Lagrange Multipliers

Date: April 12, 2001

Contents

2.1. Introduction to Lagrange Multipliers . . . . . . . . . p. 2
2.2. Enhanced Fritz John Optimality Conditions . . . . . . p. 12
2.3. Informative Lagrange Multipliers . . . . . . . . . . . p. 20
2.4. Pseudonormality and Constraint Qualifications . . . . p. 32
2.5. Exact Penalty Functions . . . . . . . . . . . . . . . . . p. 47
2.6. Using the Extended Representation . . . . . . . . . p. 55
2.7. Extensions Under Convexity Assumptions . . . . . . p. 59
2.8. Notes and Sources . . . . . . . . . . . . . . . . . . . . p. 64
The optimality conditions developed in Chapter 1 apply to any type of constraint set. On the other hand, the constraint set of an optimization problem is usually specified in terms of equality and inequality constraints. If we take into account this structure, we can obtain more refined optimality conditions, involving Lagrange multipliers. These multipliers facilitate the characterization of optimal solutions and often play an important role in computational methods. They are also central in duality theory, as will be discussed in Chapter 3.

In this chapter, we provide a new treatment of Lagrange multiplier theory. It is simple and it is also more powerful than the classical treatments. In particular, it establishes new necessary conditions that non-trivially supplement the classical Karush-Kuhn-Tucker conditions. It also deals effectively with problems where in addition to equality and inequality constraints, there is an additional abstract set constraint.

The starting point for our development is an enhanced set of necessary conditions of the Fritz John type, which is given in Section 2.2. In Section 2.3, we provide a classification of different types of Lagrange multipliers, and in Section 2.4, we introduce the notion of pseudonormality, which unifies and expands the conditions that guarantee the existence of Lagrange multipliers. In the final sections of the chapter, we discuss various topics related to Lagrange multipliers, including their connection with exact penalty functions.

2.1 INTRODUCTION TO LAGRANGE MULTIPLIERS

Let us provide some orientation by first considering a problem with equality constraints of the form

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h_i(x) = 0, & i = 1, \ldots, m.
\end{align*}
\]

We assume that \( f : \mathbb{R}^n \to \mathbb{R} \) and \( h_i : \mathbb{R}^n \to \mathbb{R}, \ i = 1, \ldots, m, \) are smooth functions. The basic Lagrange multiplier theorem for this problem states that, under appropriate assumptions, for a given local minimum \( x^* \), there exist scalars \( \lambda_1^*, \ldots, \lambda_m^* \), called Lagrange multipliers, such that

\[
\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.
\] (2.1)

These are \( n \) equations which together with the \( m \) constraints \( h_i(x^*) = 0 \), form a system of \( n + m \) equations with \( n + m \) unknowns, the vector \( x^* \) and the multipliers \( \lambda_i^* \). Thus, a constrained optimization problem can be “transformed” into a problem of solving a system of nonlinear equations.
Sec. 2.1 Introduction to Lagrange Multipliers

This is the traditional rationale for illustrating the importance of Lagrange multipliers.

To see why Lagrange multipliers may ordinarily be expected to exist, consider the case where the \( h_i \) are linear, so that

\[
h_i(x) = a_i'x - b_i, \quad i = 1, \ldots, m,
\]

and some vectors \( a_i \) and scalars \( b_i \). Then it can be seen that the tangent cone of the constraint set at the given local minimum \( x^* \) is

\[
T(x^*) = \{ y \mid a'_i y = 0, \ i = 1, \ldots, m \},
\]

and according to the necessary conditions of Chapter 1, we have

\[
-\nabla f(x^*) \in T(x^*)^\perp.
\]

Since \( T(x^*) \) is the nullspace of the \( m \times n \) matrix having as rows the \( a_i' \), \( T(x^*)^\perp \) is the range space of the matrix having as columns the \( a_i \). It follows that \(-\nabla f(x^*) \) can be expressed as a linear combination of the \( a_i \), so

\[
\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* a_i = 0
\]

for some scalars \( \lambda_i^* \).

In the general case where the \( h_i \) are nonlinear, the argument above would work if we could guarantee that the tangent cone \( T(x^*) \) can be represented as

\[
T(x^*) = \{ y \mid \nabla h_i(x^*)'y = 0, \ i = 1, \ldots, m \}. \quad (2.2)
\]

Unfortunately, additional assumptions are needed to guarantee the validity of this representation. One such assumption, called regularity of \( x^* \), is that the gradients \( \nabla h_i(x^*) \) are linearly independent; this will be shown in Section 2.4. Thus, when \( x^* \) is regular, there exist Lagrange multipliers. However, in general the equality (2.2) may fail, and there may not exist Lagrange multipliers, as shown by the following example.

**Example 2.1.1 (A Problem with no Lagrange Multipliers)**

Consider the problem of minimizing

\[
f(x) = x_1 + x_2
\]

subject to the two constraints

\[
h_1(x) = (x_1 + 1)^2 + x_2^2 - 1 = 0, \quad h_2(x) = (x_1 - 2)^2 + x_2^2 - 4 = 0.
\]
The geometry of this problem is illustrated in Fig. 2.1.1. It can be seen that at the local minimum \( x^* = (0, 0) \) (the only feasible solution), the cost gradient \( \nabla f(x^*) = (1, 1) \) cannot be expressed as a linear combination of \( \nabla h_1(x^*) \) and \( \nabla h_2(x^*) \). Thus the Lagrange multiplier condition

\[
\nabla f(x^*) + \lambda_1^* \nabla h_1(x^*) + \lambda_2^* \nabla h_2(x^*) = 0
\]

cannot hold for any \( \lambda_1^* \) and \( \lambda_2^* \). The difficulty here is that \( T(x^*) = \{0\} \), while the set \( \{ y \mid \nabla h_1(x^*)^T y = 0, \nabla h_2(x^*)^T y = 0 \} \) is equal to \( \{0, \gamma \mid \gamma \in \mathbb{R} \} \), so the characterization (2.2) fails.

**Fritz-John Conditions**

In summary, based on the preceding assertions, there are two possibilities at a local minimum \( x^* \):
Sec. 2.1 Introduction to Lagrange Multipliers

(a) The gradients $\nabla h_i(x^*)$ are linearly independent ($x^*$ is regular). Then, there exist scalars/Lagrange multipliers $\lambda_i^*$, $i = 1, \ldots, m$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.$$  

(b) The gradients $\nabla h_i(x^*)$ are linearly dependent, so there exist scalars $\lambda_1^*, \ldots, \lambda_m^*$, not all equal to 0, such that

$$\sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.$$  

These two possibilities can be lumped into a single condition: at a local minimum $x^*$ there exist scalars $\mu_0, \lambda_1^*, \ldots, \lambda_m^*$, not all equal to 0, such that $\mu_0 \geq 0$ and

$$\mu_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0. \quad (2.3)$$

Possibility (a) corresponds to the case where $\mu_0 > 0$, in which case the scalars $\lambda_i^* = \lambda_i / \mu_0$ are Lagrange multipliers. Possibility (b) corresponds to the case where $\mu_0 = 0$, in which case the condition (2.3) provides no information regarding the existence of Lagrange multipliers.

Necessary conditions that involve a nonnegative multiplier $\mu_0$ for the cost gradient, such as the one of Eq. (2.3), are known as Fritz John necessary conditions, and were first proposed in 1948 by John [Joh48]. These conditions can be extended to inequality-constrained problems, and they hold without any further assumptions on $x^*$ (such as regularity). However, this extra generality comes at a price, because the issue of whether the cost multiplier $\mu_0$ can be taken to be positive is left unresolved. Unfortunately, asserting that $\mu_0 > 0$ is nontrivial under some commonly used assumptions, and for this reason, traditionally Fritz John conditions in their classical form have played a somewhat peripheral role in the development of Lagrange multiplier theory. Nonetheless, the Fritz John conditions, when properly strengthened, can provide a simple and powerful line of analysis, as we will see in the next section.

**Sensitivity**

Lagrange multipliers frequently have an interesting interpretation in specific practical contexts. In economic applications they can often be interpreted as prices, while in other problems they represent quantities with concrete physical meaning. It turns out that within our mathematical framework, they can be viewed as rates of change of the optimal cost as
the level of constraint changes. This is fairly easy to show for the case of linear and independent constraints, as indicated in Fig. 2.1.2.

When the constraints are nonlinear, the sensitivity interpretation of Lagrange multipliers is valid, provided some assumptions are satisfied. Typically, these assumptions include the linear independence of the constraint gradients, but also additional conditions involving second derivatives (see e.g., the textbook [Ber99]).

### Inequality Constraints

The preceding discussion can be extended to the case where there are both equality and inequality constraints. Consider for example the case of linear inequality constraints:

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad a_j^T x \leq b_j, \quad j = 1, \ldots, r.
\end{align*}
\]

The tangent cone of the constraint set at a local minimum \( x^* \) is given by

\[
T(x^*) = \{ y \mid a_j^T y \leq 0, \quad j \in A(x^*) \},
\]

where \( A(x^*) \) is the set of indices of the constraints satisfied as equations at \( x^* \),

\[
A(x^*) = \{ j \mid a_j^T x^* = 0 \}.
\]

The necessary condition \(-\nabla f(x^*) \in T(x^*)^\perp\) can be shown to be equivalent to the existence of scalars \( \mu_j^* \), such that

\[
\nabla f(x^*) + \sum_{i=1}^m \mu_j^* a_j = 0,
\]

\[
\mu_j^* \geq 0, \quad \forall \ j = 1, \ldots, r, \quad \mu_j^* = 0, \quad \forall \ j \in A(x^*).
\]

This follows from the characterization of the cone \( T(x^*)^\perp \) as the cone generated by the vectors \( a_j, \ j \in A(x^*) \) (Farkas’ lemma). The above necessary conditions, properly generalized for nonlinear constraints, are known as the Karush-Kuhn-Tucker conditions.

Similar to the equality-constrained case, the typical method to ascertain existence of Lagrange multipliers for the nonlinearly-constrained general problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h_i(x) = 0, \quad i = 1, \ldots, m, \quad g_j(x) \leq 0, \quad j = 1, \ldots, r,
\end{align*}
\]

is to assume structure which guarantees that the tangent cone has the form

\[
T(x^*) = \{ y \mid \nabla h_i(x^*)^T y = 0, \quad i = 1, \ldots, m, \ \nabla g_j(x^*)^T y \leq 0, \quad j \in A(x^*) \};
\]

(2.4)
Sec. 2.1 Introduction to Lagrange Multipliers

\[ \nabla f(x^*) \quad x^* + \Delta x \quad a'x = b + \Delta b \]

**Figure 2.1.2.** Illustration of the sensitivity theorem for a problem involving a single linear constraint,

\[
\text{minimize } f(x) \\
\text{subject to } a'x = b.
\]

Here, \( x^* \) is a local minimum and \( \lambda^* \) is a corresponding Lagrange multiplier. If the level of constraint \( b \) is changed to \( b + \Delta b \), the minimum \( x^* \) will change to \( x^* + \Delta x \). Since \( b + \Delta b = a'(x^* + \Delta x) = a'x^* + a'\Delta x = b + a'\Delta x \), we see that the variations \( \Delta x \) and \( \Delta b \) are related by

\[ a'\Delta x = \Delta b. \]

Using the Lagrange multiplier condition \( \nabla f(x^*) = -\lambda^*a \), the corresponding cost change can be written as

\[ \Delta \text{cost} = f(x^* + \Delta x) - f(x^*) = \nabla f(x^*)'\Delta x + o(\|\Delta x\|) = -\lambda^*a'\Delta x + o(\|\Delta x\|). \]

By combining the above two relations, we obtain \( \Delta \text{cost} = -\lambda^*\Delta b + o(\|\Delta x\|) \), so up to first order we have

\[ \lambda^* = -\frac{\Delta \text{cost}}{\Delta b}. \]

Thus, the Lagrange multiplier \( \lambda^* \) gives the rate of optimal cost decrease as the level of constraint increases.

In the case where there are multiple constraints \( a'_i x = b_i, i = 1, \ldots, m \), the preceding argument can be appropriately modified. In particular, for any set for changes \( \Delta b_i \) for which the system \( a'_i x = b_i + \Delta b_i \) has a solution \( x + \Delta x \), we have

\[ \Delta \text{cost} = f(x^* + \Delta x) - f(x^*) \\
= \nabla f(x^*)'\Delta x + o(\|\Delta x\|) \\
= -\sum_{i=1}^{m} \lambda^*_i a'_i \Delta x + o(\|\Delta x\|), \]

and \( a'_i \Delta x = \Delta b_i \) for all \( i \), so we obtain \( \Delta \text{cost} = -\sum_{i=1}^{m} \lambda^*_i \Delta b_i + o(\|\Delta x\|) \).
a condition known as quasiregularity. This is the classical line of development of Lagrange multiplier theory: conditions implying quasiregularity (known as constraint qualifications) are established, and the existence of Lagrange multipliers is then inferred using Farkas’ lemma. This line of analysis is insightful and intuitive, but has traditionally required fairly complex proofs to show that specific constraint qualifications imply the quasiregularity condition (2.5). A more serious difficulty, however, is that the analysis based on quasiregularity does not extend to the case where, in addition to the equality and inequality constraints, there is an additional abstract set constraint $x \in X$.

**Uniqueness of Lagrange Multipliers**

For a given local minimum, the set of Lagrange multiplier vectors, call it $M$, may contain more than one element. Indeed, if $M$ is nonempty, it contains either one or an infinite number of elements. For example, for the case of equality constraints only, $M$ contains a single element if and only if the gradients $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent. Generally, if $M$ is nonempty, it is a linear manifold, which is a translation of the nullspace of the $n \times m$ matrix

$$
\nabla h(x^*) = \begin{bmatrix} \nabla h_1(x^*) & \cdots & \nabla h_m(x^*) \end{bmatrix}.
$$

For the mixed equality/inequality constrained case of problem (2.4), it can be seen that the set of Lagrange multipliers

$$
M = \left\{ (\lambda, \mu) \mid \mu \geq 0, \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) = 0 \right\}
$$

is a polyhedral set, which (if nonempty) may be bounded or unbounded.

What kind of sensitivity information do Lagrange multipliers convey when they are not unique? The classical theory does not offer satisfactory answers to this question, as indicated by the following example.

**Example 2.1.2**

Assume that $f$ and the $g_j$ are linear functions,

$$
f(x) = c^T x, \quad g_j(x) = a_j^T x - b_j, \quad j = 1, \ldots, r,
$$

and for simplicity assume that there are no equality constraints and that all the inequality constraints are active at a minimum $x^*$. The Lagrange multiplier vectors are the $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ such that $\mu^* \geq 0$ and

$$
c + \mu_1^* a_1 + \cdots + \mu_r^* a_r = 0.
$$
In a naive extension of the traditional sensitivity interpretation, a positive (or zero) multiplier $\mu_j$ would indicate that the $j$th inequality is “significant” (or “insignificant”) in that the cost can be improved (or cannot be improved) through its violation. Figure 2.1.3, however, indicates that for each $j$, one can find a Lagrange multiplier vector where $\mu_j > 0$ and another Lagrange multiplier vector where $\mu_j = 0$. Furthermore, one can find Lagrange multiplier vectors where the signs of the multipliers are “wrong” in the sense that it is impossible to improve the cost while violating only constraints with positive multipliers. Thus the traditional sensitivity interpretation fails generically when there is multiplicity of Lagrange multipliers.

