies the relationships between different types of Lagrange multipliers.

Proposition 2: Let x^* be a local minimum of problem (1.1)-(1.2). Assume that the tangent cone $T_X(x^*)$ is convex and that the set of Lagrange multipliers is nonempty. Then:
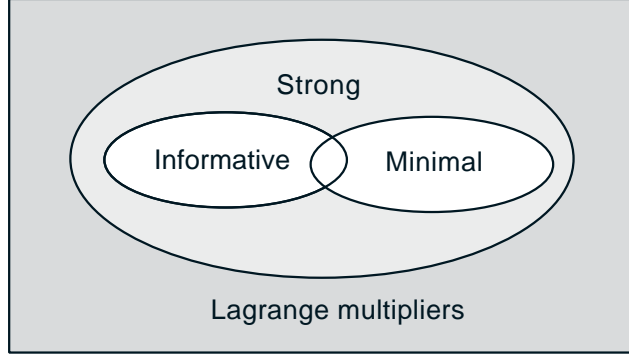
13

Figure 2. Relations of different types of Lagrange multipliers, assuming that the tangent cone $T_X(x^*)$ is convex (which is true in particular if X is regular at x^*).

(a) The set of informative Lagrange multiplier vectors is nonempty, and in fact the Lagrange multiplier vector that has minimum norm is informative.

(b) Each minimal Lagrange multiplier vector is strong.

Proof: (a) We summarize the essence of the proof argument in the following lemma (a related but different line of proof of this lemma is given in [BNO01]).

Lemma 1: Let N be a closed convex cone in $\Re^n$, and let $a_0, a_1, \ldots, a_r$ be given vectors in $\Re^n$. Suppose that the closed and convex set $M \subset \Re^r$ given by

$$M = \left\{ \mu \geq 0 \; \middle| \; -\left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right) \in N \right\},$$

is nonempty. Then there exists a sequence $\{d^k\} \subset N^*$ such that

$$a_0' d^k \to -\|\mu^*\|^2, \tag{2.13}$$

$$(a_j' d^k)^+ \to \mu_j^*, \qquad j = 1, \ldots, r, \tag{2.14}$$

where μ^* is the vector of minimum norm in M.

Proof: For any $\gamma \geq 0$, consider the function

$$L_\gamma(d, \mu) = \left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right)' d + \gamma \|d\| - \frac{1}{2}\|\mu\|^2.$$

Our proof will revolve around saddle point properties of the convex/concave function L_0, but to derive these properties, we will work with its γ-perturbed and *coercive* version L_γ for $\gamma > 0$, and then take the limit as $\gamma \to 0$. With this in mind, we first establish that if $\gamma > 0$, $L_\gamma(d, \mu)$ has a saddle point over $d \in N^*$ and $\mu \geq 0$.

Indeed, for any fixed $\overline{\mu} \geq 0$, $L_\gamma(\cdot, \overline{\mu})$ is convex over $d \in N^*$ and if $\overline{\mu} \in M$, we have $\left(a_0 + \sum_{j=1}^{r} \overline{\mu}_j a_j \right)' d \geq 0$ for all $d \in N^*$, so that

$$L_\gamma(d, \overline{\mu}) \geq \gamma \|d\| - (1/2)\|\overline{\mu}\|^2, \qquad \forall \, d \in N^*.$$

Hence $L_\gamma(\cdot, \overline{\mu})$ is coercive over N^*. Also, for any fixed $\overline{d} \in N^*$, $L_\gamma(\overline{d}, \cdot)$ is concave and $-L_\gamma(\overline{d}, \cdot)$ is coercive over $\mu \in \Re^r$. It follows from a theorem given by Hiriart-Urruty and Lemarechal [HiL93], p. 334, that for each $\gamma > 0$, there exists a saddle point (d^γ, μ^γ) of L_γ over $d \in N^*$ and $\mu \geq 0$, satisfying

$$L_\gamma(d^\gamma, \mu^\gamma) = \max_{\mu \geq 0} L_\gamma(d^\gamma, \mu) = \min_{d \in N^*} L_\gamma(d, \mu^\gamma) = \max_{\mu \geq 0} \inf_{d \in N^*} L_\gamma(d, \mu). \tag{2.15}$$

We will now calculate some of the expressions in the above equations.

We have from Eq. (2.15)

$$L_\gamma(d^\gamma, \mu^\gamma) = \max_{\mu \geq 0} L_\gamma(d^\gamma, \mu)$$

$$= a_0' d^\gamma + \gamma \|d^\gamma\| + \max_{\mu \geq 0} \left\{ \left(\sum_{j=1}^{r} \mu_j a_j \right)' d^\gamma - \frac{1}{2}\|\mu\|^2 \right\}.$$

The maximum in the right-hand side above is attained when μ_j is equal to $(a_j' d^\gamma)^+$ for all j [to maximize $\mu_j a_j' d^\gamma - (1/2)\mu_j^2$ subject to the constraint $\mu_j \geq 0$, we calculate the unconstrained maximum, which is $a_j' d^\gamma$, and if it is negative we set it to 0, so that the maximum subject to $\mu_j \geq 0$ is attained for $\mu_j = (a_j' d^\gamma)^+$]. Thus, we have

$$L_\gamma(d^\gamma, \mu^\gamma) = a_0' d^\gamma + \gamma \|d^\gamma\| + \frac{1}{2} \sum_{j=1}^{r} \left((a_j' d^\gamma)^+ \right)^2, \tag{2.16}$$

and

$$\mu^\gamma = \left((a_1' d^\gamma)^+, \dots, (a_r' d^\gamma)^+ \right)'. \tag{2.17}$$

We also have from Eq. (2.15)

$$L_\gamma(d^\gamma, \mu^\gamma) = q_\gamma(\mu^\gamma) - \frac{1}{2}\|\mu^\gamma\|^2 = \max_{\mu \geq 0} \left\{ q_\gamma(\mu) - \frac{1}{2}\|\mu\|^2 \right\}, \tag{2.18}$$

where

$$q_\gamma(\mu) = \inf_{d \in N^*} \left\{ \left(a_0 + \sum_{j=1}^{r} \mu_j a_j \right)' d + \gamma \|d\| \right\}.$$

To calculate $q_\gamma(\mu)$, we let

$$b = - \left(a_0 + \sum_{j=1}^{r} \mu_j a_j \right),$$

15

and we use the transformation $d = \alpha\xi$, where $\alpha \geq 0$ and $\|\xi\| = 1$, to write

$$q_\gamma(\mu) = \inf_{\substack{\alpha \geq 0 \\ \|\xi\| \leq 1,\, \xi \in N^*}} \left\{\alpha(\gamma - b'\xi)\right\} = \begin{cases} 0 & \text{if } \max_{\|\xi\| \leq 1,\, \xi \in N^*} b'\xi \leq \gamma, \\ -\infty & \text{otherwise.} \end{cases} \qquad (2.19)$$

We will show that

$$\max_{\|\xi\| \leq 1,\, \xi \in N^*} b'\xi \leq \gamma \qquad \text{if and only if} \qquad b \in N + S(0, \gamma), \qquad (2.20)$$

where $S(0, \gamma)$ is the closed sphere of radius γ that is centered at the origin. Indeed, if $b \in N + S(0, \gamma)$, then $b = \hat{b} + \overline{b}$ with $\hat{b} \in N$ and $\|\overline{b}\| \leq \gamma$, and it follows that for all $\xi \in N^*$ with $\|\xi\| \leq 1$, we have $\hat{b}'\xi \leq 0$ and $\overline{b}'\xi \leq \gamma$, so that

$$b'\xi = \hat{b}'\xi + \overline{b}'\xi \leq \gamma,$$

from which we obtain

$$\max_{\|\xi\| \leq 1,\, \xi \in N^*} b'\xi \leq \gamma.$$

Conversely, assume that $b'\xi \leq \gamma$ for all $\xi \in N^*$ with $\|\xi\| \leq 1$. If $b \in N$, then clearly $b \in N + S(0, \gamma)$. If $b \notin N$, let $\hat{b}$ be the projection of b onto N and let $\overline{b} = b - \hat{b}$. Because N is a convex cone, the nonzero vector $\overline{b}$ belongs to N^* and is orthogonal to $\hat{b}$. Since the vector $\xi = \overline{b}/\|\overline{b}\|$ belongs to N^* and satisfies $\|\xi\| \leq 1$, we have $\gamma \geq b'\xi$ or equivalently $\gamma \geq (\hat{b} + \overline{b})'(\overline{b}/\|\overline{b}\|) = \|\overline{b}\|$. Hence, $b = \hat{b} + \overline{b}$ with $\hat{b} \in N$ and $\|\overline{b}\| \leq \gamma$, implying that $b \in N + S(0, \gamma)$, and completing the proof of Eq. (2.20).

We have thus shown [cf. Eqs. (2.19) and Eq. (2.20)] that

$$q_\gamma(\mu) = \begin{cases} 0 & \text{if } -\left(a_0 + \sum_{j=1}^r \mu_j a_j\right) \in N + S(0, \gamma), \\ -\infty & \text{otherwise.} \end{cases} \qquad (2.21)$$

Combining this equation with Eq. (2.18), we see that μ^γ is the vector of minimum norm on the set

$$M_\gamma = \left\{ \mu \geq 0 \;\middle|\; -\left(a_0 + \sum_{j=1}^r \mu_j a_j\right) \in N + S(0, \gamma) \right\}.$$

Furthermore, from Eqs. (2.18) and (2.21), we have

$$L_\gamma(d^\gamma, \mu^\gamma) = -\frac{1}{2}\|\mu^\gamma\|^2,$$

which together with Eqs. (2.16) and (2.17), yields

$$a_0'd^\gamma + \gamma\|d^\gamma\| = -\|\mu^\gamma\|^2. \qquad (2.22)$$

We now take limit in the above equation as $\gamma \to 0$. We claim that $\mu^\gamma \to \mu^*$. Indeed, since $\mu^* \in M_\gamma$, we have $\|\mu^\gamma\| \leq \|\mu^*\|$, so that $\{\mu^\gamma \mid \gamma > 0\}$ is bounded. Let $\overline{\mu}$ be a limit point of μ^γ, and note that $\overline{\mu} \geq 0$ and $\|\overline{\mu}\| \leq \|\mu^*\|$. We have

$$-\sum_{j=1}^r \mu_j^\gamma a_j = a_0 + \nu^\gamma + s^\gamma,$$

16

for some vectors $\nu^\gamma \in N$ and $s^\gamma \in S(0, \gamma)$, so by taking limit as $\gamma \to 0$ along the relevant subsequence, it follows that ν^γ converges to some $\overline{\nu} \in N$, and we have

$$-\sum_{j=1}^{r} \overline{\mu}_j a_j = a_0 + \overline{\nu}.$$

It follows that $\overline{\mu} \in M$, and since $\|\overline{\mu}\| \le \|\mu^*\|$, we obtain $\overline{\mu} = \mu^*$. The preceding argument has shown that every limit point of μ^γ is equal to μ^*, so μ^γ converges to μ^* as $\gamma \to 0$. Thus, Eq. (2.22) yields

$$\limsup_{\gamma \to 0} a_0' d^\gamma \le -\|\mu^*\|^2. \tag{2.23}$$

Consider now the function

$$L_0(d, \mu) = \left(a_0 + \sum_{j=1}^{r} \mu_j a_j \right)' d - \frac{1}{2}\|\mu\|^2.$$

