Contents

CHAPTER 1

INTRODUCTION

Optimization theory arises in a vast variety of problems. Engineers, managers, operation researchers are constantly faced with problems that need optimal decision making. In the past, a wide range of solutions was considered acceptable. However, the rapid increase in human needs and objectives is forcing us to make more efficient use of our scarce resources. This is making optimization techniques critically important in a wide range of areas.

Mathematical models for these optimization problems can be constructed by specifying a *constraint set* C, which consists of the available decisions x, and a *cost or objective function* $f(x)$ that maps each $x \in C$ into a scalar and represents a measure of undesirability of choosing decision x. This problem can then be written as

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C. \end{aligned} \tag{0.1}$$

In this thesis, we focus on the case where each decision x is an n-dimensional vector, i.e., x is an n-tuple of real numbers $(x_1, \ldots, x_n)$. Hence, the constraint set C is a subset of $\Re^n$, the n-dimensional Euclidean space. We assume throughout the thesis (with the exception of the last chapter where we use some convexity assumptions instead) that the function $f : \Re^n \mapsto \Re$ is a smooth (continuously differentiable) function. A vector x that belongs to set C is referred to as a *feasible solution* of problem (0.1). We want to find a feasible vector x^* that satisfies

$$f(x^*) \leq f(x), \qquad \text{for all } x \in C.$$

We call such a vector a *global optimal solution* (or *global minimum*) of problem (0.1), and the corresponding cost value $f(x^*)$ the *optimal value* (or *optimal cost*) of problem (0.1). A vector x^* is called a *local optimum solution* (or *local minimum*) if it is no worse than its neighbors, i.e., if there exists some scalar $\epsilon > 0$ such that

$$f(x^*) \leq f(x), \qquad \text{for all } x \in C \text{ with } \|x - x^*\| \leq \epsilon.$$

The global or local minimum x^* is said to be *strict* if the corresponding inequalities above are strict for all $x \in C$ with $x \neq x^*$.

The optimization problem stated in (0.1) is very broad and arises in a large variety of practical applications. This problem contains as special cases several important classes of problems. In *nonlinear programming problems* either the cost function f is nonlinear or the constraint set C is specified by nonlinear equations and inequalities. In *linear programming problems*, the cost function f is linear and the constraint set C is a polyhedron. Both classes of problems have a vast range of applications, such as communication, manufacturing, production planning, scheduling, logistics, and pattern classification.

Another major class of problems is *network flow problems*. Network flow problems are one of the most important and most frequently encountered class of optimization problems. They arise naturally in the analysis and design of large systems, such as communication, transportation, and manufacturing networks. They can also be used to model important classes of combinatorial problems, such as assignment, shortest path and travelling salesman problems. Loosely speaking, network flow problems consist of supply and demand points, together with several routes that connect these points and are used to transfer the supply to the demand. Often the supply, demand, and intermediate points can be modelled by the nodes of a graph, and the routes may be modelled by the paths of the graph. Furthermore, there may be multiple types of supply/demand (or commodities) sharing the routes. For example, in communication networks, the commodities are the streams of different classes of traffic (telephone, data, video, etc.) that involve different origin-destination pairs. In such problems, roughly speaking, we try to select routes that minimize the cost of transfer of the supply to the demand.

A fundamental issue that arises in attempting to solve problem (0.1) is the characterization of optimal solutions via necessary and sufficient optimality conditions. Optimality conditions often provide the basis for the development and the analysis of algorithms. In general, algorithms iteratively improve the current solution, converging to a solution that approximately satisfy various optimality conditions. Hence, having optimality conditions that are rich in supplying information about the nature of potential solutions is important for suggesting variety of algorithmic approaches.

Necessary optimality conditions for problem (0.1) can be expressed generically in terms of the relevant conical approximations of the constraint set C. On the other hand, the constraint set of an optimization problem is usually specified in terms of equality and inequality constraints. In this work, we adopt a more general approach and assume that the constraint set C consists of equality and inequality constraints as well as an additional abstract set constraint $x \in X$:

$$C = X \cap \big\{x \mid h_1(x) = 0, \dots, h_m(x) = 0\big\} \cap \big\{x \mid g_1(x) \leq 0, \dots, g_r(x) \leq 0\big\}. \tag{0.2}$$

In this thesis, the constraint functions h_i and g_j are assumed to be smooth functions from $\Re^n$ to $\Re$ (except in the last chapter where we have various convexity assumptions instead). An abstract set constraint in this model represents constraints in the optimization problem that cannot be represented by equalities and inequalities. It is also convenient in representing simple conditions for which explicit introduction of constraint functions would be cumbersome, for instance, sign restrictions or bounds on the components of the decision vector x.

If we take into account the special structure of the constraint set, we can obtain more refined optimality conditions, involving some auxiliary variables called *Lagrange multipliers*. These multipliers facilitate the characterization of optimal solutions and often play an important role in computational methods. They provide valuable sensitivity information, quantifying up to first order the variation of the optimal cost caused by variations in problem data. They also play a significant role in duality theory, a central theme in nonlinear optimization.

1.1. ROLE OF LAGRANGE MULTIPLIERS

Lagrange multipliers have long been used in optimality conditions for problems with constraints, but recently, their role has come to be understood from many different angles. The theory of Lagrange multipliers has been one of the major research areas in nonlinear optimization and there has been a variety of different approaches. Lagrange multipliers were originally introduced for problems with equality constraints. Inequality-constrained problems were addressed considerably later. Let us first highlight the traditional rationale for illustrating the importance of Lagrange multipliers by considering a problem with equality constraints of the form

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h_i(x) = 0, \qquad i = 1, \ldots, m. \end{aligned} \tag{1.1}$$

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

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0. \tag{1.2}$$

This implies that at a local optimal solution x^*, the cost gradient $\nabla f(x^*)$ is orthogonal to the *subspace of first order feasible variations*

$$V(x^*) = \big\{ \Delta x \mid \nabla h_i(x^*)' \Delta x = 0, \ i = 1, \ldots, m \big\}. \tag{1.3}$$

This is the subspace of variations Δx from the optimal solution x^*, for which the resulting vector satisfies the constraints $h_i(x) = 0$ up to first order. Hence, condition (1.2) implies that, at the local minimum x^*, the first order cost variation $\nabla f(x^*)'\Delta x$ is zero for all variations Δx in the subspace $V(x^*)$. This is analogous to the "zero cost gradient" condition of unconstrained optimization problems. This interpretation is illustrated in Figure 1.1.1.

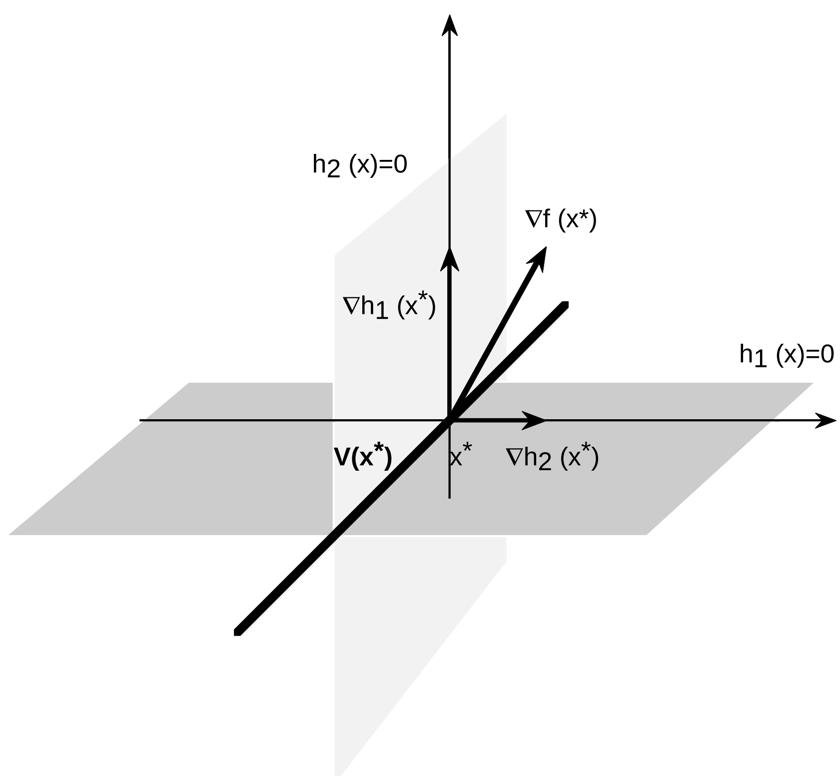


Figure 1.1.1. Illustration of the Lagrange multiplier condition (1.2) for an equality-constrained problem. The cost gradient $\nabla f(x^*)$ belongs to the subspace spanned by the constraint gradients at x^*, or equivalently, the cost gradient $\nabla f(x^*)$ is orthogonal to the subspace of first order feasible variations at x^*, $V(x^*)$.

Lagrange multiplier conditions, given in Eq. (1.2), are n equations which together with the m constraints $h_i(x^*) = 0$, constitute a system of $n + m$ equations with $n +$

m unknowns, the vector x^* and the multipliers λ_i^*. Thus, through the use of Lagrange multipliers, a constrained optimization problem can be "transformed" into a problem of solving a system of nonlinear equations. While this is the role in which Lagrange multipliers were seen traditionally, this viewpoint is certainly naive since solving a system of nonlinear equations numerically is not easier than solving an optimization problem by numerical means. In fact, nonlinear equations are often solved by converting them into nonlinear least squares problems and using optimization techniques. Still, most computational methods in nonlinear programming almost invariably depends on some use of Lagrange multipliers.

Lagrange multipliers also have interesting interpretations in different contexts.

1.1.1. Price Interpretation of Lagrange Multipliers

Lagrange multipliers can be viewed as the "equilibrium prices" of an optimization problem. This interpretation forms an important link between mathematics and theoretical economics. [1]

To illustrate this interpretation, we consider an inequality-constrained problem,

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; g_j(x) \le 0, \qquad j = 1, \dots, r, \end{aligned} \tag{1.4}$$

and assume that the functions f, g_j are smooth and convex over $\Re^n$, and that the optimal value of this problem is finite. The Lagrange multiplier condition for this problem is that, under appropriate assumptions, at a given global minimum x^*, there exist nonnegative multipliers $\mu_1^*, \dots, \mu_r^*$ such that

$$\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0, \tag{1.5}$$

where the μ_j^* satisfy the *complementary slackness* condition:

$$\mu_j^* g_j(x^*) = 0, \qquad \forall\, j = 1, \dots, r,$$

[1] According to David Gale [Gal67], Lagrange multipliers provide the "single most important tool in modern economic analysis both from the theoretical and computational point of view."

i.e., the only constraint gradients associated with nonzero multipliers in condition (1.5) correspond to constraints for which $g_j(x^*) = 0$.[2] [We call constraints for which $g_j(x^*) = 0$, the *active constraints at* x^*.]

Under the given convexity assumptions, it follows that the Lagrange multiplier condition (1.5) is a sufficient condition for x^* to be a global minimum of the function $f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x)$. Together with complementary slackness condition, this implies that

$$f(x^*) = \inf_{x \in \Re^n} \left\{ f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \right\}. \tag{1.6}$$

We next consider a perturbed version of problem (1.1) for some $u = (u_1, \ldots, u_r)$ in $\Re^r$:

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g_j(x) \le u_j, \qquad j = 1, \ldots, r. \end{aligned} \tag{1.7}$$

We denote the optimal value of the perturbed problem by $p(u)$. Clearly, $p(0)$ is the optimal value of the original problem (1.1). Considering vector $u = (u_1, \ldots, u_r)$ as perturbations of the constraint functions, we call the function p as the *perturbation function* or the *primal function.*

We interpret the value $f(x)$ as the "cost" of choosing the decision vector x. Thus, in the original problem (1.4), our objective is to minimize the cost subject to certain constraints. We also consider another scenario in which we are allowed to relax the constraints to our advantage by buying perturbations u. In particular, assume that we are allowed to change problem (1.4) to a perturbed problem (1.7) for any u that we want, with the condition that we have to pay for the change, the price being μ_j per unit of perturbation variable. Then, for any perturbation u, the minimum cost we can achieve in the perturbed problem (1.7), plus the perturbation cost, is given by

$$p(u) + \sum_{j=1}^{r} \mu_j u_j,$$

[2] The name *complementary slackness* comes from the analogy that for each j, whenever the constraint $g_j(x^*)$ is slack [meaning that $g_j(x^*) < 0$], the constraint $\mu_j^* \ge 0$ must not be slack [meaning that $\mu_j^* > 0$], and vice versa.

and we have

$$\inf_{u\in\Re^r}\left\{p(u)+\sum_{j=1}^r \mu_j u_j\right\}\le p(0)=f(x^*),$$

i.e., the minimum cost that can be achieved by a perturbation is at most as high as the optimal cost of the original unperturbed problem. A perturbation is worth buying if we have strict inequality in the preceding relation.

We claim that the Lagrange multipliers μ_j^* are the prices for which no perturbation would be worth buying, i.e., we are in an equilibrium situation such that we are content with the constraints as given. To see this, we use Eq. (1.6) to write

$$\begin{aligned}
p(0)=f(x^*)&=\inf_{x\in\Re^n}\left\{f(x)+\sum_{j=1}^r \mu_j^* g_j(x)\right\}\\
&=\inf_{\{(u,x)\mid u\in\Re^r,\ x\in\Re^n,\ g_j(x)\le u_j,\ j=1,\ldots,r\}}\left\{f(x)+\sum_{j=1}^r \mu_j^* g_j(x)\right\}\\
&=\inf_{\{(u,x)\mid u\in\Re^r,\ x\in\Re^n,\ g_j(x)\le u_j,\ j=1,\ldots,r\}}\left\{f(x)+\sum_{j=1}^r \mu_j^* u_j\right\}\\
&=\inf_{u\in\Re^r}\left\{p(u)+\sum_{j=1}^r \mu_j^* u_j\right\}\\
&\le p(0).
\end{aligned}$$

Hence, equality holds throughout in the above, proving our claim that the Lagrange multipliers are the *equilibrium prices.*

1.1.2. Game Theoretic Interpretation and Duality

Under suitable convexity assumptions, Lagrange multipliers take on a game-theoretic role, which was motivated by the creative insights of von Neumann in applying mathematics to models of social and economic conflict [Neu28], [NeM44].

To put things into perspective, let us consider the following general scenario. Let X and Z be subsets of $\Re^n$ and $\Re^r$, respectively, and let $\phi : X\times Z \mapsto \Re$ be a function. Consider a *zero sum game*, defined in terms of ϕ, X, and Z as follows: There are two players. X is the "strategy set" for the first player , Z is the "strategy set" for the second player, and ϕ is the "payoff function". The game proceeds as follows:

(1) First player selects an element $x \in X$, and second player selects an element $z \in Z$.

(2) The choices are revealed simultaneously,

(3) At the end, first player pays an amount of $\phi(x, z)$ to the second player. [1]

The following definition provides a concept that defines an equilibrium situation in this game.

Definition 1.1.1: A pair of vectors $x^* \in X$ and $z^* \in Z$ is called a saddle point of the function ϕ if

$$\phi(x^*, z) \leq \phi(x^*, z^*) \leq \phi(x, z^*), \qquad \forall\, x \in X,\ \forall\, z \in Z, \tag{1.8}$$

or equivalently,

$$\sup_{z \in Z} \phi(x^*, z) = \phi(x^*, z^*) = \inf_{x \in X} \phi(x, z^*).$$

Given a saddle point (x^*, z^*) of the function ϕ, by choosing x^*, the first player is guaranteed that no matter what player two chooses, the payment cannot exceed the amount $\phi(x^*, z^*)$ [cf. Eq. (1.8)]. Similarly, by choosing z^*, the second player is guaranteed to receive at least the same amount regardless of the choice of the first player. Hence, the saddle point concept is associated with an approach to the game in which each player tries to optimize his choice against the worst possible selection of the other player.

This idea motivates the following equivalent characterization of a saddle point in terms of two optimization problems (for the proof see [BNO02]).

[1] Although this model is very simple, a wide variety of games can be modelled this way (chess, poker etc.). The amount $\phi(x, z)$ can be negative, which corresponds to the case that the first player wins the game. The name of the game "zero sum" derives from the fact that the amount won by either player is the amount lost by the other player.

Proposition 1.1.1: A pair (x^*, z^*) is a saddle point of ϕ if and only if x^* is an optimal solution of the problem

$$\begin{aligned} &\text{minimize } \sup_{z \in Z} \phi(x, z) \\ &\text{subject to } \; x \in X, \end{aligned} \tag{1.9}$$

while z^* is an optimal solution of the problem

$$\begin{aligned} &\text{maximize } \inf_{x \in X} \phi(x, z) \\ &\text{subject to } \; z \in Z, \end{aligned} \tag{1.10}$$

and the optimal value of the two problems are equal, i.e.,

$$\sup_{z \in Z} \inf_{x \in X} \phi(x, z) = \inf_{x \in X} \sup_{z \in Z} \phi(x, z).$$

In the worst case scenario adopted above, problem (1.9) can be regarded as the optimization problem of the first player used to determine the strategy to be selected. Similarly, problem (1.10) is the optimization problem of the second player to determine its strategy.

Equipped with this general scenario, let us consider the inequality-constrained problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } \; g_j(x) \leq 0, \qquad j = 1, \ldots, r, \end{aligned} \tag{1.11}$$

and introduce the, so called, *Lagrangian function*

$$L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x),$$

for this problem. It can be seen that

$$\sup_{\mu \geq 0} L(x, \mu) = \begin{cases} f(x) & \text{if } g_j(x) \leq 0 \text{ for all } j = 1, \ldots, r, \\ \infty & \text{otherwise.} \end{cases}$$

Hence, the original problem (1.11) can be written in terms of the Lagrangian function as

$$\begin{aligned} &\text{minimize } \sup_{\mu \geq 0} L(x, \mu) \\ &\text{subject to } \; x \in \Re^n. \end{aligned} \tag{1.12}$$

In the game-theoretic setting constructed above, this problem can be regarded as the strategy problem corresponding to the first player. The strategy problem corresponding to the second player is

$$\begin{aligned} \text{maximize} \ & \inf_{x \in \Re^n} L(x, \mu) \\ \text{subject to} \ & \mu \geq 0. \end{aligned} \tag{1.13}$$

Let (x^*, μ^*) be a saddle point of the Lagrangian function $L(x, \mu)$. By Proposition 1.1.1, it follows that x^* is the optimal solution of problem (1.12), and μ^* is the optimal solution of problem (1.13), and using the equivalence of problem (1.12) with the original problem (1.11), we have

$$f(x^*) = \inf_{x \in \Re^n} L(x, \mu^*) = \inf_{x \in \Re^n} \left\{ f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \right\}.$$

Using the necessary optimality condition for unconstrained optimization, this implies that

$$\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0, \tag{1.14}$$

and

$$\mu_j^* g_j(x^*) = 0, \qquad \forall \ j = 1, \ldots, r,$$

showing that $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ is a Lagrange multiplier for problem (1.11). Hence, assuming that the Lagrangian function has a saddle point (x^*, μ^*), a game-theoretic approach provides an alternative interpretation for Lagrange multipliers, as the optimal solution of a related optimization problem, which is called the *dual problem* [cf. Eq. (1.13)]. Conditions under which the Lagrangian function has a saddle point, or under which the optimal values of the problems (1.12) and (1.13) are equal form the core of duality theory. A detailed analysis of this topic can be found in [BNO02].

1.1.3. Sensitivity Analysis

Within the mathematical model of Eqs. (0.1)-(0.2), Lagrange multipliers can be viewed as rates of change of the optimal cost as the level of constraint changes [cf. Figure 1.1.2].

To motivate the idea, let us consider a problem with a single linear equality constraint,

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad a'x = b, \end{aligned}$$

where $a \neq 0$. Here, x^* is a local minimum and λ^* 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 Δx and Δb 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 λ^* gives the rate of optimal cost decrease as the level of constraint increases. [1]

When the constraints are nonlinear, the sensitivity interpretation of Lagrange multipliers is still 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]). In this thesis, we provide a sensitivity interpretation of Lagrange multipliers for general nonlinear optimization problems under a weak set of assumptions.

---

[1] This information is very useful in engineering design applications. Suppose that we have designed a system that involves determining the values of a large number of components to satisfy certain objectives, and we are allowed to tune up some of the parameters in order to improve the performance of the system. Instead of solving the problem every time we change the value of a parameter, we can use the information provided by the Lagrange multipliers of this problem to see the resulting impact on the performance.