Exact Penalty Functions

An important analytical and algorithmic technique in nonlinear programming involves the use of penalty functions, whereby the equality and inequality constraints are discarded and are replaced by additional terms in the cost function that penalize their violation. An important example for the mixed equality and inequality constraint problem (2.4) is the quadratic penalty function

$$Q_c(x) = f(x) + \frac{c}{2} \left( \sum_{i=1}^{m} |h_i(x)|^2 + \sum_{j=1}^{r} (g^+_j(x))^2 \right),$$

where $c$ is a positive penalty parameter, and we use the notation

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

We may expect that by minimizing $Q_{c,k}(x)$ over $X$ for a sequence $\{c^k\}$ of penalty parameters with $c^k \to \infty$, we will obtain in the limit a solution of the original problem. Indeed, convergence of this type can generically be shown, and it turns out that typically a Lagrange multiplier vector can also be simultaneously obtained (assuming such a vector exists); see e.g., [Ber99]. We will use these convergence ideas in various proofs, starting with the next section.

The quadratic penalty function is not exact in the sense that a local minimum $x^*$ of the constrained minimization problem is typically not a local minimum of $Q_c(x)$ for any value of $c$. A different type of penalty function is given by

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

where $c$ is a positive penalty parameter. It can be shown that $x^*$ is typically a local minimum of $F_c$, provided $c$ is larger than some threshold value. The
Lagrange Multipliers Chap. 2

Hyperplane with normal $y$

Figure 2.1.3. Illustration of Lagrange multipliers for the case of a two-dimensional problem with linear cost and four linear inequality constraints (cf. Example 2.1.2). We assume that $-c$ and $a_1, \ldots, a_4$ lie in the same halfspace, and $-c$ lies “between” $a_1$ and $a_2$ on one side and $a_3$ and $a_4$ on the other side, as shown in the figure. For $(\mu_1, \ldots, \mu_4)$ to be a Lagrange multiplier vector, we must have $\mu_j \geq 0$ for all $j$, and

$$-c = \mu_1 a_1 + \mu_2 a_2 + \mu_3 a_3 + \mu_4 a_4.$$ 

There exist exactly four Lagrange multiplier vectors with two positive multipliers/components and the other two components equal to zero. It can be seen that it is impossible to improve the cost by moving away from $x^*$, while violating the constraints with positive multipliers but not violating any constraints with zero multipliers. For example, when $\mu_1 = 0, \mu_2 > 0, \mu_3 > 0, \mu_4 = 0$, moving from $x^*$ along a direction $y$ such that $c^' y < 0$ will violate either the 1st or the 4th inequality constraint. To see this, note that to improve the cost while violating only the constraints with positive multipliers, the direction $y$ must be the normal of a hyperplane that contains the vectors $-c$, $a_2$, and $a_3$ in one of its halfspaces, and the vectors $a_1$ and $a_4$ in the complementary subspace (see the figure).

There also exist an infinite number of Lagrange multiplier vectors where all four multipliers are positive, or any three multipliers are positive and the fourth is zero. For any one of these vectors, it is possible to find a direction $y$ such that $c^' y < 0$ and moving from $x^*$ in the direction of $y$ will violate only the constraints that have positive multipliers.

Conditions under which this is true and the threshold value for $c$ bear an intimate connection with Lagrange multipliers. Figure 2.1.4 illustrates this connection for the case of a problem with a single inequality constraint. In particular, it can be seen that usually (though not always) if the penalty parameter exceeds the value of a Lagrange multiplier, $x^*$ is also a local
Sec. 2.1 Introduction to Lagrange Multipliers

Our treatment of Lagrange multipliers

Our development of Lagrange multiplier theory in this chapter differs from the classical treatment in a number of ways:

(a) The optimality conditions of the Lagrange multiplier type that we will develop are sharper than the classical Kuhn-Tucker conditions (they include extra conditions). They are also more general in that they apply when in addition to the equality and inequality constraints, there is an additional abstract set constraint.

(b) We will simplify the proofs of various Lagrange multiplier theorems by introducing the notion of pseudonormality, which serves as a connecting link between the major constraint qualifications and quasiregularity. This analysis carries through even in the case of an additional set constraint, where the classical proof arguments based on quasiregularity fail.
(c) We will see in Section 2.3 that there may be several different types of Lagrange multipliers for a given problem, which can be characterized in terms of their sensitivity properties and the information they provide regarding the significance of the corresponding constraints. We show that one particular Lagrange multiplier vector, the one that has minimum norm, has nice sensitivity properties in that it characterizes the direction of steepest rate of improvement of the cost function for a given level of the norm of the constraint violation. Along that direction, the equality and inequality constraints are violated consistently with the signs of the corresponding multipliers.

(d) We will make a connection between pseudonormality, the existence of Lagrange multipliers, and the minimization over $X$ of the penalty function

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

where $c$ is a positive penalty parameter. In particular, we will show in Section 2.5 that pseudonormality implies that local minima of the general nonlinear programming problem (2.4) are also local minima (over just $X$) of $F$, when $c$ is large enough. We will also show that this in turn implies the existence of Lagrange multipliers.

Much of our analysis is based on an enhanced set of Fritz John necessary conditions that are introduced in the next section.

2.2 ENHANCED FRITZ JOHN OPTIMALITY CONDITIONS

We will develop general optimality conditions for optimization problems of the form

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in C,
\end{align*}$$

(2.6)

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

$$C = X \cap \{ x \mid h_1(x) = 0, \ldots, h_m(x) = 0 \} \cap \{ x \mid g_1(x) \leq 0, \ldots, g_r(x) \leq 0 \}.$$  (2.7)

Except for the last section, we assume in this chapter that $f, h_i, g_j$ are smooth functions from $\mathbb{R}^n$ to $\mathbb{R}$, and $X$ is a nonempty closed set.

We will use frequently the tangent and normal cone of $X$ at any $x \in X$, which are denoted by $T_X(x)$ and $N_X(x)$, respectively. We recall from Section 1.5 that $X$ is said to be regular at a point $x \in X$ if $N_X(x) = T_X(x)^\perp$. Furthermore, if $X$ is convex, then it is regular at all $x \in X$. 
Definition 2.2.1: We say that the constraint set $C$, as represented by Eq. (2.7), 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 (2.6) there exist vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ that satisfy the following conditions:

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

$$
\mu_j^* \geq 0, \quad \forall \ j = 1, \ldots, r, \quad (2.8)
$$

$$
\mu_j^* = 0, \quad \forall \ j \notin A(x^*), \quad (2.9)
$$

$$
\text{where } A(x^*) = \{ j \mid g_j(x^*) = 0 \} \text{ is the index set of inequality constraints that are active at } x^*. \text{ A pair } (\lambda^*, \mu^*) \text{ satisfying Eqs. (2.8)-(2.10) is called a Lagrange multiplier vector corresponding to } f \text{ and } x^*. \text{}
$$

When there is no danger of confusion, we refer to $(\lambda^*, \mu^*)$ simply as a Lagrange multiplier vector or a Lagrange multiplier, without reference to the corresponding local minimum $x^*$ and the function $f$. Figure 2.2.1 illustrates the definition of a Lagrange multiplier. Condition (2.10) is referred to as the complementary slackness condition (CS for short). Note that from Eq. (2.8), it follows that the set of Lagrange multiplier vectors corresponding to a given $f$ and $x^*$ is the intersection of a collection of closed half spaces [one for each $y \in T_X(x^*)$], so it is a (possibly empty) closed and convex set.

The condition (2.8), referred to as the Lagrangian stationarity condition, is consistent with the Lagrange multiplier theory outlined in the preceding section for the case where $X = \mathbb{R}^n$. It can be viewed as the necessary condition for $x^*$ to be a local minimum of the Lagrangian function

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

over $x \in X$. When $X$ is a convex set, Eq. (2.8) is equivalent to

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

(2.11)

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
Figure 2.2.1. Illustration of a Lagrange multiplier for the case of a single inequality constraint and the spherical set $X$ shown in the figure. The tangent cone $T_X(x^*)$ is a halfspace and its polar $T_X(x^*)_\perp$ is the halfline shown in the figure. There is a unique Lagrange multiplier $\mu^*$, and it is such that $-\nabla_x L(x^*, \mu^*)$ belongs to $T_X(x^*)_\perp$. Note that for this example, we have $\mu^* > 0$, and that there is a sequence $\{x^k\} \subset X$ that converges to $x^*$ and is such that $f(x^k) < f(x^*)$ and $g(x^k) > 0$ for all $k$, consistently with condition (iv) of the subsequent Prop. 2.2.1.

of the form $\alpha(x - x^*)$, where $\alpha > 0$ and $x \in X$. If $X = \mathbb{R}^n$, Eq. (2.11) 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 (2.9) and the CS condition (2.10), comprise the classical Karush-Kuhn-Tucker conditions.

The following proposition is central in the line of development of this chapter. It enhances the classical Fritz John optimality conditions, and forms the basis for enhancing the classical Karush-Kuhn-Tucker conditions as well. The proposition asserts that there exist multipliers corresponding to a local minimum $x^*$, including a multiplier $\mu_0^*$ for the cost function gradient. These multipliers have standard properties [conditions (i)-(iii) below], but they also have a special nonstandard property [condition (iv) below]. This condition asserts that by violating the constraints corresponding to nonzero multipliers, we can improve the optimal cost (the remaining constraints, may also need to be violated, but the degree of their violation is arbitrarily small relative to the other constraints).
Proposition 2.2.1: Let $x^*$ be a local minimum of problem (2.6)-(2.7). Then there exist scalars $\mu_0^*, \lambda_1^*, \ldots, \lambda_m^*$, and $\mu_1^*, \ldots, \mu_r^*$, satisfying the following conditions:

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

(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

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

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

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

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

where

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

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

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

subject to $x \in X \cap S$,

where we denote

$$g^+(x) = \max\{0, g_j(x)\}, \quad S = \{x \mid \|x - x^*\| \leq \epsilon\},$$

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 $k$, $F^k(x^k) \leq F^k(x^*)$, which is can be written as

$$f(x^k) + \frac{k}{2} \sum_{i=1}^m (h_i(x^k))^2 + \frac{k}{2} \sum_{j=1}^r (g_j^+(x^k))^2 + \frac{1}{2} \|x^k - x^*\|^2 \leq f(x^*).$$
Since \( f(x^k) \) is bounded over \( X \cap S \), we obtain
\[
\lim_{k \to \infty} |h_i(x^k)| = 0, \quad i = 1, \ldots, m, \quad \lim_{k \to \infty} g_j^+(x^k) = 0, \quad j = 1, \ldots, r;
\]
otherwise the left-hand side of Eq. (2.15) would become unbounded from above as \( k \to \infty \). Therefore, every limit point \( \overline{x} \) of \( \{x^k\} \) is feasible, i.e., \( \overline{x} \in C \). Furthermore, Eq. (2.15) yields \( f(x^k) + (1/2)||x^k - x^*||^2 \leq f(x^*) \) for all \( k \), so by taking the limit as \( k \to \infty \), we obtain
\[
f(\overline{x}) + \frac{1}{2}||\overline{x} - x^*||^2 \leq f(x^*).
\]
Since \( \overline{x} \in S \) and \( \overline{x} \) is feasible, we have \( f(x^*) \leq f(\overline{x}) \), which when combined with the preceding inequality yields \( ||\overline{x} - x^*|| = 0 \) so that \( \overline{x} = x^* \). Thus the sequence \( \{x^k\} \) converges to \( x^* \), and it follows that \( x^k \) is an interior point of the closed sphere \( S \) for all \( k \) greater than some \( \overline{k} \).

For \( k \geq \overline{k} \), we have the necessary optimality condition \( -\nabla F^k(x^k) \in T_X(x^k)^\perp \), which [by using the formula \( \nabla (g_j^+(x))^2 = 2g_j^+(x)\nabla g_j(x) \)] is written as
\[
- \left( \nabla f(x^k) + \sum_{i=1}^m \xi_i^k \nabla h_i(x^k) + \sum_{j=1}^r \zeta_j^k \nabla g_j(x^k) + (x^k - x^*) \right) \in T_X(x^k)^\perp, \tag{2.16}
\]
where
\[
\xi_i^k = kh_i(x^k), \quad \zeta_j^k = kg_j^+(x^k). \tag{2.17}
\]
Denote
\[
\delta^k = \sqrt{1 + \sum_{i=1}^m (\xi_i^k)^2 + \sum_{j=1}^r (\zeta_j^k)^2}, \tag{2.18}
\]
\[
\mu_0^k = \frac{1}{\delta^k}, \quad \lambda_i^k = \frac{\xi_i^k}{\delta^k}, \quad i = 1, \ldots, m, \quad \mu_j^k = \frac{\zeta_j^k}{\delta^k}, \quad j = 1, \ldots, r. \tag{2.19}
\]
Then by dividing Eq. (2.16) with \( \delta^k \), we obtain
\[
- \left( \mu_0^k \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^r \mu_j^k \nabla g_j(x^k) + \frac{1}{\delta^k}(x^k - x^*) \right) \in T_X(x^k)^\perp \tag{2.20}
\]
Since by construction we have
\[
(\mu_0^k)^2 + \sum_{i=1}^m (\lambda_i^k)^2 + \sum_{j=1}^r (\mu_j^k)^2 = 1, \tag{2.21}
\]
the sequence \( \{ \mu_0^k, \lambda_1^k, \ldots, \lambda_n^k, \mu_1^k, \ldots, \mu_k^k \} \) is bounded and must contain a subsequence that converges to some limit \( \{ \mu_0^*, \lambda_1^*, \ldots, \lambda_n^*, \mu_1^*, \ldots, \mu_k^* \} \).

From Eq. (2.20) and the defining property of the normal cone \( N_X(x^*) \)
\( [x^k \to x^*, z^k \to z^*] \), and \( z^k \in T_X(x^k) \perp \) for all \( k \), imply that \( z^* \in N_X(x^*) \), we see that \( \mu_0^*, \lambda_1^*, \) and \( \mu_k^* \) must satisfy condition (i). From Eqs. (2.17) and (2.19), \( \mu_0^* \) and \( \mu_j^* \) must satisfy condition (ii), and from Eq. (2.21), \( \mu_0^*, \lambda_1^*, \) and \( \mu_k^* \) must satisfy condition (iii). Finally, to show that condition (iv) is satisfied, assume that \( I \cup J \) is nonempty [otherwise, by conditions (i) and (iii), we are done], and note that for all sufficiently large \( k \) within the index set \( K \) of the convergent subsequence, we must have \( \lambda_i^k \lambda_j^k > 0 \) for all \( i \in I \) and \( \mu_j^k \mu_j^k > 0 \) for all \( j \in J \). Therefore, for these \( k \), from Eqs. (2.17) and (2.19), we must have \( \lambda_i^k \lambda_j^k h_i(x^k) > 0 \) for all \( i \in I \) and \( \mu_j^k g_j(x^k) > 0 \) for all \( j \in J \), while from Eq. (2.15), 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

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

and

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

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

Note that condition (iv) of Prop. 2.2.1 is stronger than the CS condition (2.10). [If \( \mu_j^* > 0 \), then according to condition (iv), the corresponding \( j \)th inequality constraint must be violated arbitrarily close to \( x^* \) [cf. Eq. (2.12)], implying that \( g_j(x^*) = 0 \).] For ease of reference, we refer to condition (iv) as the complementarity violation condition (CV for short).\( \dagger \) This condition can be visualized in the examples of Fig. 2.2.2. It will turn out to be of crucial significance in the following sections.