We have

$$a_0' d^\gamma + \frac{1}{2}\sum_{j=1}^{r}\left((a_j' d^\gamma)+\right)^2 = \sup_{\mu \ge 0} L_0(d^\gamma, \mu)$$
$$\ge \sup_{\mu \ge 0} \inf_{d \in N^*} L_0(d, \mu)$$
$$\ge \inf_{d \in N^*} L_0(d, \mu^*).$$

It can be seen that

$$\inf_{d \in N^*} L_0(d, \mu) = \begin{cases} -\frac{1}{2}\|\mu\|^2 & \text{if } -\left(a_0 + \sum_{j=1}^{r}\mu_j a_j\right) \in N, \\ -\infty & \text{otherwise.} \end{cases}$$

Combining the last two equations, we have

$$a_0' d^\gamma + \frac{1}{2}\sum_{j=1}^{r}\left((a_j' d^\gamma)+\right)^2 \ge -\frac{1}{2}\|\mu^*\|^2,$$

and since $(a_j' d^\gamma)+ = \mu_j^\gamma$ [cf. Eq. (2.17)],

$$a_0' d^\gamma \ge -\frac{1}{2}\|\mu^*\|^2 - \frac{1}{2}\|\mu^\gamma\|^2.$$

Taking the limit as $\gamma \to \infty$, we obtain

$$\liminf_{\gamma \to 0} a_0' d^\gamma \ge -\|\mu^*\|^2,$$

which together with Eq. (2.23), shows that $a_0' d^\gamma \to -\|\mu^*\|^2$. Since we have also shown that $(a_j' d^\gamma)+ = \mu_j^\gamma \to \mu_j^*$, the proof is complete.   **Q.E.D.**

We now return to the proof of Prop. 2(a). For simplicity we assume that all the constraints are inequalities that are active at x^* (equality constraints can be handled by conversion to two

inequalities, and inactive inequality constraints are inconsequential in the subsequent analysis). We will use Lemma 1 with the following identifications:

$$N = T_X(x^*)^*, \qquad a_0 = \nabla f(x^*), \qquad a_j = \nabla g_j(x^*), \quad j = 1, \ldots, r,$$

$$M = \text{ set of Lagrange multipliers,}$$

$$\mu^* = \text{ Lagrange multiplier of minimum norm.}$$

If $\mu^* = 0$, then μ^* is an informative Lagrange multiplier and we are done. If $\mu^* \neq 0$, by Lemma 1, for any $\epsilon > 0$, there exists a $\overline{d} \in N^* = T_X(x^*)$ such that

$$a_0'\overline{d} < 0, \tag{2.24}$$

$$a_j'\overline{d} > 0, \quad \forall\, j \in J^*, \qquad a_j'\overline{d} \le \epsilon \min_{l \in J^*} a_l'\overline{d}, \quad \forall\, j \notin J^*, \tag{2.25}$$

where $J^* = \{j \mid \mu_j^* > 0\}$. By suitably scaling the vector $\overline{d}$, we can assume that $\|\overline{d}\| = 1$. Let $\{x^k\} \subset X$ be such that $x^k \neq x^*$ for all k and

$$x^k \to x^*, \qquad \frac{x^k - x^*}{\|x^k - x^*\|} \to \overline{d}.$$

Using Taylor's theorem for the cost function f, we have for some vector sequence ξ^k converging to 0

$$\begin{aligned}
f(x^k) - f(x^*) &= \nabla f(x^*)'(x^k - x^*) + o(\|x^k - x^*\|) \\
&= \nabla f(x^*)' \left(\overline{d} + \xi^k\right) \|x^k - x^*\| + o(\|x^k - x^*\|) \\
&= \|x^k - x^*\| \left(\nabla f(x^*)'\overline{d} + \nabla f(x^*)'\xi^k + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} \right).
\end{aligned}$$

From Eq. (2.24), we have $\nabla f(x^*)'\overline{d} < 0$, so we obtain $f(x^k) < f(x^*)$ for k sufficiently large. Using also Taylor's theorem for the constraint functions g_j, we have for some vector sequence ξ^k converging to 0,

$$\begin{aligned}
g_j(x^k) - g_j(x^*) &= \nabla g_j(x^*)'(x^k - x^*) + o(\|x^k - x^*\|) \\
&= \nabla g_j(x^*)' \left(\overline{d} + \xi^k\right) \|x^k - x^*\| + o(\|x^k - x^*\|) \\
&= \|x^k - x^*\| \left(\nabla g_j(x^*)'\overline{d} + \nabla g_j(x^*)'\xi^k + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} \right).
\end{aligned}$$

This, combined with Eq. (2.25), shows that for k sufficiently large, $g_j(x^k)$ is bounded from below by a constant times $\|x^k - x^*\|$ for all j such that $\mu_j^* > 0$ [and hence $g_j(x^*) = 0$], and satisfies $g_j(x^k) \le o(\|x^k - x^*\|)$ for all j such that $\mu_j^* = 0$ [and hence $g_j(x^*) \le 0$]. Thus, the sequence $\{x^k\}$ can be used to establish the CV condition for μ^*, and it follows that μ^* is an informative Lagrange multiplier.

(b) We summarize the essence of the proof argument of this part in the following lemma.

18

Lemma 2: Let N be a closed convex cone in $\Re^n$, let $a_0, a_1, \ldots, a_r$ be given vectors in $\Re^n$. Suppose that the closed and convex set $M \subset \Re^r$ given by

$$M = \left\{ \mu \geq 0 \,\Big|\, -\left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right) \in N \right\},$$

is nonempty. Among index subsets $J \subset \{1, \ldots, r\}$ such that for some $\mu \in M$ we have $J = \{j \mid \mu_j > 0\}$, let $\overline{J} \subset \{1, \ldots, r\}$ have a minimal number of elements. Then if $\overline{J}$ is nonempty, there exists a vector $\overline{d} \in N^*$ such that

$$a_0' \overline{d} < 0, \qquad a_j' \overline{d} > 0, \qquad \text{for all } j \in \overline{J}. \tag{2.26}$$

Proof: We apply Lemma 1 with the vectors $a_1, \ldots, a_r$ replaced by the vectors a_j, $j \in \overline{J}$. The subset of M given by

$$\overline{M} = \left\{ \mu \geq 0 \,\Big|\, -\left(a_0 + \sum_{j \in \overline{J}} \mu_j a_j\right) \in N, \ \mu_j = 0, \ \forall\, j \notin \overline{J} \right\}$$

is nonempty by assumption. Let $\overline{\mu}$ be the vector of minimum norm on $\overline{M}$. Since $\overline{J}$ has a minimal number of indices, we must have $\overline{\mu}_j > 0$ for all $j \in \overline{J}$. If $\overline{J}$ is nonempty, Lemma 1 implies that there exists a $\overline{d} \in N^*$ such that Eq. (2.26) holds. **Q.E.D.**

Given Lemma 2, the proof of Prop. 2(b) is very similar to the corresponding part of the proof of Prop. 2(a). **Q.E.D.**

Sensitivity and the Lagrange Multiplier of Minimum Norm

Let us first introduce an interesting variation of Lemma 1:

Lemma 3: Let N be a closed convex cone in $\Re^n$, and let $a_0, \ldots, a_r$ be given vectors in $\Re^n$. Suppose that the closed and convex set $M \subset \Re^r$ given by

$$M = \left\{ \mu \geq 0 \,\Big|\, -\left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right) \in N \right\},$$

is nonempty, and let μ^* be the vector of minimum norm on M. Then

$$-\|\mu^*\|^2 \leq a_0' d + \frac{1}{2} \sum_{j=1}^{r} \left((a_j' d)^+\right)^2, \qquad \forall\, d \in N^*.$$

Furthermore, if $\overline{d}$ is an optimal solution of the problem

$$\text{minimize} \quad a_0'd + \frac{1}{2}\sum_{j=1}^{r}\left((a_j'd)^+\right)^2 \tag{2.27}$$
$$\text{subject to} \quad d \in N^*,$$

we have

$$a_0'\overline{d} = -\|\mu^*\|^2, \qquad (a_j'\overline{d})^+ = \mu_j^*, \quad j = 1,\ldots,r. \tag{2.28}$$

Proof: From the proof of Lemma 1, we have for all $\gamma > 0$

$$
\begin{aligned}
-\frac{1}{2}\|\mu^*\|^2 &= \sup_{\mu \geq 0} \inf_{d \in N^*} L_0(d,\mu) \\
&\leq \inf_{d \in N^*} \sup_{\mu \geq 0} L_0(d,\mu) \\
&= \inf_{d \in N^*} \left\{ a_0'd + \frac{1}{2}\sum_{j=1}^{r}\left((a_j'd)^+\right)^2 \right\}.
\end{aligned} \tag{2.29}
$$

If $\overline{d}$ is an optimal solution of problem (2.27), we obtain

$$
\begin{aligned}
\inf_{d \in N^*} \left\{ a_0'd + \frac{1}{2}\sum_{j=1}^{r}\left((a_j'd)^+\right)^2 \right\} &= a_0'\overline{d} + \frac{1}{2}\sum_{j=1}^{r}\left((a_j'\overline{d})^+\right)^2 \\
&\leq a_0'd^\gamma + \frac{1}{2}\sum_{j=1}^{r}\left((a_j'd^\gamma)^+\right)^2.
\end{aligned}
$$

Since (according to the proof of Lemma 1) $a_0'd^\gamma \to -\|\mu^*\|^2$ and $(a_j'd^\gamma)^+ \to \mu_j^*$ as $\gamma \to 0$, by taking the limit above as $\gamma \to 0$, we see that equality holds throughout in the two above inequalities. Thus $(\overline{d},\mu^*)$ is a saddle point of the function $L_0(d,\mu)$ over $d \in N^*$ and $\mu \geq 0$. It follows that μ^* maximizes $L_0(\overline{d},\mu)$ over $\mu \geq 0$, so that $\mu_j^* = (a_j'\overline{d})^+$ for all j and $-\|\mu^*\|^2 = \alpha_0'\overline{d}$.   **Q.E.D.**