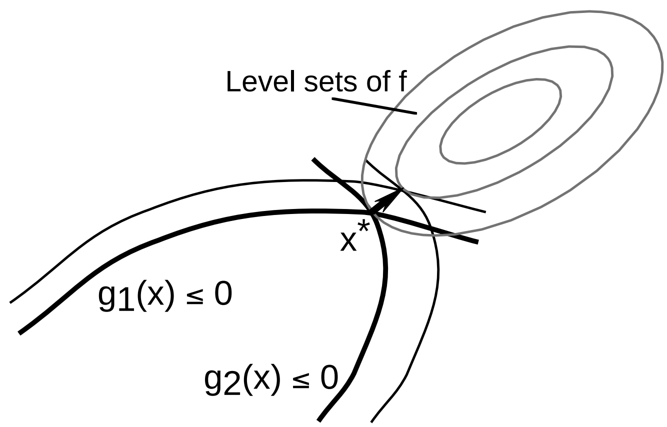


Figure 1.1.2. Sensitivity interpretation of Lagrange multipliers. Suppose that we have a problem with two inequality constraints, $g_1(x) \leq 0$ and $g_2(x) \leq 0$, where the optimal solution is denoted by x^*. If the constraints are perturbed a little, the optimal solution of the problem changes. Under certain conditions, Lagrange multipliers can be shown to give the rates of change of the optimal cost as the level of constraint changes.

1.2. CONSTRAINT QUALIFICATIONS

As we have seen in the previous section, Lagrange multipliers hold fundamental significance in a variety of different areas in optimization theory. However, not every optimization problem can be treated using Lagrange multipliers and additional assumptions on the problem structure are required to guarantee their existence, as illustrated by the following example.

Example 1.2.1: (A Problem with No Lagrange Multipliers)

Consider the problem of minimizing

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

subject to two equality constraints

$$\begin{aligned} h_1(x) &= x_1^2 - x_2 = 0, \\ h_2(x) &= x_1^2 + x_2 = 0. \end{aligned}$$

The geometry of this problem is illustrated in Figure 1.2.3. The only feasible solution is $x^* = (0, 0)$, which is therefore the optimal solution of this problem. It can be seen that at the local minimum $x^* = (0,0)$, the cost gradient $\nabla f(x^*) = (1,1)$ cannot be expressed as a linear combination of the constraint gradients $\nabla h_1(x^*) = (0,-1)$ and $\nabla h_2(x^*) = (0,1)$. 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 λ_1^* and λ_2^*.

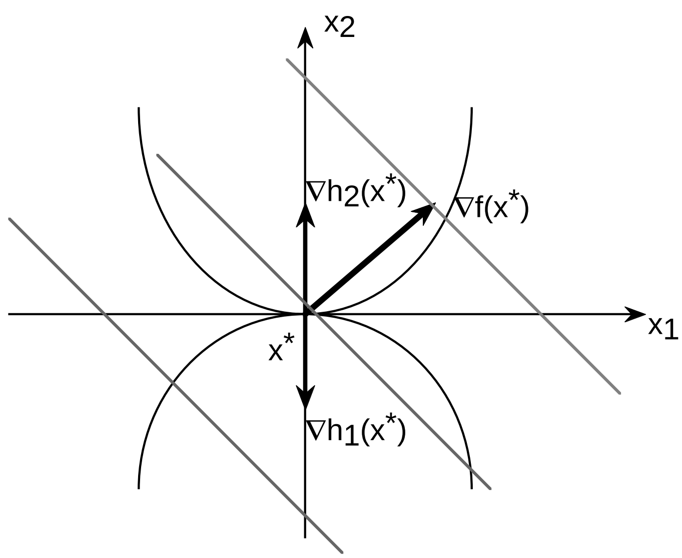


Figure 1.2.3. Illustration of how Lagrange multipliers may not exist for some problems (cf. Example 1.2.1). Here the cost gradient can not be expressed as a linear combination of the constraint gradients, so there are no Lagrange multipliers.

The difficulty in this example is that the subspace of first order feasible variations

$$V(x^*) = \big\{ y \mid \nabla h_1(x^*)'y = 0,\ \nabla h_2(x^*)'y = 0 \big\}$$

[cf. Eq. (1.3)], which is $\{y \mid y_1 \in \Re,\ y_2 = 0\}$, has larger dimension than the true set of feasible variations $\{y \mid y = 0\}$. The optimality of x^* implies that $\nabla f(x^*)$ is orthogonal to the true set of feasible variations, but for a Lagrange multiplier to exist, $\nabla f(x^*)$ must be orthogonal to the subspace of first order of feasible variations. This problem would not have occurred if the constraint gradients $\nabla h_1(x^*)$ and $\nabla h_2(x^*)$ were linearly independent, since then there would not be a mismatch between the set of feasible variations and the set of first order feasible variations.

A fundamental research question in nonlinear optimization is to determine the type of qualifications that are needed to be satisfied by a problem so that Lagrange multipliers can be of use in its analysis. Such conditions can be meaningful if they are independent of the cost function, so that when they hold, the same results can be inferred for any other cost function with the same optimality properties. Hence, it is the constraint set of an optimization problem that needs to have additional structure for the existence of Lagrange multipliers.

There has been much interest in developing general and easily verifiable conditions that guarantee the existence of Lagrange multipliers for a problem. There are a large number of such conditions developed in the 60s and early 70s, for problems with smooth equality and inequality constraint functions, which are often referred to as *constraint qualifications*. Modern applications require using more general optimization models with more complicated side conditions [cf. Eqs. (0.1)-(0.2)]. Analysis of such optimization problems demands a more sophisticated and deeply understood theory of Lagrange multipliers. Developing such a unified and extended theory is one of the main themes of this thesis.

1.2.1. Linear Equality Constraints

To see why Lagrange multipliers may be expected to exist for some problems, let us consider a simple equality-constrained problem where the equality constraint functions h_i are linear so that

$$h_i(x) = a_i' x = 0, \qquad i = 1, \ldots, m,$$

for some vectors a_i [cf. problem (1.1)]. To analyze this problem, we make use of the well-known necessary optimality condition for optimization over a convex set (for the proof, see [Ber99]).

Proposition 1.2.2: Let X be a convex set. If x^* is a local minimum of f over X, then

$$\nabla f(x^*)'(x - x^*) \geq 0, \qquad \forall\, x \in X.$$

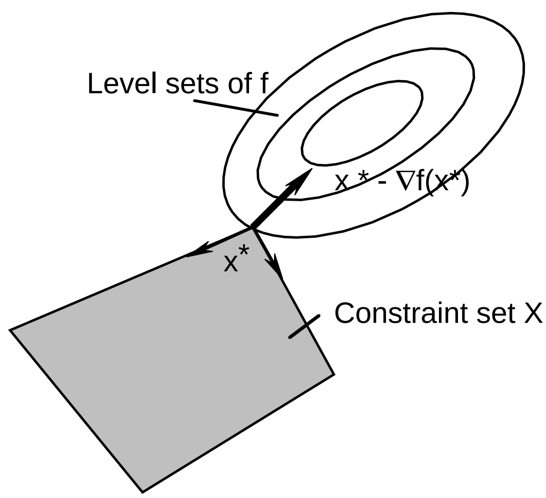


Figure 1.2.4. Illustration of the necessary optimality condition

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

for x^* to be a local minimum of f over X.

Geometric interpretation of this result is illustrated in Figure 1.2.4. Hence, at a given local minimum x^* of the above linear equality-constrained problem, we have

$$\nabla f(x^*)'(x - x^*) \geq 0, \qquad \forall\, x \text{ such that } a_i'x = 0,\ \forall\, i = 1, \ldots, m.$$

The feasible set of this problem is given by the nullspace of the $m \times n$ matrix A having as rows the a_i, which we denote by $N(A)$. By taking $x = 0$ and $x = 2x^*$ in the preceding relation, it is seen that

$$\nabla f(x^*)'x^* = 0.$$

Combining the last two relations, we obtain $\nabla f(x^*)'x \geq 0$ for all $x \in N(A)$. Since for all $x \in N(A)$, we also have $-x \in N(A)$, it follows that $\nabla f(x^*)'x = 0$ for all $x \in N(A)$.

Therefore, $\nabla f(x^*)$ belongs to the range space of the matrix having as columns the a_i, and can be expressed as a linear combination of the a_i. Hence, we can write

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

for some scalars λ_i, which implies the existence of Lagrange multipliers.

In the general case, where the constraint functions h_i are nonlinear, additional assumptions are needed to guarantee the existence of Lagrange multipliers. One such condition, called *regularity* of x^*, is that the gradients $\nabla h_i(x^*)$ are linearly independent, as hinted in the discussion following Example 1.2.1. We will digress into this topic in more detail in Chapter 4.

1.2.2. Fritz John Conditions

Over the years, there has been considerable research effort in deriving optimality conditions involving Lagrange multipliers under different constraint qualifications. Necessary optimality conditions for constrained problems that involve Lagrange multipliers were first presented in 1948 by John [Joh48]. These conditions are known as *Fritz John necessary conditions*. These conditions assume no qualification, instead involves an additional multiplier for the cost gradient in their statement. (An excellent historical review of optimality conditions for nonlinear programming can be found in [Kuh76].[1])

[1] The following quotation from Takayama [Tak74] gives an accurate account of the history of these conditions. “Linear programming aroused interest in constraints in the form of inequalities and in the theory of linear inequalities and convex sets. The Kuhn-Tucker study appeared in the middle of this interest with a full recognition of such developments. However, the theory of nonlinear programming when the constraints are all in the form of equalities has been known for a long time– in fact since Euler and Lagrange. The inequality constraints were treated in a fairly satisfactory manner already in 1939 by Karush. Karush’s work is apparently under the influence of a similar work in the calculus of variations by Valentine. Unfortunately, Karush’s work has been largely ignored. Next to Karush, but

To get a sense of the main idea of Fritz John conditions, we consider the equality-constrained problem

$$\text{minimize } f(x)$$
$$\text{subject to } \; h_i(x) = 0, \qquad i = 1, \ldots, m.$$

There are two possibilities at a local minimum x^*:

(a) The gradients $\nabla h_i(x^*)$ are linearly independent (x^* is regular). Then, there exist scalars (Lagrange multipliers) $\lambda_1^*, \ldots, \lambda_m^*$ 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

$$\mu_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0. \tag{2.1}$$

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 condition (2.1) provides no information regarding the existence of Lagrange multipliers.

Fritz John conditions can also be extended to inequality-constrained problems, and they hold without any further assumptions on x^* (such as regularity). However, this extra

---

still prior to Kuhn and Tucker, Fritz John considered the nonlinear programming problem with inequality constraints. He assumed no qualification except that all functions are continuously differentiable. Here the Lagrangian expression looks like $\mu_0 f(x) + \mu' g(x)$ instead of $f(x) + \mu' g(x)$ and μ_0 can be zero in the first order conditions. The Karush-Kuhn-Tucker constraint qualification amounts to providing a condition which guarantees $\mu_0 > 0$ (that is, a normality condition)."

generality comes at a price, because the issue of whether the cost multiplier μ_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 of Lagrange multiplier theory, as we will see in Chapter 3.

1.3. EXACT PENALTY FUNCTIONS

An important analytical and algorithmic technique in nonlinear programming to solve problem (0.1)-(0.2) involves the use of penalty functions. The basic idea in penalty methods is to eliminate the equality and inequality constraints and add to the cost function a penalty term that prescribes a high cost for their violation. Associated with the penalty term is a parameter c that determines the severity of the penalty and as a consequence, the extent to which the "penalized" problem approximates the original. An important example is the quadratic penalty function

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

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

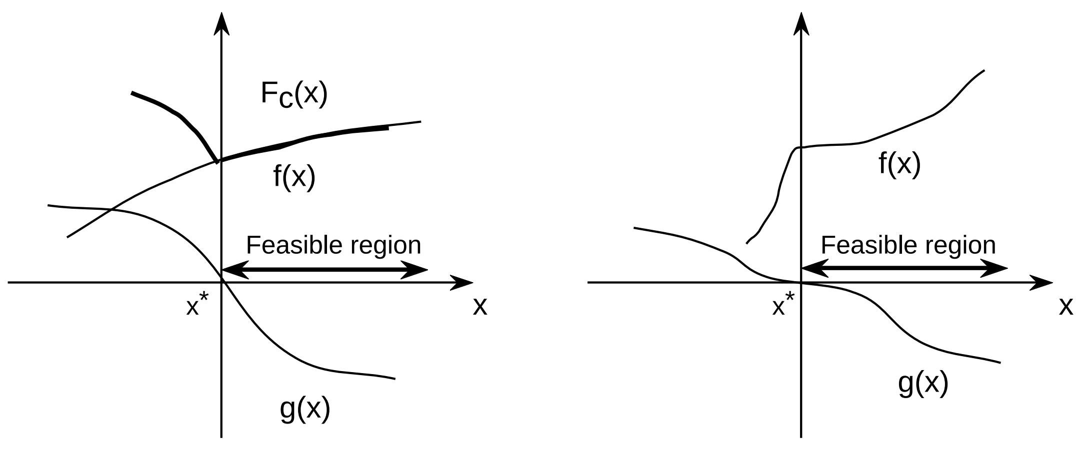
$$g_j^+(x) = \max\{0, g_j(x)\}.$$

Instead of the original optimization problem (0.1)-(0.2), consider minimizing this function over the set constraint X. For large values of c, a high penalty is incurred for infeasible points. Therefore, 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 throughout the thesis.

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 for certain problems, x^* is also a local minimum of F_c, provided that c is larger than some threshold value. This idea is depicted in Figure 1.3.5.



(a) (b)

Figure 1.3.5. Illustration of an exact penalty function for the case of one-dimensional problems with a single inequality constraint and an optimal solution at x^*. Figure (a) illustrates the case in which x^* is also a local minimum of $F_c(x) = f(x) + cg^+(x)$, hence the penalty function is "exact". Figure (b) illustrates an exceptional case where the penalty function is not exact. In this case, $\nabla g(x^*) = 0$, thus violating the condition of constraint gradient linear independence (we will show later that one condition guaranteeing the exactness of F_c is that the constraint gradients at x^* are linearly independent). For this constraint set, it is possible that $F_c(x)$ does not have a local minimum at x^* for any $c > 0$ (as for the cost function depicted in the figure, where the downward order of growth of f exceeds the upward order of growth of g at x^* when moving from x^* towards smaller values).

Hence, through the use of penalty functions, the constrained optimization problem can be solved via unconstrained optimization techniques. The conditions under which a problem admits an exact penalty have been an important research topic since 70s. It is very interesting to note that such conditions (as well as the threshold value for c) bear an intimate connection with constraint qualification theory of Lagrange multipliers. The line of analysis we adopt in this thesis clearly depicts how exact penalty functions fit in a theoretical picture with Lagrange multipliers.

1.4. A NEW THEORY OF LAGRANGE MULTIPLIERS

In this work, we present a new theory of Lagrange multipliers, which is simple and more powerful than the classical treatments. Our objective is to generalize, unify, and streamline the theory of constraint qualifications, which are conditions on the constraint set that guarantee the existence of Lagrange multipliers. The diversity of these conditions motivated researchers to examine their interrelations and try to come up with a central notion that places these conditions in a larger theoretical picture. For problems that have smooth equality and inequality constraint functions, but no abstract set constraint, the notion called *quasiregularity*, acts as the unifying concept that relates constraint qualifications. In the presence of an abstract set constraint, quasiregularity fails to provide the required unification. Our development introduces a new notion, called *pseudonormality*, as a substitute for quasiregularity for the case of an abstract set constraint. Even without an abstract set constraint, pseudonormality simplifies the proofs of Lagrange multiplier theorems and provides information about special Lagrange multipliers that carry sensitivity information. Our analysis also yields a number of interesting related results. In particular, our contributions can be summarized as follows:

(a) The optimality conditions of the Lagrange multiplier type that we develop are sharper than the classical Karush-Kuhn-Tucker conditions (they include extra conditions, which may narrow down the set of candidate local minima). 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 introduce the notion of pseudonormality, which serves to unify the major constraint qualifications and forms a connecting link between the constraint qualifications and existence of Lagrange multipliers. This analysis carries through even in the case of an additional abstract set constraint, where the classical treatments of the theory fail.

(c) We develop 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 investigate the relations between different types of Lagrange multipliers. We show that one particular Lagrange multiplier vector, called the *informative Lagrange multiplier*, 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. We show that, under mild convexity assumptions, an informative Lagrange multiplier always exists when the set of Lagrange multipliers is nonempty.

(d) There is another equally powerful approach to Lagrange multipliers, based on exact penalty functions, which has not received much attention thus far. In particular, let us say that the constraint set C *admits an exact penalty* at the feasible point x^* if for every smooth function f for which x^* is a strict local minimum of f over C, there is a scalar $c > 0$ such that x^* is also a local minimum of the function

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

over $x \in X$, where we denote

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

Exact penalty functions have traditionally been viewed as a device used in computational methods. In this work, we use exact penalty functions as a vehicle towards asserting the existence of Lagrange multipliers. In particular, we make a connection between pseudonormality, the existence of Lagrange multipliers, and the exact penalty functions. We show that pseudonormality implies the admittance of exact penalty functions, which in turn implies the existence of Lagrange multipliers.

(e) We extend the theory developed for the case where the functions f, h_i and g_j are assumed to be smooth, to the case where these functions are nondifferentiable, but are instead assumed convex, using the theory of subgradients.

(f) We consider problems that do not necessarily have an optimal solution. For this purpose, we adopt a different approach based on tools from convex analysis. We consider certain types of multipliers, called *geometric*, that are not tied to a specific local or global minimum and do not assume differentiability of the cost and constraint functions. Geometric multipliers admit insightful visualization through the use of hyperplanes and the related convex set support/separation arguments. Under convexity assumptions, geometric multipliers are strongly related to Lagrange multipliers. Geometric multipliers can also be viewed as the optimization variables of a related auxiliary optimization problem, called the *dual problem.* We develop necessary optimality conditions for problems without an optimal solution under various assumptions. In particular, under convexity assumptions, we derive Fritz John-type conditions, which provides a pathway that highlights the relations between the original and the dual problems. Under additional closedness assumptions, we develop Fritz John optimality conditions that involve sensitivity-type conditions.

(g) We introduce a special geometric multiplier, called *informative*, that provides similar sensitivity information regarding the constraints to violate to effect a cost reduction, as the informative Lagrange multipliers. We show that an informative geometric multiplier always exists when the set of geometric multipliers is nonempty.

(h) We derive Fritz John-type optimality conditions for the dual problem. Based on these optimality conditions, we introduce a special type of dual optimal solution, called *informative*, which is analogous to informative geometric multipliers. We show that such a dual optimal solution always exists, when the dual problem has an optimal solution.

The outline of the thesis is as follows: In Chapter 2, we provide basic definitions and results that will be used throughout this thesis. We also study the geometry of constraint sets of optimization problems in detail in terms of conical approximations and present general optimality conditions. In Chapter 3, we develop enhanced necessary optimality conditions of the Fritz John-type for problems that involve smooth equality and inequality constraints and an abstract (possibly nonconvex) set constraint. We also provide a classification of different types of Lagrange multipliers, based on the sensitivity information they provide; investigate their properties and relations. In Chapter 4, we introduce the notion of pseudonormality and show that it plays a central role within the taxonomy of interesting constraint characteristics. In particular, pseudonormality unifies and expands classical constraint qualifications that guarantee the existence of Lagrange multipliers. We also show that, for optimization problems with additional set constraints, the classical treatment of the theory based on the notion of *quasiregularity* fails, whereas pseudonormality still provides the required connections. Moreover, the relation of exact penalty functions and the Lagrange multipliers is well understood through the notion of pseudonormality. In Chapter 5, we extend the theory regarding pseudonormality to problems in which continuity/differentiability assumptions are replaced by convexity assumptions. We consider problems without an optimal solution and derive optimality conditions for such problems. Finally, Chapter 6 summarizes our results and points out future research directions.

CHAPTER 2

CONSTRAINT GEOMETRY

We consider finite dimensional optimization problems of the form

$$\text{minimize } f(x)$$
$$\text{subject to } \; x \in C,$$

where $f : \Re^n \mapsto \Re$ is a function and C is a subset of $\Re^n$.

Necessary optimality conditions for equality and inequality-constrained problems were presented by Karush, Kuhn, and Tucker with a *constraint qualification* (cf. Chapter 1). However, these conditions do not cover the case where there is an additional abstract set constraint X, and therefore is limited to applications where it is possible and convenient to represent all constraints explicitly by a finite number of equalities and inequalities. Moreover, necessary optimality conditions are often presented with the assumption of "linear independence of constraint gradients". This is unnecessarily restrictive especially for problems with inequality constraints. Therefore, the key to understanding Lagrange multipliers is through a closer study of the constraint geometry in optimization problems. For this purpose, at first, we do not insist on any particular representation for C; we just assume that C is some subset of $\Re^n$.

The problem of minimizing f over C leads to the possibility that points of interest may lie on the boundary of C. Therefore, an in-depth understanding of the properties of the boundary of C is crucial in characterizing optimal solutions. The boundary of C may be very complicated due to all kinds of curvilinear faces and corners.

In this chapter, we first study the local geometry of C in terms of "tangent vectors" and "normal vectors", which are useful tools in studying variational properties of set C despite boundary complications. This type of analysis is called *nonsmooth analysis* due to one-sided nature of the geometry as well as kinks and corners in set boundaries. We next use this analysis in connection with optimality conditions.