If \( X \) is regular at \( x^* \), i.e., \( N_X(x^*) = T_X(x^*) \perp \), condition (i) of Prop. 2.2.1 becomes

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

or equivalently

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

\( \dagger \) 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 \( x^* \).
Lagrange Multipliers  Chap. 2

If in addition, the scalar \( \mu_0^* \) can be shown to be strictly positive, then by normalization we can choose \( \mu_0^* = 1 \), and condition (i) of Prop. 2.2.1 becomes equivalent to the Lagrangian stationarity condition (2.8). 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, which satisfies the CV condition (a stronger condition than the CS condition, as mentioned above).

As an example, suppose that there is no abstract set constraint (\( X = \mathbb{R}^n \)), and the gradients \( \nabla h_i(x^*) \), \( i = 1, \ldots, m \), and \( \nabla g_j(x^*) \), \( j \in A(x^*) \) are linearly independent, then we cannot have \( \mu_0^* = 0 \), since then condition (i) of Prop. 2.2.1 would be violated. It follows that there exists a Lagrange multiplier vector, which in this case is unique in view of the linear independence assumption. We thus obtain the Lagrange multiplier theorem alluded to in Section 2.1. This is a classical result, found in almost all nonlinear programming textbooks, but it is obtained here in a stronger form, which includes the assertion that the multipliers satisfy the CV condition in place of the CS condition.

To illustrate the use of the generalized Fritz John conditions of Prop. 2.2.1 and the CV condition in particular, consider the following example.

**Example 2.2.1**

Suppose that we convert the problem \( \min_{h(x) = 0} f(x) \), involving a single equal-
Sec. 2.3  Informative Lagrange Multipliers

ity constraint, to the inequality constrained problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h(x) \leq 0, \quad -h(x) \leq 0.
\end{align*}
\]

The Fritz John conditions, in their classical form, assert the existence of nonnegative \( \mu^0, \lambda^+, \lambda^- \), not all zero, such that

\[
\mu^0 \nabla f(x^*) + \lambda^+ \nabla h(x^*) - \lambda^- \nabla h(x^*) = 0. \tag{2.22}
\]

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. 2.2.1, 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. (2.22), so it follows that \( \mu^0 > 0 \). Thus, by dividing Eq. (2.22) by \( \mu^0 \), we recover the familiar first order condition \( \nabla f(x^*) + \lambda^* \nabla h(x^*) = 0 \) with \( \lambda^* = (\lambda^+ - \lambda^-)/\mu^0 \), under the assumption \( \nabla h(x^*) \neq 0 \). Note that this deduction would not have been possible without the CV condition.

We will explore further the CV condition as a vehicle for characterizing Lagrange multipliers in the next section. Subsequently, in Section 1.4, we will derive conditions that guarantee that one can take \( \mu^0 = 1 \) in Prop. 2.2.1, by using the notion of pseudonormality.

---

**EXERCISES**

2.2.1 (Lagrange Multipliers Under Convexity)

Consider problem (2.6) assuming that \( X \) is convex, the functions \( f, g_1, \ldots, g_r \) are convex, and the functions \( h_1, \ldots, h_m \) are linear.

(a) Show that \( x^* \) is an optimal solution and \( (\lambda^*, \mu^*) \) is a Lagrange multiplier associated with \( x^* \) if and only if \( x^* \) is feasible, \( \mu^* \geq 0, \mu^* g_j(x^*) = 0 \) for all \( j \), and \( f(x^*) = \min_{x \in X} L(x, \lambda^*, \mu^*) \), where \( L \) is the Lagrangian function.

(b) Show that all the optimal solutions of the problem have the same set of Lagrange multipliers.
2.3 INFORMATIVE LAGRANGE MULTIPLIERS

The Lagrange multipliers whose existence is guaranteed by Prop. 2.2.1 (assuming that $\mu^*_0 = 1$) are of a special type: they satisfy the stronger CV condition in place of the CS condition. These multipliers provide a significant amount of sensitivity information by in effect indicating which constraints to violate in order to effect a cost reduction. In view of this interpretation, we refer to a Lagrange multiplier vector $(\lambda^*, \mu^*)$ that satisfies, in addition to Eqs. (2.8)-(2.10), the CV condition [condition (iv) of Prop. 2.2.1] as being informative.

The salient property of informative Lagrange multipliers is consistent with the classical sensitivity interpretation of a Lagrange multiplier as the rate of cost reduction when 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, at least one informative Lagrange multiplier vector exists if $X$ is regular and we can guarantee that we can take $\mu^*_0 = 1$ in Prop. 2.2.1. In fact we will show shortly a stronger and more definitive property: if $X$ is regular and there exists at least one Lagrange multiplier vector, there exists one that is 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, as we will show in Chapter 3, 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 discarding constraints is of analytical or computational interest and constraints with 0 Lagrange multipliers can be discarded as indicated above, we are motivated to find 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. Another interesting example is provided by Fig. 2.1.3. There are four minimal Lagrange multiplier vectors in this example, and each of them has two nonzero components. However, none of these multipliers is informative. In fact all the other Lagrange
multiplier vectors, those involving three out of four, or all four components positive, are informative. Thus, in the example of Fig. 2.1.3, the sets of minimal and informative Lagrange multipliers are nonempty and disjoint.

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. 2.3.1. In particular, let us say that a Lagrange multiplier \( (\lambda^*, \mu^*) \) is strong if in addition to Eqs. (2.8)-(2.10), it satisfies the condition

\[
(iv') \quad \text{If the set } I \cup J \text{ is nonempty, where } I = \{ i \mid \lambda^*_i \neq 0 \} \text{ and } J = \{ j \neq 0 \mid \mu^*_j > 0 \}, \text{ there exists a sequence } \{ x^k \} \subset X \text{ that converges to } x^* \text{ and is such that for all } k, \\
f(x^k) < f(x^*), \quad \lambda^*_i h_i(x^k) > 0, \quad \forall \ i \in I, \quad \mu^*_j g_j(x^k) > 0, \quad \forall \ j \in J. 
\]

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. (2.13) and (2.14). As a result, informative Lagrange multipliers are also strong, but not reversely.

The following proposition, illustrated in Fig. 2.3.1, clarifies the relationships between different types of Lagrange multipliers. It requires that the tangent cone \( T_X(x^*) \) be convex, which is true in particular if \( X \) is regular. In fact it turns out that regularity of \( X \) at \( x^* \) is sufficient to guarantee that \( T_X(x^*) \) is convex, although we have not proved this fact here (see Rockafellar and Wets [RoW98]).

![Figure 2.3.1. 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 \( x^* \)).](image)

**Proposition 2.3.1:** Let \( x^* \) be a local minimum of problem (2.6)-(2.7). Assume that the tangent cone \( T_X(x^*) \) is convex and that the set of Lagrange multipliers is nonempty. Then:
(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.

Lemma 2.3.1: Let $N$ be a closed convex cone in $\mathbb{R}^n$, and let $a_0, \ldots, a_r$ be given vectors in $\mathbb{R}^n$. Suppose that the closed and convex set $M \subset \mathbb{R}^r$ given by

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

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

$$a'_0 d^k \to -\|\mu^*\|^2,$$

$$\left(a'_j d^k\right)^+ \to \mu^*,$$ \hspace{1em} \(j = 1, \ldots, r,\)

where $\mu^*$ is the vector of minimum norm in $M$ and we use the notation $(a'_j d^k)^+ = \max\{0, a'_j d^k\}$. Furthermore, we have

$$-\frac{1}{2} \|\mu^*\|^2 = \inf_{d \in N^\perp} \left\{ a'_0 d + \frac{1}{2} \sum_{j=1}^{r} \left( (a'_j d)^+ \right)^2 \right\}$$

$$= \lim_{k \to \infty} \left\{ a'_0 d^k + \frac{1}{2} \sum_{j=1}^{r} \left( (a'_j d^k)^+ \right)^2 \right\}.$$ \hspace{1em} (2.26)

In addition, if the problem

$$\text{minimize } a'_0 d + \frac{1}{2} \sum_{j=1}^{r} \left( (a'_j d)^+ \right)^2$$

subject to \(d \in N^\perp,\)

$$\text{subject to } d \in N^\perp,$$ \hspace{1em} (2.27)
has an optimal solution, denoted \(d^*\), we have
\[
a_0'd^* = -\|\mu^*\|^2, \quad (a_j'd^*)^+ = \mu_j^*, \quad j = 1, \ldots, r.
\] (2.28)

**Proof:** For any \(\gamma \geq 0\), consider the function
\[
L_\gamma(d, \mu) = \left( a_0 + \sum_{j=1}^r \mu_j a_j \right)'d + \frac{\gamma}{2} \|d\|^2 - \frac{1}{2} \|\mu\|^2.
\] (2.29)

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 \(\gamma\)-perturbed and coercive version \(L_\gamma\) for \(\gamma > 0\), and then take the limit as \(\gamma \to 0\).

From the saddle point theorem of Section 1.4, for all \(\gamma > 0\), the coercive convex/concave quadratic function \(L_\gamma\) has a saddle point, denoted \((d_\gamma, \mu_\gamma)\), over \(d \in N^\perp\) and \(\mu \geq 0\). This saddle point is unique and can be easily characterized, taking advantage of the quadratic nature of \(L_\gamma\). In particular, the maximization over \(\mu \geq 0\) when \(d = d_\gamma\) yields
\[
\mu_j^\gamma = (a_j'd_\gamma^\gamma)^+ + \frac{1}{2} \|\mu_j^\gamma\|^2, \quad j = 1, \ldots, r
\] (2.30)
[to maximize \(\mu_j a_j (a_j'd_\gamma^\gamma) - \frac{1}{2} \|\mu_j^\gamma\|^2\) subject to the constraint \(\mu_j \geq 0\), we calculate the unconstrained maximum, which is \(a_j'd_\gamma^\gamma\), and if it is negative we set it to 0, so that the maximum subject to \(\mu_j \geq 0\) is attained for \(\mu_j = (a_j'd_\gamma^\gamma)^+\).]

The minimization over \(d \in N^\perp\) when \(\mu = \mu^\gamma\) is equivalent to projecting the vector
\[
s^\gamma = -\frac{a_0 + \sum_{j=1}^r \mu_j^\gamma a_j}{\gamma}
\]
on the cone \(N^\perp\), since by adding and subtracting \(-\frac{\gamma}{2} \|s^\gamma\|^2\), we can write \(L_\gamma(d, \mu^\gamma)\) as
\[
L_\gamma(d, \mu^\gamma) = \frac{\gamma}{2} \|d - s^\gamma\|^2 - \frac{\gamma}{2} \|s^\gamma\|^2 - \frac{1}{2} \|\mu^\gamma\|^2.
\] (2.31)

It follows that
\[
d^\gamma = P_{N^\perp}(s^\gamma),
\] (2.32)
where \(P_{N^\perp}(\cdot)\) denotes projection on \(N^\perp\). Furthermore, using the projection theorem and the fact that \(N^\perp\) is a cone, we have
\[
\|s^\gamma\|^2 - \|d^\gamma - s^\gamma\|^2 = \|d^\gamma\|^2 = \|P_{N^\perp}(s^\gamma)\|^2,
\]
so Eq. (2.31) yields
\[
L_\gamma(d^\gamma, \mu^\gamma) = -\frac{\gamma}{2} \|P_{N^\perp}(s^\gamma)\|^2 - \frac{1}{2} \|\mu^\gamma\|^2,
\] (2.33)
or equivalently using the definition of $s_\gamma$:

$$L_\gamma(d\gamma, \mu\gamma) = -\frac{\|P_{N^\perp} \left( -\left( a_0 + \sum_{j=1}^r \mu_j^\gamma a_j \right) \right) \|^2}{2\gamma} - \frac{1}{2}\|\mu\gamma\|^2. \quad (2.34)$$

We will use the preceding facts to show that as $\gamma \to 0$, $\mu\gamma \to \mu^*$, while $d\gamma$ yields the desired sequence $d_k$, which satisfies Eqs. (2.24)-(2.26). Furthermore, we will show that if $d^*$ is an optimal solution of problem (2.27), then $(d^*, \mu^*)$ is a saddle point of the function $L_0$.

We note that

$$\inf_{d \in N^\perp} L_0(d, \mu) = \begin{cases} -\frac{1}{2}\|\mu\|^2 & \text{if } \mu \in M, \\ -\infty & \text{otherwise}, \end{cases} \quad (2.35)$$

so since $\mu^*$ is the vector of minimum norm in $M$, we obtain for all $\gamma > 0$,

$$-\frac{1}{2}\|\mu^*\|^2 = \sup_{\mu \geq 0} \inf_{d \in N^\perp} L_0(d, \mu) \leq \inf_{d \in N^\perp} \sup_{\mu \geq 0} L_0(d, \mu) \leq \inf_{d \in N^\perp} \sup_{\mu \geq 0} L_\gamma(d, \mu) = L_\gamma(d^*, \mu^*).$$

Combining Eqs. (2.34) and (2.35), we obtain

$$L_\gamma(d^*, \mu^*) = -\frac{\|P_{N^\perp} \left( -\left( a_0 + \sum_{j=1}^r \mu_j^* a_j \right) \right) \|^2}{2\gamma} - \frac{1}{2}\|\mu^*\|^2 \geq -\frac{1}{2}\|\mu^*\|^2. \quad (2.36)$$

From this, we see that $\|\mu\gamma\| \leq \|\mu^*\|$, so that $\mu\gamma$ remains bounded as $\gamma \to 0$. By taking the limit above as $\gamma \to 0$, we see that

$$\lim_{\gamma \to 0} P_{N^\perp} \left( -\left( a_0 + \sum_{j=1}^r \mu_j^\gamma a_j \right) \right) = 0,$$

so (by the continuity of the projection operation) any limit point of $\mu\gamma$, call it $\overline{\mu}$, satisfies $P_{N^\perp} \left( -\left( a_0 + \sum_{j=1}^r \overline{\mu}_j a_j \right) \right) = 0$, or $-\left( a_0 + \sum_{j=1}^r \overline{\mu}_j a_j \right) \in N$. Since $\mu^\gamma \geq 0$, it follows that $\overline{\mu} \geq 0$, so $\overline{\mu} \in M$. We also have $\|\overline{\mu}\| \leq \|\mu^\gamma\|$ (since $\|\mu\gamma\| \leq \|\mu^*\|$), so by using the minimum norm property of $\mu^*$, we conclude that any limit point $\overline{\mu}$ of $\mu\gamma$ must be equal to $\mu^*$. Thus, $\mu\gamma \to \mu^*$. From Eq. (2.36) we then obtain

$$L_\gamma(d^*, \mu^*) \to -\frac{1}{2}\|\mu^*\|^2, \quad (2.37)$$
while from Eqs. (2.32) and (2.33), we have
\[ \frac{\gamma}{2} \|d^\gamma\|^2 \to 0. \] (2.38)

We also have
\[ L_\gamma(d^\gamma, \mu^\gamma) = \sup_{\mu \geq 0} L_\gamma(d^\gamma, \mu) \]
\[ = a'_0 d^\gamma + \frac{\gamma}{2} \|d^\gamma\|^2 + \frac{1}{2} \sum_{j=1}^r \left((a'_j d^\gamma)^+\right)^2 \]
\[ = a'_0 d^\gamma + \frac{\gamma}{2} \|d^\gamma\|^2 + \frac{1}{2} \mu^\gamma \|d\|^2. \] (2.39)

Taking the limit as \( \gamma \to 0 \) and using the fact \( \mu^\gamma \to \mu^* \) and \( (\gamma/2)\|d^\gamma\|^2 \to 0 \) [cf. Eq. (2.38)], it follows that
\[ \lim_{\gamma \to 0} L_\gamma(d^\gamma, \mu^\gamma) = \lim_{\gamma \to 0} a'_0 d^\gamma + \frac{1}{2} \|\mu^*\|^2. \]

Combining this equation with Eq. (2.37), we obtain
\[ \lim_{\gamma \to 0} a'_0 d^\gamma = -\|\mu^*\|^2, \]
which together with the fact \( a'_j d^\gamma = \mu^*_j \to \mu^*_j \) shown earlier, proves Eqs. (2.24) and (2.25).