The difference between Lemmas 1 and 3 is that in Lemma 3, there is the extra assumption that problem (2.27) has an optimal solution (otherwise the lemma is vacuous). It can be shown that, assuming the set M is nonempty, *problem (2.27) is guaranteed to have at least one solution when N^* is a polyhedral cone*. To see this, note that problem (2.27) can be written as

$$\text{minimize} \quad a_0'd + \frac{1}{2}\sum_{j=1}^{r} z_j^2$$
$$\text{subject to} \quad d \in N^*, \quad 0 \leq z_j, \; a_j'd \leq z_j, \quad j = 1,\ldots,r,$$

where the z_j are auxiliary variables. Thus, if N^* is polyhedral, then problem (2.27) is a quadratic program with a cost function that is bounded below by Eq. (2.29), and hence it has an optimal solution (see Bonnans and Shapiro [BeS00], Th. 3.128). Thus, when N^* is polyhedral, Lemma 3

applies. An important context where this is relevant is when $X = \Re^n$ in which case $N_X(x^*)^* = T_X(x^*) = \Re^n$, or more generally when X is polyhedral, in which case $T_X(x^*)$ is polyhedral. Another condition that guarantees the existence of an optimal solution of problem (2.27) is that there exists a vector $\overline{\mu}$ in the set

$$M = \left\{ \mu \geq 0 \ \middle| \ -\left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right) \in N \right\}$$

such that

$$-\left(a_0 + \sum_{j=1}^{r} \overline{\mu}_j a_j\right) \in \mathrm{ri}(N),$$

where $\mathrm{ri}(N)$ denotes the relative interior of N. The relevant analysis, which is due to Xin Chen (private communication), is given in [BNO01].

When problem (2.27) can be guaranteed to have an optimal solution and Lemma 3 applies, the line of proof of Prop. 2(a) can be used to show that if the Lagrange multiplier that has minimum norm, denoted by (λ^*, μ^*), is nonzero, there exists a sequence $\{x^k\} \subset X$ and a positive constant c such that

$$f(x^k) = f(x^*) - \sum_{i=1}^{m} \lambda_i^* h_i(x^k) - \sum_{j=1}^{r} \mu_j^* g_j(x^k) + o(\|x^k - x^*\|), \tag{2.30}$$

$$h_i(x^k) = c\lambda_i^* \|x^k - x^*\| + o(\|x^k - x^*\|), \qquad i = 1, \ldots, m, \tag{2.31}$$

$$g_j(x^k) = c\mu_j^* \|x^k - x^*\| + o(\|x^k - x^*\|), \qquad \text{if } \mu_j^* > 0, \tag{2.32}$$

$$g_j(x^k) \leq o(\|x^k - x^*\|), \qquad \text{if } \mu_j^* = 0. \tag{2.33}$$

These equations suggest that the minimum-norm Lagrange multiplier has a sensitivity interpretation. In particular, the sequence $\{x^k\}$ above corresponds to the vector $\overline{d} \in T_X(x^*)$ of Eq. (2.28), which solves problem (2.27). From this it can be seen that a positive multiple of $\overline{d}$ solves the problem

$$\text{minimize} \ \ \nabla f(x^*)'d$$

$$\text{subject to} \ \ \sum_{i=1}^{m} \left(\nabla h_i(x^*)'d\right)^2 + \sum_{j \in A(x^*)} \left(\nabla g_j(x^*)'d)^+\right)^2 = \beta, \qquad d \in T_X(x^*),$$

for any given positive scalar β. Thus, $\overline{d}$ is the tangent direction that maximizes the cost function improvement (calculated up to first order) for a given value of the norm of the constraint violation (calculated up to first order). From Eq. (2.30), this first order cost improvement is equal to

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

Thus, the multipliers λ_i^* and μ_j^* express the rate of improvement per unit constraint violation, along the maximum improvement (or steepest descent) direction $\overline{d}$. This is consistent with the traditional sensitivity interpretation of Lagrange multipliers.

An Alternative Definition of Lagrange Multipliers

Finally, let us make the connection with Rockafellar's treatment of Lagrange multipliers from [Roc93] and [RoW98]. Consider vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ that satisfy the conditions

$$-\left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*), \qquad (2.34)$$

$$\mu_j^* \geq 0, \quad \forall \, j = 1, \ldots, r, \qquad \mu_j^* = 0, \quad \forall \, j \notin A(x^*). \qquad (2.35)$$

Such vectors are called "Lagrange multipliers" by Rockafellar, but in this paper we will refer to them as *R-multipliers*, to distinguish them from Lagrange multipliers as we have defined them [cf. Eqs. (1.4)-(1.6)].

When X is regular at x^*, Rockafellar's definition and our definition coincide. In general, however, the set of Lagrange multipliers is a (possibly strict) subset of the set of R-multipliers, since $T_X(x^*)^* \subset N_X(x^*)$ with inequality holding when X is not regular at x^*. Also, the existence of R-multipliers does not guarantee the existence of Lagrange multipliers. Furthermore, even if Lagrange multipliers exist, none of them may be informative or strong, unless the tangent cone is convex (cf. Prop. 2 and Example 2 given below).

Note that multipliers satisfying the enhanced Fritz John conditions of Prop. 1 with $\mu_0^* = 1$ are R-multipliers, and they still have the extra sensitivity-like property embodied in the CV condition. Furthermore, Lemma 1 can be used to show that assuming $N_X(x^*)$ is convex, if the set of R-multipliers is nonempty, it contains an R-multiplier with the sensitivity-like property of the CV condition. However, if X is not regular at x^*, an R-multiplier may not render the Lagrangian function stationary. The following is an illustrative example.
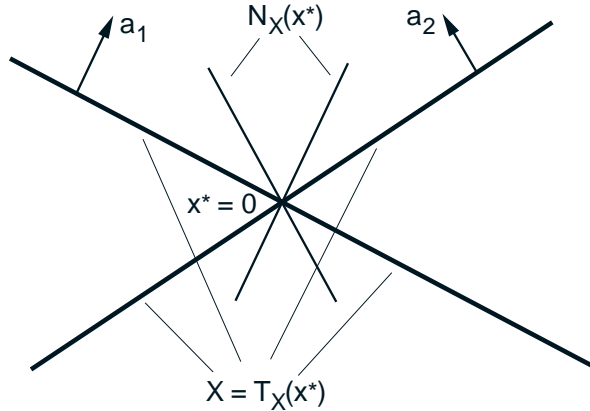

Figure 3. Constraints of Example 2. We have $T_X(x^*) = X = \{x \mid (a_1'x)(a_2'x) = 0\}$ and $N_X(x^*)$ is the nonconvex set consisting of the two rays of vectors that are colinear to either a_1 or a_2.

22

Example 2

In this 2-dimensional example there are two linear constraints $a_1' x \leq 0$ and $a_2' x \leq 0$ with the vectors a_1 and a_2 linearly independent. The set X is the (nonconvex) cone

$$X = \{x \mid (a_1' x)(a_2' x) = 0\}.$$

Consider the vector $x^* = (0,0)$. Here $T_X(x^*) = X$ and $T_X(x^*)^* = \{0\}$. However, it can be seen that $N_X(x^*)$ consists of the two rays of vectors that are colinear to either a_1 or a_2:

$$N_X(x^*) = \{\gamma a_1 \mid \gamma \in \Re\} \cup \{\gamma a_2 \mid \gamma \in \Re\}$$

(see Fig. 3).

Because $N_X(x^*) \neq T_X(x^*)^*$, X is not regular at x^*. Furthermore, both $T_X(x^*)$ and $N_X(x^*)$ are not convex. For any f for which x^* is a local minimum, there exists a unique Lagrange multiplier (μ_1^*, μ_2^*) satisfying Eqs. (1.4)-(1.6). The scalars μ_1^*, μ_2^* are determined from the requirement

$$\nabla f(x^*) + \mu_1^* a_1 + \mu_2^* a_2 = 0. \tag{2.36}$$

Except in the cases where $\nabla f(x^*)$ is equal to 0 or to $-a_1$ or to $-a_2$, we have $\mu_1^* > 0$ and $\mu_2^* > 0$, but the Lagrange multiplier (μ_1^*, μ_2^*) is neither informative nor strong, because there is no $x \in X$ that simultaneously violates both inequality constraints. The R-multipliers here are the vectors (μ_1^*, μ_2^*) such that $\nabla f(x^*) + \mu_1^* a_1 + \mu_2^* a_2$ is either equal to a multiple of a_1 or to a multiple of a_2. Except for the Lagrange multipliers, which satisfy Eq. (2.36), all other R-multipliers are such that the Lagrangian function has negative slope along some of the feasible directions of X.

3. PSEUDONORMALITY, QUASINORMALITY, 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.

Definition 1: We say that a feasible vector x^* of problem (1.1)-(1.2) is *quasinormal* if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$, and a sequence $\{x^k\} \subset X$ 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 $j = 1, \ldots, r$.

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

(iv) $\{x^k\}$ converges to x^* and for all k, $\lambda_i h_i(x^k) > 0$ for all i with $\lambda_i \neq 0$ and $\mu_j g_j(x^k) > 0$ for all j with $\mu_j \neq 0$.

If x^* is a quasinormal 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. Quasinormality was introduced for the special case where $X = \Re^n$ by Hestenes [Hes75], who showed how it can be used to unify various constraint qualifications. The extension for the case where $X \neq \Re^n$ is investigated here for the first time. A related notion, also introduced here for the first time, is given in the following definition.

Definition 2: We say that a feasible vector x^* of problem (1.1)-(1.2) is *pseudonormal* if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$, and a sequence $\{x^k\} \subset X$ 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) $\{x^k\}$ converges to x^* and

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

It can be seen that pseudonormality implies quasinormality. The following example shows that the reverse is not true. We will show later in this section (Prop. 4), that under the assumption that $N_X(x^*)$ is convex (which is true in particular if X is regular at x^*), quasinormality is in fact equivalent to a slightly weaker version of pseudonormality.