2.1. NOTATION AND TERMINOLOGY

In this section, we present some basic definitions and results that will be used throughout this thesis.

We first provide some notation. All of the vectors are column vectors and a prime denotes transposition. We write $x \geq 0$ or $x > 0$ when a vector x has nonnegative or positive components, respectively. Similarly, we write $x \leq 0$ or $x < 0$ when a vector x has nonpositive or negative components, respectively. We use throughout the thesis the standard Euclidean norm in $\Re^n$, $\|x\| = (x'x)^{1/2}$, where $x'y$ denotes the inner product of any $x, y \in \Re^n$. We denote by $\text{cl}(C)$ and $\text{int}(C)$ the closure and the interior of a set C, respectively.

We also use some of the standard notions of convex analysis. In particular, for a set X, we denote by $\text{conv}(X)$ the convex hull of X, i.e., the intersection of all convex sets containing X, or equivalently the set of all convex combinations of elements of X. For a convex set C, we denote by $\text{aff}(C)$ the affine hull of C, i.e., the smallest affine set containing C, and by $\text{ri}(C)$ the relative interior of C, i.e., its interior relative to $\text{aff}(C)$. The epigraph $\big\{(x, w) \mid f(x) \leq w,\, x \in X,\, w \in \Re\big\}$ of a function $f : X \mapsto \Re$ is denoted by $\text{epi}(f)$.

Given any set X, the set of vectors that are orthogonal to all elements of X is a subspace denoted by $X^\perp$:

$$X^\perp = \{y \mid y'x = 0,\ \forall\, x \in X\}.$$

If S is a subspace, $S^\perp$ is called the *orthogonal complement* of S. A set C is said to be a *cone* if for all $x \in C$ and $\lambda > 0$, we have $\lambda x \in C$.

We next give an important duality relation between cones. Given a set C, the cone given by

$$C^* = \{y \mid y'x \leq 0,\ \forall\, x \in C\},$$

is called the *polar cone* of C. Clearly, the polar cone C^*, being the intersection of a collection of closed halfspaces, is always closed and convex (regardless of whether C is closed and/or convex). If C is a subspace, it can be seen that the polar cone C^* is equal to the orthogonal subspace $C^\perp$. The following basic result generalizes the equality $C = (C^\perp)^\perp$, which holds in the case where C is a subspace (for the proof see [BNO02]).

Proposition 2.1.1: (Polar Cone Theorem) For any cone C, we have

$$(C^*)^* = \mathrm{cl}\big(\mathrm{conv}(C)\big).$$

In particular, if C is closed and convex, we have $(C^*)^* = C$.

We next give some basic results regarding cones and their polars that will be useful in our analysis (for the proofs, see [BNO02]).

Proposition 2.1.2:

(a) Let C_1 and C_2 be two cones. If $C_1 \subset C_2$, then $C_2^* \subset C_1^*$.

(b) Let C_1 and C_2 be two cones. Then,

$$(C_1 + C_2)^* = C_1^* \cap C_2^*,$$

and

$$C_1^* + C_2^* \subset (C_1 \cap C_2)^*.$$

In particular if C_1 and C_2 are closed and convex, $(C_1 \cap C_2)^* = \mathrm{cl}(C_1^* + C_2^*)$.

Existence of Optimal Solutions

A basic question in optimization problems is whether an optimal solution exists. This question can often be resolved with the aid of the classical theorem of Weierstrass, given in the following proposition. To this end, we introduce some terminology. Let X be a nonempty subset of $\Re^n$. We say that a function $f : X \mapsto (-\infty, \infty]$ is *coercive* if

$$\lim_{k \to \infty} f(x_k) = \infty$$

for every sequence $\{x_k\}$ of elements of X such that $\|x_k\| \to \infty$. Note that as a consequence of the definition, a nonempty level set $\big\{x \mid f(x) \leq a\big\}$ of a coercive function f is bounded.

Proposition 2.1.3: (Weierstrass' Theorem) Let X be a nonempty closed subset of $\Re^n$, and let $f : X \mapsto \Re$ be a lower semicontinuous function over X. Assume that one of the following three conditions holds:

(1) X is bounded.

(2) There exists a scalar a such that the level set

$$\big\{x \in X \mid f(x) \leq a\big\}$$

is nonempty and bounded.

(3) f is coercive.

Then the set of minima of f over X is nonempty and compact.

Separation Results

In Chapter 5, our development will require tools from convex analysis. For the purpose of easy reference, we list here some of the classical supporting and separating hyperplane results that we will use in our analysis. Recall that a *hyperplane* in $\Re^n$ is a set of the form $\{x \mid a'x = b\}$, where $a \in \Re^n$, $a \neq 0$, and $b \in \Re$. The sets

$$\{x \mid a'x \geq b\}, \qquad \{x \mid a'x \leq b\},$$

are called the *closed halfspaces* associated with the hyperplane.

Proposition 2.1.4: (Supporting Hyperplane Theorem) Let C be a nonempty convex subset of $\Re^n$ and let $\overline{x}$ be a vector in $\Re^n$. If either C has empty interior or, more generally, if $\overline{x}$ is not an interior point of C, there exists a hyperplane that passes through $\overline{x}$ and contains C in one of its closed halfspaces, i.e., there exists a vector $a \neq 0$ such that

$$a'\overline{x} \leq a'x, \qquad \forall\, x \in C. \tag{1.1}$$

Proposition 2.1.5: (Proper Separation Theorem) Let C_1 and C_2 be nonempty convex subsets of $\Re^n$ such that

$$\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset.$$

Then there exists a hyperplane that properly separates C_1 from C_2, i.e., a vector a such that

$$\sup_{x \in C_2} a'x \leq \inf_{x \in C_1} a'x, \qquad \inf_{x \in C_2} a'x < \sup_{x \in C_1} a'x.$$

Proposition 2.1.6: (Polyhedral Proper Separation Theorem) Let C_1 and C_2 be nonempty convex subsets of $\Re^n$ such that C_2 is polyhedral and

$$\mathrm{ri}(C_1) \cap C_2 = \emptyset.$$

Then there exists a hyperplane that properly separates them and does not contain C_1, i.e., a vector a such that

$$\sup_{x \in C_2} a'x \leq \inf_{x \in C_1} a'x, \qquad \inf_{x \in C_1} a'x < \sup_{x \in C_1} a'x.$$

Saddle Points

Our analysis also requires the following result regarding the existence of saddle points of functions, which is a slight extension of the classical theorem of von Neumann (for the proof, see [BNO02]).

Proposition 2.1.7: (Saddle Point Theorem) Let X be a nonempty convex subset of $\Re^n$, let Z be a nonempty convex subset of $\Re^m$, and let $\phi : X \times Z \mapsto \Re$ be a function such that either

$$-\infty < \sup_{z \in Z} \inf_{x \in X} \phi(x, z),$$

or

$$\inf_{x \in X} \sup_{z \in Z} \phi(x, z) < \infty.$$

Assume that for each $z \in Z$, the function $t_z : \Re^n \mapsto (-\infty, \infty]$ defined by

$$t_z(x) = \begin{cases} \phi(x, z), & \text{if } x \in X, \\ \infty, & \text{if } x \notin X, \end{cases}$$

is closed and convex, and that for each $x \in X$, the function $r_x : \Re^m \mapsto (-\infty, \infty]$ defined by

$$r_x(z) = \begin{cases} -\phi(x, z) & \text{if } z \in Z, \\ \infty & \text{otherwise,} \end{cases}$$

is closed and convex. The set of saddle points of ϕ is nonempty and compact under any of the following conditions:

(1) X and Z are compact.

(2) Z is compact and there exists a vector $\overline{z} \in Z$ and a scalar γ such that the level set

$$\{x \in X \mid \phi(x, \overline{z}) \leq \gamma\}$$

is nonempty and compact.

(3) X is compact and there exists a vector $\overline{x} \in X$ and a scalar γ such that the level set

$$\{z \in Z \mid \phi(\overline{x}, z) \geq \gamma\}$$

is nonempty and compact.

(4) There exist vectors $\overline{x} \in X$ and $\overline{z} \in Z$, and a scalar γ such that the level sets

$$\{x \in X \mid \phi(x, \overline{z}) \leq \gamma\}, \qquad \{z \in Z \mid \phi(\overline{x}, z) \geq \gamma\}.$$

2.2. CONICAL APPROXIMATIONS

The analysis of constrained optimization problems is centered around characterizing how the cost function behaves as we move from a local minimum to neighboring feasible points. In optimizing a function f over a set C, since the local minima may very well lie on the boundary, properties of the boundary of C can be crucial in characterizing an optimal solution. The difficulty is that the boundary may have all kinds of weird curvilinear facets, edges, and corners. In such a lack of smoothness, an approach is needed through which main variational properties of set C can be characterized. The relevant variational properties can be studied in terms of various tangential and normal cone approximations to the constraint set at each point.

Many different definitions of tangent and normal vectors have been offered over the years. It turns out that two of these are particularly useful in characterizing local optimality of feasible solutions, and are actually sufficient to go directly into the heart of the issues about Lagrange multipliers.

2.2.1. Tangent Cone

A simple notion of variation at a point x that belongs to the constraint set C can be defined by taking a vector $y \in \Re^n$ and considering the vector $x + \alpha y$ for a small positive scalar α. (For instance, directional derivatives are defined in terms of such variations). This idea gives rise to the following definition.

Definition 2.2.1: Given a subset C of $\Re^n$ and a vector $x \in C$, a *feasible direction* of C at x is a vector $y \in \Re^n$ such that there exists an $\overline{\alpha} > 0$ with $x + \alpha y \in C$ for all $\alpha \in [0, \overline{\alpha}]$. The set of all feasible directions of C at x is a cone denoted by $F_C(x)$.

It can be seen that if C is convex, the feasible directions at x are the vectors of the form $\alpha(\overline{x} - x)$ with $\alpha > 0$ and $\overline{x} \in C$ [cf. Figure 2.2.1(a)].

However, when C is nonconvex, straight line variations of the preceding sort may not be appropriate to characterize the local structure of the set C near the point x. [For example, often there is no nonzero feasible direction at x when C is nonconvex, think of the set $C = \{x \mid h(x) = 0\}$, where $h : \Re^n \mapsto \Re$ is a nonlinear function, see Figure 2.2.1(b)]. Nonetheless, the concept of direction can still be utilized in terms of sequences that converge to the point of interest without violating the set constraint. The next definition introduces a cone that illustrates this idea.

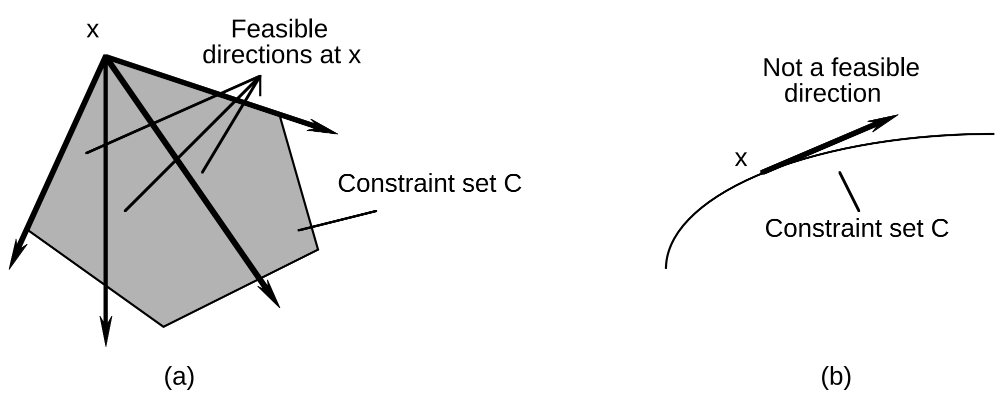


Figure 2.2.1. Feasible directions at a vector x. By definition, y is a feasible direction if changing x in the direction y by a small amount maintains feasibility.

Definition 2.2.2: Given a subset C of $\Re^n$ and a vector $x \in C$, a vector y is said to be a *tangent* of C at x if either $y = 0$ or there exists a sequence $\{x_k\} \subset C$ such that $x_k \neq x$ for all k and

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

The set of all tangents of C at x is a cone called the *tangent cone* of C at x, and is denoted by $T_C(x)$.

Thus a nonzero vector y is a tangent at x if it is possible to approach x with a feasible sequence $\{x_k\}$ such that the normalized direction sequence $(x_k - x)/\|x_k - x\|$ converges to $y/\|y\|$, the normalized direction of y, cf. Figure 2.2.2(a). The tangent vectors to a set C at a point x are illustrated in Figure 2.2.2(b). It can be seen that $T_C(x)$ is a cone, hence the name "tangent cone". The following proposition provides an equivalent definition of a tangent, which is sometimes more convenient in analysis.

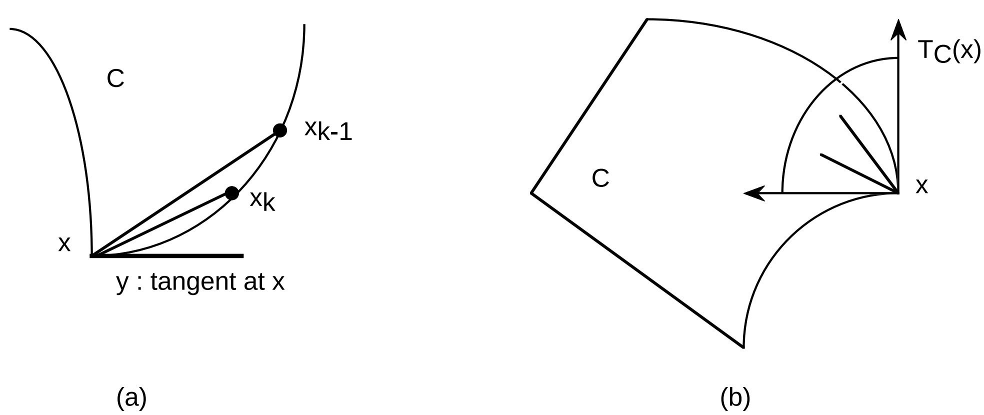


Figure 2.2.2. Part (a) illustrates a tangent y at a vector $x \in C$. Part (b) illustrates the tangent cone to set C at a vector x.

Proposition 2.2.8: Given a subset C of $\Re^n$ and a vector $x \in C$, a vector y is a tangent of C at x if and only if there exists a sequence $\{x_k\} \subset C$ with $x_k \to x$, and a positive sequence $\{\alpha_k\}$ such that $\alpha_k \to 0$ and

$$\frac{(x_k - x)}{\alpha_k} \to y.$$

Proof: Let y be a tangent of set C at the vector x. If $y = 0$, take $x_k = x$ for all k and α_k any positive sequence that converges to 0, and we are done. Therefore, assume that $y \neq 0$. Then, we take x_k to be the sequence in the definition of a tangent, and $\alpha_k = \|x_k - x\|/\|y\|$.

Conversely, assume that y is such that sequences $\{x_k\}$ and $\{\alpha_k\}$ with the above properties exist. If $y = 0$, then y is a tangent of C at x. If $y \neq 0$, then since $(x_k - x)/\alpha_k \to y$, we have

$$\frac{x_k - x}{\|x_k - x\|} = \frac{(x_k - x)/\alpha_k}{\|(x_k - x)/\alpha_k\|} \to \frac{y}{\|y\|},$$

so $\{x_k\}$ satisfies the definition of a tangent. **Q.E.D.**

Figure 2.2.3 illustrates the cones $F_C(x)$ and $T_C(x)$, and hints at their relation with examples. The following proposition gives some of the properties of the cones $F_C(x)$ and $T_C(x)$ (for the proofs, see [BNO02]).

Proposition 2.2.9: Let C be a nonempty subset of $\Re^n$ and let x be a vector in C. The following hold regarding the cone of feasible directions $F_C(x)$ and the tangent cone $T_C(x)$.

(a) $T_C(x)$ is a closed cone.

(b) $\mathrm{cl}\big(F_C(x)\big) \subset T_C(x)$.

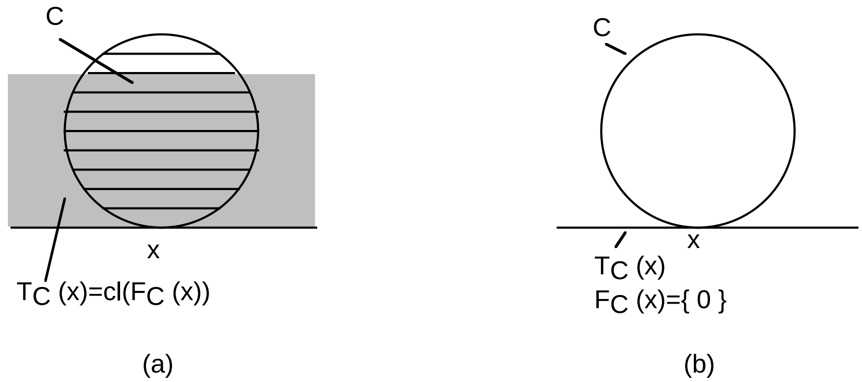


Figure 2.2.3. Illustration of feasible cone of directions and tangent cone. In part (a), set C is convex, and the tangent cone of C at x is equal to the closure of the cone of feasible directions. In part (b), the cone of feasible directions consists of just the zero vector.

(c) If C is convex, then $F_C(x)$ and $T_C(x)$ are convex, and we have

$$\mathrm{cl}\big(F_C(x)\big) = T_C(x).$$

2.2.2. Normal Cone

In addition to the cone of feasible directions and the tangent cone, there is one more conical approximation that is of special interest in relation to optimality conditions in this thesis.

Definition 2.2.3: Given a subset C of $\Re^n$ and a vector $x \in C$, a vector z is said to be a *normal* of C at x if there exist sequences $\{x_k\} \subset X$ and $\{z_k\}$ such that

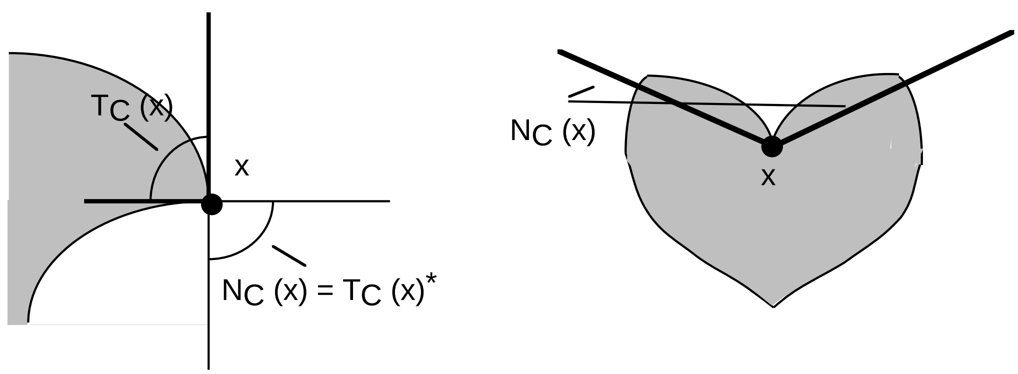
$$x_k \to x, \qquad z_k \to z, \qquad z_k \in T_C(x_k)^*, \quad \text{for all } k.$$

The set of all normals of C at x is called the *normal cone* of C at x, and is denoted by $N_C(x)$.

Hence, the normal cone $N_C(x)$ is obtained from the polar cone $T_C(x)^*$ by means of a closure operation. Equivalently, the graph of $N_C(\cdot)$, viewed as a point-to-set mapping, is the intersection of the closure of the graph of $T_C(\cdot)^*$ with the set $\{(x,z) \mid x \in C\}$. In the case where C is a closed set, the set $\{(x,z) \mid x \in C\}$ contains the closure of the graph of $T_C(\cdot)^*$, so the graph of $N_C(\cdot)$ is equal to the closure of the graph of $T_C(\cdot)^*$[1] :

$$\big\{(x,z) \mid x \in C,\, z \in N_C(x)\big\} = \mathrm{cl}\Big(\big\{(x,z) \mid x \in C,\, z \in T_C(x)^*\big\}\Big) \qquad \text{if } C \text{ is closed.}$$

In general, it can be seen that $T_C(x)^* \subset N_C(x)$ for any $x \in C$. However, $N_C(x)$ may not be equal to $T_C(x)^*$, and in fact it may not even be a convex set (see the examples of Figure 2.2.4).



(a) (b)

Figure 2.2.4. Examples of normal cones. In the case of part (a), we have $N_C(x) = T_C(x^*)$, hence C is regular at x. In part (b), $N_C(x)$ is the union of two lines. In this case $N_C(x)$ is not equal to $T_C(x)$ and is nonconvex, i.e., C is not regular at x.

[1] The normal cone, introduced by Mordukhovich [Mor76], has been studied by several authors, and is of central importance in nonsmooth analysis (see the books by Aubin and Frankowska [AuF90], Rockafellar and Wets [RoW98], and Borwein and Lewis [BoL00]).

Definition 2.2.4: A set C is said to be *regular* at some vector $x \in C$ if

$$T_C(x)^* = N_C(x).$$

The term "regular at x in the sense of Clarke" is also used in the literature.