The maximum of \( L_0 \) over \( \mu \geq 0 \) for fixed \( d \) is attained at \( \mu_j = (a'_j d)^+ \) [compare with Eq. (2.30)], so that
\[ \sup_{\mu \geq 0} L_0(d, \mu) = a'_0 d + \frac{1}{2} \sum_{j=1}^r \left((a'_j d)^+\right)^2. \]

Hence
\[ \inf_{d \in \mathbb{N}^r} \sup_{\mu \geq 0} L_0(d, \mu) = \inf_{d \in \mathbb{N}^r} \left\{ a'_0 d + \frac{1}{2} \sum_{j=1}^r \left((a'_j d)^+\right)^2 \right\}. \] (2.40)

Combining this with the equation
\[ \inf_{d \in \mathbb{N}^r} \sup_{\mu \geq 0} L_0(d, \mu) = \lim_{\gamma \to 0} L_\gamma(d^\gamma, \mu^\gamma) = -\frac{1}{2} \|\mu^*\|^2 \]
[cf. Eqs. (2.35) and (2.37)], and Eqs. (2.38) and (2.39), we obtain the desired relation
\[ -\frac{1}{2} \|\mu^*\|^2 = \inf_{d \in \mathbb{N}^r} \left\{ a'_0 d + \frac{1}{2} \sum_{j=1}^r \left((a'_j d)^+\right)^2 \right\} \]
\[ = \lim_{\gamma \to 0} \left\{ a'_0 d^\gamma + \frac{1}{2} \sum_{j=1}^r \left((a'_j d^\gamma)^+\right)^2 \right\}. \]
[cf. Eq. (2.26)].

Finally, if \( \alpha^* \) attains the infimum in the right-hand side above, Eqs. (2.35), (2.37), and (2.40) show that \((\alpha^*, \mu^*)\) is a saddle point of \( L_0 \), and that
\[
\alpha_0^* \alpha^* = -||\mu^*||^2, \quad (\alpha_j^* \alpha^*) = \mu_j^*, \quad j = 1, \ldots, r.
\]

Q.E.D.

We now return to the proof of Prop. 2.3.1(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 2.3.1 with the following identifications:
\[
N = T_X(x^*)^\perp, \quad a_0 = \nabla f(x^*), \quad a_j = \nabla g_j(x^*), \quad j = 1, \ldots, r,
\]
\[
M = \text{set of Lagrange multipliers},
\]
\[
\mu^* = \text{Lagrange multiplier of minimum norm}.
\]

If \( \mu^* = 0 \), then \( \mu^* \) is an informative Lagrange multiplier and we are done. If \( \mu^* \neq 0 \), by Lemma 2.3.1 [cf. (2.24) and (2.25)], for any \( \epsilon > 0 \), there exists a \( \bar{d} \in N^\perp = T_X(x^*) \) such that
\[
a_0' \bar{d} < 0, \tag{2.41}
\]
\[
a_j' \bar{d} > 0, \quad \forall \ j \in J^*, \quad a_j' \bar{d} \leq \epsilon \min_{l \in J^*} a_l' \bar{d}, \quad \forall \ j \notin J^*, \tag{2.42}
\]
where \( J^* = \{ j \mid \mu_j^* > 0 \} \). By suitably scaling the vector \( \bar{d} \), we can assume that \( ||\bar{d}|| = 1 \). Let \( \{ x_k \} \subset X \) be such that \( x_k \neq x^* \) for all \( k \) and
\[
x_k \to x^*, \quad \frac{x_k - x^*}{||x_k - x^*||} \to \frac{\bar{d}}{||\bar{d}||}.
\]

Using Taylor’s theorem for the cost function \( f \), we have for some vector sequence \( \xi_k \) converging to 0
\[
f(x_k) - f(x^*) = \nabla f(x^*)' (x_k - x^*) + o(||x_k - x^*||)
\]
\[
= \nabla f(x^*)' (\bar{d} + \xi_k) ||x_k - x^*|| + o(||x_k - x^*||)
\]
\[
= ||x_k - x^*|| \left( \nabla f(x^*)' \bar{d} + \nabla f(x^*)' \xi_k + \frac{o(||x_k - x^*||)}{||x_k - x^*||} \right).
\]

From Eq. (2.41), we have \( \nabla f(x^*)' \bar{d} < 0 \), 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 \( \xi_k \) converging to 0,
\[
g_j(x_k) - g_j(x^*) = \nabla g_j(x^*)' (x_k - x^*) + o(||x_k - x^*||)
\]
\[
= \nabla g_j(x^*)' (\bar{d} + \xi_k) ||x_k - x^*|| + o(||x_k - x^*||)
\]
\[
= ||x_k - x^*|| \left( \nabla g_j(x^*)' \bar{d} + \nabla g_j(x^*)' \xi_k + \frac{o(||x_k - x^*||)}{||x_k - x^*||} \right).
\]
Sec. 2.3  Informative Lagrange Multipliers

This, combined with Eq. (2.42), 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 \( g_j(x^k) \leq o(\|x^k - x^*\|) \) for all \( j \) such that \( \mu_j^* = 0 \) [and hence \( g_j(x^*) \leq 0 \)]. Thus, the sequence \( \{x^k\} \) can be used to establish the CV condition for \( \mu^* \), and it follows that \( \mu^* \) is an informative Lagrange multiplier.

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

**Lemma 2.3.2:** Let \( N \) be a closed convex cone in \( \mathbb{R}^n \), let \( a_0, a_1, \ldots, a_r \) be given vectors in \( \mathbb{R}^n \). Suppose that the closed and convex set \( M \subset \mathbb{R}^r \) given by
\[
M = \left\{ \mu \geq 0 \mid - \left( a_0 + \sum_{j=1}^{r} \mu_j a_j \right) \in N \right\}
\]
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 \( \overline{d} \in N^\perp \) such that
\[
a_0^\prime \overline{d} < 0, \quad a_j^\prime \overline{d} > 0, \quad \text{for all } j \in \overline{J}. \tag{2.45}
\]

**Proof:** We apply Lemma 2.3.1 with the vectors \( a_1, \ldots, a_r \) replaced by the vectors \( a_j, \ j \in \overline{J} \). The subset of \( M \) given by
\[
\overline{M} = \left\{ \mu \geq 0 \mid - \left( a_0 + \sum_{j \in \overline{J}} \mu_j a_j \right) \in N, \mu_j = 0, \forall j \notin \overline{J} \right\}
\]
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 2.3.1 implies that there exists a \( \overline{d} \in N^\perp \) such that Eq. (2.45) holds. \( \text{Q.E.D.} \)

Given Lemma 2.3.2, the proof of Prop. 2.3.1(b) is very similar to the corresponding part of the proof of Prop. 2.3.1(a). \( \text{Q.E.D.} \)

**Sensitivity**

Let us consider now the special direction \( d^* \) that appears in Lemma 2.3.1, and is a solution of problem (2.27) (assuming this problem has an optimal
solution). Let us note that this problem is guaranteed to have at least one solution when $N^\perp$ is a polyhedral cone. This is because problem (2.27) can be written as

$$\begin{align*}
\minimize & \quad a'_0d + \frac{1}{2} \sum_{j=1}^{r} z_j^2 \\
\text{subject to} & \quad d \in N^\perp, \quad 0 \leq z_j, \quad a'_j d \leq z_j, \quad j = 1, \ldots, r,
\end{align*}$$

where the $z_j$ are auxiliary variables. Thus, if $N^\perp$ is polyhedral, then problem (2.27) is a quadratic program with a cost function that is bounded below by Eq. (2.26), and as shown in Section 1.3, it has an optimal solution. An important context where this is relevant is when $X = \mathbb{R}^n$ in which case $N_X(x^*)^\perp = T_X(x^*) = \mathbb{R}^n$, or more generally when $X$ is polyhedral, in which case $T_X(x^*)$ is polyhedral.

It can also be shown that problem (2.27) has an optimal solution if there exists a vector $\mu \in M$ such that

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

where $\text{ri}(N)$ denotes the relative interior of the cone $N$. We will show this in the next section after we develop the theory of pseudonormality and constraint qualifications.

Assuming now that $T_X(x^*)$ is polyhedral (or more generally, that problem (2.27) has an optimal solution), the line of proof of Prop. 2.3.1(a) (combine Eqs. (2.43) and (2.44)) can be used to show that if the Lagrange multiplier that has minimum norm, denoted by $(\lambda^*, \mu^*)$, is nonzero, there exists a sequence \( \{x^k\} \subset X \), corresponding to the vector $d^* \in T_X(x^*)$ of Eq. (2.28) such that

$$f(x^k) = f(x^*) - \sum_{i=1}^{m} \lambda^*_i h_i(x^k) - \sum_{j=1}^{r} \mu^*_j g_j(x^k) + o(||x^k - x^*||). \quad (2.46)$$

Furthermore, the vector $d^*$ solves problem (2.27), from which it can be seen that $d^*$ solves the problem

$$\begin{align*}
\minimize & \quad \nabla f(x^*)'d \\
\text{subject to} & \quad \sum_{i=1}^{m} (\nabla h_i(x^*)'d)^2 + \sum_{j \in A(x^*)} (\nabla g_j(x^*)'d^+)^2 = \beta, \quad d \in T_X(x^*),
\end{align*}$$

where $\beta$ is given by

$$\beta = \sum_{i=1}^{m} (\nabla h_i(x^*)'d^*)^2 + \sum_{j \in A(x^*)} (\nabla g_j(x^*)'d^*)^2.$$
More generally, it can be seen that for any given positive scalar $\beta$, a positive multiple of $d^*$ solves the problem

\[
\text{minimize } \nabla f(x^*)'d \\
\text{subject to } \sum_{i=1}^{m} (\nabla h_i(x^*)'d)^2 + \sum_{j \in A(x^*)} (\nabla g_j(x^*)'d^+)^2 = \beta, \quad d \in T_X(x^*),
\]

Thus, $d^*$ 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. (2.46), 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 $\lambda^*_i$ and $\mu^*_j$ express the rate of improvement per unit constraint violation, along the maximum improvement (or steepest descent) direction $d^*$. This is consistent with the traditional sensitivity interpretation of Lagrange multipliers.

**An Alternative Definition of Lagrange Multipliers**

Finally, let us make the connection with another treatment of Lagrange multipliers, due to Rockafellar [Roc93]. Consider vectors $\lambda^* = (\lambda^*_1, \ldots, \lambda^*_m)$ and $\mu^* = (\mu^*_1, \ldots, \mu^*_r)$ that satisfy the conditions

\[
- \left( \nabla f(x^*) + \sum_{i=1}^{m} \lambda^*_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu^*_j \nabla g_j(x^*) \right) \in N_X(x^*), \quad (2.47)
\]

\[
\mu^*_j \geq 0, \quad \forall \ j = 1, \ldots, r, \quad \mu^*_j = 0, \quad \forall \ j \notin A(x^*). \quad (2.48)
\]

Such vectors are called “Lagrange multipliers” by Rockafellar, but we will here refer to them as \textit{R-multipliers}, to distinguish them from Lagrange multipliers as we have defined them [cf. Eqs. (2.8)-(2.10)].

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 $T_X(x^*)^\perp \subset N_X(x^*)$ with inequality holding when $X$ is not regular at $x^*$. Note that multipliers satisfying the enhanced Fritz John conditions of Prop. 2.2.1 with $\mu^*_0 = 1$ are R-multipliers, and they still have the extra sensitivity-like property embodied in the CV condition. Furthermore, Lemma 2.3.1 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 be such that the Lagrangian function can decrease along some tangent directions. Furthermore, the existence of R-multipliers does not guarantee the existence of Lagrange multipliers. Even if Lagrange multipliers exist, none of them may be informative or strong, unless the tangent cone is convex. Thus regularity of $X$ is the property that separates problems that possess satisfactory Lagrange multiplier theory and problems that do not.

The following example illustrates the above.

**Example 2.3.1**

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

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

![Figure 2.3.2. Constraints of Example 2.3.1. We have](image)

Here $T_X(x^*) = X$ and $N_X(x^*)$ is the nonconvex set consisting of the two rays of vectors that are colinear to either $a_1$ or $a_2$.

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

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

(see Fig. 2.3.2).
Because $N_X(x^*) \neq T_X(x^*) \perp$, $X$ 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 $(\mu^*_1, \mu^*_2)$ satisfying Eqs. (2.8)-(2.10). The scalars $\mu^*_1, \mu^*_2$ are determined from the requirement
\[
\nabla f(x^*) + \mu^*_1 a_1 + \mu^*_2 a_2 = 0. \quad (2.49)
\]
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 $(\mu^*_1, \mu^*_2)$ 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 $(\mu^*_1, \mu^*_2)$ 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. (2.49), all other R-multipliers are such that the Lagrangian function has negative slope along some of the feasible directions of $X$.

\section*{Exercises}

\subsection*{2.3.1}

The purpose of this exercise is to work out a simplified proof of Lemma 2.3.1, assuming that $N = \{0\}$ [which corresponds to the case where there is no abstract set constraint ($X = \mathbb{R}^n$)]. Let $a_0, \ldots, a_r$ be given vectors in $\mathbb{R}^n$. Suppose that the set
\[
M = \left\{ \mu \geq 0 \mid a_0 + \sum_{j=1}^r \mu_j a_j = 0 \right\}
\]
is nonempty, and let $\mu^*$ be the vector of minimum norm on $M$. For any $\gamma \geq 0$, consider the function
\[
L_\gamma(d, \mu) = \left( a_0 + \sum_{j=1}^r \mu_j a_j \right)' d + \frac{\gamma}{2} \|d\|^2 - \frac{1}{2} \|\mu\|^2.
\]
(a) Show that
\[
-\frac{1}{2} \|\mu^*\|^2 = \sup_{\mu \geq 0} \inf_{d \in \mathbb{R}^n} L_0(d, \mu)
\]
\[
\leq \inf_{d \in \mathbb{R}^n} \sup_{\mu \geq 0} L_0(d, \mu)
\]
\[
= \inf_{d \in \mathbb{R}^n} \left\{ a_0 d + \frac{1}{2} \sum_{j=1}^r \left( (a_j d)' \right)^2 \right\}. \quad (2.50)
\]
(b) Use the lower bound of part (a) and the theory of Section 1.3 on the existence of solutions of quadratic programs to conclude that the infimum in the right-hand side above is attained for some $d^* \in \mathbb{R}^n$.