Example 3

Let the constraint set be specified by $C = \{x \in X \mid g_1(x) \leq 0, g_2(x) \leq 0, g_3(x) \leq 0\}$, where $X = \Re^2$ and
$$
\begin{aligned}
g_1(x) &= x_1^2 + (x_2 - 1)^2 - 1, \\
g_2(x) &= \left(x_1 - \cos(\pi/6)\right)^2 + \left(x_2 + \sin(\pi/6)\right)^2 - 1, \\
g_3(x) &= \left(x_1 + \cos(\pi/6)\right)^2 + \left(x_2 + \sin(\pi/6)\right)^2 - 1.
\end{aligned}
$$
(see Fig. 4). Consider the feasible vector $x^* = (0, 0)$. Because there is no x that simultaneously violates all three constraints, quasinormality is satisfied. However, a straightforward calculation shows that we have $\sum_{j=1}^3 \nabla g_j(x^*) = 0$, while
$$g_1(x) + g_2(x) + g_3(x) = 3(x_1^2 + x_2^2) > 0, \qquad \forall\, x \neq x^*,$$
so by using $\mu = (1, 1, 1)$, the conditions for pseudonormality of x^* are violated. Thus, even when $X = \Re^n$, quasinormality does not imply pseudonormality.
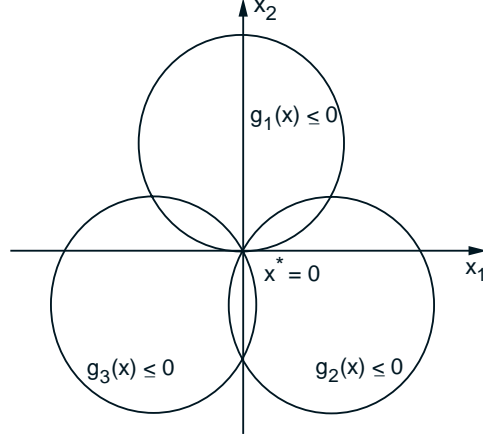
Figure 4. Constraints of Example 3.

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 $\bar{r} < r$, the following superset $\overline{C}$ of the constraint set C,

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

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

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

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

Since CQ1-CQ3 imply pseudonormality, a fact to be shown in the subsequent Prop. 3, 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.

25

(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 $N_X(x^*)^*$.

(d) There exists a $y \in N_X(x^*)^*$ 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. The constraint qualification CQ5 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 $y \in N_X(x^*)^*$ 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 $N_X(x^*)^*$.

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

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

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 $N_X(x^*)^*$, 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: The set
$$W = \big\{(\lambda, \mu) \mid \lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r \text{ satisfy conditions (i) and (ii)}$$
$$\text{of the definition of pseudonormality}\big\} \tag{3.3}$$
consists of just the vector 0.

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 CS replaces CV in Prop. 1. Clearly CQ6 is a more restrictive condition than pseudonormality, since the vectors in W are not required to satisfy condition (iii) of the definition of pseudonormality. If the set of R-multipliers [Eqs. (2.34) and (2.35)] is a nonempty closed convex set, its recession cone is the set W of Eq. (3.3) (this is shown in a less general context by Bonnans and Shapiro [BoS00], Prop. 3.14, but their proof applies to the present context as well). Since compactness of a closed, convex set is equivalent to its recession cone containing just the 0 vector (Rockafellar [Roc70], Th. 8.4), it follows that if the set of R-multipliers is nonempty convex and compact, then CQ6 holds. In view of Prop. 1, the reverse is also true, provided the set of R-multipliers is guaranteed to be convex, which is true in particular if $N_X(x^*)$ is convex. Thus, *if $N_X(x^*)$ is convex, CQ6 is equivalent to the set of R-multipliers being nonempty and compact.* It can also be shown that if X is regular at x^*, then CQ6 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] (in fact it is well-known that for $X = \Re^n$, CQ5a is equivalent to nonemptiness and compactness of the set of Lagrange multipliers; this is a result of Gauvin [Gau77]). However, CQ3, CQ4, and CQ5 do not preclude unboundedness of the set of Lagrange multipliers and hence 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 3: 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 in each case we arrive at a contradiction.

CQ1: Since $X = \Re^n$, implying that $N_X(x^*) = \{0\}$, and we also 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 $j \notin A(x^*)$, contradicts condition (iii).

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

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

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

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

$$\sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x) \leq \sum_{i=1}^{m} \lambda_i h_i(x^*) + \sum_{j=1}^{r} \mu_j g_j(x^*)$$
$$+ \left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \right)' (x - x^*) \tag{3.4}$$
$$= 0,$$

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 $\sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x) > 0$, which contradicts Eq. (3.4).

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

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

Let $\overline{y}$ be the interior point of $N_X(x^*)^*$ 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 $\overline{y} + d \in N_X(x^*)^*$ for all $d \in S$. We have from Eq. (3.5),

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

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

$$-\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*). \tag{3.6}$$

By what has been proved above, the multipliers $\lambda_1, \ldots, \lambda_{\overline{m}}$ of the nonlinear constraints cannot be all zero, so Eq. (3.6) 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 $N_X(x^*)^*$ that appears in assumption (d) of CQ5. Thus,

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

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

A consequence of Prop. 3 is that if any one of the constraint qualifications CQ1-CQ6 holds and X is regular at x^*, by Prop. 1, the constraint set C admits informative Lagrange multipliers at x^*. Without the regularity assumption on X, CQ5 and CQ6 similarly imply the admittance of an R-multiplier vector. In the next section, we will also show similar implications regarding the admittance of an exact penalty at x^*. To this end, we establish a relation between quasinormality and a weaker version of pseudonormality.

Proposition 4:   Let x^* be a feasible vector of problem (1.1)-(1.2), and assume that the normal cone $N_X(x^*)$ is convex. Then x^* is quasinormal if and only if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$ satisfying conditions (i)-(iii) of the definition of quasinormality together with the following condition:

(iv′) $\{x^k\}$ converges to x^* and for all k, $\lambda_i h_i(x^k) \geq 0$ for all i, $\mu_j g_j(x^k) \geq 0$ for all j, and

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

Proof: For simplicity we assume that all the constraints are inequalities that are active at x^*. First we note that if there are no scalars $\mu_1, \ldots, \mu_r$ with the properties described in the proposition, then there are no scalars $\mu_1, \ldots, \mu_r$ satisfying the more restrictive conditions (i)-(iv) in the definition of quasinormality, so x^* is not quasinormal.

To show the converse, suppose that there exist scalars $\mu_1, \ldots, \mu_r$ satisfying conditions (i)-(iii) of the definition of quasinormality together with condition (iv′), i.e., there exist scalars $\mu_1, \ldots, \mu_r$ such that:

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

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

(iii) $\{x^k\}$ converges to x^* and for all k, $g_j(x^k) \geq 0$ for all j, and

$$\sum_{j=1}^{r} \mu_j g_j(x^k) > 0.$$

Condition (iii) implies that $g_j(x^k) \geq 0$ for all j, and $g_{\bar{j}}(x^k) > 0$ for some $\bar{j}$ such that $\mu_{\bar{j}} > 0$. Without loss of generality, we can assume $\bar{j} = 1$, so that we have $g_1(x^k) > 0$ for all k. Let $a_j = \nabla g_j(x^*)$, $j = 1, \ldots, r$. Then by appropriate normalization, we can assume that $\mu_1 = 1$, so that

$$-\left(a_1 + \sum_{j=2}^{r} \mu_j a_j\right) \in N_X(x^*). \tag{3.7}$$

30

If $-a_1 \in N_X(x^*)$, the choice of scalars $\overline{\mu}_1 = 1$ and $\overline{\mu}_j = 0$ for all $j = 2, \ldots, r$, satisfies conditions (i)-(iv) in the definition of quasinormality, hence x^* is not quasinormal and we are done. Assume that $-a_1 \notin N_X(x^*)$. The assumptions of Lemma 2 are satisfied, so it follows that there exist scalars $\overline{\mu}_2, \ldots, \overline{\mu}_r$, not all 0, such that

$$-\left(a_1 + \sum_{j=2}^{r} \overline{\mu}_j a_j\right) \in N_X(x^*), \tag{3.8}$$

and a vector $\overline{d} \in N_X(x^*)^*$ with $a'_j \overline{d} > 0$, for all $j = 2, \ldots, r$ such that $\overline{\mu}_j > 0$. Thus

$$\nabla g_j(x^*)' \overline{d} > 0, \qquad \forall \; j = 2, \ldots, r \text{ with } \overline{\mu}_j > 0, \tag{3.9}$$

while by Eq. (3.8), the $\overline{\mu}_j$ satisfy

$$-\left(\nabla g_1(x^*) + \sum_{j=2}^{r} \overline{\mu}_j \nabla g_j(x^*)\right) \in N_X(x^*). \tag{3.10}$$

Next, we show that the scalars $\overline{\mu}_1 = 1$ and $\overline{\mu}_2, \ldots, \overline{\mu}_r$ satisfy condition (iv) in the definition of quasinormality, completing the proof. We use Thm. 6.26 and Thm. 6.28 of Rockafellar and Wets [RoW98] to argue that for the vector $\overline{d} \in N_X(x^*)^*$ and the sequence x^k constructed above, there is a sequence $d^k \in T_X(x^k)$ such that $d^k \to \overline{d}$. Since $x^k \to x^*$ and $d^k \to \overline{d}$, by Eq. (3.9), we obtain for all sufficiently large k,

$$\nabla g_j(x^k)' d^k > 0, \qquad \forall \; j = 2, \ldots, r \text{ with } \overline{\mu}_j > 0.$$

Since $d^k \in T_X(x^k)$, there exists a sequence $\{x_\nu^k\} \subset X$ such that, for each k, we have $x_\nu^k \neq x^k$ for all ν and

$$x_\nu^k \to x^k, \qquad \frac{x_\nu^k - x^k}{\|x_\nu^k - x^k\|} \to \frac{d^k}{\|d^k\|}, \qquad \text{as } \nu \to \infty. \tag{3.11}$$

For each $j = 2, \ldots, r$ such that $\overline{\mu}_j > 0$, we use Taylor's theorem for the constraint function g_j. We have, for some vector sequence ξ^ν converging to 0,

$$\begin{aligned}
g_j(x_\nu^k) &= g_j(x^k) + \nabla g_j(x^k)'(x_\nu^k - x^k) + o(\|x_\nu^k - x^k\|) \\
&\geq \nabla g_j(x^k)' \left(\frac{d^k}{\|d^k\|} + \xi^\nu\right) \|x_\nu^k - x^k\| + o(\|x_\nu^k - x^k\|) \\
&= \|x_\nu^k - x^k\| \left(\nabla g_j(x^k)' \frac{d^k}{\|d^k\|} + \nabla g_j(x^k)' \xi^\nu + \frac{o(\|x_\nu^k - x^k\|)}{\|x_\nu^k - x^k\|}\right),
\end{aligned}$$

where the inequality above follows from Eq. (3.11) and the assumption that $g_j(x^k) \geq 0$, for all j and x^k. It follows that for ν and k sufficiently large, there exists $x_\nu^k \in X$ arbitrarily close to x^k such that $g_j(x_\nu^k) > 0$, for all $j = 2, \ldots, r$ with $\overline{\mu}_j > 0$.