2.2.3. Tangent-Normal Cone Relations

The relationships between tangent and normal cones defined in the previous sections play a central role in our development of enhanced optimality conditions in Chapter 3. It turns out that these cones are nicely connected through polarity relations. Furthermore, these relations reveal alternative characterizations of "Clarke regularity", which will be useful for our purposes. These polarity relations were given in [RoW98] as a result of a series of exercises. Here, we provide a streamlined development of these results together with detailed proofs. These proofs make use of concepts related to sequences of sets and their convergence properties, which we summarize in the following section.

2.2.3.1. Sequences of Sets and Set Convergence:

Let $\{C_k\}$ be a sequence of nonempty subsets of $\Re^n$. The *outer limit* of $\{C_k\}$, denoted $\limsup_{k\to\infty} C_k$, is the set of all $x \in \Re^n$ such that every neighborhood of x has a nonempty intersection with infinitely many of the sets C_k, $k = 1, 2, \ldots$. Equivalently, $\limsup_{k\to\infty} C_k$ is the set of all limits of subsequences $\{x_k\}_{\mathcal{K}}$ such that $x_k \in C_k$ for all $k \in \mathcal{K}$.

The *inner limit* of $\{C_k\}$, denoted $\liminf_{k\to\infty} C_k$, is the set of all $x \in \Re^n$ such that every neighborhood of x has a nonempty intersection with all except finitely many of the sets C_k, $k = 1, 2, \ldots$. Equivalently, $\liminf_{k\to\infty} C_k$ is the set of all limits of sequences $\{x_k\}$ such that $x_k \in C_k$ for all $k = 1, 2, \ldots$. These definitions are illustrated in Figure 2.2.5.

The sequence $\{C_k\}$ is said to converge to a set C if

$$C = \liminf_{k\to\infty} C_k = \limsup_{k\to\infty} C_k.$$

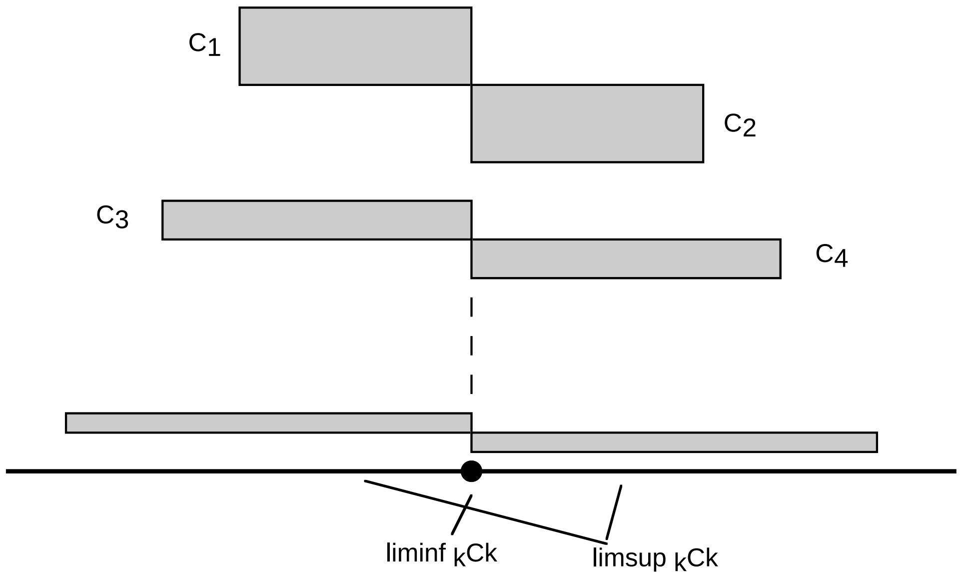


Figure 2.2.5. Inner and outer limits of a nonconvergent sequence of sets.

In this case, C is called the *limit of* $\{C_k\}$, and is denoted by $\lim_{k\to\infty} C_k$.[1]

The inner and outer limits are closed (possibly empty) sets. It is clear that we always have $\liminf_{k\to\infty} C_k \subset \limsup_{k\to\infty} C_k$. If each set C_k consists of a single point x_k, $\limsup_{k\to\infty} C_k$ is the set of limit points of $\{x_k\}$, while $\liminf_{k\to\infty} C_k$ is just the limit of $\{x_k\}$ if $\{x_k\}$ converges, and otherwise it is empty.

The next proposition provide a major tool for checking results about inner and outer limits.

[1] Set convergence in this sense is known more specifically as *Painleve-Kuratowski convergence.*

Proposition 2.2.10 (Set Convergence Criteria): Let $\{C_k\}$ be a sequence of nonempty closed subsets of $\Re^n$ and C be a nonempty closed subset of $\Re^n$. Let $B(x,\epsilon)$ denote the closed ball centered at x with radius ϵ.

(a)

(i) $C \subset \liminf_{k\to\infty} C_k$ if and only if for every ball $B(x,\epsilon)$ with $C \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$, we have $C_k \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$ for all sufficiently large k.

(ii) $C \supset \limsup_{k\to\infty} C_k$ if and only if for every ball $B(x,\epsilon)$ with $C \cap B(x,\epsilon) = \emptyset$, we have $C_k \cap B(x,\epsilon) = \emptyset$ for all sufficiently large k.

(b) In part (a), it is sufficient to consider the countable collection of balls $B(x,\epsilon)$, where ϵ and the coordinates of x are rational numbers.

Proof:

(a)

(i) Assume that $C \subset \liminf_{k\to\infty} C_k$ and let $B(x,\epsilon)$ be a ball such that $C \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$. Let x be a vector that belongs to $C \cap \text{int}\big(B(x,\epsilon)\big)$. By assumption, it follows that $x \in \liminf_{k\to\infty} C_k$, which by definition of the inner limit of a sequence of sets, implies the existence of a sequence $\{x_k\}$ with $x_k \in C_k$ such that $x_k \to x$. Since $x \in \text{int}\big(B(x,\epsilon)\big)$, we have that $x_k \in \text{int}\big(B(x,\epsilon)\big)$ for all sufficiently large k, which proves that $C_k \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$ for all sufficiently large k.

Conversely, assume that for every ball $B(x,\epsilon)$ with $C \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$, we have $C_k \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$ for all sufficiently large k. Consider any $x \in C$ and $\epsilon > 0$. By assumption, there exists some x_k that belongs to $C_k \cap \text{int}\big(B(x,\epsilon)\big)$ for sufficiently large k, thereby implying the existence of a sequence $\{x_k\}$ with $x_k \in C_k$ such that $x_k \to x$, and hence proving that $x \in \liminf_{k\to\infty} C_k$.

(ii) Assume that $C \supset \limsup_{k\to\infty} C_k$ and let $B(x,\epsilon)$ be a ball such that $C \cap B(x,\epsilon) = \emptyset$. Hence, for any $\overline{x} \in B(x,\epsilon)$, we have $\overline{x} \notin C$, which by assumption implies that $\overline{x} \notin$

$\limsup_{k\to\infty C_k}$. By definition of the outer limit of a sequence of sets, it follows that $\overline{x} \notin C_k$ for all sufficiently large k, proving that $C_k \cap B(x,\epsilon) = \emptyset$ for all sufficiently large k.

Conversely, assume that for every ball $B(x,\epsilon)$ with $C \cap B(x,\epsilon) = \emptyset$, we have $C_k \cap B(x,\epsilon) = \emptyset$ for all sufficiently large k. Let $x \notin C$. Since C is closed, there exists some $\epsilon > 0$ such that $C \cap B(x,\epsilon) = \emptyset$, which implies by assumption that $C_k \cap B(x,\epsilon) = \emptyset$ for all sufficiently large k, thereby proving that $x \notin \limsup_{k\to\infty} C_k$.

(b) Since this condition is a special case of the condition given in part (a), the implications " $\Rightarrow$ " hold trivially. We now show the reverse implications.

(i) Assume that for every ball $B(x,\epsilon)$, where ϵ and the coordinates of x are rational numbers with $C \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$, we have $C_k \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$ for all sufficiently large k. Consider any $x \in C$ and any rational $\epsilon > 0$. There exists a point $\overline{x} \in B(x,\epsilon/2)$ whose coordinates are rational. For such a point, we have $C \cap B(\overline{x},\epsilon/2) \neq \emptyset$, which by assumption implies $C_k \cap \text{int}\big(B(\overline{x},\epsilon/2)\big) \neq \emptyset$ for all sufficiently large k. In particular, we have $\overline{x} \in C_k + \epsilon/2B$ [B denotes the closed ball $B(0,1)$], so that $x \in C_k + \epsilon B$ for all sufficiently large k. This implies the existence of a sequence $\{x_k\}$ with $x_k \in C_k$ such that $x_k \to x$, and hence proving that $x \in \liminf_{k\to\infty} C_k$.

(ii) Assume that for every ball $B(x,\epsilon)$, where ϵ and the coordinates of x are rational numbers with $C \cap B(x,\epsilon) = \emptyset$, we have $C_k \cap B(x,\epsilon) = \emptyset$ for all sufficiently large k. Let $x \notin C$. Since C is closed, there exists some rational $\epsilon > 0$ such that $C \cap B(x,2\epsilon) = \emptyset$. A point $\overline{x}$ with rational coordinates can be selected from $\text{int}\big(B(x,\epsilon)\big)$. Then, we have $x \in \text{int}\big(B(\overline{x},\epsilon)\big)$ and $C \cap B(\overline{x},\epsilon) = \emptyset$. By assumption, we get $C_k \cap B(\overline{x},\epsilon) = \emptyset$ for all sufficiently large k. Since $x \in \text{int}\big(B(\overline{x},\epsilon)\big)$, this implies that $x \notin \limsup_{k\to\infty} C_k$, proving the desired claim. **Q.E.D.**

We next provide alternative characterizations for set convergence through distance functions and projections.

Proposition 2.2.11 (Set Convergence through Distance Functions): Let $\{C_k\}$ be a sequence of nonempty closed subsets of $\Re^n$ and C be a nonempty closed subset of $\Re^n$. Let $d(x, C)$ denote the distance of a vector $x \in \Re^n$ to set C, i.e., $d(x, C) = \min_{y \in C} \|x - y\|$.

(a)

(i) $C \subset \liminf_{k\to\infty} C_k$ if and only if $d(x, C) \geq \limsup_{k\to\infty} d(x, C_k)$ for all $x \in \Re^n$.

(ii) $C \supset \limsup_{k\to\infty} C_k$ if and only if $d(x, C) \leq \liminf_{k\to\infty} d(x, C_k)$ for all $x \in \Re^n$.

In particular, we have $C_k \to C$ if and only if $d(x, C_k) \to d(x, C)$ for all $x \in \Re^n$.

(b) The result of part (a) can be extended as follows: $C_k \to C$ if and only if $d(x_k, C_k) \to d(x, C)$ for all sequences $\{x_k\} \to x$ and all $x \in \Re^n$.

Proof:

(a)

(i) Assume that $C \subset \liminf_{k\to\infty} C_k$. Consider any $x \in \Re^n$. It can be seen that for a closed set C,

$$d(x, C) < \alpha \text{ if and only if } C \cap \operatorname{int}\big(B(x, \alpha)\big) \neq \emptyset, \tag{2.1}$$

(cf. Weierstrass' Theorem). Let $\alpha = \limsup_{k\to\infty} d(x, C_k)$. Since C is closed, $d(x, C)$ is finite (cf. Weierstrass' Theorem), and therefore, by Proposition 2.2.10(a)-(i) and relation (2.1), it follows that α is finite. Suppose, to arrive at a contradiction, that $d(x, C) < \alpha$. Let $\epsilon > 0$ be such that $d(x, C) < \alpha - \epsilon$. It follows from Proposition 2.2.10(a)-(i) and relation (2.1) that

$$\limsup_{k\to\infty} d(x, C_k) \leq \alpha - \epsilon,$$

which is a contradiction.

Conversely, assume that

$$d(x, C) \geq \limsup_{k\to\infty} d(x, C_k), \qquad \forall\, x \in \Re^n. \tag{2.2}$$

Let $B(x,\epsilon)$ be a closed ball with $C \cap \text{int}\big(B(x,\epsilon)\big) \neq \emptyset$. By Eq. (2.1), this implies that $d(x,C) < \epsilon$, which by assumption (2.2) yields $d(x,C_k) < \epsilon$ for all sufficiently large k. Using Proposition 2.2.10(a)-(i) and relation (2.1), it follows that $C \subset \liminf_{k\to\infty} C_k$.

(ii) Assume that $C \supset \limsup_{k\to\infty} C_k$. Consider any $x \in \Re^n$. It can be seen that for a closed set C,

$$d(x,C) > \beta \text{ if and only if } C \cap B(x,\beta) = \emptyset, \tag{2.3}$$

(cf. Weierstrass' Theorem). Let $\beta = \liminf_{k\to\infty} d(x,C_k)$. Since C is closed, $d(x,C)$ is finite (cf. Weierstrass' Theorem), and therefore, by Proposition 2.2.10(a)-(ii) and relation (2.3), it follows that β is finite. Suppose, to arrive at a contradiction, that $d(x,C) > \beta$. Let $\epsilon > 0$ be such that $d(x,C) > \beta + \epsilon$. It follows from Proposition 2.2.10(a)-(ii) and relation (2.3) that

$$\liminf_{k\to\infty} d(x,C_k) \geq \beta + \epsilon,$$

which is a contradiction.

Conversely, assume that

$$d(x,C) \leq \liminf_{k\to\infty} d(x,C_k), \qquad \forall\, x \in \Re^n. \tag{2.4}$$

Let $B(x,\epsilon)$ be a closed ball with $C \cap B(x,\epsilon) = \emptyset$. By Eq. (2.3), this implies that $d(x,C) > \epsilon$, which by assumption (2.4) yields $d(x,C_k) > \epsilon$ for all sufficiently large k. Using Proposition 2.2.10(a)-(ii) and relation (2.3), it follows that $C \supset \limsup_{k\to\infty} C_k$.

(b) This part follows from part (a) and the fact that for any closed set C, $d(x,C)$ is a continuous function of x. In particular, for any sequence $\{x_i\}$ that converges to x and any closed set C_k, we have

$$\lim_{i\to\infty} d(x_i,C_k) = d(x,C_k),$$

from which we get

$$\limsup_{k\to\infty} d(x_k,C_k) = \limsup_{k\to\infty} d(x,C_k),$$

and

$$\liminf_{k\to\infty} d(x_k,C_k) = \liminf_{k\to\infty} d(x,C_k),$$

which together with part (a) proves the desired result. **Q.E.D.**

Proposition 2.2.12 (Set Convergence through Projections): Let $\{C_k\}$ be a sequence of nonempty closed subsets of $\Re^n$ and C be a nonempty closed subset of $\Re^n$. Let $P_C(x)$ denote the projection set of a vector $x \in \Re^n$ to set C, i.e., $P_C(x) = \arg\min_{y\in C} \|x - y\|$.

(a) We have $C_k \to C$ if and only if $\limsup_{k\to\infty} d(0, C_k) < \infty$ and

$$\limsup_{k\to\infty} P_{C_k}(x) \subset P_C(x), \qquad \text{for all } x \in \Re^n.$$

(b) The result of part (a) can be extended as follows: $C_k \to C$ if and only if $\limsup_{k\to\infty} d(0, C_k) < \infty$ and

$$\limsup_{k\to\infty} P_{C_k}(x_k) \subset P_C(x), \quad \text{for all sequences } \{x_k\} \to x \text{ and all } x \in \Re^n.$$

(c) Define the graph of the projection mapping P_C as a subset of $\Re^{2n}$ given by

$$\mathrm{gph}(P_C) = \big\{(x, u) \mid x \in \Re^n, u \in P_C(x)\big\}.$$

$C_k \to C$ if and only if the corresponding sequence of graphs of projection mappings $\{P_{C_k}\}$ converges to the graph of P_C.

Proof:

(a) Assume that $C_k \to C$. By Proposition 2.2.11(a), this implies that $d(x, C_k) \to d(x, C)$ for all $x \in \Re^n$. In particular, for $x = 0$, we have $\limsup_{k\to\infty} d(0, C_k) = d(0, C) < \infty$ (by closedness of C and Weierstrass' Theorem). For any $x \in \Re^n$, let $\overline{x} \in \limsup_{k\to\infty} P_{C_k}(x)$. By definition of the outer limit of a sequence of sets, this implies the existence of a subsequence $\big\{P_{C_k}(x)\big\}_{k\in\mathcal{K}}$ and vectors $x_k \in P_{C_k}(x)$ for all $k \in \mathcal{K}$, such that $\lim_{k\to\infty,\ k\in\mathcal{K}} x_k = \overline{x}$. Since

$x_k \in P_{C_k}(x)$, we have

$$\|x_k - x\| = d(x, C_k), \qquad \forall\ k \in \mathcal{K}.$$

Taking the limit in the preceding relation along the relevant subsequence and using Proposition 2.2.11(a), we get

$$\|\overline{x} - x\| = d(x, C).$$

Since by assumption $C_k \to C$ and $\overline{x} = \lim_{k\to\infty,\ k\in\mathcal{K}} x_k$ with $x_k \in C_k$, we also have that $\overline{x} \in C$, from which, using the preceding relation, we get $\overline{x} \in P_C(x)$, thereby proving that $\limsup_{k\to\infty} P_{C_k}(x) \subset P_C(x)$ for all $x \in \Re^n$.

Conversely, assume that $\limsup_{k\to\infty} d(0, C_k) < \infty$ and

$$\limsup_{k\to\infty} P_{C_k}(x) \subset P_C(x), \qquad \text{for all } x \in \Re^n. \tag{2.5}$$

To show that $C_k \to C$, using Proposition 2.2.11(a), it suffices to show that for all $x \in \Re^n$, $d(x, C_k) \to d(x, C)$. Since the set C_k is closed for all k, it follows that the set $P_{C_k}(x)$ is nonempty for all k. Therefore, for all k, we can choose a vector $x_k \in P_{C_k}(x)$, i.e., $x_k \in C_k$ and

$$\|x_k - x\| = d(x, C_k).$$

From the triangle inequality, we have

$$\|x - y\| \leq \|x\| + \|y\|, \qquad \forall y\ \in C_k,$$

for all k. By taking the minimum over all $y \in C_k$ of both sides in this relation, we get

$$d(x, C_k) \leq \|x\| + d(0, C_k).$$

In view of the assumption that $\limsup_{k\to\infty} d(0, C_k) < \infty$, and the preceding relation, it follows that $\big\{d(x, C_k)\big\}$ forms a bounded sequence. Therefore, using the continuity of the norm, any limit point of this sequence must be of the form $\|\overline{x} - x\|$ for some limit point $\overline{x}$ of the sequence $\{x_k\}$. But, by assumption (2.5), such a limit point $\overline{x}$ belongs to $P_C(x)$, and therefore we have

$$\|\overline{x} - x\| = d(x, C).$$

Hence, the bounded sequence $\{d(x, C_k)\}$ has a unique limit point, $d(x, C)$, implying that $d(x, C_k) \to d(x, C)$ for all $x \in \Re^n$, and proving by Proposition 2.2.11 that $C_k \to C$.

(b) The proof of this part is nearly a verbatim repetition of the proof of part (a), once we use the result of Proposition 2.2.11(b) instead of the result of Proposition 2.2.11(a) in part (a).

(c) We first assume that $C_k \to C$. From part (b), this is equivalent to the conditions,

$$\limsup_{k\to\infty} d(0, C_k) < \infty, \tag{2.6}$$

$$\limsup_{k\to\infty} P_{C_k}(x_k) \subset P_C(x), \quad \text{for all } \{x_k\} \to x \text{ and all } x \in \Re^n. \tag{2.7}$$

It can be seen that condition (2.7) is equivalent to

$$\limsup_{k\to\infty} \text{gph}(P_{C_k}) \subset \text{gph}(P_C). \tag{2.8}$$

We want to show that

$$\lim_{k\to\infty} \text{gph}(P_{C_k}) = \text{gph}(P_C).$$

We will show that $\text{gph}(P_C) \subset \liminf_{k\to\infty} \text{gph}(P_{C_k})$, which together with Eq. (2.8) and the relation $\liminf_{k\to\infty} \text{gph}(P_{C_k}) \subset \limsup_{k\to\infty} \text{gph}(P_{C_k})$ proves the desired result.