(c) Show that for every $\gamma > 0$, $L_\gamma$ has a saddle point $(d^\gamma, \mu^\gamma)$ such that

$$\mu^\gamma_j = (a'_j d^\gamma)^+, \quad j = 1, \ldots, r.$$ 

Furthermore,

$$L_\gamma(d^\gamma, \mu^\gamma) = -\frac{\|a_0 + \sum_{j=1}^r \mu^\gamma_j a_j\|^2}{2\gamma} - \frac{1}{2}\|\mu^\gamma\|^2 \geq -\frac{1}{2}\|\mu^*\|^2.$$ 

(d) Use part (c) to show that $\|\mu^\gamma\| \leq \|\mu^*\|$, and use the minimum norm property of $\mu^*$ to conclude that as $\gamma \to 0$, we have $\mu^\gamma \to \mu^*$ and $L_\gamma(d^\gamma, \mu^\gamma) \to -(1/2)\|\mu^*\|^2$.

(e) Use part (d) and Eq. (2.50) to show that $(d^\gamma, \mu^\gamma)$ is a saddle point of $L_0$, and that

$$a'_0 d^\gamma = -\|\mu^*\|^2; \quad (a'_j d^\gamma)^+ = \mu^*_j, \quad j = 1, \ldots, r.$$ 

2.4 PSEUDONORMALITY AND CONSTRAINT QUALIFICATIONS

The enhanced Fritz John conditions lead to the introduction of the following general constraint qualification under which the scalar $\mu^*_0$ in Prop. 2.2.1 cannot be zero.

**Definition 2.4.1:** We say that a feasible vector $x^*$ of problem (2.6)-(2.7) 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 $j \notin A(x^*)$, where

$$A(x^*) = \{j \mid g_j(x^*)\}.$$ 

(iii) $\{x^k\}$ converges to $x^*$ and

$$\sum_{i=1}^m \lambda_i h_i(x^k) + \sum_{j=1}^r \mu_j g_j(x^k) > 0, \quad \forall k. \quad (2.51)$$
Sec. 2.4 Pseudonormality and Constraint Qualifications

If \( x^* \) is a pseudonormal local minimum, the enhanced Fritz John conditions of Prop. 2.2.1 cannot be satisfied with \( \mu_0^* = 0 \), so that \( \mu_0^* \) can be taken equal to 1. Then, if \( X \) is regular at \( x^* \), the vector \((\lambda^*, \mu^*) = (\lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*)\) obtained from the enhanced Fritz John conditions is an informative Lagrange multiplier.

We will provide a geometric interpretation of pseudonormality in Section 2.7, where we discuss the special case where \( X \) and the functions \( g_j \) are convex, and the functions \( h_i \) are linear.

We now focus on deriving conditions, also referred to as constraint qualifications (CQ for short), which will be shown in this section to imply pseudonormality of a feasible vector \( x^* \) and hence also existence of informative Lagrange multiplier (assuming also regularity of \( X \) at \( x^* \)).

**CQ1:** \( X = \mathbb{R}^n \) and \( x^* \) is a regular point, meaning 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:** \( X = \mathbb{R}^n \), the equality constraint gradients \( \nabla h_i(x^*), \ i = 1, \ldots, m \), are linearly independent, and there exists a \( y \in \mathbb{R}^n \) such that

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

For the case where there are no equality constraints, CQ2 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:** \( X = \mathbb{R}^n \), the functions \( h_i \) are linear and the functions \( g_j \) are concave.

**CQ4:** \( X = \mathbb{R}^n \) and for some integer \( r < r \), the following superset \( \overline{C} \) of the constraint set \( C \),

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

is pseudonormal at \( x^* \). Furthermore, there exists a \( y \in \mathbb{R}^n \) such that
\[ \nabla h_i(x^*)'y = 0, \quad i = 1, \ldots, m, \quad \nabla g_j(x^*)'y \leq 0, \quad \forall \ j \in A(x^*), \]
\[ \nabla g_j(x^*)'y < 0, \quad \forall \ j \in \{1, \ldots, r\} \cap A(x^*). \]

Since CQ1-CQ3 imply pseudonormality, a fact to be shown in the subsequent Prop. 2.4.1, we see that CQ4 generalizes CQ1-CQ3.

**CQ5:**

(a) The equality constraints with index above some \( m \leq \overline{m} \):
\[ h_i(x) = 0, \quad i = \overline{m} + 1, \ldots, m, \]
are linear.

(b) There does not exist a vector \( \lambda = (\lambda_1, \ldots, \lambda_m) \) such that
\[ \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*) \]  \hspace{1cm} (2.52)
and at least one of the scalars \( \lambda_1, \ldots, \lambda_{\overline{m}} \) is nonzero.

(c) The subspace
\[ V_L(x^*) = \{ y | \nabla h_i(x^*)'y = 0, \ i = \overline{m} + 1, \ldots, m \} \]
has a nonempty intersection with the interior of \( N_X(x^*)^\perp \).

(d) There exists a \( y \in N_X(x^*)^\perp \) such that
\[ \nabla h_i(x^*)'y = 0, \quad i = 1, \ldots, m, \quad \nabla g_j(x^*)'y < 0, \quad \forall \ j \in A(x^*). \]

We refer to CQ5 as the *generalized Mangasarian-Fromovitz constraint qualification*, since it reduces to CQ2 when \( X = \mathbb{R}^n \) and none of the equality constraints is linear. 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 $y \in N_{X}(x^*)^\perp$ such that
\[ \nabla h_i(x^*)'y = 0, \quad i = 1, \ldots, m, \quad \nabla g_j(x^*)'y < 0, \quad \forall j \in A(x^*). \]

CQ5b: There are no inequality constraints, the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, are linearly independent, and the subspace
\[ V(x^*) = \{ y \mid \nabla h_i(x^*)'y = 0, \quad i = 1, \ldots, m \} \]
contains a point in the interior of $N_{X}(x^*)^\perp$.

CQ5c: $X$ is convex, there are no inequality constraints, the functions $h_i$, $i = 1, \ldots, m$, are linear, and the linear manifold
\[ \{ x \mid h_i(x) = 0, \quad i = 1, \ldots, m \} \]
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 $\bar{x}$ satisfying
\[ g_j(\bar{x}) < 0, \quad \forall j \in A(x^*). \]

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

CQ6: The set

\[ W = \{ (\lambda, \mu) \mid \lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r \text{ satisfy conditions (i) and (ii)} \} \]

of the definition of pseudonormality \[(2.53)\]

consists of just the vector 0.

It can be shown that the set \( W \) of Eq. (2.53) is the recession cone of the set of R-multipliers, provided the set of R-multipliers is a nonempty closed convex set (so we can talk about its recession cone). To see this, note that for any R-multiplier \((\lambda^*, \mu^*)\), and any \((\lambda, \mu) \in W\), we have for all \(\alpha \geq 0\),

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

since \( N_X(x^*) \) is a cone. Thus \((\lambda, \mu)\) is a direction of recession. Conversely, if \((\lambda, \mu)\) is a direction of recession, then for all R-multipliers \((\lambda^*, \mu^*)\), we have for all \(\alpha > 0\),

\[ -\left( \frac{1}{\alpha} \nabla f(x^*) + \sum_{i=1}^{m} \left( \frac{1}{\alpha} \lambda_i^* + \lambda_i \right) \nabla h_i(x^*) \right. \]

\[ + \sum_{j=1}^{r} \left( \frac{1}{\alpha} \mu_j^* + \mu_j \right) \nabla g_j(x^*) \left. \right) \in N_X(x^*). \]

Taking the limit as \(\alpha \to 0\) and using the closedness of \( N_X(x^*) \), we see that \((\lambda, \mu) \in W\).

Since compactness of a closed, convex set is equivalent to its recession cone consisting of just the 0 vector, it follows that if the set of R-multipliers is nonempty, convex, and compact, then CQ6 holds. In view of Prop. 2.2.1, 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 (see the exercises).

Clearly CQ6 implies pseudonormality, since the vectors in \( W \) are not required to satisfy condition (iii) of the definition of pseudonormality. However, CQ3, CQ4, and CQ5 do not preclude unboundedness of the set of Lagrange multipliers and hence do not imply CQ6.
Proposition 2.4.1: A feasible point $x^*$ of problem (2.6)-(2.7) is pseu-
onormal if any one of the constraint qualifications CQ1-CQ6 is sat-
ished.

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
$\lambda_i, i = 1, \ldots, m, \text{ and } \mu_j, j = 1, \ldots, r$, 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 $X = \mathbb{R}^n$, 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 $\nabla h_i(x^*), i = 1, \ldots, m, \text{ and } \nabla g_j(x^*), j \in A(x^*)$,
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 $x \in \mathbb{R}^n$,
\[
h_i(x) = h_i(x^*) + \nabla h_i(x^*)'(x - x^*), \quad i = 1, \ldots, m,
g_j(x) \leq g_j(x^*) + \nabla g_j(x^*)'(x - x^*), \quad j = 1, \ldots, r.
\]
By multiplying these two relations with $\lambda_i$ and $\mu_j$, and by adding over $i$ and $j$, respectively, we obtain
\[
\sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x) \leq \sum_{i=1}^{m} \lambda_i h_i(x^*) + \sum_{j=1}^{r} \mu_j g_j(x^*)
+ \left( \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \right)'(x - x^*)
= 0,
\]
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. (2.54).

**CQ4:** It is not possible that \( \mu_j = 0 \) for all \( j \in \{1, \ldots, r\} \), since if this were so, the pseudonormality assumption for \( C \) would be violated. Thus we have \( \mu_j > 0 \) for some \( j \in \{1, \ldots, r\} \cap A(x^*) \). It follows that for the vector \( y \) appearing in the statement of CQ4, we have \( \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)'y < 0 \), 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 \( \lambda_1, \ldots, \lambda_m \) and \( \mu_j, j \in A(x^*) \) 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 \( a_i'x = 0, i = m+1, \ldots, m \). Using condition (i) we have

\[
- \sum_{i=m+1}^{m} \lambda_i a_i \in N_X(x^*). \tag{2.55}
\]

Let \( y \) be the interior point of \( N_X(x^*)' \) that satisfies \( a_i'y = 0 \) for all \( i = m+1, \ldots, m \), and let \( S \) be an open sphere centered at the origin such that \( y + d \in N_X(x^*)' \) for all \( d \in S \). We have from Eq. (2.55),

\[
\sum_{i=m+1}^{m} \lambda_i a_i' d \geq 0, \quad \forall \ d \in S,
\]

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

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

\[
- \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*). \tag{2.56}
\]

By what has been proved above, the multipliers \( \lambda_1, \ldots, \lambda_m \) of the nonlinear constraints cannot be all zero, so Eq. (2.56) 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 \( j \not\in A(x^*) \), we obtain
\[
\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)' y + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)' y < 0,
\]
for the vector \( y \) of \( N_X(x^*)^\perp \) that appears in assumption (d) of CQ5. Thus,
\[
-\left( \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \right) \notin (N_X(x^*)^\perp)^\perp.
\]
Since \( N_X(x^*) \subset (N_X(x^*)^\perp)^\perp \), this contradicts condition (i). \( \mathbf{Q.E.D.} \)

As an example of application of Prop. 2.4.1, let us analyze the existence of a solution to problem (2.27) in Lemma 2.3.1 (this analysis is due to Xin Chen). As discussed in the preceding section, the solution to this problem gives the direction of steepest descent, along which the Lagrange multiplier of minimum norm specifies the rate of cost improvement.

**Example 2.4.1 (Existence of the Steepest Descent Direction)**

Consider the problem
\[
\begin{align*}
\text{minimize} & \quad a_0' d + \frac{1}{2} \sum_{j=1}^{r} ((a_j')^+)^2 \\
\text{subject to} & \quad d \in N^\perp,
\end{align*}
\]
(2.57)
of Lemma 2.3.1, where \( a_0, a_1, \ldots, a_r \) are given vectors and \( N \) is a closed convex cone. We will assume that there exists a vector \( \overline{\mu} \) in the set
\[
M = \left\{ \mu \geq 0 \mid -\left( a_0 + \sum_{j=1}^{r} \overline{\mu}_j a_j \right) \in N \right\}
\]
such that
\[
-\left( a_0 + \sum_{j=1}^{r} \overline{\mu}_j a_j \right) \in \text{ri}(N),
\]
(2.58)
where \( \text{ri}(N) \) denotes the relative interior of \( N \), and we will show that problem (2.57) has at least one optimal solution \( d^* \).

Consider the problem
\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| \mu \|^2 \\
\text{subject to} & \quad y = -\left( a_0 + \sum_{j=1}^{r} \mu_j a_j \right) \\
& \quad y \in N, \quad \mu \geq 0,
\end{align*}
\]
(2.59)
where \((y, \mu)\) is the vector of optimization variables, and note that its unique optimal solution is the vector \((y^*, \mu^*)\), where \(\mu^*\) is the vector of minimum norm on \(M\) and

\[
y^* = -\left(a_0 + \sum_{j=1}^{r} \mu_j^* a_j\right).
\]

Let us consider the set \(\{ (y, \mu) \mid y \in N, \mu \geq 0 \}\) as defining the abstract set constraint of problem (2.59), and view the equation

\[
-y + a_0 + \sum_{j=1}^{r} \mu_j a_j = 0
\]

as a linear equality constraint. Then we can use the constraint qualification CQ5c and Prop. 2.4.1 to assert existence of a Lagrange multiplier corresponding to the linear constraint. In particular, if we reformulate problem (2.59) over the affine hull of \(N\) via a linear transformation argument, the relative interior assumption (2.58) guarantees that CQ5c is satisfied for the reformulated problem. Thus, there exists a Lagrange multiplier vector corresponding to \((y^*, \mu^*)\), which is denoted by \(d^*\). We will show that \(d^*\) solves problem (2.57).

Applying the definition of Lagrange multiplier, the Lagrangian stationarity condition for problem (2.59) can be written as

\[
-d^*(y - y^*) \geq 0, \quad \forall y \in N,
\]

\[
(\mu_j^* - a_j^* d^*)(\mu_j - \mu_j^*) \geq 0, \quad \forall \mu_j \geq 0, j = 1, \ldots, r.
\]

The second of these two relations yields

\[
\mu_j^* = (a_j^* d^*)^+, \quad \forall j = 1, \ldots, r, \tag{2.60}
\]

while the first yields \(d^* \in N^\perp\) and \(y^* d^* = 0\), or equivalently since \(y^* = -\left(a_0 + \sum_{j=1}^{r} \mu_j^* a_j\right)\),

\[
d^* \in N^\perp, \quad \left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right)' d^* = 0. \tag{2.61}
\]

Consider (as in the proof of Lemma 2.3.1) the function

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

From Eqs. (2.60) and (2.61), we have

\[
L_0(d^*, \mu^*) = -\frac{1}{2} \|\mu^*\|^2.
\]
Sec. 2.4 Pseudonormality and Constraint Qualifications

as well as
\[ L_0(d^*, \mu^*) = a'_0 d^* + \frac{1}{2} \sum_{j=1}^r \left( (a'_j d^*)^+ \right)^2. \]

On the other hand, as shown in the proof of Lemma 2.3.1, we have
\[ -\frac{1}{2} \|\mu\|^2 = \sup_{\mu \in M} \inf_{\delta \in N^\perp} L_0(d, \mu) = \inf_{d \in N^\perp} \sup_{\mu \in M} L_0(d, \mu) = \inf_{d \in N^\perp} \left\{ a'_0 d + \frac{1}{2} \sum_{j=1}^r \left( (a'_j d)^+ \right)^2 \right\}. \]

By combining the last three equations, we see that \( d^* \) solves problem (2.57).