Since $g_1(x^k) > 0$ and g_1 is a continuous function, we have that $g_1(\tilde{x}) > 0$ for all $\tilde{x}$ in some neighborhood V_k of x^k. Since $x^k \to x^*$ and $x^k_\nu \to x^k$ for each k, by choosing ν and k sufficiently large, we get $g_j(x^k_\nu) > 0$ for $j = 1$ and each $j = 2, \ldots, r$ with $\overline{\mu}_j > 0$. This together with Eq. (3.10), violates the quasinormality assumption of x^*, which completes the proof.   **Q.E.D.**

The following example shows that convexity of $N_X(x^*)$ is an essential assumption for the conclusion of Prop. 4.

Example 4

Here X is the subset of $\Re^2$ given by

$$X = \left\{ x_2 \geq 0 \mid \left((x_1 + 1)^2 + (x_2 + 1)^2 - 2\right) \left((x_1 - 1)^2 + (x_2 + 1)^2 - 2\right) \leq 0 \right\}$$

(see Fig. 5). The normal cone $N_X(x^*)$ consists of the three rays shown in Fig. 5, and is not convex. Let there be two inequality constraints with

$$g_1(x) = -(x_1 + 1)^2 - (x_2)^2 + 1, \qquad g_2(x) = -x_2.$$

In order to have $-\sum_j \mu_j \nabla g_j(x^*) \in N_X(x^*)$, we must have $\mu_1 > 0$ and $\mu_2 > 0$. There is no $x \in X$ such that $g_2(x) > 0$, so x^* is quasinormal. However, for $-2 \leq x_1 \leq 0$ and $x_2 = 0$, we have $x \in X$, $g_1(x) > 0$, and $g_2(x) = 0$. Hence x^* does not satisfy the weak form of pseudonormality given in Prop. 4.
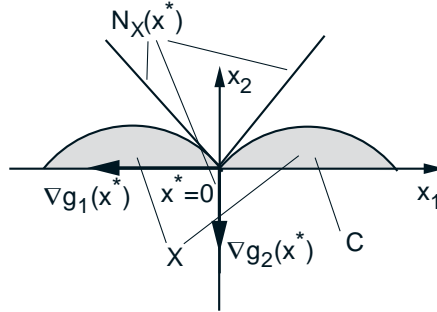

Figure 5.   Constraints of Example 4.

4. PSEUDONORMALITY AND ADMITTANCE OF AN EXACT PENALTY

We will show that pseudonormality implies that the constraint set 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 5:   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

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

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

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. 3, this local minimum is pseudonormal, and hence there exist multipliers satisfying the conditions of Prop. 1 with $\mu_0^* = 1$. With straightforward calculation, these conditions yield scalars $\lambda_1^*, \ldots, \lambda_m^*$ and $\mu_1^*, \ldots, \mu_r^*$, satisfying the desired conditions.   **Q.E.D.**

Proposition 6:   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 smooth 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

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

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. 5):

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

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.**

A cursory examination shows that the proof of Prop. 6 goes through if we substitute pseudonormality with the weaker version of pseudonormality introduced in Prop. 4. Thus by using also Prop. 4, we obtain the following:

Proposition 7: If x^* is a feasible vector of problem (1.1)-(1.2), which is quasinormal, and the normal cone $N_X(x^*)$ is convex, then the constraint set admits an exact penalty at x^*.

The following proposition establishes the connection between admittance of an exact penalty and admittance of Lagrange multipliers. Regularity of X is an important condition for this connection.

Proposition 8: 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 smooth 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. 5. (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.**

As an illustration of the above propositions, consider Example 3. Here, since x^* is quasinormal but not pseudonormal, Prop. 6 cannot be used. However, since $X = \Re^n$ and $N_X(x^*) = \{0\}$ is convex, Prop. 7 applies and shows that the constraint set admits an exact penalty at x^*. By Prop. 8, since X is regular, the constraint set admits Lagrange multipliers at x^*. [This can also be shown using the fact $T_C(x^*) = V(x^*) = \{0\}$, which implies that X^* is quasiregular.]

We will show in Example 5 in Section 7 that the converses of Props. 6 and 7 do not hold; i.e., the admittance of an exact penalty function at a point x^* does not imply pseudonormality or quasinormality. Furthermore, we will also show in Example 8 that the regularity assumption on X in Prop. 8 cannot be dispensed with. On the other hand, because Prop. 5 does not require regularity of X, the proof of Prop. 8 can be used to establish that *admittance of an exact penalty implies the admittance of R-multipliers*, as defined in Section 2. The relations shown thus far are summarized in Fig. 6, which illustrates the unifying role of pseudonormality and quasinormality. In this figure, unless indicated otherwise, the implications cannot be established in the opposite direction without additional assumptions (Section 7 provides the necessary examples and counterexamples).

X = R^n

```
┌─────────────────────────┐
│ Constraint Qualifications │
│         CQ1-CQ4          │
└─────────────────────────┘
            │
            ▼
   ┌──────────────────┐
   │ Pseudonormality  │
   └──────────────────┘
            │
            ▼
   ┌──────────────────┐
   │  Quasinormality  │
   └──────────────────┘
         ╱      ╲
        ▼        ▼
┌──────────────┐  ┌─────────────────────┐
│Quasiregularity│  │ Admittance of an Exact │
│              │  │       Penalty        │
└──────────────┘  └─────────────────────┘
        ╲          ╱
         ▼        ▼
┌─────────────────────────┐
│ Admittance of Informative and │
│ Strong  Lagrange Multipliers │
└─────────────────────────┘
         │    ▲
         ▼    │
┌─────────────────────────┐
│  Admittance of Lagrange  │
│       Multipliers        │
└─────────────────────────┘
```

X ≠ R^n and Regular

```
┌─────────────────────────┐
│ Constraint Qualifications │
│        CQ5, CQ6          │
└─────────────────────────┘
            │
            ▼
   ┌──────────────────┐
   │ Pseudonormality  │
   └──────────────────┘
            │
            ▼
   ┌──────────────────┐
   │  Quasinormality  │
   └──────────────────┘
            │
            ▼
┌─────────────────────┐
│ Admittance of an Exact │
│       Penalty        │
└─────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Admittance of Informative and │
│ Strong  Lagrange Multipliers │
└─────────────────────────┘
         │    ▲
         ▼    │
┌─────────────────────────┐
│  Admittance of Lagrange  │
│       Multipliers        │
└─────────────────────────┘
```

X ≠ R^n

```
┌─────────────────────────┐
│ Constraint Qualifications │
│        CQ5, CQ6          │
└─────────────────────────┘
            │
            ▼
   ┌──────────────────┐
   │ Pseudonormality  │
   └──────────────────┘
            │
            ▼
┌─────────────────────┐
│ Admittance of an Exact │
│       Penalty        │
└─────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Admittance of R-multipliers │
└─────────────────────────┘
```
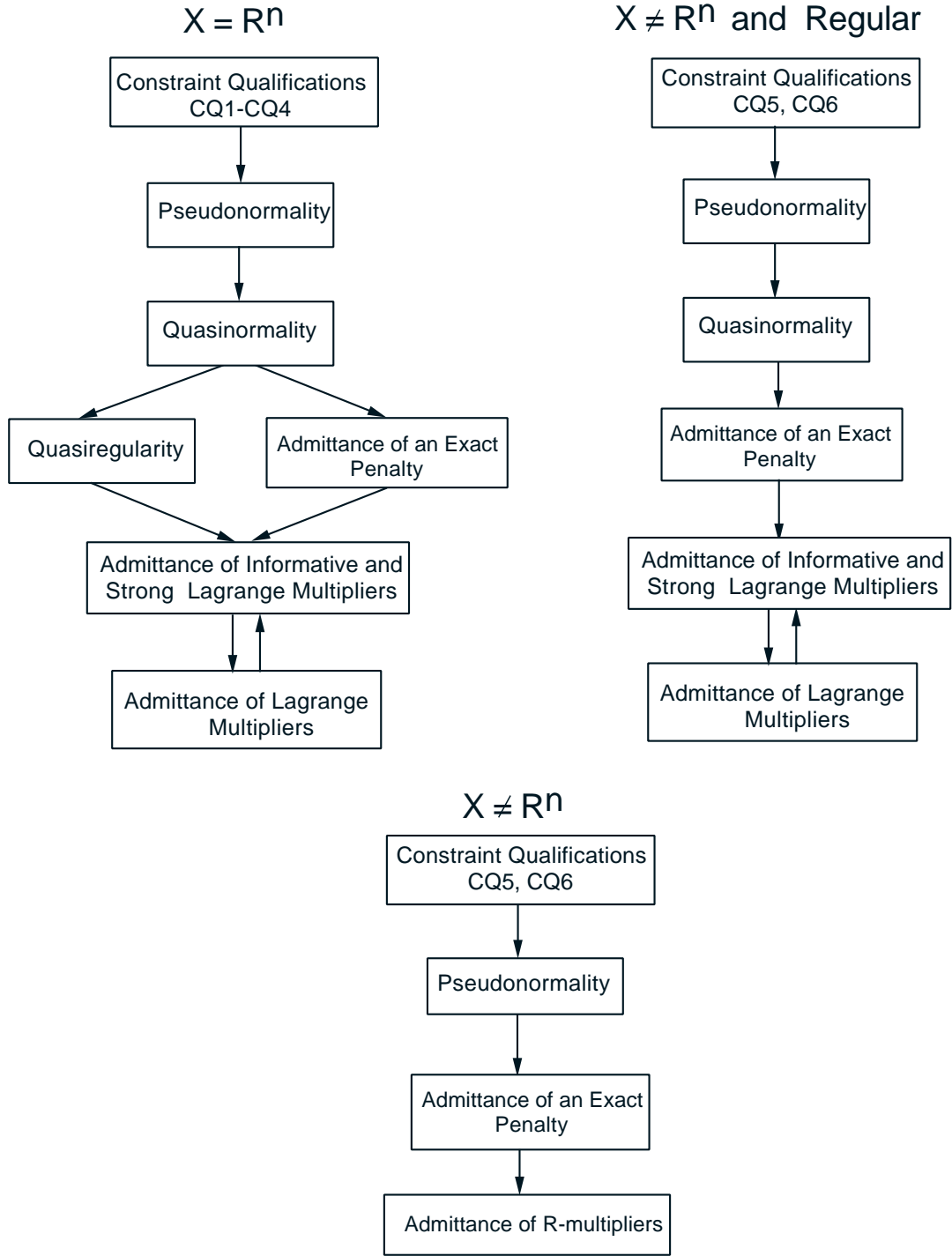
Figure 6.   Relations between various conditions, which when satisfied at a local minimum x^*, guarantee the admittance of an exact penalty and corresponding multipliers. In the case where X is regular, the tangent and normal cones are convex. Hence, by Prop. 2(a), the admittance of Lagrange multipliers implies the admittance of an informative Lagrange multiplier, while by Prop. 7, quasinormality implies the admittance of an exact penalty.