Let $x_0 \in \Re^n$ and $\overline{x}_0 \in P_C(x_0)$. For any $\epsilon > 0$, we define $x_\epsilon = (1-\epsilon)x_0 + \epsilon\overline{x}_0$. It can be verified using triangle inequality that the set $P_C(x_\epsilon)$ consists of a single vector $\overline{x}_0$. For each k, we select any $x_k \in P_{C_k}(x_\epsilon)$. (This can be done since the set $P_{C_k}(x)$ is nonempty for all k.) Using the triangle inequality, we get

$$\|x_k - x_\epsilon\| = d(x_\epsilon, C_k) \le d(0, C_k) + \|x_\epsilon\|.$$

By assumptions (2.6) and (2.7)and the preceding relation, we have that the sequence $\{x_k\}$ is bounded and all its limit points belong to $P_C(x_\epsilon)$, which only contains the vector $\overline{x}_0$. Hence, the vectors $(x_\epsilon, x_k) \in \text{gph}(P_{C_k})$ converge to $(x_\epsilon, \overline{x}_0) \in \text{gph}(P_C)$, which by definition of the inner limit of a sequence of sets implies that $(x_\epsilon, \overline{x}_0) \in \liminf_{k\to\infty} \text{gph}(P_{C_k})$. This being true for arbitrary $\epsilon > 0$, we get $(x_0, \overline{x}_0) \in \liminf_{k\to\infty} \text{gph}(P_{C_k})$, proving the desired claim.

Conversely, assume that $\lim_{k\to\infty} \mathrm{gph}(P_{C_k}) = \mathrm{gph}(P_C)$. This implies that $\limsup_{k\to\infty} \mathrm{gph}(P_{C_k}) \subset \mathrm{gph}(P_C)$, which is equivalent to condition (2.7). It remains to show condition (2.6). By assumption, for any $(x, \overline{x}) \in \mathrm{gph}(P_C)$, there exists a sequence of points $(x_k, \overline{x}_k) \in \mathrm{gph}(P_{C_k})$ that converge to $(x, \overline{x})$. From triangle inequality, we have

$$d(0, C_k) \leq \|x_k\| + \|x_k - \overline{x}_k\|.$$

Taking the limit in the preceding relation, we get $\limsup_{k\to\infty} d(0, C_k) < \infty$, hence proving that $C_k \to C$. **Q.E.D.**

A remarkable feature of set convergence is that there is an associated "compactness property," i.e., for any sequence of sets $\{C_k\}$, there exists a convergent subsequence, as shown in the next proposition.

Proposition 2.2.13 (Extraction of Convergent Subsequences): Let $\{C_k\}$ be a sequence of nonempty subsets of $\Re^n$. Let $\overline{x}$ be a vector in the outer limit set $\limsup_{k\to\infty} C_k$. The sequence $\{C_k\}$ has a subsequence converging to a nonempty subset C of $\Re^n$ that contains $\overline{x}$.

Proof: Let $\overline{x}$ be a vector in the set $\limsup_{k\to\infty} C_k$. Then, there exists an index set N_0 and a corresponding subsequence $\{x_k\}_{k\in N_0}$ such that $x_k \in C_k$ for all $k \in N_0$and $\lim_{k\to\infty,\ k\in N_0} x_k = \overline{x}$. Consider the countable collection of open balls given in Proposition 2.2.10 [i.e., balls $\mathrm{int}\big(B(x,\epsilon)\big)$ where ϵ and the coordinates of x are rational] and arrange them in a sequence $\{O_k\}$. We construct a nest of index sets $N_0 \supset N_1 \supset N_2 \supset \cdots$ by defining

$$N_j = \begin{cases} \{k \in N_{j-1} \mid C_k \cap O_j \neq \emptyset\} & \text{if this set of indices is infinite,} \\ \{k \in N_{j-1} \mid C_k \cap O_j = \emptyset\} & \text{otherwise.} \end{cases}$$

Finally, we form another index set N by taking the first index in N_0, and at each step, letting the j^{th} element of N to be the first index in N_j larger than all the indices previously added to N. Then, N has infinitely many elements, and for each j, either $C_k \cap O_j \neq \emptyset$ for all but finitely many $k \in N$ or $C_k \cap O_j = \emptyset$ for all but finitely many $k \in N$.

Let $C = \limsup_{k\in N} C_k$. The set C contains $\overline{x}$, and is therefore nonempty. For each of the balls O_j intersecting C, it can't be true that $C_k \cap O_j = \emptyset$ for all but finitely many k, so such balls O_j must be in the other category in the above construction scheme: we must have $C_k \cap O_j \neq \emptyset$ for all but finitely many $k \in N$. Therefore, using Proposition 2.2.10, it follows that $C \subset \liminf_{k\in N} C_k$, proving that $\lim_{k\to\infty,\ k\in N} C_k = C$. **Q.E.D.**

2.2.3.2. Polars of Tangent and Normal Cones

In this section, we derive polar cones corresponding to tangent and normal cones. We will use these results to relate tangent and normal cones to each other and obtain alternative characterizations of regularity. These will also be useful later on in our analysis in Chapter 4.

Proposition 2.2.14 (Polar of the Tangent Cone): Let C be a subset of $\Re^n$. A vector $v \in \Re^n$ belongs to $T_C(x)^*$ if and only if

$$v'(\overline{x} - x) \le o\big(\|\overline{x} - x\|\big), \qquad \forall\, \overline{x} \in C.$$

Proof: Let v be a vector such that

$$v'(\overline{x} - x) \le o\big(\|\overline{x} - x\|\big), \qquad \forall\, \overline{x} \in C. \tag{2.9}$$

Let y be an arbitrary vector in $T_C(x)$. By Prop. 2.2.8, this implies the existence of sequences $\{x_k\} \subset C$ and $\alpha_k \downarrow 0$ with

$$y_k = \frac{x_k - x}{\alpha_k} \to y.$$

We have from Eq. (2.9) that

$$v'(x_k - x) \le o\big(\|x_k - x\|\big),$$

from which using the definition of y_k, we get

$$v'\frac{(x_k - x)}{\alpha_k} = v'y_k \le \frac{o\big(\|x_k - x\|\big)}{\alpha_k} = \frac{o\big(\alpha_k\|y_k\|\big)}{\alpha_k}.$$

Taking the limit as $k \to \infty$ in the preceding relation together with the fact that $y_k \to y$, we obtain

$$v'y \leq 0.$$

Since $y \in T_C(x)$ is arbitrary, this shows that $v \in T_C(x)^*$.

To show the converse, we first note that the property given in Eq. (2.9) is equivalent to the condition that for all sequences $\{x_k\} \subset C$ such that $x_k \to x$ with $x_k \neq x$ for all k, we have

$$\limsup_{k\to\infty} v' \frac{x_k - x}{\|x_k - x\|} \leq 0.$$

Suppose that vector v does not satisfy this condition, i.e., there exists a sequence $\{x_k\} \subset C$ such that $x_k \to x$ with $x_k \neq x$ for all k and

$$\limsup_{k\to\infty} v' \frac{x_k - x}{\|x_k - x\|} > 0.$$

We show that $v \notin T_C(x)^*$. Denote

$$y_k = \frac{x_k - x}{\|x_k - x\|}.$$

By passing to the appropriate subsequence if necessary and using the continuity of the inner product, we assume without loss of generality that $y_k \to y$ with

$$v'y > 0.$$

We also have by definition of a tangent that $y \in T_C(x)$. Hence, it follows that $v \notin T_C(x)^*$, concluding the proof. **Q.E.D.**

We next define a special kind of normal vector, which can be used to approximate normal vectors and is easier to use in analysis.

Proposition 2.2.15 (Proximal Normals): Let C be a closed subset of $\Re^n$ and for any $x \in \Re^n$, let $\overline{x}$ be the projection of x on C, i.e., $\overline{x} \in P_C(x)$. Any vector v of the form $v = \lambda(x - \overline{x})$ for some $\lambda \geq 0$ is called a *proximal normal* to C at $\overline{x}$. [The proximal vectors to C at $\overline{x}$ are thus the vectors v such that $\overline{x} \in P_C(\overline{x} + \tau v)$ for some $\tau > 0$.]

(a) Every proximal normal vector is a normal vector.

(b) Assume that C is convex. Every normal vector is a proximal normal vector.

Proof:

(a) Let v be a proximal normal vector to C at some $\overline{x} \in C$. By definition, this implies that $v = \lambda(x - \overline{x})$ for some $\lambda \geq 0$ with $\overline{x} \in P_C(x)$. Hence, $\overline{x} \in \arg\min_{y \in C} \frac{1}{2}\|y - x\|^2$, which, using the necessary optimality condition, implies that

$$(\overline{x} - x)'y \geq 0, \qquad \forall\, y \in T_C(\overline{x}),$$

and therefore $(x - \overline{x}) \in T_C(\overline{x}) \subset N_C(\overline{x})$. Since $N_C(x)$ is a cone, it follows that $\lambda(x - \overline{x}) \in N_x(\overline{x})$ for all $\lambda \geq 0$, showing that $v \in N_C(\overline{x})$.

(b) Let $v \in N_C(\overline{x})$ for some $\overline{x} \in C$. Consider the function $f(x) = \frac{1}{2}\|x - (\overline{x} + v)\|^2$. The gradient of this function at $\overline{x}$ is $\nabla f(\overline{x}) = -v$, which by assumption satisfies $-\nabla f(\overline{x}) \in N_C(\overline{x})$. Since C is a convex set and f is a strictly convex function, this condition is necessary and sufficient for $\overline{x}$ to minimize $f(x)$ over C. Hence, we have $\overline{x} \in P_C(\overline{x} + v)$, which by definition implies that v is a proximal normal to C at $\overline{x}$, and concludes the proof.
Q.E.D.

For a nonconvex set C, there can be normal vectors that are not proximal normal vectors. Consider the set C given by

$$C = \big\{(x_1, x_2) \mid x_2 \geq x_1^{3/5},\ x_2 \geq 0\big\}.$$

The vector $v = (1, 0)$ is a normal vector to C at $\overline{x} = (0, 0)$. However, no point of $\{\overline{x} + \tau v \mid \tau > 0\}$ projects onto $\overline{x}$, implying that v is not a proximal normal to C at $\overline{x}$. The next proposition illustrates how proximal normal vectors can be used to approximate normal vectors.

Proposition 2.2.16 (Approximation of Normal Vectors): Let C be a nonempty closed subset of $\Re^n$ and let $\overline{x} \in C$ and $\overline{v} \in N_C(\overline{x})$.

(a) There exist a feasible sequence $x_k \to \overline{x}$ and a sequence $v_k \to \overline{v}$ such that v_k is a proximal normal to C at x_k (see Proposition 2.2.15), and therefore $v_k \in N_C(x_k)$.

(b) Let $\{C_k\}$ be a sequence of nonempty closed sets with $\limsup_{k\to\infty} C_k = C$. There exists a subsequence $\{C_k\}_{k\in\mathcal{K}}$ together with vectors $x_k \in C_k$ and proximal normals $v_k \in N_{C_k}(x_k)$ such that

$$\{x_k\}_{k\in\mathcal{K}} \to \overline{x}, \qquad \{v_k\}_{k\in\mathcal{K}} \to \overline{v}.$$

(c) Let I be an arbitrary index set and for each $i \in I$, let $\{C_k^i\}$ be a sequence of nonempty closed sets with $\bigcup_{i\in I} \limsup_{k\to\infty} C_k^i = C$. For each $i \in I$ such that $\overline{x} \in \limsup_{k\to\infty} C_k^i$, there exists a subsequence $\{C_k^i\}_{k\in\mathcal{K}}$ together with vectors $x_k \in C_k^i$ and proximal normals $v_k \in N_{C_k^i}(x_k)$ such that

$$\{x_k\}_{k\in\mathcal{K}} \to \overline{x}, \qquad \{v_k\}_{k\in\mathcal{K}} \to \overline{v}.$$

Proof:

(a) It is sufficient to treat the case when $\overline{v} \in T_C(\overline{x})^*$. (The more general case follows straightforwardly from the definition of a normal vector.) Assume without loss of generality that $\|\overline{v}\| = 1$. For a sequence of values $\epsilon_k \downarrow 0$, consider the vectors

$$\tilde{x}_k = \overline{x} + \epsilon_k \overline{v}, \tag{2.10}$$

and their projection set $P_C(\tilde{x}_k)$, which is nonempty. For each k, let $x_k \in P_C(\tilde{x}_k)$. It follows from Eq. (2.10) that $x_k \to \overline{x}$. consider the proximal normals to C at x_k defined by

$$v_k = \frac{\tilde{x}_k - x_k}{\epsilon_k} = \overline{v} + \frac{\overline{x} - x_k}{\epsilon_k}. \tag{2.11}$$

We now show that $v_k \to \overline{v}$. Since $x_k \in P_C(\tilde{x}_k)$ and $\|\tilde{x}_k - \overline{x}\| = \epsilon_k$, we have $\epsilon_k \geq \|\tilde{x}_k - x_k\|$, so that $\|v_k\| \leq 1$. We get from Eq. (2.11)

$$\begin{aligned}
\|\overline{v} - v_k\|^2 &\leq 2 - 2\overline{v}'v_k \\
&= 2 - 2\left(1 + \overline{v}'\frac{\overline{x} - x_k}{\epsilon_k}\right) \\
&= 2\overline{v}'\frac{x_k - \overline{x}}{\epsilon_k} \\
&\leq 2\frac{o(\|x_k - \overline{x}\|)}{\|x_k - \overline{x}\|}.
\end{aligned}$$

In the last inequality, we used the assumption that $\overline{v} \in T_C(\overline{x})^*$ together with the characterization of $T_C(\overline{x})^*$ given in Proposition 2.2.14. We also used the fact that $\epsilon_k \geq \|x_k - \overline{x}\|$, which follows from elementary geometry and the fact that $(\tilde{x}_k - x_k) \in T_C(x_k)^*$ (cf. Proposition 2.2.15). Taking the limit in the preceding relation as $k \to \infty$, we see that $v_k \to \overline{v}$, hence proving our claim.

(b) Using part (a), it is sufficient to consider the case when $\overline{v}$ is a proximal normal to C at $\overline{x}$, i.e., for some $\tau > 0$, we have $\overline{x} \in P_C(\overline{x} + \tau\overline{v})$. From Proposition 2.2.13, we have that there exists a subsequence $\{C_k\}_{k\in\mathcal{K}}$ with $\lim_{k\to\infty,\ k\in\mathcal{K}} C_k = Y$, such that $\overline{x}$ belongs to Y. Since $Y \subset C$ and $\overline{x} \in Y$, it follows that $\overline{x} \in P_Y(\overline{x} + \tau\overline{v})$. Then, using Proposition 2.2.12(c), we have that there exist sequences

$$\{x_k\}_{k\in\mathcal{K}} \to \overline{x}, \qquad \{u_k\}_{k\in\mathcal{K}} \to \overline{x} + \tau\overline{v},$$

with $x_k \in P_{C_k}(u_k)$. Equivalently, there exist sequences

$$\{x_k\}_{k\in\mathcal{K}} \to \overline{x}, \qquad \{v_k\}_{k\in\mathcal{K}} \to \overline{v}$$

with $x_k \in P_{C_k}(x_k + \tau v_k)$, which implies that v_k is a proximal normal to C_k at x_k, and concludes the proof.

(c) The proof of this part is nearly a verbatim repetition of the proof of part (b) once we focus on the outer limit sets $\limsup_{k\to\infty} C_k^i$ that contain $\overline{x}$. **Q.E.D.**

We next characterize the polar of the normal cone. The estimate obtained in part (c) of the following proposition is crucial for this purpose.

Proposition 2.2.17 (Normals to Tangent Cones): Let C be a closed nonempty subset of $\Re^n$ and let $x \in C$. Denote the tangent cone of C at x by T for notational convenience.

(a) For any vector $w \in T$ and scalar $\lambda > 0$,

$$T_T(\lambda w) = T_T(w).$$

(b) $N_T(0) = \bigcup_{w \in T} N_T(w) \subset N_C(x)$.

(c) For any vector $w \notin T$, there is a vector $v \in N_C(x)$ with $\|v\| = 1$ such that $d(w, T) = v'w$, where $d(w, T)$ denotes the distance of the vector w to set T, i.e., $d(w, T) = \min_{y \in T} \|y - w\|$.

Proof:

(a) Let y be a vector in $T_T(w)$. By definition, this implies the existence of sequences $\{w_k\} \subset T$ with $w_k \to w$, and $\alpha_k \downarrow 0$ such that

$$\frac{w_k - w}{\alpha_k} \to y,$$

or equivalently for some $\lambda > 0$,

$$\frac{\lambda w_k - \lambda w}{\alpha_k} \to \lambda y.$$

Since T is a cone, the sequence $\{\lambda w_k\} \subset T$ with $\lambda w_k \to \lambda w$. This implies that $\lambda y \in T_T(\lambda w)$. Because $T_T(\lambda w)$ is also a cone and $\lambda > 0$, we also have $y \in T_T(\lambda w)$.

Conversely, let $y \in T_T(\lambda w)$ for some $\lambda > 0$. By definition, this implies the existence of sequences $\{w_k\} \subset T$ with $w_k \to \lambda w$, and $\alpha_k \downarrow 0$ such that

$$\frac{w_k - \lambda w}{\alpha_k} \to y,$$

or equivalently,

$$\frac{w_k/\lambda - w}{\alpha_k/\lambda} \to y.$$

Since T is a cone and $\lambda > 0$, the sequence $\left\{\frac{w_k}{\lambda}\right\} \subset T$ with $\frac{w_k}{\lambda} \to w$. Moreover, $\frac{\alpha_k}{\lambda} \downarrow 0$, implying together with the preceding relation that $y \in T_T(w)$ and concluding the proof.

(b) We first show that $N_T(0) = \bigcup_{w\in T} N_T(w)$. Clearly $N_T(0) \subset \bigcup_{w\in T} N_T(w)$. Next, we show that for all $w \in T$, $N_T(w) \subset N_T(0)$. Let $y \in N_T(w)$. By definition, this implies that there exist sequences $\{w_k\} \subset T$ with $w_k \to w$ and $y_k \to y$ with $y_k \in T_T(w_k)^*$. Consider next the sequence $\{\lambda_k w_k\}$ for an arbitrary sequence $\lambda_k \downarrow 0$. Since T is a cone, it follows that

$$\{\lambda_k w_k\} \subset T, \qquad \lambda_k w_k \to 0.$$

We also have from part (a) that

$$T_T(\lambda_k w_k) = T_T(w_k), \qquad \forall\ k.$$

Hence, there exist sequences $\{\lambda_k w_k\} \subset T$ with $\lambda_k w_k \to 0$ and $y_k \to y$ with $y_k \in T_T(\lambda_k w_k)^*$, which by definition of a tangent implies that $y \in N_T(0)$, and thus proving that $\bigcup_{w\in T} N_T(w) \subset N_T(0)$.

Next, we show that $N_T(0) \subset N_C(x)$. Consider any vector $v \in N_T(0)$. By definition, the tangent cone can equivalently be represented as

$$T = \bigcup \left\{ \limsup_{k\to\infty} \frac{C - x}{\tau_k} \;\Big|\; \tau_k \downarrow 0 \right\},$$

i.e., the union is taken over all sequences $\tau_k \downarrow 0$. Hence, from Proposition 2.2.16, there exists a sequence $\tau_k \downarrow 0$ along with points $w_k \in T_k = \frac{C-x}{\tau_k}$ and vectors $v_k \in N_{T_k}(w_k)$ such that $w_k \to 0$ and $v_k \to v$. For each k, define

$$x_k = x + \tau_k w_k.$$

We have that $\{x_k\} \subset C$ with $x_k \to x$. It also follows that $N_{T_k}(w_k) = N_C(x_k)$. Hence, there exist sequences $\{x_k\} \subset C$ with $x_k \to x$ and $v_k \to v$ with $v_k \in N_C(x_k)$. Using the definition of the normal cone, this implies that $v \in N_C(x)$ and concludes the proof.

(c) Consider the set $P_T(w) = \arg\min_{y\in T} \|y - w\|$, which is nonempty since T is closed. Let $\overline{w} \in P_T(w)$, and define

$$v = \frac{w - \overline{w}}{\|w - \overline{w}\|},$$

which is a proximal normal to T at $\overline{w}$ and therefore, by Proposition 2.2.15, $v \in N_T(\overline{w})$, which implies by part (a) that $v \in N_C(x)$. This establishes the first part of the assertion. We now show that

$$d(w, T) = \|w - \overline{w}\| = v'w.$$

For this purpose, consider the function $f(\tau) = \frac{1}{2}\|w - \tau\overline{w}\|^2$. Since T is a closed cone, we have $\tau\overline{w} \in T$ for all $\tau \geq 0$, and because $\overline{w} \in P_T(w)$, the minimum of $f(\tau)$ over $\tau \geq 0$ is attained at $\tau = 1$. This implies

$$\nabla f(1) = -(w - \overline{w})'\overline{w} = 0.$$

Hence, it follows that $v'\overline{w} = 0$ and

$$v'(w - \overline{w}) = \|w - \overline{w}\| = v'w,$$

which is the desired result. **Q.E.D.**

Proposition 2.2.18 (Polar of the Normal Cone): Let C be a closed subset of $\Re^n$. A vector $w \in \Re^n$ belongs to $N_C(x)^*$ if and only if for every sequence $\{x_k\} \subset C$ with $x_k \to x$, there are vectors $w_k \in T_C(x_k)$ such that $w_k \to w$.