As another example of application of Prop. 2.4.1, let us show a strengthened version of Farkas’ lemma.

**Proposition 2.4.2: (Enhanced Farkas’ Lemma)** Let \( a_1, \ldots, a_r \) and \( c \) be given vectors in \( \mathbb{R}^n \), and assume that \( c \neq 0 \). We have
\[ c'y \leq 0, \quad \text{for all } y \text{ such that } a_j'y \leq 0, \quad \forall \; j = 1, \ldots, r, \]
if and only if there exist nonnegative scalars \( \mu_1, \ldots, \mu_r \) and a vector \( \overline{y} \in \mathbb{R}^n \) such that \( c'\overline{y} > 0, \quad a_j'\overline{y} > 0 \) for all \( j \) such that \( \mu_j > 0 \), \( a_j'\overline{y} \leq 0 \) for all \( j \) such that \( \mu_j = 0 \), and
\[ c = \mu_1 a_1 + \cdots + \mu_r a_r. \]

**Proof:** If \( c = \sum_{j=1}^r \mu_j a_j \) for some scalars \( \mu_j \geq 0 \), then for every \( y \) satisfying \( a_j'y \leq 0 \) for all \( j \), we have \( c'y = \sum_{j=1}^r \mu_j a_j'y \leq 0 \).

Conversely, if \( c \) satisfies \( c'y \leq 0 \) for all \( y \) such that \( a_j'y \leq 0 \) for all \( j \), then \( y^* = 0 \) minimizes \(-c'y\) subject to \( a_j'y \leq 0, \; j = 1, \ldots, r \). Since the constraint qualification CQ3 holds for this problem, by Prop. 2.4.1, there is a nonempty set of Lagrange multipliers \( M \), which has the form
\[ M = \{ \mu \geq 0 \mid c = \mu_1 a_1 + \cdots + \mu_r a_r \}. \]

Lemma 2.3.1 applies with \( N = \{0\} \), so we have
\[ -\|\mu^*\|^2 \leq -c'y + (1/2) \sum_{j=1}^r (a'_j y^+)^2, \quad \forall \; y \in \mathbb{R}^n, \]
where $\mu^*$ is the Lagrange multiplier vector of minimum norm. Minimization of the right-hand side above over $y \in \mathbb{R}^n$ is a quadratic program, which is bounded below, so it follows from the results of Section 1.3.3 that there is a vector $\bar{y} \in \mathbb{R}^n$ that minimizes this right-hand side. Applying Lemma 2.3.1 again, we have

$$ -c^\top \bar{y} = -\|\mu^*\|^2, \quad (a_j^\top \bar{y})^+ = \mu_j^*, \quad j = 1, \ldots, r. $$

Since by assumption $c \neq 0$, we have $\mu^* \neq 0$ and the result follows. \textbf{Q.E.D.}

The classical version of Farkas’ lemma does not include the assertion on the existence of a $y$ that satisfies $c^\top y > 0$ while violating precisely those inequalities that correspond to positive multipliers. A more general version of the enhanced Farkas’ lemma is given in the exercises following Section 2.6.

Proposition 2.4.1 does not apply to the important class of problems where the constraint functions $h_i$ and $g_j$ are linear, and the set $X$ is polyhedral but is a strict subset of $\mathbb{R}^n$. Unfortunately, in this case pseudonormality may be violated, as will be shown by example later. However, it turns out that in this case there always exists a Lagrange multiplier vector. We can show this from first principles, by using Farkas’ lemma, but we can also show it by applying Prop. 2.4.1 to a problem involving an “extended” representation of the problem constraints. In particular, we introduce an auxiliary and equivalent problem where the linear constraints embodied in $X$ are added to the other linear constraints corresponding to $h_i$ and $g_j$, so that there is no abstract set constraint. For the auxiliary problem, CQ3 is satisfied and Prop. 2.4.1 guarantees the existence of Lagrange multipliers. We can then show that the multipliers corresponding to the $h_i$ and $g_j$ form a Lagrange multiplier vector for the original problem. We defer to Section 2.6 a fuller and more general discussion of this approach.

**Pseudonormality and Quasiregularity**

We recall that for the case where $X = \mathbb{R}^n$, a point $x \in C$ said to be a quasiregular point of the constraint set $C$ if

$$ T_C(x) = V(x), \quad (2.62) $$

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

$$ V(x) = \{ y \mid \nabla h_i(x)^\top y = 0, \; i = 1, \ldots, m, \nabla g_j(x)^\top y \leq 0, \; j \in A(x) \}, \quad (2.63) $$

where $A(x) = \{ j \mid g_j(x) = 0 \}$.

At a local minimum $x^*$ of $f$ over $C$, we have $-\nabla f(x^*) \in T_C(x^*)^\perp$, so if $x^*$ is quasiregular, we have

$$ \nabla f(x^*)^\top y \geq 0, \quad \forall \; y \in V(x^*). $$
By Farkas’ lemma, we either have that $\nabla f(x^*) = 0$ or else there exists a nonzero Lagrange multiplier vector $(\lambda^*, \mu^*)$. This is the classical approach to the existence of Lagrange multipliers.

An interesting converse was shown by Gould and Tolle [GoT71], namely that every vector in $T_C(x^*)^\perp$ can be described as the negative of the gradient of a function having $x^*$ as a local minimum over $C$. Rockafellar and Wets [RoW98] showed that this function can be taken to be smooth on $\mathbb{R}^n$. We give their result with an abbreviated version of their proof (see [RoW98], pp. 205-206 for a more detailed argument).

**Proposition 2.4.3:** For every $z \in T_C(x^*)^\perp$ there is a smooth function $F$ with $-\nabla F(x^*) = z$, which achieves a strict global minimum over $C$ at $x^*$.

**Proof:** Consider the scalar function $\theta_0 : [0, \infty) \mapsto \mathbb{R}$ defined by

$$\theta_0(r) = \sup_{x \in C, \|x - x^*\| \leq r} z'(x - x^*), \quad r \geq 0.$$  

Clearly $\theta_0(r)$ is nondecreasing and satisfies

$$0 = \theta_0(0) \leq \theta_0(r), \quad \forall \ r \geq 0.$$  

Furthermore, since $z \in T_C(x^*)^\perp$, we have $z'(x - x^*) \leq o(||x - x^*||)$ for $x \in C$, so that $\theta_0(r) = o(r)$, which implies that $\theta_0$ is differentiable at $r = 0$ with $\nabla \theta_0(0) = 0$. Thus, the function $F_0$ defined by

$$F_0(x) = \theta_0(||x - x^*||) - z'(x - x^*)$$  

is differentiable at $x^*$, attains a global minimum over $C$ at $x^*$, and satisfies

$$- \nabla F_0(x^*) = z.$$  

If $F_0$ were smooth we would be done, but since it need not even be continuous, we will successively perturb it into a smooth function. We first define the function $\theta_1 : [0, \infty) \mapsto \mathbb{R}$ by

$$\theta_1(r) = \begin{cases} \frac{1}{r} \int_r^{2r} \theta_0(s)ds & \text{if } r > 0, \\ 0 & \text{if } r = 0. \end{cases}$$  

This function is seen to be nondecreasing and continuous, and satisfies

$$0 \leq \theta_0(r) \leq \theta_1(r), \quad \forall \ r \geq 0,$$
\( \theta_1(0) = 0 \), and \( \nabla \theta_1(0) = 0 \). Thus the function
\[
F_1(x) = \theta_1(||x - x^*||) - z'(x - x^*)
\]
has the same significant properties for our purposes as \( F_0 \) [attains a global minimum over \( C \) at \( x^* \), and has \( -\nabla F_1(x^*) = z \)], and is in addition continuous.

We next define the function \( \theta_2 : [0, \infty) \to \mathbb{R} \) by
\[
\theta_2(r) = \begin{cases} 
\frac{1}{r} \int_r^{2r} \theta_1(s)ds & \text{if } r > 0, \\
0 & \text{if } r = 0.
\end{cases}
\]
Again \( \theta_2 \) is seen to be nondecreasing, and satisfies
\[
0 \leq \theta_1(r) \leq \theta_2(r), \quad \forall \ r \geq 0,
\]
\( \theta_2(0) = 0 \), and \( \nabla \theta_2(0) = 0 \). Also, because \( \theta_1 \) is continuous, \( \theta_2 \) is smooth, and so is the function \( F_2 \) given by
\[
F_2(x) = \theta_2(||x - x^*||) - z'(x - x^*).
\]
The function \( F_2 \) fulfills all the requirements of the proposition, except that it may have global minima other than \( x^* \). To ensure the uniqueness of \( x^* \) we modify \( F_2 \) as follows:
\[
F(x) = F_2(x) + ||x - x^*||^2.
\]
The function \( F \) is smooth, attains a strict global minimum over \( C \) at \( x^* \), and satisfies \( -\nabla F(x^*) = z \). Q.E.D.

We now show that pseudonormality implies quasiregularity, so that by Prop. 2.4.1, if any one of the constraint qualifications CQ1-CQ4 holds then quasiregularity also holds.

**Proposition 2.4.4:** Consider the case where \( X = \mathbb{R}^n \). Then a pseudo-normal point of \( C \) is quasiregular.

**Proof:** For simplicity we assume that all the constraints are inequalities that are active at \( x^* \). We will first show that \( T_C(x^*) \subset V(x^*) \) and then show the reverse inclusion. Indeed, let \( y \) be a nonzero tangent of \( C \) at \( x^* \). Then there exists a sequence \( \{\xi_k\} \) and a sequence \( \{x^k\} \subset C \) such that \( x^k \neq x^* \) for all \( k \),
\[
\xi_k \to 0, \quad x^k \to x^*.
\]
and
\[ \frac{x^k - x^*}{\|x^k - x^*\|} = \frac{y}{\|y\|} + \xi^k. \]

By the mean value theorem, we have for all \( j \) and \( k \)
\[ 0 \geq g_j(x^k) = g_j(x^*) + \nabla g_j(\tilde{x}^k)'(x^k - x^*) = \nabla g_j(\tilde{x}^k)'(x^k - x^*), \]
where \( \tilde{x}^k \) is a vector that lies on the line segment joining \( x^k \) and \( x^* \). This
relation can be written as
\[ \frac{\|x^k - x^*\|}{\|y\|} \nabla g_j(\tilde{x}^k)'y^k \leq 0, \]
where \( y^k = y + \xi^k\|y\| \), or equivalently
\[ \nabla g_j(\tilde{x}^k)'y^k \leq 0, \quad y^k = y + \xi^k\|y\|. \]

Taking the limit as \( k \to \infty \), we obtain \( \nabla g_j(x^*)'y \leq 0 \) for all \( j \), thus proving
that \( y \in V(x^*) \). Hence, \( T_C(x^*) \subseteq V(x^*) \).

To show that \( V(x^*) \subseteq T_C(x^*) \), it is sufficient to show that \( T_C(x^*) \subseteq V(x^*) \). Let \( z \in T_C(x^*) \). By Lemma 2.4.3, there exists a smooth function
\( F \) attaining a strict global minimum over \( C \) at \( x^* \) with \( -\nabla F(x^*) = z \). For
each \( k = 1, 2, \ldots \), choose an \( \epsilon > 0 \), and consider the “penalized” problem

\[
\begin{align*}
\text{minimize} & \quad F_k(x) \equiv F(x) + \frac{k}{2} \sum_{j=1}^{r} (g_j^+(x))^2 \\
\text{subject to} & \quad \|x - x^*\| \leq \epsilon.
\end{align*}
\]

By Weierstrass’ theorem, there exists an optimal solution \( x^k \) for the above
problem. An argument similar to the one used for Prop. 2.2.1 (that also
uses the fact that \( x^* \) is a strict local minimum of \( F \) over \( C \)) shows that
\( x^k \to x^* \) and that \( x^k \) is an interior point of \( S \) for sufficiently large \( k \). For
such \( k \), we have
\[ \nabla F(x^k) + \sum_{j=1}^{r} \zeta_j^k \nabla g_j(x^k) = 0, \quad (2.64) \]
where
\[ \zeta_j^k = kg_j^+(x^k). \quad (2.65) \]
Denote,
\[ \delta_k^k = \sqrt{1 + \sum_{j=1}^{r} (\zeta_j^k)^2}, \quad (2.66) \]
\[ \mu_j^k = \frac{1}{\delta_k^k}, \quad \mu_j^k = \frac{\zeta_j^k}{\delta_k^k}, \quad j = 1, \ldots, r. \quad (2.67) \]
Then by dividing Eq. (2.64) with $\delta^k$, we obtain

$$
\mu_k^0 \nabla F(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k) = 0. \tag{2.68}
$$

Since by construction the sequence $\{\mu_k^0, \mu_k^1, \ldots, \mu_k^r\}$ satisfies $\sum_{j=0}^{r} (\mu_j^k)^2 = 1$ for all $k$, it must contain a subsequence that converges to some nonzero limit $\{\mu^*_0, \mu^*_1, \ldots, \mu^*_r\}$ with $\mu^*_j \geq 0$, $\mu^*_j = 0$, for all $j \notin A(x^*)$, and

$$
\mu^*_0 \nabla F(x^*) + \sum_{j=1}^{r} \mu^*_j \nabla g_j(x^*) = 0.
$$

Furthermore, from Eq. (2.65), we have $g_j(x^k) > 0$ for all $j$ such that $\mu^*_j > 0$ and $k$ sufficiently large, so that $\sum_{j=1}^{r} \mu^*_j g_j(x^k) > 0$ for $k$ sufficiently large, provided the vector $(\mu^*_1, \ldots, \mu^*_r)$ is nonzero. By using the pseudonormality of $x^*$, it follows that we cannot have $\mu^*_0 = 0$, and by appropriately normalizing, we can take $\mu^*_0 = 1$ and obtain

$$
\nabla F(x^*) + \sum_{j=1}^{r} \mu^*_j \nabla g_j(x^*) = 0,
$$

or

$$
z = \sum_{j \in A(x^*)} \mu^*_j \nabla g_j(x^*). \tag{2.69}
$$

By Farkas’ lemma, $V(x^*)^\perp$ is the cone generated by $\nabla g_j(x^*)$, $j \in A(x^*)$. Hence, $z \in V(x^*)^\perp$ and we conclude that $T_C(x^*)^\perp \subset V(x^*)^\perp$. Q.E.D.

---

**EXERCISES**

2.4.1

Consider the problem

minimize $\sum_{i=1}^{n} f_i(x_i)$

subject to $x \in S$, $x_i \in X_i$, $i = 0, 1, \ldots, n$. 