36

5. USING THE EXTENDED REPRESENTATION

In practice, the set X can often be described in terms of smooth equality and inequality constraints:

$$X = \left\{ x \mid h_i(x) = 0, \ i = m+1, \ldots, \overline{m}, \ g_j(x) \leq 0, \ j = r+1, \ldots, \overline{r} \right\}.$$

Then the constraint set C can alternatively be described without an abstract set constraint, in terms of all of the constraint functions

$$h_i(x) = 0, \quad i = 1, \ldots, \overline{m}, \qquad g_j(x) \leq 0, \quad j = 1, \ldots, \overline{r}.$$

We call this the *extended representation* of C, to contrast it with the representation (1.2), which we call the *original representation*. Issues relating to exact penalty functions and Lagrange multipliers can be investigated for the extended representation and results can be carried over to the original representation by using the following proposition.

Proposition 9:

(a) If the constraint set admits Lagrange multipliers in the extended representation, it admits Lagrange multipliers in the original representation.

(b) If the constraint set admits an exact penalty in the extended representation, it admits an exact penalty in the original representation.

Proof: (a) The hypothesis implies that for every smooth cost function f for which x^* is a local minimum there exist scalars $\lambda_1^*, \ldots, \lambda_{\overline{m}}^*$ and $\mu_1^*, \ldots, \mu_{\overline{r}}^*$ satisfying

$$\nabla f(x^*) + \sum_{i=1}^{\overline{m}} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{\overline{r}} \mu_j^* \nabla g_j(x^*) = 0, \tag{5.1}$$

$$\mu_j^* \geq 0, \qquad \forall \ j = 0, 1, \ldots, \overline{r},$$

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

where

$$\overline{A}(x^*) = \{ j \mid g_j(x^*) = 0, \ j = 1, \ldots, \overline{r} \}.$$

For $y \in T_X(x^*)$, we have $\nabla h_i(x^*)'y = 0$ for all $i = m+1, \ldots, \overline{m}$, and $\nabla g_j(x^*)'y \leq 0$ for all $j = r+1, \ldots, \overline{r}$ with $j \in \overline{A}(x^*)$. Hence Eq. (5.1) implies that

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

37

and it follows that λ_i^*, $i = 1, \ldots, m$, and μ_j^*, $j = 1, \ldots, r$, are Lagrange multipliers for the original representation.

(b) Consider the exact penalty function for the extended representation:

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

We have $F_c(x) = \overline{F}_c(x)$ for all $x \in X$. Hence if x^* is an unconstrained local minimum of $\overline{F}_c(x)$, it is also a local minimum of $F_c(x)$ over $x \in X$. Thus, for a given $c > 0$, if x^* is both a strict local minimum of f over C and an unconstrained local minimum of $\overline{F}_c(x)$, it is also a local minimum of $F_c(x)$ over $x \in X$.   **Q.E.D.**

Note that part (a) of the above proposition does not guarantee the existence of informative Lagrange multipliers in the original representation, and indeed in Example 9 given in Section 7, there exist some informative Lagrange multipliers in the extended representation, but there exists none in the original representation.

As an example where Prop. 9 is useful, consider the important special case where all constraints are linear and X is a polyhedron. Then the constraint set need not satisfy quasinormality, as will be shown in Example 10 in Section 7. However, by Prop. 3, it satisfies quasinormality in the extended representation, so using Prop. 9, it admits Lagrange multipliers and an exact penalty at any feasible point in the original representation.

6. EXTENSIONS UNDER CONVEXITY ASSUMPTIONS

In this section, we extend the theory of preceding sections to the case where the functions f and g_j may be nondifferentiable, but are all instead assumed convex (a slightly more general development is given in [BNO01], where each of the functions f and g_j may be either smooth or convex). We thus consider the problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in X, \qquad g(x) \leq 0, \end{aligned} \tag{6.1}$$

where $g(x) = \big(g_1(x), \ldots, g_r(x)\big)$ is the constraint function vector, and we assume that X is a nonempty, closed, and convex set, and each of the functions f and g_j is convex over $\Re^n$. For simplicity, we assume no equality constraints. The extension of the following analysis to the case

where there are linear equality constraints is straightforward: we simply replace each equality constraint into two linear (and hence convex) inequality constraints.

For a convex function F, we use the notation $\partial F(x)$ to denote the set of all subgradients of F at $x \in \Re^n$. To generalize the enhanced Fritz-John conditions under these assumptions, we use the following classical condition for the problem of minimizing $F(x)$ over X, where F is convex: if x^* is a local minimum of this problem and the tangent cone $T_X(x^*)$ is convex, then

$$0 \in \partial F(x^*) + T_X(x^*)^*. \tag{6.2}$$

By a nearly verbatim repetition of the proof of Prop. 1, while using this necessary condition in place of $-\nabla F^k(x^k) \in T_X(x^k)^*$, together with the convexity assumptions on X, f, and g_j, we obtain the following extension of enhanced Fritz-John conditions:

Proposition 10: Consider problem (6.1), assuming that X is convex, and the functions f and g_j are convex over $\Re^n$, and let x^* be a global minimum. Then there exists a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, satisfying the following conditions:

(i) $\mu_0^* f(x^*) = \min_{x \in X} \{\mu_0^* f(x) + \mu^{*\prime} g(x)\}$.

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

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

(iv) If the index set $J = \{j \neq 0 \mid \mu_j^* > 0\}$ is nonempty, there exists a sequence $\{x^k\} \subset X$ that converges to x^* and is such that for all k,

$$f(x^k) < f(x^*), \qquad \mu_j^* g_j(x^k) > 0, \quad \forall \, j \in J,$$

$$g_j(x^k) = o\left(\min_{l \in J} g_l(x^k)\right), \qquad \forall \, j \notin J.$$

The theory of the preceding sections can now be generalized using Prop. 10. We first extend the definition of a Lagrange multiplier. If x^* is a global minimum of problem (6.1), we say that a vector $\mu^* \geq 0$ is a *Lagrange multiplier vector corresponding to f and x^** if

$$f(x^*) = \min_{x \in X}\{f(x) + \mu^{*\prime} g(x)\}, \qquad \mu^{*\prime} g(x^*) = 0. \tag{6.3}$$

Note here that since $g(x^*) \leq 0$ and $\mu^* \geq 0$, the condition $\mu^{*\prime} g(x^*) = 0$ of Eq. (6.3) is equivalent to the condition $\mu_j^* g_j(x^*) = 0$ for all j, which is the CS condition.

Similarly, the definition of pseudonormality is extended as follows:

39

Definition 3:   Consider problem (6.1), assuming that X is convex, and the functions f and g_j are convex over $\Re^n$. A feasible vector x^* is said to be *pseudonormal* if there exist no vector $\mu = (\mu_1, \dots, \mu_r) \geq 0$, and a sequence $\{x^k\} \subset X$ such that:

(i) $0 = \mu'g(x^*) = \inf_{x \in X} \mu'g(x)$.

(ii) $\{x^k\}$ converges to x^* and $\mu'g(x^k) > 0$ for all k.

If a global minimum x^* is pseudonormal, by Prop. 10, there exists a Lagrange multiplier vector, which also satisfies the extra CV condition (iv) of that proposition. Furthermore, the analysis of Section 3 is easily extended to show that x^* is pseudonormal under either one of the following two criteria:

(a) *Polyhedral criterion*: $X = \Re^n$ and the functions g_j are linear.

(b) *Slater criterion*: There exists a feasible vector $\bar{x}$ such that

$$g_j(\bar{x}) < 0, \qquad j = 1, \dots, r.$$

Thus, under either one of these criteria, a Lagrange multiplier vector is guaranteed to exist.

If X is a polyhedron (rather than $X = \Re^n$) and the functions g_j are linear, we can also prove existence of at least one Lagrange multiplier, by combining the linearity criterion above with the extended representation of the problem as in the preceding section. Also the Slater criterion can be extended to the case where there are additional linear equality constraints. Then in addition to the condition $g_j(\bar{x}) < 0$ for all j, for existence of a Lagrange multiplier, there should exist a feasible vector in the relative interior of X.

Finally, let us provide a geometric interpretation of pseudonormality under convexity assumptions. Consider the set

$$G = \{g(x) \mid x \in X\}$$

and hyperplanes that support this set at $g(x^*)$. As Fig. 7 illustrates, pseudonormality of the feasible point x^* means that there is no hyperplane H with a normal $\mu \geq 0$ such that:

(1) H supports G at $g(x^*)$ and passes through 0, i.e.,

$$H = \{z \mid \mu'z = \mu'g(x^*) = 0\}.$$

(2) $g(x^*)$ can be approached by a sequence $\{g(x^k)\} \subset G \cap \text{int}(\overline{H})$, where $\overline{H}$ is the upper halfspace defined by the hyperplane H,

$$\overline{H} = \{z \mid \mu'z \geq 0\}.$$

40

Figure 7 also indicates the type of constraint qualifications that guarantee pseudonormality. The Slater condition can be rephrased to mean that the set G intersects the interior of the nonpositive orthant. Clearly, if this is so, there cannot exist a hyperplane with a normal $\mu \geq 0$ that simultaneously supports G at $g(x^*)$ and passes through 0. Similarly, if $X = \Re^n$ and the g_j are linear, the set G is an affine set, and if this is so, G is fully contained in the hyperplane H and cannot intersect the interior of the upper halfspace $\overline{H}$. Thus the polyhedral and Slater criteria imply pseudonormality of all feasible points.
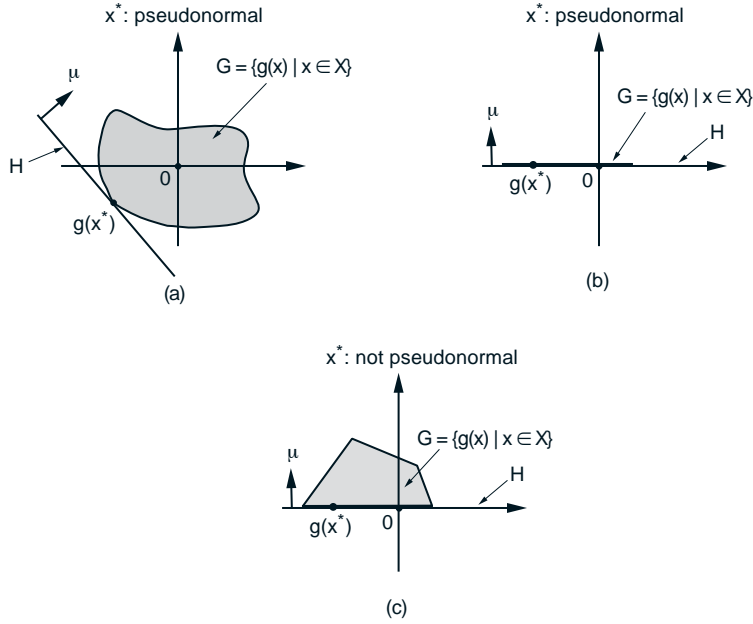
Figure 7. Geometric interpretation of pseudonormality. Consider the set

$$G = \{g(x) \mid x \in X\}$$

and hyperplanes that support this set at $g(x^*)$. For feasibility, G should intersect the nonpositive orthant $\{z \mid z \leq 0\}$. The first condition $[0 = \mu'g(x^*) = \inf_{x \in X} \mu'g(x)]$ in the definition of pseudonormality means that there is a hyperplane with normal μ, which simultaneously supports G at $g(x^*)$ and passes through 0 [note that, as illustrated in figure (a), this cannot happen if G intersects the interior of the nonpositive orthant; cf. the Slater criterion]. The second condition $[\{x^k\}$ converges to x^* and $\mu'g(x^k) > 0$ for all $k]$ means that $g(x^*)$ can be approached by a sequence $\{g(x^k)\} \subset G \cap \operatorname{int}(\overline{H})$, where $\overline{H}$ is the upper halfspace defined by the hyperplane,

$$\overline{H} = \{z \mid \mu'z \geq 0\};$$

[cf. figures (b) and (c)]. Pseudonormality of x^* means that there is no $\mu \geq 0$ and $\{x^k\} \subset X$ satisfying both of these conditions. If the Slater criterion holds, the first condition cannot be satisfied. If the polyhedral criterion holds, the set G is an affine set and the second condition cannot be satisfied (this depends critically on $X = \Re^n$ rather than X being a general polyhedron).

41

7. EXAMPLES AND COUNTEREXAMPLES

In this section, we provide examples and counterexamples that clarify some of the earlier developments. In particular, we explore the relations between various conditions given in Fig. 6, and we show (at least in the less obvious cases) that the one-directional arrows cannot be reversed in the absence of additional assumptions.

Example 5

Here we show that even with $X = \Re^n$, the admittance of an exact penalty does not imply pseudonormality and quasinormality. Let $C = \left\{ x \in \Re^2 \mid g_1(x) \leq 0,\ g_2(x) \leq 0,\ h_1(x) = 0 \right\}$, where

$$
\begin{aligned}
g_1(x) &= (x_1 - 1)^2 + x_2^2 - 1, \\
g_2(x) &= (x_1 + 1)^2 + x_2^2 - 1, \\
h_1(x) &= x_2,
\end{aligned}
$$

(see Fig. 8). The only feasible solution is $x^* = (0,0)$ and the constraint gradients are given by

$$
\nabla g_1(x^*) = (-2, 0), \quad \nabla g_2(x^*) = (2, 0), \quad \nabla h_1(x^*) = (0, 1).
$$

Let $\mu_1 = \mu_2 = 1$ and $\lambda = 0$. With this choice of multipliers we have

$$
\mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) + \lambda \nabla h_1(x^*) = 0.
$$

In addition it can be seen that arbitrarily close to x^*, there exists some x such that $g_1(x) > 0$ and $g_2(x) > 0$ simultaneously. Thus x^* is not quasinormal, and hence not pseudonormal.
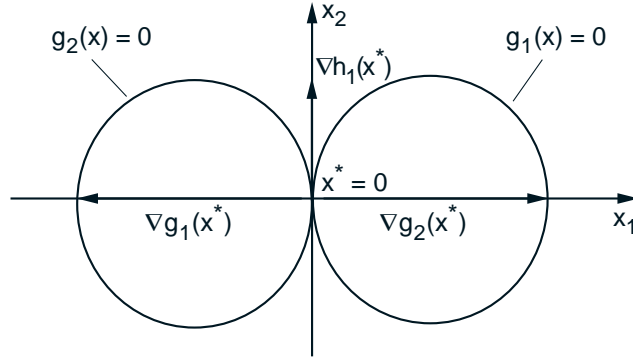

Figure 8. Constraints of Example 5. The only feasible point is $x^* = (0,0)$.

On the other hand, the directional derivative of the function

$$
P(x) = |h_1(x)| + \sum_{j=1}^{2} g_j^+(x)
$$

42

at x^* is positive in all directions. This is because the directional derivative of $\sum_{j=1}^2 g_j^+(x)$ is positive everywhere except in directions $d_1 = (0,1)$ and $d_2 = (0,-1)$, and with the introduction of the equality constraint, the function $P(x)$ acquires a positive directional derivative in these directions as well. By choosing a sufficiently large penalty parameter c, we can guarantee that x^* is a local minimum of the function $F_c(x)$. Hence the constraint set admits an exact penalty at x^*.

Example 6

Here we show that even with $X = \Re^n$, the admittance of an exact penalty function does not imply quasiregularity. Let $C = \{x \in \Re^2 \mid g_1(x) \le 0,\ g_2(x) \le 0,\ g_3(x) \le 0\}$, where

$$
\begin{aligned}
g_1(x) &= -(x_1 + 1)^2 - (x_2)^2 + 1, \\
g_2(x) &= x_1^2 + (x_2 + 1)^2 - 1, \\
g_3(x) &= -x_2,
\end{aligned}
$$

(see Fig. 9). The only feasible solution is $x^* = (0,0)$ and the constraint gradients are given by

$$
\nabla g_1(x^*) = (-2,0), \quad \nabla g_2(x^*) = (0,2), \quad \nabla g_3(x^*) = (0,-1).
$$

At $x^* = (0,0)$, the cone of first order feasible variations $V(x^*)$ is equal to the nonnegative x_1 axis and strictly contains $T(x^*)$, which is equal to x^* only. Therefore x^* is not a quasiregular point.
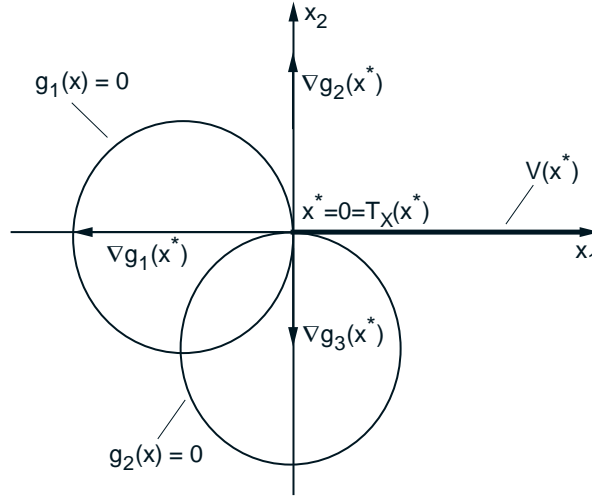

Figure 9. Constraints of Example 6. The only feasible point is $x^* = (0,0)$. The tangent cone $T(x^*)$ and the cone of first order feasible variations $V(x^*)$ are also illustrated in the figure.

However, it can be seen that the directional derivative of the function $P(x) = \sum_{j=1}^3 g_j^+(x)$ at x^* is positive in all directions. This implies we can choose a sufficiently large penalty parameter

c, so that x^* is a local minimum of the function $F_c(x)$. The constraint set admits an exact penalty function at x^*.

Example 7

Here we show that, when $X = \Re^n$, the admittance of Lagrange multipliers with quasiregularity (but not quasinormality) holding, does not imply the admittance of an exact penalty. Let $C = \left\{x \in \Re^2 \mid g_1(x) \leq 0, g_2(x) \leq 0\right\}$, where

$$
\begin{aligned}
g_1(x) &= x_2, \\
g_2(x) &= x_1^6 + x_2^3.
\end{aligned}
$$

At $x^* = (0,0)$, the tangent cone is equal to the cone of first order feasible variations. Hence x^* is a quasiregular point, which implies that the constraint set admits Lagrange multipliers at x^*. However, it is not true that the constraint set admits an exact penalty function at x^*. For this purpose, we consider the function $f(x) = -x_1^4 - x_2$, which is a smooth function with a strict local minimum at x^*. By contrast, x^* is not a local minimum of the function $F_c(x)$ no matter how large c is chosen. To illustrate this we define the function $l(x_1) = F_c(x_1,0) = -x_1^4 + cx_1^6$, which has a local maximum at x^* for any $c > 0$. Hence the existence of Lagrange multipliers does not guarantee the local exactness of the penalty function $F_c(x)$.

Example 8

Here we show that if X is not regular, the admittance of an exact penalty does not imply the admittance of Lagrange multipliers (although it does imply the admittance of R-multipliers). Consider the set $X \subset \Re^2$ depicted in Fig. 10, and let there be a single linear equality constraint $h(x) = x_1 = 0$. For $x^* = (0,0)$, we have $T_X(x^*)^* = \{0\}$, while $N_X(x^*)$ consists of the two rays shown in Fig. 10. Because $\nabla h(x^*) = (1,0) \notin N_X(x^*)$, pseudonormality is satisfied, and hence by Prop. 6, the constraint set admits an exact penalty at x^*. On the other hand for the cost function $f(x) = -x_2$, we have $\nabla f(x^*) + \lambda \nabla h(x^*) \neq 0$ for all λ, so there is no Lagrange multiplier. The non-admittance of Lagrange multipliers can also be verified in this example by noting that $V(x^*)^* + T_X(x^*)^* \neq T_C(x^*)^*$.