Proof: Let w be a vector such that for every sequence $\{x_k\} \subset C$ with $x_k \to x$, there are vectors $w_k \in T_C(x_k)$ such that $w_k \to w$. Let v be an arbitrary vector that belongs to $N_C(x)$. By definition of the normal cone, this implies that there exist sequences $\{\overline{x}_k\} \subset C$ with $\overline{x}_k \to x$ and $v_k \to v$ with $v_k \in T_C(\overline{x}_k)^*$. Hence, for each k we have

$$v_k'w_k \leq 0,$$

which taking the limit as $k \to \infty$ yields

$$v'w \leq 0.$$

Since $v \in N_C(x)$ is arbitrary, we have $w \in N_C(x)^*$.

Conversely, assume that w does not satisfy the condition given in this exercise, i.e., there exists a sequence $\{x_k\} \subset C$ with $x_k \to x$ and $\epsilon > 0$ such that $d\big(w, T_C(x_k)\big) \geq \epsilon$. Using Proposition 2.2.17(c), this implies that there exist vectors $v_k \in N_C(x_k)$ with $\|v_k\| = 1$ and

$$v_k' w = d\big(w, T_C(x_k)\big) \geq \epsilon.$$

Any limit point v of the sequence $\{v_k\}$ belongs to $N_C(x)$, by definition of the normal cone, and satisfies

$$v' w \geq \epsilon > 0,$$

thereby implying that $w \notin N_C(x)^*$, and proving the desired result. **Q.E.D.**

Proposition 2.2.19 (Alternative Characterization of Regularity): Assume that C is closed. An equivalent definition of regularity at x (cf. Definition 2.2.4) is

$$T_C(x) = N_C(x)^*.$$

In particular, if C is regular at x, the cones $T_C(x)$ and $N_C(x)$ are convex.

Proof: Assume that $N_C(x) = T_C(x)^*$. Using the Polar Cone Theorem, this implies

$$N_C(x)^* = \big(T_C(x)^*\big)^* = \mathrm{conv}\big(T_C(x)\big).$$

From Proposition 2.2.18, it follows that $N_C(x)^* \subset T_C(x)$. Together with the preceding relation, this implies that the cone $T_C(x)$ is convex and $N_C(x)^* = T_C(x)$.

Conversely, assume that $N_C(x)^* = T_C(x)$. Using the Polar Cone Theorem, this implies

$$T_C(x)^* = \big(N_C(x)^*\big)^* = \mathrm{conv}\big(N_C(x)\big).$$

By definition of the normal cone, it follows that $T_C(x)^* \subset N_C(x)$. Together with the preceding relation, this implies that the cone $N_C(x)$ is convex and $T_C(x)^* = N_C(x)$. **Q.E.D.**

We have already seen in Chapter 1 necessary optimality conditions for optimizing arbitrary cost functions over convex constraint sets. In this section we present classical optimality conditions for different types of constrained optimization problems. Here, we do not assume any structure on the constraint set C. For the proofs of these results, see [BNO02]. In particular, we consider problems involving:

(a) A smooth cost function and an arbitrary constraint set.

(b) A convex (not necessarily smooth) cost function and a convex constraint set.

(c) A convex (not necessarily smooth)cost function and an arbitrary constraint set.

When the constraint set is nonconvex, the tangent cone defined in the preceding section is used as a suitable approximation to the constraint set, as illustrated in the following basic necessary condition for local optimality.

Proposition 2.3.20: Let $f : \Re^n \mapsto \Re$ be a smooth function, and let x^* be a local minimum of f over a set $C \subset \Re^n$. Then

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

If C is convex, this condition can be equivalently written as

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

and in the case where $C = \Re^n$, reduces to $\nabla f(x^*) = 0$.

The necessary condition of Prop. 2.3.20 for a vector $x^* \in C$ to be a local minimum of the function f over the set C can be written as

$$-\nabla f(x^*) \in T_C(x^*)^*. \tag{3.1}$$

An interesting converse was given by Gould and Tolle [GoT71], namely that every vector in $T_C(x^*)^*$ can be described as the negative of the gradient of a function having x^* as a local minimum over C. The following version of this result was given by Rockafellar and Wets ([RoW98], p. 205).

Proposition 2.3.21: If x^* is a vector in C, then for every $z \in T_C(x^*)^*$, there is a smooth function f with $-\nabla f(x^*) = z$, which achieves a strict global minimum over C at x^*.

We will return to this result and to the subject of conical approximations when we discuss Lagrange multipliers and conical approximations in Chapters 3 and 4.

Proposition 2.3.22: Let $f : \Re^n \mapsto \Re$ be a convex function. A vector x^* minimizes f over a convex set $C \subset \Re^n$ if and only if there exists a subgradient $d \in \partial f(x^*)$ such that

$$d'(x - x^*) \geq 0, \qquad \forall\, x \in C.$$

Equivalently, x^* minimizes f over a convex set $C \subset \Re^n$ if and only if

$$0 \in \partial f(x^*) + T_C(x^*)^*,$$

where $T_C(x^*)^*$ is the polar of the tangent cone of C at x^*.

We finally extend the optimality conditions of Props. 2.3.20 and 2.3.22 to the case where the cost function is convex (possibly nondifferentiable).

Proposition 2.3.23: Let x^* be a local minimum of a function $f : \Re^n \mapsto \Re$ over a subset C of $\Re^n$. Assume that the tangent cone $T_C(x^*)$ is convex and f is convex. Then

$$-\partial f_1(x^*) \in T_C(x^*)^*.$$

CHAPTER 3

ENHANCED OPTIMALITY CONDITIONS AND DIFFERENT TYPES OF LAGRANGE MULTIPLIERS

We focus on optimization problems of the form

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C, \end{aligned} \tag{0.1}$$

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

$$C = X \cap \{x \mid h_1(x) = 0, \dots, h_m(x) = 0\} \cap \{x \mid g_1(x) \le 0, \dots, g_r(x) \le 0\}. \tag{0.2}$$

We assume in Chapters 3 and 4 that f, h_i, g_j are continuously differentiable (smooth) functions from $\Re^n$ to $\Re$, and X is a nonempty closed set. More succinctly, we can write problem (0.1)-(0.2) as

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad h(x) = 0, \quad g(x) \le 0, \end{aligned}$$

where $h : \Re^n \mapsto \Re^m$ and $g : \Re^n \mapsto \Re^r$ are functions

$$h = (h_1, \dots, h_m), \qquad g = (g_1, \dots, g_r).$$

We have seen in Chapter 2 that a classical necessary condition for a vector $x^* \in C$ to be a local minimum of f over C is

$$\nabla f(x^*)' y \ge 0, \qquad \forall\, y \in T_C(x^*), \tag{0.3}$$

where $T_C(x^*)$ is the tangent cone of C at x^*. In this chapter, our objective is to develop optimality conditions for problem (0.1) that take into account the specific representation of set C in terms of constraint functions h_i and g_j. These optimality conditions are sharper than the classical Karush-Kuhn-Tucker conditions in that they include extra conditions, which may narrow down the set of candidate local minima. 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. For this purpose, we make use of the variational geometry concepts that we have developed in Chapter 2. These optimality conditions motivate the introduction of new types of Lagrange multipliers that differ in the amount of sensitivity information they provide. In this chapter, we also investigate existence of such multipliers and their relations.

3.1. CLASSICAL THEORY OF LAGRANGE MULTIPLIERS

Necessary optimality conditions for optimization problems with equality constraints have been known for a long time, in fact since Euler and Lagrange. Lagrange multiplier theorems for inequality constraints come considerably later. Important works in this area were done by Karush [Kar39], and Kuhn and Tucker[KuT50], who essentially proved the same result under different assumptions. The next proposition presents this result under a "regularity" assumption (for the proof, see [Ber99]). For any feasible vector x, the *set of active inequality constraints* is denoted by

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

We say that a feasible vector of problem (0.1)-(0.2) is *regular* when $X = \Re^n$, and 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.

Proposition 3.1.1 (Karush-Kuhn-Tucker Necessary Optimality Conditions): Let x^* be a local minimum of problem (0.1)-(0.2), where $X = \Re^n$, and assume that x^* is regular. Then there exist unique Lagrange multiplier vectors $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)$, $\mu^* = (\mu_1^*, \dots, \mu_r^*)$, such that

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

$$\mu_j^* \geq 0, \qquad j = 1, \dots, r,$$

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

The preceding is an important, widely-used result; however it is limited by the regularity assumption. Although this assumption is natural for equality constraints, it is somewhat restrictive for inequality constraints. The reason is that in many types of problems; for instance linear programming problems, there may be many inequality constraints that are satisfied as equalities at a local minimum, but the corresponding gradients are linearly dependent because of inherent symmetries in the problem's structure. Therefore, we would like to have a development of the Lagrange multiplier theory that is not based on regularity-type assumptions.

Moreover, the preceding proposition does not apply to the case where the constraint set description includes an additional abstract set constraint X. We would like to build up a theory that handles such constraint sets. With this motivation, we give the following definition of Lagrange multipliers.

Definition 3.1.1: We say that the constraint set C, as represented by Eq. (0.2), *admits Lagrange multipliers* at a point $x^* \in C$ if for every smooth cost function f for which x^* is a local minimum of problem (0.1) there exist vectors $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)$ and $\mu^* = (\mu_1^*, \dots, \mu_r^*)$ that satisfy the following conditions:

$$\left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right)' y \geq 0, \quad \forall\, y \in T_X(x^*), \tag{1.1}$$

$$\mu_j^* \geq 0, \qquad \forall\, j = 1, \ldots, r, \tag{1.2}$$

$$\mu_j^* = 0, \qquad \forall\, j \notin A(x^*). \tag{1.3}$$

A pair (λ^*, μ^*) satisfying Eqs. (1.1)-(1.3) is called a *Lagrange multiplier vector corresponding to* f *and* x^*.

When there is no danger of confusion, we refer to (λ^*, μ^*) simply as a *Lagrange multiplier vector* or a *Lagrange multiplier*, without reference to the corresponding local minimum x^* and the function f. Figure 3.1.1 illustrates the definition of a Lagrange multiplier. Condition (1.3) is referred to as the *complementary slackness* condition (CS for short). Note that from Eq. (1.1), 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 or unbounded) closed and convex set.

The condition (1.1) 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 $x \in X$ (cf. Prop. 2.3.22). This is consistent with the traditional characteristic property of Lagrange multipliers: *rendering the Lagrangian function stationary at* x^*. When X is a convex set, Eq. (1.1) is equivalent to

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

This is because when X is convex, $T_X(x^*)$ is equal to the closure of the set of feasible directions $F_X(x^*)$ (cf. Proposition 2.2.9), which is in turn equal to the set of vectors of the form $\alpha(x - x^*)$, where $\alpha > 0$ and $x \in X$. If $X = \Re^n$, Eq. (1.4) becomes

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

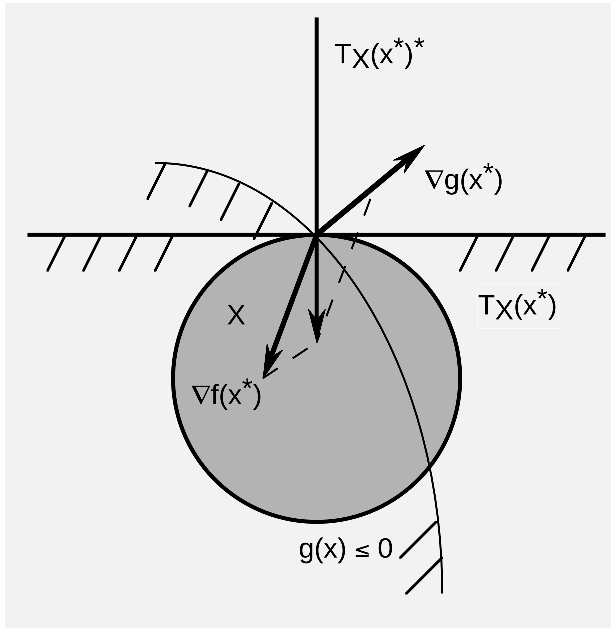


Figure 3.1.1. Illustration of a Lagrange multiplier for the case of a single inequality constraint and the spherical set X shown in the figure. The tangent cone $T_X(x^*)$ is a closed halfspace and its polar $T_X(x^*)^*$ is the halfline shown in the figure. There is a unique Lagrange multiplier μ^*, and it is such that $-\big(\nabla f(x^*) + \mu^* \nabla g(x^*)\big)$ belongs to $T_X(x^*)^*$.

which together with the nonnegativity condition (1.2) and the CS condition (1.3), comprise the classical Karush-Kuhn-Tucker conditions (cf. Proposition 3.1.1).

3.2. ENHANCED FRITZ JOHN CONDITIONS

The analysis in this thesis is strongly motivated by an enhanced set of optimality conditions, which will be the focus of this section. Weaker versions of these conditions were shown in a largely overlooked analysis by Hestenes [Hes75] for the case where $X = \Re^n$, and in [Ber99] for the case where X is a closed convex set. They are strengthened here [cf. condition (iv) of the following proposition] and further generalized for the case where X is a closed but not necessarily convex set.

The following proposition presents these optimality conditions. It enhances the classical Fritz John optimality conditions by providing additional necessary conditions through a penalty function-based proof. These conditions will turn out to be crucial in our analysis in the next chapter. They also form the basis for enhancing the classical Karush-Kuhn-Tucker conditions. The proposition asserts that there exist multipliers corresponding to a local minimum x^*, including a multiplier μ_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 3.2.3: Let x^* be a local minimum of problem (0.1)-(0.2). Then there exist scalars μ_0^*, $\lambda_1^*, \ldots, \lambda_m^*$, and $\mu_1^*, \ldots, \mu_r^*$, satisfying the following conditions:

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

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

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

(iv) If the index set $I \cup J$ is nonempty where

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

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

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

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

where we denote $g^+(x) = \max\big\{0,\, g_j(x)\big\}$ and

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

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

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

$$\text{subject to} \;\; x \in X \cap S,$$

where we denote

$$S = \{x \mid ||x - x^*|| \leq \epsilon\},$$

and ϵ is positive and 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 can be written as

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

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

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

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

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

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

For $k \geq \overline{k}$, we have the necessary optimality condition of Prop. 2.3.22: $-\nabla F^k(x^k) \in T_X(x^k)^*$, which [by using the formula $\nabla\big(g_j^+(x)\big)^2 = 2g_j^+(x)\nabla g_j(x)$] is written as

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

where

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

Denote

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

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

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

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

Since by construction we have

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

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

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

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

and

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

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

Condition (iv) of Prop. 3.2.3 resembles the "complementary slackness" (CS) condition (1.3) of Karush-Kuhn-Tucker optimality conditions. Recall that, this name derives from the fact that for each j, whenever the constraint $g_j(x^*) \leq 0$ is slack [meaning that $g_j(x^*) < 0$], the constraint $\mu_j^* \geq 0$ must not be slack (meaning that $\mu_j^* = 0$). In analogy with this interpretation, we refer to condition (iv) as the *complementary violation condition* (CV for short), signifying the fact that for all j, $\mu_j^* > 0$ implies $g_j(x) > 0$ for some x arbitrarily close to x^*. This condition can be visualized in the examples of Fig. 3.2.2. It will turn out to be of crucial significance in our development. The next proposition clarifies the relation between CS and CV conditions.

Proposition 3.2.4: Let μ^* be a vector that satisfies CV condition. Then μ^* also satisfies CS condition.

Proof: Let μ^* be a vector that satisfies CV condition. This implies that, if $\mu_j^* > 0$ for some j, then the corresponding jth inequality constraint must be violated arbitrarily close to x^* [cf. Eq. (2.1)]. Hence, we must have $g_j(x^*) = 0$, showing that μ^* satisfies CS condition. **Q.E.D.**

The following example shows that the converse of the preceding statement is not true.

Example 3.2.1:

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

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

Assume that $\nabla f(x^*) = \nabla h(x^*)$ and consider the multipliers $\mu_1^* = 1$, $\mu_2^* = 2$ in Definition 3.1.1. These multipliers satisfy the Lagrangian stationary condition as well as the CS condition. However, they fail the CV condition, since both constraints have positive multipliers and it is not possible to find a vector $x \in \Re^n$ that violates both constraints simultaneously.

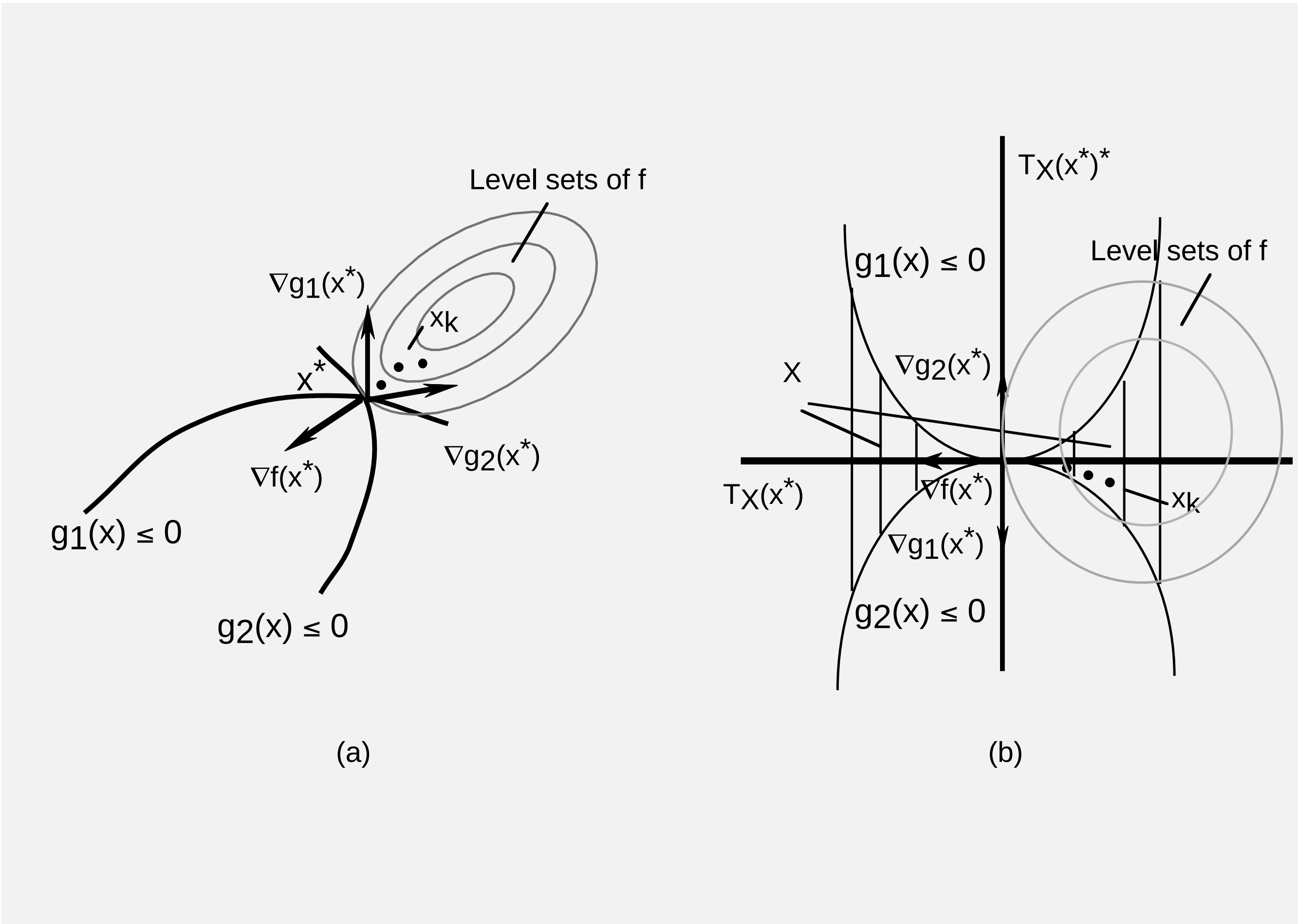


Figure 3.2.2. Illustration of the CV condition. In the example of (a), where $X = \Re^2$, the multipliers that satisfy the enhanced Fritz John conditions are the positive multiples of a unique vector of the form $(1, \mu_1^*, \mu_2^*)$ where $\mu_1^* > 0$ and $\mu_2^* > 0$. It is possible to violate both constraints simultaneously by approaching x^* along the sequence $\{x^k\}$ shown, which has a lower cost value than x^*. In the example of (b) X is the shaded region shown rather than $X = \Re^2$. Origin is the only feasible solution, therefore is optimal for the cost function depicted in the figure. An example of a multiplier that satisfy the enhanced Fritz John conditions is the vector $(0, \mu_1^*, \mu_2^*)$, where $\mu_1^* > 0$ and $\mu_2^* > 0$. It is possible to violate both constraints simultaneously by approaching x^* along the sequence $\{x^k\}$ shown.

If X is regular at x^*, i.e., $N_X(x^*) = T_X(x^*)^*$, condition (i) of Prop. 3.2.3 becomes

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

or equivalently

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

If in addition, the scalar μ_0^* can be shown to be strictly positive, then by normalization we can choose $\mu_0^* = 1$, and condition (i) of Prop. 3.2.3 becomes equivalent to the Lagrangian stationarity condition (1.1). 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 stronger CV condition.