---
where \( f_i : \mathbb{R} \to \mathbb{R} \) are smooth functions, \( X_i \) are closed intervals of real numbers of \( \mathbb{R}^n \), and \( S \) is a subspace of \( \mathbb{R}^n \). Let \( x^* \) be a local minimum. Introduce artificial optimization variables \( z_1, \ldots, z_n \) and the linear constraints \( x_i = z_i, \ i = 1, \ldots, n \), while replacing the constraint \( x \in S \) with \( z \in S \), so that the problem becomes

\[
\begin{align*}
\min & \quad \sum_{i=1}^{n} f_i(x_i) \\
\text{subject to} & \quad z \in S, \quad x_i \in X_i, \quad x_i = z_i, \quad i = 0, 1, \ldots, n.
\end{align*}
\]

Show that there exists a Lagrange multiplier vector \( \lambda^* = (\lambda_1^*, \ldots, \lambda_n^*) \) such that \( \lambda^* \in S^\perp \) and

\[
(\nabla f_i(x_i^*) + \lambda_i^*) (x_i - x_i^*) \geq 0, \quad \forall x_i \in X_i, \ i = 1, \ldots, n.
\]

### 2.4.2

Show that if \( X \) is regular at \( x^* \) the constraint qualifications CQ5a and CQ6 are equivalent.

### 2.5 EXACT PENALTY FUNCTIONS

In this section, we relate the problem

\[
\begin{align*}
\min & \quad f(x) \\
\text{subject to} & \quad x \in C,
\end{align*}
\]

where

\[
C = X \cap \{ x \mid h_1(x) = 0, \ldots, h_m(x) = 0 \} \cap \{ x \mid g_1(x) \leq 0, \ldots, g_r(x) \leq 0 \},
\]

with another problem that involves minimization over \( X \) of the cost function

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

where \( c \) is a positive scalar, and as earlier, we use the notation

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

Here the equality and inequality constraints are eliminated, and instead the cost is augmented with a term that penalizes the violation of these
constraints. The severity of the penalty is controlled by \( c \), which is called the **penalty parameter**.

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 \( F_c \) over \( X \). In the absence of additional assumptions, it is essential for our analysis to require that \( x^* \) be a strict local minimum in the definition of admittance of an exact penalty. This restriction may not be important in analytical studies, 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 ones of the preceding section, the admittance of an exact penalty, and the admittance of Lagrange multipliers. This is illustrated in the following example.

**Example 2.5.1**

Consider the 2-dimensional constraint set specified by

\[
h_1(x) = \frac{x_2}{x_1^2 + 1} = 0, \quad x \in X = \mathbb{R}^2.
\]

The feasible points are of the form \( x = (x_1, 0) \) with \( x_1 \in \mathbb{R} \), and at each of them the gradient \( \nabla h_1(x^*) \) is nonzero, so \( x^* \) is regular (CQ1 holds). If \( f(x) = x_2 \), every feasible point is a local minimum, yet for any \( c > 0 \), we have

\[
\inf_{x \in \mathbb{R}^2} \left\{ x_2 + c \frac{|x_2|}{x_1^2 + 1} \right\} = -\infty
\]

(take \( x_1 = x_2 \) as \( x_2 \to -\infty \)). Thus, the penalty function is not exact for any \( c > 0 \). It follows that regularity of \( x^* \) 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.

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 2.5.1: 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

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

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

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

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

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

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

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

subject to $x \in X, \ h_i(x) \leq w_i, \ -h_i(x) \leq w_i, \ i = 1, \ldots, m,$

$$g_j(x) \leq v_j, \ 0 \leq v_j, \ j = 1, \ldots, r,$$

which involves the auxiliary variables $w_i$ and $v_j$. It can be seen that at the local minimum of this problem that corresponds to $x^*$ the constraint qualification CQ5 is satisfied. Thus, by Prop. 2.4.1, this local minimum is pseudonormal, and hence there exist multipliers satisfying the enhanced Fritz-John conditions (Prop. 2.2.1) 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 2.5.2: If $x^*$ is a feasible vector of problem (2.70)-(2.71), 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

$$f(x^*) < f(x), \quad \forall x \in C \text{ with } x \neq x^* \text{ and } \|x - x^*\| \leq \epsilon. \quad (2.72)$$

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 (2.71), i.e.,

$$\sum_{i=1}^{m} |h_i(x^k)| + \sum_{j=1}^{r} g_j^+(x^k) > 0. \quad (2.73)$$

We have

$$F_k(x^k) = f(x^k) + k \left( \sum_{i=1}^{m} |h_i(x^k)| + \sum_{j=1}^{r} g_j^+(x^k) \right) \leq f(x^*), \quad (2.74)$$

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. (2.72) and (2.74) 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. 2.5.1):

$$- \left( \frac{1}{k} \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k) \right) \in N_X(x^k), \quad (2.75)$$

where

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

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

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

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

In view of Eq. (2.73), 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...
Sec. 2.5 Exact Penalty Functions

$g_j(x^k) > 0$ for all $k \in \mathcal{K}$. Let $(\lambda, \mu)$ 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. (2.75) yields

$$-\left( \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \right) \in N_X(x^*). \tag{2.76}$$

Finally, for all $k \in \mathcal{K}$, we have $\lambda_k h_i(x^k) \geq 0$ for all $i$, $\mu_k g_j(x^k) \geq 0$ for all $j$, so that, for all $k \in \mathcal{K}$, $\lambda_k h_i(x^k) \geq 0$ for all $i$, $\mu_k g_j(x^k) \geq 0$ 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 $k \in \mathcal{K}$,

$$\sum_{i=1}^{m} \lambda_i h_i(x^k) + \sum_{j=1}^{r} \mu_j g_j(x^k) > 0. \tag{2.77}$$

Thus, Eqs. (2.76) and (2.77) violate the hypothesis that $x^*$ is pseudonormal. Q.E.D.

The following example shows that the converse of Prop. 2.5.2 does not hold. In particular, the admittance of an exact penalty function at a point $x^*$ does not imply pseudonormality.

**Example 2.5.2**

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

$$g_1(x) = -(x_1 + 1)^2 - (x_2)^2 + 1, \quad g_2(x) = x_1^2 + (x_2 + 1)^2 - 1, \quad g_3(x) = -x_2,$$

(see Fig. 2.5.1). 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.

However, it can be seen that the directional derivative of the function $P(x) = \sum_{j=1}^{3} g_j^3(x)$ at $x^*$ is positive in all directions. This implies that we can choose a sufficiently large penalty parameter $c$, so that $x^*$ is a local
Figure 2.5.1. Constraints of Example 2.5.2. 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.

minimum of the function \( F_c(x) \). Therefore, the constraint set admits an exact penalty function 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 2.5.3: Let \( x^* \) be a feasible vector of problem (2.70)-(2.71), 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 \( \lambda_i^* \) and \( \mu_j^* \) satisfying the conditions of Prop. 2.5.1. (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 \( \lambda_i^* \) and \( \mu_j^* \) are Lagrange multipliers. Q.E.D.

Note that because Prop. 2.5.1 does not require regularity of \( X \), the proof of Prop. 2.5.3 can be used to establish that admittance of an exact penalty implies the admittance of \( R \)-multipliers, as defined in Section 2.3. On the other hand, the following example shows that the regularity
assumption on $X$ in Prop. 2.5.3 cannot be dispensed with.

**Example 2.5.3**

Consider the set $X \subset \mathbb{R}^2$ given by

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

(Fig. 2.5.2), and let there be a single linear equality constraint $h(x) = x_1 = 0$. For $x^* = (0,0)$, we have $T_X(x^*)^\bot = \{0\}$, while $N_X(x^*)$ consists of the two rays shown in Fig. regexact. Because $\nabla h(x^*) = (1,0) \notin N_X(x^*)$, pseudonormality is satisfied, and hence by Prop. 2.5.2, 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 $\lambda$, so there is no Lagrange multiplier. Note here that the admittance of an exact penalty does imply the admittance of $R$-multipliers.

**Figure 2.5.2.** Constraints of Example 2.5.3.

The relations shown thus far are summarized in Fig. 2.5.3, which illustrates the unifying role of pseudonormality. In this figure, unless indicated otherwise, the implications cannot be established in the opposite direction without additional assumptions (the exercises provide the necessary examples and counterexamples).

---

**EXERCISES**
Figure 2.5.3. 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. 2.3.1(a), the admittance of Lagrange multipliers implies the admittance of an informative Lagrange multiplier, while by Prop. 2.5.1, pseudonormality implies the admittance of an exact penalty.
Section 2.6 Using the Extended Representation

2.5.1

Consider the problem

\[
\begin{align*}
\text{minimize } & f(x) \\
\text{subject to } & x \in C,
\end{align*}
\]

where

\[
C = X \cap \{x \mid h_i(x) = 0, \ldots, h_m(x) = 0\} \cap \{x \mid g_1(x) \leq 0, \ldots, g_r(x) \leq 0\},
\]

and the exact penalty function

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

where \(c\) is a positive scalar.

(a) Suppose that \(x^*\) is a local minimum of problem (2.78), and that for some given \(c > 0\), \(x^*\) is also a local minimum of \(F_c\) over \(X\). Show that there exists an R-multiplier vector \((\lambda^*, \mu^*)\) for problem (2.78) such that

\[
|\lambda_i^*| \leq c, \quad i = 1, \ldots, m, \quad \mu_j^* \in [0, c], \quad j = 1, \ldots, r.
\]

(b) Derive conditions that guarantee that if \(x^*\) is a local minimum of problem (2.78) and \((\lambda^*, \mu^*)\) is a corresponding Lagrange multiplier vector, then \(x^*\) is also a local minimum of \(F_c\) over \(X\) when Eq. (2.79) holds.

2.6 USING THE EXTENDED REPRESENTATION

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

\[
X = \{x \mid h_i(x) = 0, \ i = m + 1, \ldots, m, \ g_j(x) \leq 0, \ j = r + 1, \ldots, r\}.
\]

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

\[
h_i(x) = 0, \quad i = 1, \ldots, m, \quad g_j(x) \leq 0, \quad j = 1, \ldots, r.
\]

We call this the \textit{extended representation} of \(C\), to contrast it with the representation (2.71), which we call the \textit{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 2.6.1:

(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 \( \lambda_1^*, \ldots, \lambda_m^* \) and \( \mu_1^*, \ldots, \mu_r^* \) satisfying

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

(2.80)

\( \mu_j^* \geq 0, \quad \forall j = 0, 1, \ldots, r, \)

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

where

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

For \( y \in T_X(x^*) \), we have \( \nabla h_i(x^*)'y = 0 \) for all \( i = m + 1, \ldots, m \), and \( \nabla g_j(x^*)'y \leq 0 \) for all \( j = r + 1, \ldots, r \) with \( j \notin \overline{A}(x^*) \). Hence Eq. (2.80) implies that

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

and it follows that \( \lambda_i^*, i = 1, \ldots, m, \) and \( \mu_j^*, j = 1, \ldots, r, \) are Lagrange multipliers for the original representation.

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

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

We have \( F_c(x) = \mathcal{F}_c(x) \) for all \( x \in X \). Hence if \( x^* \) is an unconstrained local minimum of \( \mathcal{F}_c(x) \), 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 \( \mathcal{F}_c(x) \), it is also a local minimum of \( F_c(x) \) over \( x \in X \). Q.E.D.
A generic and important special case where Prop. 2.6.1 is useful is when all the constraints are linear and \( X \) is a polyhedron. Here, the constraint set need not satisfy pseudonormality (see the following example). However, by Prop. 2.4.1, it satisfies pseudonormality in the extended representation, so using Prop. 2.6.1, it admits Lagrange multipliers and an exact penalty at any feasible point in the original representation.

**Example 2.6.1**

Let \( C = \{ x \in X \mid a'x \leq 0, b'x \leq 0 \} \), where \( a = (1, -1) \), \( b = (-1, -1) \), and \( X = \{ x \in \mathbb{R}^2 \mid a'x \geq 0, b'x \geq 0 \} \). The constraint set is depicted in Fig. 2.6.1.

The only feasible point is \( x^* = (0, 0) \). By choosing \( \mu = (1, 1) \), we get 
\[ -(a + b) \in T_X(x^*)^\perp, \]
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 pseudonormal. This constraint set, however, admits Lagrange multipliers at \( x^* = (0, 0) \) with respect to its extended representation (cf. Prop. 2.6.1), and hence it admits Lagrange multipliers at \( x^* = (0, 0) \) with respect to the original representation.

![Figure 2.6.1. Constraints of Example 2.6.1. The only feasible point is \( x^* = (0, 0) \). The tangent cone \( T_X(x^*)^\perp \) and its polar \( T_X(x^*) \) are shown in the figure.](image)

Note that part (a) of Prop. 2.6.1 does not guarantee the existence of informative Lagrange multipliers in the original representation, and indeed
in the following example, there exists an informative Lagrange multiplier in the extended representation, but there exists none in the original representation. For this to happen, of course, the tangent cone $T_X(x^*)$ must be nonconvex, for otherwise Prop. 2.3.1(a) applies.

**Example 2.6.2**

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

$$C = \{ x \in \mathbb{R}^2 \mid a_1'x \leq 0, a_2'x \leq 0, (a_1'x)(a_2'x) = 0 \},$$

where $a_1 = (-1,0)$ and $a_2 = (0,-1)$. Consider the vector $x^* = (0,0)$. It can be verified that this constraint set admits Lagrange multipliers in the extended representation, so it also admits informative Lagrange multipliers, as shown by Prop. 2.3.1(a).

![Figure 2.6.2. Constraints and relevant cones for different representations of the problem in Example 2.6.2.](image)

Now let the same constraint set be specified by the two linear constraint functions $a_1'x \leq 0$ 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 $T_X(x^*)^\perp = \{0\}$. The normal cone $N_X(x^*)$ consists of the coordinate axes. Since $N_X(x^*) \neq T_X(x^*)^\perp$, $X$ is not regular at $x^*$. Furthermore, $T_X(x^*)$ is not convex, so Prop. 2.3.1(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 $-\nabla f(x^*) \in T_C(x^*)^\perp$ (see Fig. 2.6.2). The candidate multipliers are determined from the requirement that

$$-\left( \nabla f(x^*) + \sum_{j=1}^{2} \mu_j a_j \right) \in T_X(x^*)^\perp = \{0\},$$
which uniquely determines $\mu_1$ and $\mu_2$. 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 $a_j^t x \leq 0$ and $a_k^t x \leq 0$, so the multipliers do not qualify as informative. Thus, the constraint set does not admit informative Lagrange multipliers in the original representation.