Example 9

Here we show that there may exist informative Lagrange multipliers in the extended representation of the constraint set, but not in the original representation. This suggests that the extended representation of a constraint set cannot be used to infer the admittance of informative Lagrange multipliers as opposed to other characteristics of the constraint set given by Prop. 9.
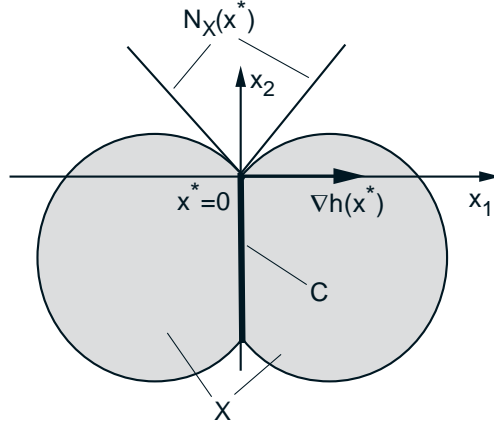
44

Figure 10. Constraints of Example 8.

Let the constraint set be represented in extended form without an abstract set constraint as

$$C = \big\{ x \in \Re^2 \mid a_1' x \le 0,\ a_2' x \le 0,\ (a_1' x)(a_2' x) = 0 \big\},$$

where $a_1 = (-1, 0)$ and $a_2 = (0, -1)$. Consider the vector $x^* = (0, 0)$. Here $T_C(x^*)$ consists of the nonnegative coordinate axes, whereas $V(x^*)$ is equal to the nonnegative orthant in the extended representation, as shown in Fig. 11. Although x^* is not a quasiregular point, we have $T_C(x^*)^* = V(x^*)^*$, which (as mentioned in Section 1) is a necessary and sufficient condition for the admittance of Lagrange multipliers at x^* when $X = \Re^n$. By Prop. 2, this implies that the constraint set admits informative Lagrange multipliers in the extended representation.
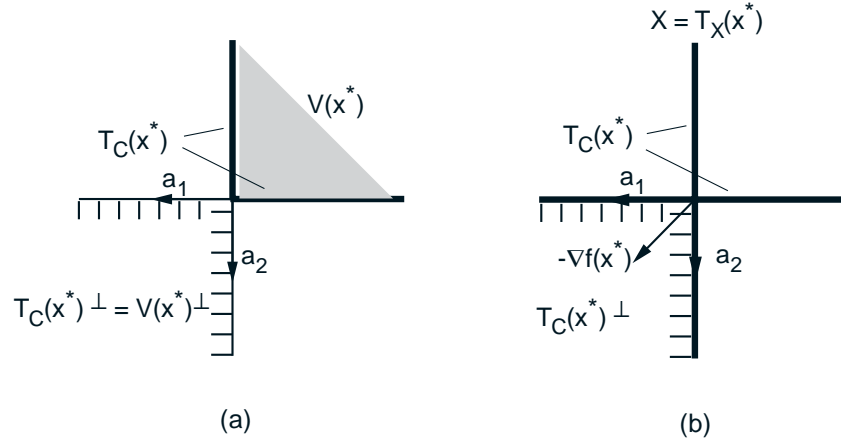

Figure 11. Constraints and relevant cones for different representations of the problem in Example 9.

Now let the same constraint set be specified by the two linear constraint functions $a_1' x \le 0$

and $a'_2 x \leq 0$ together with the abstract constraint set

$$X = \{x \mid (a'_1 x)(a'_2 x) = 0\}$$

Here $T_X(x^*) = X$ and $T_X(x^*)^* = \{0\}$. The normal cone $N_X(x^*)$ consists of the coordinate axes. Since $N_X(x^*) \neq T_X(x^*)^*$, X is not regular at x^*. Furthermore, $T_X(x^*)$ is not convex, so Prop. 2(a) cannot be used to guarantee the admittance of an informative Lagrange multiplier. For any f for which x^* is a local minimum, we must have $-\nabla f(x^*) \in T_C(x^*)^*$ (see Fig. 11). The candidate multipliers are determined from the requirement that

$$-\left(\nabla f(x^*) + \sum_{j=1}^{2} \mu_j a_j\right) \in T_X(x^*)^* = \{0\},$$

which uniquely determines μ_1 and μ_2. If $\nabla f(x^*)$ lies in the interior of the positive orthant, we need to have $\mu_1 > 0$ and $\mu_2 > 0$. However, there exists no $x \in X$ that violates both constraints $a'_1 x \leq 0$ and $a'_2 x \leq 0$, so the multipliers do not qualify as informative. Thus, the constraint set does not admit informative Lagrange multipliers in the original representation.

Example 10

Here we show that if the constraint set involves linear constraint functions and a polyhedral set X, then x^* need not be quasinormal (even though it admits Lagrange multipliers). Let $C = \{x \in X \mid a'x \leq 0, \ b'x \leq 0\}$, where $a = (1, -1)$, $b = (-1, -1)$, and $X = \{x \in \Re^2 \mid a'x \geq 0, \ b'x \geq 0\}$. The constraint set is depicted in Fig. 12.
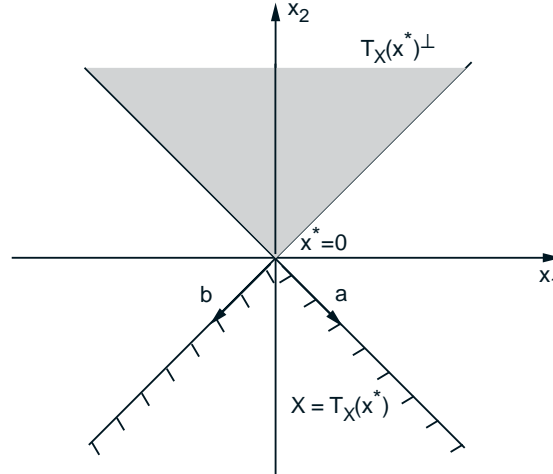

Figure 12. Constraints of Example 10. The only feasible point is $x^* = (0, 0)$. The tangent cone $T_X(x^*)$ and its polar $T_X(x^*)^*$ are shown in the figure.

The only feasible point is $x^* = (0,0)$. By choosing $\mu = (1,1)$, we get

$$-(a+b) \in T_X(x^*)^*,$$

while in every neighborhood N of x^* there is an $x \in X \cap N$ such that $a'x > 0$ and $b'x > 0$ simultaneously. Hence x^* is not quasinormal. Note that this constraint set admits Lagrange multipliers at $x^* = (0,0)$ with respect to its extended representation (cf. Prop. 3), and hence it admits Lagrange multipliers at $x^* = (0,0)$ with respect to the original representation (cf. the discussion at the end of Section 5).

Example 11

Here we show why in the definition of admittance of an exact penalty, it is necessary to restrict attention to functions f with a strict local minimum at x^*. Consider the 2-dimensional constraint set specified by

$$h(x) = \frac{x_2}{x_1^2 + 1} = 0, \qquad x \in X = \Re^2.$$

The feasible points are of the form $x = (x_1, 0)$ with $x_1 \in \Re$, and at each of them the gradient $\nabla h(x^*)$ is nonzero, so CQ1 holds. If $f(x) = x_2$, every feasible point is a local minimum, yet for any $c > 0$, we have

$$\inf_{x \in \Re^2} \left\{ x_2 + c\frac{|x_2|}{|x_1^2 + 1|} \right\} = -\infty$$

(take $x_1 = x_2$ as $x_2 \to -\infty$). Thus, the penalty function is not exact for any $c > 0$. It follows that CQ1 would not imply the admittance of an exact penalty if we were to change the definition of the latter to allow cost functions with nonstrict local minima. Note that for the cost function $f(x) = x_2$, it can be shown that the problem is not calm as per the definition of Clarke [Cla76], [Cla83]. In particular, the primal function of this problem is given by

$$p(u) = \inf_{\frac{x_2}{x_1^2+1}=u} x_2 = \begin{cases} u & \text{if } u \geq 0, \\ -\infty & \text{if } u < 0, \end{cases}$$

and violates the definition of calmness as defined in [Cla83]. Thus CQ1 does not imply calmness, illustrating a fundamental difference between the notions of calmness and of admittance of an exact penalty.

REFERENCES

[AHU61] Arrow, K. J., Hurwicz, L., and Uzawa, H., 1961. "Constraint Qualifications in Maximization Problems," Naval Research Logistics Quarterly, Vol. 8, pp. 175-191.

[Aba67] Abadie, J., 1967. "On the Kuhn-Tucker Theorem," in Nonlinear Programming, Abadie, J., (Ed.), North Holland, Amsterdam.

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

[BNO01] Bertsekas, D. P., with Nedić, A., and Ozdaglar, A. E., 2001. Lectures Notes on Convexity, Duality, and Lagrange Multipliers, M.I.T., Cambridge, MA.

[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.

[Cla76] Clarke, F. H., 1976. Math. of Operations Reserach, Vol. 1 pp. 165-174.

[Cla83] Clarke, F. H., 1983. Optimization and Nonsmooth Analysis, Wiley, N. Y.

[Gau77] Gauvin, J., 1977. "A Necessary and Sufficient Condition to Have Bounded Multipliers in Convex Programming," Math. Programming., Vol. 12, pp. 136-138.

[GoT71] Gould, F. J., and Tolle, J., 1971. "A Necessary and Sufficient Condition for Constrained Optimization," SIAM J. Applied Math., Vol. 20, pp. 164-172.

[GoT72] Gould, F. J., and Tolle, J., 1972. "Geometry of Optimality Conditions and Constraint Qualifications," Math. Programming, Vol. 2, pp. 1-18.

[Gui69] Guignard, M., 1969. "Generalized Kuhn-Tucker Conditions for Mathematical Programming Problems in a Banach Space," SIAM J. on Control, Vol. 7, pp. 232-241.

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

[HiL93] Hiriart-Urruty, J.-B., and Lemarechal, C., 1993. Convex Analysis and Minimization Algorithms, Vol. I, Springer-Verlag, Berlin and N. Y.

[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.

[MaF67] Mangasarian, O. L., and Fromovitz, S., 1967. "The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints," J. Math. Anal. and Appl., Vol. 17, pp. 37-47.

[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.

[Pet73] Peterson, D. W., 1973. "A Review of Constraint Qualifications in Finite-Dimensional Spaces," SIAM Review, Vol. 15, pp. 639-654.

[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.

[Roc70] Rockafellar, R. T., 1970. Convex Analysis, Princeton Univ. Press, Princeton, N. J.

[Roc73] Rockafellar, R. T., 1973. "A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Minimization," Math. Programming, Vol. 5, pp. 354-373.

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

[Sla50] Slater, M., 1950. "Lagrange Multipliers Revisited: A Contribution to Non-Linear Programming," Cowles Commission Discussion paper, Math. 403.

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