As an example, if there is no abstract set constraint ($X = \Re^n$), and the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, and $\nabla g_j(x^*)$, $j \in A(x^*)$, are linearly independent, we cannot have $\mu_0^* = 0$, since then condition (i) of Prop. 3.2.3 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 presented in Proposition 3.1.1. This is a classical result, found in almost all nonlinear programming textbooks, but it is obtained here through a simple argument and in a stronger form, since it includes the assertion that the multipliers satisfy the stronger CV condition in place of the CS condition.

To illustrate the use of the generalized Fritz John conditions of Prop. 3.2.3 and the CV condition in particular, consider the following example.

Example 3.2.2:

Consider the problem of Example 3.2.1 and let x^* be a local minimum. The Fritz John conditions, in their classical form, assert the existence of nonnegative $\mu_0^*, \lambda^+, \lambda^-$, not all zero, such that

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

The candidate multipliers that satisfy the above condition as well as the CS condition $\lambda^+ h(x^*) = \lambda^- h(x^*) = 0$, include those of the form $\mu_0^* = 0$ and $\lambda^+ = \lambda^- > 0$, which provide no relevant information about the problem. However, these multipliers fail the stronger CV condition of Prop. 3.2.3, showing that if $\mu_0^* = 0$, we must have either $\lambda^+ \neq 0$ and $\lambda^- = 0$ or $\lambda^+ = 0$ and $\lambda^- \neq 0$. Assuming $\nabla h(x^*) \neq 0$, this violates Eq. (2.11), so it follows that $\mu_0^* > 0$. Thus, by dividing Eq. (2.11) by μ_0^*, we recover the familiar first order condition $\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$ with $\lambda^* = (\lambda^+ - \lambda^-)/\mu_0^*$, under the assumption $\nabla h(x^*) \neq 0$. Note that this deduction would not have been possible without the CV condition.

We will further explore the CV condition as a vehicle for characterizing Lagrange multipliers in the next section.

3.3. DIFFERENT TYPES OF LAGRANGE MULTIPLIERS

Motivated by the complementary violation condition of the preceding section, we introduce different types of Lagrange multipliers in this section, which carry different amount of information about sensitivity to constraints of the problem, and investigate their relations.

3.3.1. Minimal Lagrange Multipliers

In some applications, it may be of analytical or computational interest to deal with Lagrange multipliers 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.

In the next proposition, we will show that under some convexity assumptions regarding the abstract set constraint X, every minimal Lagrange multiplier possesses significant amount of sensitivity information. For this purpose, we first make the following definition. In particular, let us say that a Lagrange multiplier (λ^*, μ^*) is *strong* if in addition to Eqs. (1.1)-(1.3), it satisfies the condition

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

$$f(x^k) < f(x^*), \quad \lambda_i^* h_i(x^k) > 0, \;\; \forall\, i \in I, \quad \mu_j^* g_j(x^k) > 0, \;\; \forall\, j \in J. \tag{3.1}$$

Proposition 3.3.5: Let x^* be a local minimum of problem (0.1)-(0.2). Assume that the tangent cone $T_X(x^*)$ is convex and that the set of Lagrange multipliers is nonempty. Then, each minimal Lagrange multiplier vector is strong.

Proof: We first show the following lemma, which is of independent interest.

Lemma 3.3.1: Let N be a closed convex cone in $\Re^n$. Let $a_0, a_1, \ldots, a_r$ be given vectors in $\Re^n$, and A be the cone generated by $a_1, \ldots, a_r$:

$$A = \left\{ \sum_{j=1}^{r} \mu_j a_j \;\middle|\; \mu_j \geq 0,\, j = 1, \ldots, r \right\}.$$

Assume that the sets $-a_0 + A$ and N have nonempty intersection. Among index subsets $J \subset \{1, \ldots, r\}$ such that, for some positive μ_j, $j \in J$, we have $\left(-a_0 + \sum_{j \in J} \mu_j a_j\right) \in N$, let $\overline{J} \subset \{1, \ldots, r\}$ have a minimal number of elements. If the set $\overline{J}$ is nonempty, then there exists a vector $y \in N^*$ such that

$$a_j' y < 0, \qquad \forall\, j \in \overline{J} \cup \{0\}.$$

Proof: Consider the index set $\overline{J}$ defined in the Lemma, and let $\overline{\mu}_j > 0$, $j \in \overline{J}$, be such that

$$\left(-a_0 + \sum_{j \in \overline{J}} \overline{\mu}_j a_j \right) \in N. \tag{3.2}$$

Let $\overline{A}$ be the finitely generated cone

$$\overline{A} = \left\{ y \;\middle|\; y = \sum_{j \in \overline{J}} \mu_j a_j,\; \mu_j \geq 0,\, j \in \overline{J} \right\},$$

and its polar

$$\overline{A}^* = \{ y \mid a_j' y \leq 0,\; j \in \overline{J} \}.$$

It can be seen that the vectors a_j, $j \in \overline{J}$, are linearly independent, since otherwise for some λ_j, $j \in \overline{J}$, all of which are not equal to 0, we would have

$$\sum_{j \in \overline{J}} \lambda_j a_j = 0.$$

Since N is a convex cone, this implies

$$\left(-a_0 + \sum_{j \in \overline{J}} (\overline{\mu}_j - \gamma \lambda_j) a_j \right) \in N,$$

for all scalars γ, and an appropriate value of γ would yield an index subset J with

$$\left(-a_0 + \sum_{j \in J} \mu_j a_j \right) \in N,$$

$\mu_j > 0$, $j \in J$, and a smaller number of elements than those of $\overline{J}$. Thus, we can find a vector y such that $a_j' y < 0$ for all $j \in \overline{J}$, and it follows that the interior of $\overline{A}^*$, given by

$$\mathrm{int}\big(\overline{A}^*\big) = \{y \mid a_j' y < 0,\ j \in \overline{J}\},$$

is nonempty.

Let F be the cone generated by a_0,

$$F = \left\{ \mu_0 a_0 \;\middle|\; \mu_0 \geq 0 \right\}$$

We claim that $\mathrm{int}(\overline{A}^* \cap F^*)$ is nonempty. Note that a_0 cannot be represented as a nonnegative combination of vectors a_j, $j \in J'$, where J' is a strict subset of $\overline{J}$, since this would violate the minimal support assumption in Eq. (3.2). It is possible to have $a_0 = \sum_{j \in \overline{J}} \beta_j a_j$, $\beta_j > 0$ for all j, in which case $F \subset \overline{A}$, which implies by Proposition 2.1.2 of Chapter 2 that $\overline{A}^* \subset F^*$. Hence, we have $\mathrm{int}(\overline{A}^* \cap F^*) = \mathrm{int}(\overline{A}^*)$, which is nonempty. Otherwise, the vectors a_0, and a_j, $j \in \overline{J}$ are all linearly independent and we can find a vector y such that $a_0' y < 0$ and $a_j' y < 0$, for all $j \in \overline{J}$. This implies that $\mathrm{int}(\overline{A}^* \cap F^*)$ is nonempty.

We next show that there exists a $y \in N^* \cap \mathrm{int}\big(\overline{A}^* \cap F^*\big)$. Assume, to arrive at a contradiction, that N^* and $\mathrm{int}\big(\overline{A}^* \cap F^*\big)$ are disjoint. These sets are nonempty [cf. preceding discussion] and convex (since the intersection of two convex sets is convex and the interior of a convex set is convex). Hence, by the Separating Hyperplane Theorem, there exists a nonzero vector p such that

$$p'x_1 \leq p'x_2, \qquad \forall\ x_1 \in \text{int}\big(\overline{A}^* \cap F^*\big),\ \forall\ x_2 \in N^*.$$

Using the Line Segment Principle and the preceding relation, it can be seen that

$$p'x_1 \leq p'x_2, \qquad \forall\ x_1 \in \big(\overline{A}^* \cap F^*\big),\ \forall\ x_2 \in N^*. \tag{3.3}$$

In particular, taking $x_1 = 0$ in this equation, we obtain

$$0 \leq p'x_2, \qquad \forall\ x_2 \in N^*,$$

which implies, by the Polar Cone Theorem, that

$$-p \in (N^*)^* = N.$$

Similarly, taking $x_2 = 0$ in Eq. (3.3), we obtain

$$p'x_1 \leq 0, \qquad \forall\ x_1 \in \big(\overline{A}^* \cap F^*\big),$$

which implies that

$$p \in \big(\overline{A}^* \cap F^*\big)^*.$$

Since both $\overline{A}$ and F are finitely generated cones, their polars are polyhedral cones (see [BNO02]). Therefore, the set $\overline{A}^* \cap F^*$ is polyhedral and closed, and it follows from Proposition 2.1.2 that

$$\big(\overline{A}^* \cap F^*\big)^* = \overline{A} + F,$$

Hence, the vector p belongs to $\overline{A} + F$ and can be expressed as

$$p = \xi_0 a_0 + \sum_{j \in \overline{J}} \xi_j a_j\ , \qquad \xi_j \geq 0,\ \forall\ j \in \overline{J} \cup \{0\}. \tag{3.4}$$

Since $-p \in N$, we have from the preceding relation that

$$-\xi_0 a_0 + \sum_{j \in \overline{J}} (-\xi_j) a_j\ \in N,$$

and therefore

$$-\xi_0\gamma a_0 + \sum_{j\in\overline{J}}(-\xi_j\gamma)a_j \ \in N,$$

for all $\gamma \geq 0$. Together with Eq. (3.2), this implies that

$$-(1+\xi_0\gamma)a_0 + \sum_{j\in\overline{J}}(\overline{\mu}_j - \gamma\xi_j)a_j \in N, \qquad \forall\ \gamma \geq 0. \tag{3.5}$$

Thus we can find $\gamma > 0$ such that $\overline{\mu}_j - \gamma\xi_j \geq 0$, for all j and $\overline{\mu}_{\bar{j}} - \gamma\xi_{\bar{j}} = 0$ for at least one index $\bar{j} \in \overline{J}$. Since $(1+\xi_0\gamma) > 0$, we can divide Eq. (3.5) by $(1+\xi_0\gamma)$ and the resulting coefficient vector has less support than $(\overline{\mu}_1,\ldots,\overline{\mu}_r)$. This contradicts the assumption that the index set $\overline{J}$ has a minimal number of elements. Hence, there must exist a vector y that is in both N^* and $\text{int}\big(\overline{A}^* \cap F^*\big)$, implying that $a_j' y < 0$ for all $j \in \overline{J}\cup\{0\}$. **Q.E.D.**

We now return to the proof of Proposition 3.3.5. 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 apply Lemma 3.3.1 with the following identifications:

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

If $-\nabla f(x^*) \in T_X(x^*)^*$, the scalars $\overline{\mu}_j = 0$, $j = 1,\ldots,r$, form a strong Lagrange multiplier vector, and we are done. So assume that $-\nabla f(x^*) \notin T_X(x^*)^*$. Then, since by assumption, there exist Lagrange multipliers corresponding to x^*, the sets $-a_0 + A$ and N have nonempty intersection [cf. Eq. (1.1)], and by Lemma 3.3.1, there exists a nonempty set $\overline{J} \subset \{1,\ldots,r\}$ and positive scalars $\overline{\mu}_j$, $j \in \overline{J}$, such that

$$-\left(\nabla f(x^*) + \sum_{j\in\overline{J}}\overline{\mu}_j\nabla g_j(x^*)\right) \in T_X(x^*)^*,$$

and a vector $y \in \big(T_X(x^*)^*\big)^* = T_X(x^*)$ (by the Polar Cone Theorem, since by assumption $T_X(x^*)$ is convex) such that

$$\nabla f(x^*)'y < 0, \qquad \nabla g_j(x^*)'y > 0, \qquad \forall\ j \in \overline{J}. \tag{3.6}$$

Let $\{x^k\} \subset X$ be a sequence such that $x^k \neq x^*$ for all k and

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

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

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

From Eq. (3.6), we have $\nabla f(x^*)'y < 0$, so we obtain $f(x^k) < f(x^*)$ for k sufficiently large. Using also Taylor's Theorem for the constraint functions g_j, $j \in \overline{J}$, we have, for some vector sequence ξ^k converging to 0,

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

from which it follows that for k sufficiently large, we have $g_j(x^k) > 0$. It follows that the scalars $\overline{\mu}_j$, $j \in \overline{J}$, together with the scalars $\overline{\mu}_j = 0$, $j \notin \overline{J}$, form a strong Lagrange multiplier vector. **Q.E.D.**

The next example shows that the converse of the preceding result is not true, i.e., there may exist Lagrange multipliers that are strong but are not minimal.

Example 3.3.3:

Let the constraint set C be specified by two identical inequality constraints, $g_1(x) = g_2(x) \leq 0$, and consider the vector x^*. The tangent cone at x^*, $T_C(x^*)$ and its polar $T_C(x^*)^*$ are depicted in Fig. 3.3.3. Let f be a cost function that has a local minimum at x^*. By the necessary condition for optimality, this implies that $-\nabla f(x^*) \in T_C(x^*)^*$. The Lagrange multipliers are determined from the requirement

$$\nabla f(x^*) + \mu_1^* \nabla g_1(x^*) + \mu_2^* \nabla g_2(x^*) = 0. \tag{3.7}$$

Note that by appropriate normalization, we can select $\mu_1^* = \mu_2^* > 0$. These multipliers are strong. However, they are not minimal since we can set only one of the multipliers to be positive and still satisfy Eq. (3.7).

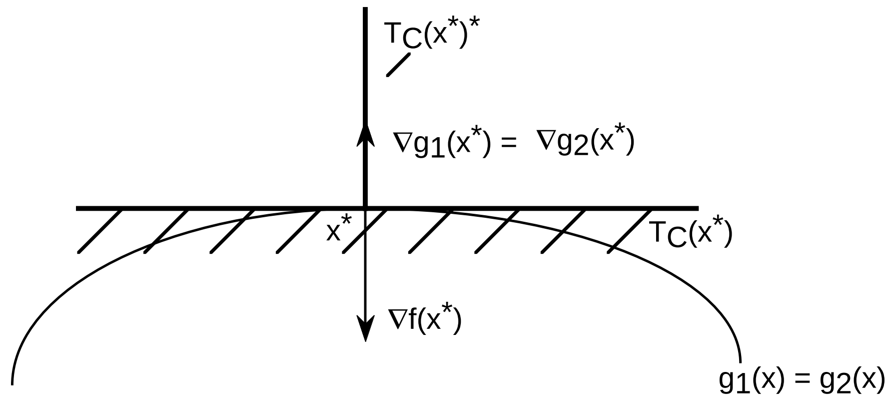


Figure 3.3.3. Constraint set of Example 3.3.5. The tangent cone of the feasible set $T_C(x^*)$ and its polar $T_C(x^*)^*$ are illustrated in the figure.

3.3.2. Informative Lagrange Multipliers

The Lagrange multipliers whose existence is guaranteed by Prop. 3.2.3 (assuming that $\mu_0^* = 1$) are special: 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 (λ^*, μ^*) that satisfies, in addition to Eqs. (1.1)-(1.3), the CV condition [condition (iv) of Prop. 3.2.3] as being *informative*.

Since CV condition is stronger than condition (iv') in the definition of a strong multiplier [cf. Eq. (3.1)], it follows that informative Lagrange multipliers are also strong. We have seen in the previous section that minimal Lagrange multipliers are strong. However, it is not true that minimal Lagrange multipliers are 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. The relations between different types of Lagrange multipliers are illustrated in Fig. 3.3.4.

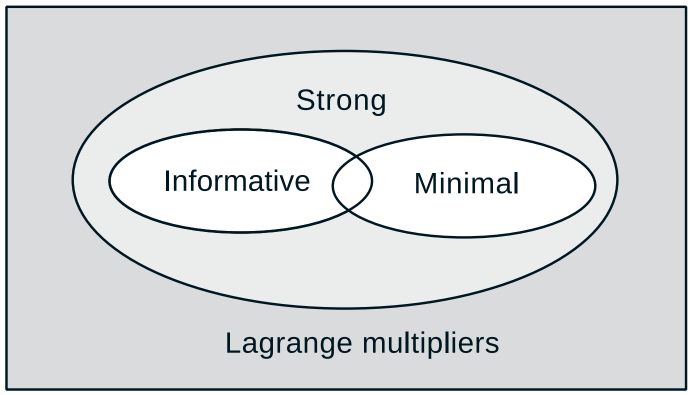


Figure 3.3.4. Relations of different types of Lagrange multipliers, assuming that the tangent cone $T_X(x^*)$ is convex (which is true in particular if X is regular at x^*).

The salient defining property of informative Lagrange multipliers is consistent with the classical sensitivity interpretation of a Lagrange multiplier as the rate of cost improvement 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. 3.2.3. In fact we will show in the next proposition a stronger and more definitive property: *if the tangent cone $T_X(x^*)$ is convex (which is true if X is convex or regular, cf. Proposition 2.2.19 of Chapter 2) and there exists at least one Lagrange multiplier vector, there exists one that is informative.*

Proposition 3.3.6: Let x^* be a local minimum of problem (0.1)-(0.2). Assume that the tangent cone $T_X(x^*)$ is convex and that the set of Lagrange multipliers is nonempty. Then, the set of informative Lagrange multiplier vectors is nonempty, and in fact the Lagrange multiplier vector that has minimum norm is informative.

Proof: We summarize the essence of the proof argument in the following lemma.

Lemma 3.3.2: Let N be a closed convex cone in $\Re^n$, and let $a_0, \ldots, a_r$ be given vectors in $\Re^n$. Suppose that the closed and convex subset of $\Re^r$ given by

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

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

$$a_0' d^k \to -\|\mu^*\|^2, \tag{3.8}$$

$$(a_j' d^k)^+ \to \mu_j^*, \qquad j = 1, \ldots, r, \tag{3.9}$$

where μ^* 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

$$\begin{aligned} -\frac{1}{2}\|\mu^*\|^2 &= \inf_{d \in N^*} \left\{ a_0' d + \frac{1}{2} \sum_{j=1}^{r} \big((a_j' d)^+ \big)^2 \right\} \\ &= \lim_{k \to \infty} \left\{ a_0' d^k + \frac{1}{2} \sum_{j=1}^{r} \big((a_j' d^k)^+ \big)^2 \right\}. \end{aligned} \tag{3.10}$$

In addition, if the problem

$$\begin{aligned} &\text{minimize} \;\; a_0' d + \frac{1}{2} \sum_{j=1}^{r} \big((a_j' d)^+ \big)^2 \\ &\text{subject to} \;\; d \in N^*, \end{aligned} \tag{3.11}$$

has an optimal solution, denoted d^*, we have

$$a_0'd^* = -\|\mu^*\|^2, \qquad (a_j'd^*)^+ = \mu_j^*, \quad j = 1, \ldots, r. \tag{3.12}$$

Proof: We consider the function

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

and we note that L is convex in d, and concave and coercive in μ. For any $\gamma > 0$, we consider saddle points of L over $d \in N^* \cap B(0, 1/\gamma)$ and $\mu \geq 0$, where $B(0, 1/\gamma)$ denotes the closed unit ball centered at 0 with radius $1/\gamma$. From the Saddle Point Theorem, L has a saddle point for each $\gamma > 0$, denoted by (d^γ, μ^γ).