---

**EXERCISES**

**2.6.1 (Generalized Farkas’ Lemma)**

Let $P$ be a polyhedral cone in $\mathbb{R}^n$, let $a_1, \ldots, a_r$ be given vectors, and let $c$ be a given nonzero vector. Show that we have

$$c^t y \leq 0, \quad \text{for all } y \in P \text{ such that } a_j^t y \leq 0, \; \forall \; j = 1, \ldots, r,$$

if and only if there exist nonnegative scalars $\mu_1, \ldots, \mu_r$ and a vector $\overline{y} \in P$ such that $c^t \overline{y} > 0$, $a_j^t \overline{y} > 0$ for all $j$ such that $\mu_j > 0$, $a_j^t \overline{y} \leq 0$ for all $j$ such that $\mu_j = 0$, and

$$c \in \mu_1 a_1 + \cdots + \mu_r a_r + P^\perp.$$

---

**2.7 EXTENSIONS UNDER CONVEXITY ASSUMPTIONS**

In this section, we extend the theory of the preceding sections to the case where some or all of the functions $f$ and $g_j$ are nondifferentiable but are instead assumed convex. We thus consider the problem

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

subject to $x \in C$, \hfill (2.81)

where

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

(2.82)

and $X$ is a nonempty closed set, the $h_i$ are smooth, and each of the functions $f$ and $g_j$ is either smooth or is convex over $\mathbb{R}^n$. (The following analysis also extends straightforwardly to the case where each of the $f$ and $g_j$ is the sum of a smooth function and a function that is convex over $\mathbb{R}^n$.)
The theory of the preceding sections can be generalized under the assumptions of the present section, once we substitute the gradients of the convex but nondifferentiable functions with subgradients. To see this, we recall the necessary optimality condition given in Section 1.9 for the problem of minimizing a sum $F_1(x) + F_2(x)$ over $X$, where $F_1$ is convex and $F_2$ is smooth: if $x^*$ is a local minimum and the tangent cone $T_X(x^*)$ is convex, then
\[-\nabla F_2(x^*) \in \partial F_1(x^*) + T_X(x^*)^\perp.\] (2.83)

For a smooth convex function $F$, we will use the notation
\[\partial F(x) = \{\nabla F(x)\}\]
even if $F$ is not convex, so the necessary condition (2.83) can be written as
\[0 \in \partial F_1(x^*) + \partial F_2(x^*) + T_X(x^*)^\perp.\] (2.84)

By a nearly verbatim repetition of the proof of Prop. 2.2.1, while using this necessary condition in place of $-\nabla F_k(x_k) \in T_X(x_k)^\perp$, together with the closedness of $N_X(x^*)$, we obtain the following extension.

**Proposition 2.7.1:** Let $x^*$ be a local minimum of problem (2.81)-(2.82). Then, assuming the tangent cone $T_X(x^*)$ is convex, there exist scalars $\mu_0^*, \lambda_1^*, \ldots, \lambda_m^*$, and $\mu_1^*, \ldots, \mu_r^*$, satisfying the following conditions:

(i) $0 \in \mu_0^* \partial f(x^*) + \sum_{i=1}^m \lambda_i^* \partial h_i(x^*) + \sum_{j=1}^r \mu_j^* \partial g_j(x^*) + N_X(x^*)$.

(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
\[I = \{i \mid \lambda_i^* \neq 0\}, \quad J = \{j \neq 0 \mid \mu_j^* > 0\},\]
there exists a sequence $\{x^k\} \subset X$ that converges to $x^*$ and is such that for all $k$,
\[f(x^k) < f(x^*), \quad \lambda_i^* h_i(x^k) > 0, \quad \forall i \in I, \quad \mu_j^* g_j(x^k) > 0, \quad \forall j \in J,\]
\[|h_i(x^k)| = o(w(x^k)), \quad \forall i \notin I, \quad g_j(x^k) = o(w(x^k)), \quad \forall j \notin J,\]
where
\[w(x) = \min \left\{ \min_{i \in I} |h_i(x)|, \min_{j \in J} g_j(x) \right\}.\]
The theory of the preceding sections, regarding pseudonormality, informative Lagrange multipliers, and exact penalty functions, can now be generalized. We first extend the definitions of Lagrange multiplier and pseudonormality.

**Definition 2.7.1:** Consider problem (2.70) under the convexity and smoothness assumptions of this section, and let \( x^* \) be a local minimum. A pair \((\lambda^*, \mu^*)\) is called a *Lagrange multiplier vector corresponding to \( f \) and \( x^* \) if

\[
0 \in \partial f(x^*) + \sum_{i=1}^{m} \lambda_i^* \partial h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \partial g_j(x^*) + T_X(x^*)^\perp, \quad (2.85)
\]

\[
\mu^* \geq 0, \quad \mu^*'g(x^*) = 0. \quad (2.86)
\]

Note here that since \( g(x^*) \leq 0 \) and \( \mu^* \geq 0 \), the condition \( \mu^*'g(x^*) = 0 \) of Eq. (2.86) is equivalent to \( \mu_j^*g_j(x^*) = 0 \) for all \( j \), which is the CS condition.

**Definition 2.7.2:** Consider problem (2.70) under the convexity and smoothness assumptions of this section. A feasible vector \( x^* \) is said to be *pseudonormal* if there exist no \( \lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r \), and a sequence \( \{x^k\} \subset X \) such that:

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

(ii) \( \mu \geq 0 \) and \( \mu'g(x^*) = 0 \).

(iii) \( \{x^k\} \) converges to \( x^* \) and

\[
\sum_{i=1}^{m} \lambda_i h_i(x^k) + \sum_{j=1}^{r} \mu_j g_j(x^k) > 0, \quad \forall \ k.
\]

If a local minimum \( x^* \) is pseudonormal and the tangent cone \( T_X(x^*) \) is convex, by Prop. 2.7.1, there exists a Lagrange multiplier vector, which also satisfies the extra condition (iv) of that proposition.
We now consider the problem
\[
\text{minimize } \ f(x) \\
\text{subject to } \ x \in X, \ g(x) \leq 0,
\]
where \(g(x) = (g_1(x), \ldots, g_r(x))\) is the constraint function vector, and we assume that \(X\) is convex, and the functions \(g_j\) are convex over \(\mathbb{R}^n\). For simplicity, we assume no equality constraints. The extension of the following analysis to the case where there are equality constraints, and the corresponding functions \(h_i\) are linear, is straightforward: we simply replace each equality constraint into two linear (and hence convex) inequality constraints. The enhanced Fritz-John of Prop. 2.7.1 take the following form:

**Proposition 2.7.2:** Consider problem (2.87), assuming that \(X\) is convex, and the functions \(g_j\) are convex over \(\mathbb{R}^n\), and let \(x^*\) be a global minimum. Then there exists a scalar \(\mu^*_0\) and a vector \(\mu^* = (\mu^*_1, \ldots, \mu^*_r)\), satisfying the following conditions:

(i) \(\mu^*_0 f(x^*) = \min_{x \in X} \{\mu^*_0 f(x) + \mu^*_j g(x)\}\).

(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 \(k\),
\[
f(x^k) < f(x^*), \quad \mu^*_j g_j(x^k) > 0, \quad \forall j \in J,
\]
\[
g_j(x^k) = o \left( \min_{l \in J} g_l(x^k) \right), \quad \forall j \notin J.
\]

The proposition above follows from Prop. 2.7.1 and the convexity assumptions on \(X\) and \(g_j\). In particular, condition (i) of Prop. 2.7.1 and the convexity assumptions imply that \(x^*\) minimizes \(\mu^*_0 f(x) + \mu^*_j g(x)\) over \(x \in X\). Furthermore, condition (iv) of Prop. 2.7.1 implies the CS condition \(\mu^*_j g_j(x^*) = 0\) for all \(j\). Hence, we have \(\mu^*_0 f(x^*) = \mu^*_0 f(x^*) + \mu^*_j g(x^*) = \min_{x \in X} \{\mu^*_0 f(x) + \mu^*_j g(x)\}\), which is condition (i) of Prop. 2.7.2.

The definition of a Lagrange multiplier is similarly specialized as follows.
**Definition 2.7.3:** Consider problem (2.87), assuming that $X$ is convex, and the functions $g_j$ are convex over $\mathbb{R}^n$, and let $x^*$ be a global minimum. A vector $\mu^* \geq 0$ is called a *Lagrange multiplier vector* corresponding to $f$ and $x^*$ if

$$
 f(x^*) = \min_{x \in X} \left\{ f(x) + \mu^* g(x) \right\}, \quad \mu^* g(x^*) = 0. \tag{2.88}
$$

Note that given the feasibility of $x^*$, which implies $g(x^*) \leq 0$, and the nonnegativity of $\mu^*$, the condition $\mu^* g(x^*) = 0$ of Eq. (2.88) is equivalent to the complementary slackness condition $\mu^*_j = 0$ for $j \in A(x^*)$ [cf. Eq. (2.86)]. Furthermore, Eq. (2.88) implies that $x^*$ minimizes $f(x) + \mu^* g(x)$ over $X$, which in view of the convexity assumptions, is equivalent to the condition $0 \in \partial f(x^*) + \sum_{j=1}^r \mu^*_j \partial g_j(x^*) + T_X(x^*)^\perp$ [cf. Eq. (2.85)].

Similarly, the definition of pseudonormality is specialized as follows.

**Definition 2.7.4:** Consider problem (2.87), assuming that $X$ is convex, and the functions $g_j$ are convex over $\mathbb{R}^n$. A feasible vector $x^*$ is said to be *pseudonormal* if there exist no vector $\mu = (\mu_1, \ldots, \mu_r) \geq 0$, 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. 2.7.2, there exists a Lagrange multiplier vector, which also satisfies the extra CV condition (iv) of that proposition.

The analysis of Section 2.4 is easily extended to show that $x^*$ is pseudonormal under either one of the following two criteria:

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

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

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

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

If $X$ is a polyhedron (rather than $X = \mathbb{R}^n$) 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 $X$ (see also the analysis of the next chapter).

Finally, let us provide an intuitive interpretation of pseudonormality in the convex case. Consider the set

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

and hyperplanes that support this set at $g(x^*)$. As Fig. 2.7.1 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 $H$,
   $$\overline{H} = \{z \mid \mu'z \geq 0\}.$$

Figure 2.7.1 also indicates the type of constraint qualifications that guarantee pseudonormality. The Slater condition can be rephrased to mean that the set $G$ does not intersect the interior of the negative 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 $X = \mathbb{R}^n$ and the $g_j$ are linear, the set $G$ is a linear manifold, 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.

## 2.8 NOTES AND SOURCES

Lagrange multipliers were originally introduced for problems with equality constraints, while inequality constrained problems were addressed considerably later. Important early works are those of Karush [Kar39] (an unpublished MS thesis), John [Joh48], and Kuhn and Tucker [KuT51]. The survey by Kuhn [Kuh76] gives a historical view of the development of the subject.

There has been much work in the 60s and early 70s on deriving constraint qualifications. Important examples are those of Arrow, Hurwicz, and Uzawa [AHU61], Abadie [Aba67], Mangasarian and Fromovitz, and
Sec. 2.8 Notes and Sources

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

\( g(x^*) \) is pseudonormal if \( \mu^* \) is pseudonormal and \( x^* \) is not pseudonormal.

**Figure 2.7.1.** 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 hyperplane with normal \( \mu \), 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 \( \{ g(x_k) \} \subset G \cap \text{int}(\overline{H}) \), 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 a linear manifold and the second condition cannot be satisfied (this depends critically on \( X = \mathbb{R}^n \) rather than \( X \) being a general polyhedron).

[MaF67], and Guignard [Gui69]. For textbook treatments see Mangasarian [Man69] and Hestenes [Hes75]. There has been much subsequent work on the subject, some of which addresses nondifferentiable problems, e.g., Gould and Tolle [GoT71], [GoT72], Bazaraa, Goode, and Shetty [BGS72],
Clarke [Cla83], Demjanov and Vasil’ev [DeV85], Mordukhovich [Mor88], and Rockafellar [Roc93].

Most of the preceding works consider problems with equality and inequality constraints only. Abstract set constraints (in addition to equality and inequality constraints) have been considered along two different lines:

(a) For convex programs (convex $f$, $g_j$, and $X$, and linear $h_i$), and in the context of the geometric multiplier theory to be developed in Chapter 3. Here the abstract set constraint does not cause significant complications, because for convex $X$, the tangent cone is conveniently defined in terms of feasible directions, and nonsmooth analysis issues of nonregularity do not arise.

(b) For the nonconvex setting of this chapter, where the abstract set constraint causes significant difficulties because the classical approach that is based on quasiregularity is not fully satisfactory.

This has motivated alternative approaches. The work of Rockafellar [Roc93], [RoW98] was instrumental in introducing nonsmooth analysis concepts, such as the normal cone of Mordukhovich [Mor76] and related work by Clarke (see e.g., the book [Cla83]), in Lagrange multiplier theory. Rockafellar used the Lagrange multiplier definition given in Section 2.3 (what we have called R-multiplier), but he did not develop or use the main ideas of this chapter, i.e., the enhanced Fritz-John conditions, informative Lagrange multipliers, and pseudonormality. Instead he assumed the constraint qualification CQ6, which, as discussed in Section 2.4, is restrictive because, when $X$ is regular, it implies that the set of Lagrange multipliers is not only nonempty but also compact.

The material of this chapter is based on the author’s joint work with Asuman Koksal [BeK00a], [BeK00b]. This work derived the enhanced Fritz-John conditions, introduced the notion of an informative Lagrange multiplier and the notion of pseudonormality, and established the constraint qualification results of Section 2.3 and the exact penalty results of Section 2.4.

The penalty-based line of proof of the enhanced Fritz-John conditions originated with McShane [McS73]. Hestenes [Hes75] observed that McShane’s proof can be used to strengthen the CS condition to assert the existence, within any neighborhood $B$ of $x^*$, of an $x \in B \cap X$ such that

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

which is slightly weaker than CV as defined here [there is no requirement that $x$, simultaneously with violation of the constraints with nonzero multipliers, satisfies $f(x) < f(x^*)$ and Eq. (2.13)]. McShane and Hestenes considered only the case where $X = \mathbb{R}^n$. 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) were given in Bertsekas and Koksal [BeK00a], [BeK00b].

A notion that is related to pseudonormality, called *quasinormality* and implied by pseudonormality, is given by Hestenes [Hes75] (for the case where $X = \mathbb{R}^n$) and Bertsekas [Ber99] (for the case where $X$ is a closed convex set). The relation between pseudonormality and quasinormality is discussed in [BeK00a], where it is argued that pseudonormality is better suited as a unifying vehicle for Lagrange multiplier theory.

The material on the relation between pseudonormality and quasiregularity (Section 2.4) is from Koksal and Bertsekas [KoB01]. This work is based on the following extension of the notion of quasiregularity in a more general setting where $X$ may be a strict subset of $\mathbb{R}^n$:

$$T_C(x^*) = V(x^*) \cap T_X(x^*), \quad (2.89)$$

where $V(x^*)$ is the cone of first order feasible variations [cf. Eq. (2.63)]. It is shown in [KoB01] that under a regularity assumption on $X$, quasiregularity is implied by pseudonormality (cf. Prop. 2.4.4). It is also shown in [KoB01] that contrary to the case where $X = \mathbb{R}^n$, quasiregularity is not sufficient to guarantee the existence of a Lagrange multiplier. Thus the importance of quasiregularity, the classical pathway to Lagrange multipliers when $X = \mathbb{R}^n$, diminishes when $X \neq \mathbb{R}^n$. By contrast, pseudonormality provides satisfactory unification of the theory. What is happening here is that the constraint set admits Lagrange multipliers at $x^*$ if and only if

$$T_C(x^*)^\perp = T_X(x^*)^\perp + V(x^*)^\perp; \quad (2.90)$$

this is a classical result derived in various forms by Gould and Tolle [GoT72], Guignard [Gui69], and Rockafellar [Roc93]. When $X = \mathbb{R}^n$, this condition reduces to $T_C(x^*)^\perp = V(x^*)^\perp$, and is implied by quasiregularity [$T_C(x^*) = V(x^*)$]. However, when $X \neq \mathbb{R}^n$ the natural definition of quasiregularity, given by Eq. (2.89), does not imply the Lagrange multiplier condition (2.90), unless substantial additional assumptions are imposed.

REFERENCES