By making a change of variable $d = y/\gamma$ and using the fact that N^* is a cone (and therefore $N^* = N^*/\gamma$), we have

$$\begin{aligned}
\inf_{d \in N^* \cap B(0,1/\gamma)} L(d, \mu^\gamma) &= \inf_{y \in N^* \cap B(0,1)} L\left(\frac{y}{\gamma}, \mu^\gamma \right) \\
&= \inf_{y \in N^* \cap B(0,1)} - \left(\frac{s^{\gamma\prime} y}{\gamma} + \frac{1}{2}\|\mu^\gamma\|^2 \right),
\end{aligned}$$

where

$$s^\gamma = -\left(a_0 + \sum_{j=1}^{r} \mu_j^\gamma a_j \right).$$

Hence, it follows that $d^\gamma = y^\gamma/\gamma$, where

$$y^\gamma \in \arg \min_{y \in N^* \cap B(0,1)} -\frac{s^{\gamma\prime} y}{\gamma}.$$

The necessary optimality condition for the above minimization problem yields

$$s^{\gamma\prime}(y - y^\gamma) \leq 0, \qquad \forall\, y \in N^* \cap B(0, 1),\ \forall\, \gamma > 0. \tag{3.13}$$

In particular, letting $y = 0$ in the preceding relation, we obtain

$$s^{\gamma\prime} y^\gamma \geq 0, \qquad \forall\, \gamma > 0. \tag{3.14}$$

We now note that, for any $\mu \in M$, we have by the definition of M,

$$\left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right)' d \geq 0, \qquad \forall\, d \in N^*,$$

so that

$$\inf_{d \in N^* \cap B(0,1/\gamma)} L(d,\mu) = \inf_{d \in N^* \cap B(0,1/\gamma)} \left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right)' d - \frac{1}{2}\|\mu\|^2 = -\frac{1}{2}\|\mu\|^2.$$

Therefore,

$$\begin{aligned} L(d^\gamma, \mu^\gamma) &= \sup_{\mu \geq 0} \inf_{d \in N^* \cap B(0,1/\gamma)} L(d,\mu) \\ &\geq \sup_{\mu \in M} \inf_{d \in N^* \cap B(0,1/\gamma)} L(d,\mu) \\ &= \sup_{\mu \in M} \left(-\frac{1}{2}\|\mu\|^2\right) \\ &= -\frac{1}{2}\|\mu^*\|^2. \end{aligned}$$

It follows that

$$L(d^\gamma, \mu^\gamma) = L\left(\frac{y^\gamma}{\gamma}, \mu^\gamma\right) = -\frac{s^{\gamma\prime} y^\gamma}{\gamma} - \frac{1}{2}\|\mu^\gamma\|^2 \geq -\frac{1}{2}\|\mu^*\|^2. \tag{3.15}$$

Since $s^{\gamma\prime} y^\gamma \geq 0$ for all $\gamma > 0$ [cf. Eq. (3.14)], we see from the preceding relation that

$$\lim_{\gamma \to 0} s^{\gamma\prime} y^\gamma = 0, \tag{3.16}$$

and that $\|\mu^\gamma\| \leq \|\mu^*\|$, so that μ^γ remains bounded as $\gamma \to 0$. Let $\overline{\mu}$ be a limit point of $\{\mu^\gamma\}$. Taking the limit along the relevant subsequence in Eq. (3.13), and using Eq. (3.16), we get

$$0 \leq \lim_{\gamma \to 0} s^{\gamma\prime}(y - y^\gamma) = -\left(a_0 + \sum_{j=1}^{r} \overline{\mu}_j a_j\right)' y, \qquad \forall\, y \in N^* \cap B(0,1).$$

This implies that $-\left(a_0 + \sum_{j=1}^{r} \overline{\mu}_j a_j\right) \in \left(N^* \cap B(0,1)\right)^* = N + B(0,1)^* = N$. Hence $\overline{\mu} \in M$, and since $\|\overline{\mu}\| \leq \|\mu^*\|$ (in view of $\|\mu^\gamma\| \leq \|\mu^*\|$), by using the minimum norm property of μ^*, we conclude that any limit point $\overline{\mu}$ of μ^γ must be equal to μ^*. Thus $\mu^\gamma \to \mu^*$.

To show Eqs. (3.8) and (3.9), we note that since L is quadratic in μ, the supremum in $\sup_{\mu \geq 0} L(d, \mu)$ is attained at

$$\mu_j = (a_j' d)^+, \qquad j = 1, \ldots, r, \tag{3.17}$$

so that

$$\sup_{\mu \geq 0} L(d, \mu) = a_0' d + \frac{1}{2} \sum_{j=1}^{r} \big((a_j' d)^+\big)^2, \tag{3.18}$$

and

$$L(d^\gamma, \mu^\gamma) = \sup_{\mu \geq 0} L(d^\gamma, \mu) = a_0' d^\gamma + \frac{1}{2} \|\mu^\gamma\|^2. \tag{3.19}$$

From Eq. (3.15) and the facts $\mu^\gamma \to \mu^*$ and $s^{\gamma\prime} y^\gamma \geq 0$ for all $\gamma > 0$, we obtain $s^{\gamma\prime} y^\gamma / \gamma \to 0$ and

$$\lim_{\gamma \to 0} L(d^\gamma, \mu^\gamma) = -\frac{1}{2} \|\mu^*\|^2. \tag{3.20}$$

Equations (3.17), (3.19), and (3.20), together with the fact $\mu^\gamma \to \mu^*$, yield

$$a_0' d^\gamma \to -\|\mu^*\|^2, \qquad (a_j' d^\gamma)^+ \to \mu_j^*, \quad j = 1, \ldots, r.$$

To show Eq. (3.10), we note that we have

$$\inf_{d \in N^*} \sup_{\mu \geq 0} L(d, \mu) = \inf_{d \in N^*} \left\{ a_0' d + \frac{1}{2} \sum_{j=1}^{r} \big((a_j' d)^+\big)^2 \right\}. \tag{3.21}$$

We also have

$$\begin{aligned} \inf_{d \in N^*} \sup_{\mu \geq 0} L(d, \mu) &= \lim_{\gamma \to 0} \inf_{d \in N^* \cap B(0, 1/\gamma)} \sup_{\mu \geq 0} L(d, \mu) \\ &= \lim_{\gamma \to 0} \inf_{d \in N^* \cap B(0, 1/\gamma)} L(d, \mu^\gamma) \\ &= \lim_{\gamma \to 0} -\frac{1}{2} \|\mu^\gamma\|^2 \\ &= -\frac{1}{2} \|\mu^*\|^2. \end{aligned}$$

Combining the last two equations, we obtain the desired relation

$$\begin{aligned} -\frac{1}{2} \|\mu^*\|^2 &= \inf_{d \in N^*} \left\{ a_0' d + \frac{1}{2} \sum_{j=1}^{r} \big((a_j' d)^+\big)^2 \right\} \\ &= \lim_{\gamma \to 0} \left\{ a_0' d^\gamma + \frac{1}{2} \sum_{j=1}^{r} \big((a_j' d^\gamma)^+\big)^2 \right\} \end{aligned}$$

[cf. Eq. (3.10)].

Finally, if d^* attains the infimum in the right-hand side above, it is seen that (d^*, μ^*) is a saddle point of L over $d \in N^*$ and $\mu \geq 0$, and that

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

Q.E.D.

We now return to the proof of Prop. 3.3.6(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 3.3.2 with the following identifications:

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

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

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

If $\mu^* = 0$, then μ^* is an informative Lagrange multiplier and we are done. If $\mu^* \neq 0$, by Lemma 3.3.2 [cf. (3.8) and (3.9)], for any $\epsilon > 0$, there exists a $\overline{d} \in N^* = T_X(x^*)$ such that

$$a_0' \overline{d} < 0, \tag{3.22}$$

$$a_j' \overline{d} > 0, \quad \forall\, j \in J^*, \qquad a_j' \overline{d} \leq \epsilon \min_{l \in J^*} a_l' \overline{d}, \quad \forall\, j \notin J^*, \tag{3.23}$$

where

$$J^* = \{j \mid \mu_j^* > 0\}.$$

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

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

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

$$\begin{aligned} f(x^k) - f(x^*) &= \nabla f(x^*)'(x^k - x^*) + o(\|x^k - x^*\|) \\ &= \nabla f(x^*)' \left(\overline{d} + \xi^k\right) \|x^k - x^*\| + o(\|x^k - x^*\|) \\ &= \|x^k - x^*\| \left(\nabla f(x^*)'\overline{d} + \nabla f(x^*)'\xi^k + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|}\right). \end{aligned} \tag{3.24}$$

From Eq. (3.22), we have $\nabla f(x^*)'\overline{d} < 0$, so we obtain $f(x^k) < f(x^*)$ for k sufficiently large. Using also Taylor's theorem for the constraint functions g_j, we have for some vector sequence ξ^k converging to 0,

$$\begin{aligned} g_j(x^k) - g_j(x^*) &= \nabla g_j(x^*)'(x^k - x^*) + o(\|x^k - x^*\|) \\ &= \nabla g_j(x^*)' \left(\overline{d} + \xi^k\right) \|x^k - x^*\| + o(\|x^k - x^*\|) \\ &= \|x^k - x^*\| \left(\nabla g_j(x^*)'\overline{d} + \nabla g_j(x^*)'\xi^k + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} \right). \end{aligned} \tag{3.25}$$

This, combined with Eq. (3.23), shows that for k sufficiently large, $g_j(x^k)$ is bounded from below by a positive constant times $\|x^k - x^*\|$ for all $j \in J^*$, and satisfies $g_j(x^k) \leq o(\|x^k - x^*\|)$ for all $j \notin J^*$. Thus, the sequence $\{x^k\}$ can be used to establish the CV condition for μ^*, and it follows that μ^* is an informative Lagrange multiplier. **Q.E.D.**

Lemma 3.3.2 also provides an alternative proof for Proposition 3.3.5, as shown in the following.

Alternative Proof for Proposition 3.3.5:

The essence of the proof argument can be summarized in the following lemma.

Lemma 3.3.3: Let N be a closed convex cone in $\Re^n$, let $a_0, a_1, \ldots, a_r$ be given vectors in $\Re^n$. Suppose that the closed and convex set $M \subset \Re^r$ given by

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

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

$$a_0'\overline{d} < 0, \qquad a_j'\overline{d} > 0, \qquad \text{for all } j \in \overline{J}. \tag{3.26}$$

Proof: We apply Lemma 3.3.2 with the vectors $a_1, \ldots, a_r$ replaced by the vectors a_j, $j \in \overline{J}$. The subset of M given by

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

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 3.3.2 implies that there exists a $\overline{d} \in N^*$ such that Eq. (3.26) holds. **Q.E.D.**

Given Lemma 3.3.3, the proof is very similar to the corresponding part of the proof of Proposition 3.3.5. **Q.E.D.**

3.3.3. Sensitivity

Let us consider now the special direction d^* that appears in Lemma 3.3.2, and is a solution of problem (3.11) (assuming this problem has an optimal solution). Let us note that *this problem is guaranteed to have at least one solution when N^* is a polyhedral cone*. This is because problem (3.11) can be written as

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

where the z_j are auxiliary variables. Thus, if N^* is polyhedral, then problem (3.11) is a convex quadratic program with a cost function that is bounded below by Eq. (3.10), and it has an optimal solution. An important context where this is relevant is when $X = \Re^n$ in which case $N_X(x^*)^* = T_X(x^*) = \Re^n$, or more generally when X is polyhedral, in which case $T_X(x^*)$ is polyhedral.

Assuming now that problem (3.11) has an optimal solution, the line of proof of Prop. 3.3.6(a) [combine Eqs. (3.24) and (3.25)] can be used to show that if the Lagrange multiplier that has minimum norm, denoted by (λ^*, μ^*), is nonzero, there exists a sequence $\{x^k\} \subset X$, corresponding to the vector $d^* \in T_X(x^*)$ of Eq. (3.12), such that

$$f(x^k) = f(x^*) - \sum_{i=1}^{m} \lambda_i^* h_i(x^k) - \sum_{j=1}^{r} \mu_j^* g_j(x^k) + o(\|x^k - x^*\|). \tag{3.27}$$

Furthermore, the vector d^* solves problem (3.11), from which it can be seen that d^* solves the problem

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

where $\overline{\beta}$ is given by

$$\overline{\beta} = \sum_{i=1}^{m} \big(\nabla h_i(x^*)'d^*\big)^2 + \sum_{j \in A(x^*)} \big(\big(\nabla g_j(x^*)'d^*\big)^+\big)^2.$$

More generally, it can be seen that for any given positive scalar β, a positive multiple of d^* solves the problem

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

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. (3.27), 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 minimum norm multipliers λ_i^* and μ_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.

3.4. AN ALTERNATIVE DEFINITION OF LAGRANGE MULTIPLIERS

In this section, we 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

$$-\left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*), \tag{4.1}$$

$$\mu_j^* \geq 0, \quad \forall\ j = 1, \ldots, r, \qquad \mu_j^* = 0, \quad \forall\ j \notin A(x^*). \tag{4.2}$$

Such vectors are called "Lagrange multipliers" by Rockafellar, but here we will refer to them as *R-multipliers*, to distinguish them from Lagrange multipliers as we have defined them [cf. Eqs. (1.1)-(1.3)]. It can be seen that the set of R-multipliers is a closed set [since $N_X(x^*)$ is closed], and is convex when $N_X(x^*)$ is convex [if $N_X(x^*)$ is not convex, the set of R-multipliers need not be convex].

When X is regular at x^*, the sets of Lagrange multipliers and R-multipliers coincide. In general, however, the set of Lagrange multipliers is a (possibly strict) subset of the set of R-multipliers, since $T_X(x^*)^* \subset N_X(x^*)$ with inequality holding when X is not regular at x^*. Note that multipliers satisfying the enhanced Fritz John conditions of Prop. 3.2.3 with $\mu_0^* = 1$ are R-multipliers, and they still have the extra sensitivity-like property embodied in the CV condition. Furthermore, Lemma 3.3.2 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. This is in sharp contrast with Lagrange multipliers, whose salient defining property is that they render the Lagrangian function stationary at x^*. The following example illustrates this.

Example 3.4.4:

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

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

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

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

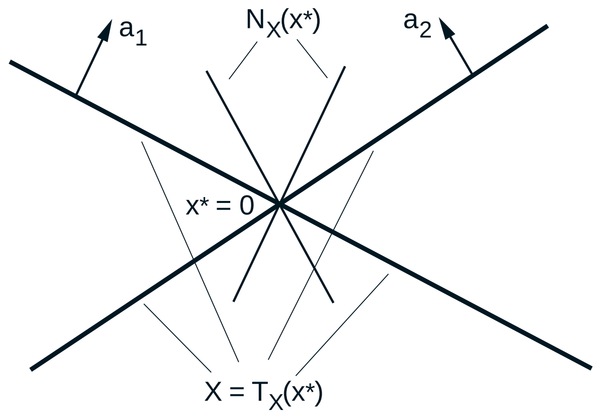


Figure 3.4.5. Constraints of Example 3.4.4. We have

$$T_X(x^*) = X = \{x \mid (a_1'x)(a_2'x) = 0\}$$

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

(see Fig. 3.4.5).

Because $N_X(x^*) \neq T_X(x^*)^*$, X is not regular at x^*. Furthermore, both $T_X(x^*)$ and $N_X(x^*)$ are not convex. For any f for which x^* is a local minimum, there exists a unique Lagrange multiplier (μ_1^*, μ_2^*) satisfying Eqs. (1.1)-(1.3). The scalars μ_1^*, μ_2^* are determined from the requirement

$$\nabla f(x^*) + \mu_1^* a_1 + \mu_2^* a_2 = 0. \tag{4.3}$$

Except in the cases where $\nabla f(x^*)$ is equal to 0 or to $-a_1$ or to $-a_2$, we have $\mu_1^* > 0$ and $\mu_2^* > 0$, but the Lagrange multiplier (μ_1^*, μ_2^*) is neither informative nor strong, because there is no $x \in X$ that simultaneously violates both inequality constraints. The R-multipliers here are the vectors (μ_1^*, μ_2^*) such that $\nabla f(x^*) + \mu_1^* a_1 + \mu_2^* a_2$ is either equal to a multiple of a_1 or to a multiple of a_2. Except for the Lagrange multipliers, which satisfy Eq. (4.3), all other R-multipliers are such that the Lagrangian function has negative slope along some of the feasible directions of X.

The existence of R-multipliers does not guarantee the existence of Lagrange multipliers. Furthermore, as shown in the previous example, even if Lagrange multipliers exist, none of them may be informative or strong, unless the tangent cone is convex (which is guaranteed if the set X is regular, cf. Proposition 2.2.19 of Chapter 2). Thus regularity of X at the given local minimum is the property that separates problems that possess a satisfactory Lagrange multiplier theory and problems that do not.

Example 3.4.5:

In this 2-dimensional example, there exists an R-multiplier for every smooth cost function f, but the constraint set does not admit Lagrange multipliers. Let X be the subset of $\Re^2$ given by

$$X = \{(x_1, x_2) \mid (x_2 - x_1)(x_2 + x_1) = 0\},$$

and let there be a single equality constraint

$$h(x) = x_2 = 0$$

(see Fig. 3.4.6). There is only one feasible point $x^* = (0, 0)$, which is optimal for any cost function f. Here we have $T_X(x^*) = X$ and $T_X(x^*)^* = \{0\}$, so for λ^* to be a Lagrange multiplier, we must have

$$\nabla f(x^*) + \lambda^*(0, 1) = 0.$$

Thus, there exists a Lagrange multiplier if and only if $\partial f(x^*)/\partial x_1 = 0$. On the other hand, it can be seen that we have

$$N_X(x^*) = X,$$

and that there exists an R-multiplier for every smooth cost function f.

3.5. NECESSARY AND SUFFICIENT CONDITIONS FOR ADMITTANCE OF LAGRANGE AND R-MULTIPLIERS

In this section, we provide necessary and sufficient conditions for the constraint set C of Eq. (0.2) to admit Lagrange and R-multipliers. Conditions of this kind related to Lagrange multipliers were dealt with in various forms by Gould and Tolle [GoT72], Guignard [Gui69], and Rockafellar [Roc93]. To show this result, we make use of the extended result on the gradient characterization of vectors in $T_C(x^*)^*$, given in Proposition 2.3.21 in Chapter 2.

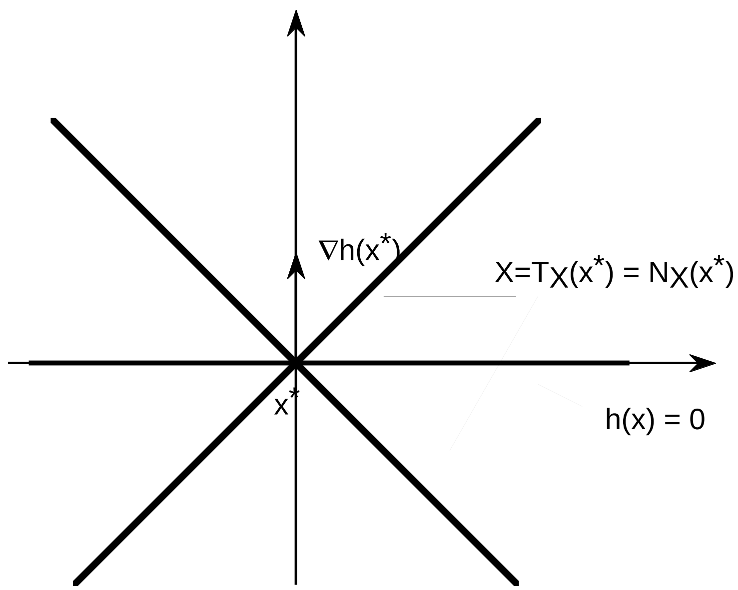


Figure 3.4.6. Constraints of Example 3.4.5. Here,

$$X = \left\{ (x_1, x_2) \mid (x_2 - x_1)(x_2 + x_1) = 0 \right\},$$

and X is not regular at $x^* = (0,0)$, since we have $T_X(x^*) = X$, $T_X(x^*)^* = \{0\}$, but $N_X(x^*) = X$. For

$$h(x) = x_2 = 0$$

the constraint set admits no Lagrange multipliers at x^*, yet there exist R-multipliers for every smooth cost function f, since for any f, there exists a λ^* such that $-\big(\nabla f(x^*) + \lambda^* \nabla h(x^*)\big)$ belongs to $N_X(x^*)$.

Proposition 3.5.7: Let x^* be a feasible vector of problem (0.1)-(0.2). Then:

(a) The constraint set admits Lagrange multipliers at x^* if and only if

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

(b) The constraint set admits R-multipliers at x^* if and only if

$$T_C(x^*)^* \subset N_X(x^*) + V(x^*)^*.$$

Proof:

(a) Denote by $D(x^*)$ the set of gradients of all smooth cost functions for which x^* is a local minimum of problem . We claim that $-D(x^*) = T_C(x^*)^*$. Indeed by the necessary condition for optimality, we have

$$-D(x^*) \subset T_C(x^*)^*.$$

To show the reverse inclusion, let $y \in T_C(x^*)^*$. By Proposition 2.3.21, there exists a smooth function F with $-\nabla F(x^*) = y$, which achieves a strict global minimum over C at $\overline{x}$. Thus, $y \in -D(x^*)$, showing that

$$-D(x^*) = T_C(x^*)^*. \tag{5.1}$$

We now note that by definition, the constraint set C admits Lagrange multipliers at x^* if and only if

$$-D(x^*) \subset T_X(x^*)^* + V(x^*)^*.$$

In view of Eq. (5.1), this implies that the constraint set C admits Lagrange multipliers at x^* if and only if

$$T_C(x^*)^* \subset T_X(x^*)^* + V(x^*)^*. \tag{5.2}$$

We next show that

$$T_C(x^*) \subset T_X(x^*) \cap V(x^*).$$

Since $C \subset X$, it can be seen by the definition of the tangent cone that

$$T_C(x^*) \subset T_X(x^*). \tag{5.3}$$

Next, we show that $T_C(x^*) \subset V(x^*)$. Let y be a nonzero tangent of C at x^*. Then there exist sequences $\{\xi^k\}$ and $\{x^k\} \subset C$ such that $x^k \neq x^*$ for all k,

$$\xi^k \to 0, \qquad x^k \to x^*,$$

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

$$\nabla g_j(\tilde{x}^k)' y^k \leq 0, \qquad y^k = y + \xi^k \|y\|.$$

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, $T_C(x^*) \subset V(x^*)$. Together with Eq. (5.3), this shows that

$$T_C(x^*) \subset T_X(x^*) \cap V(x^*). \tag{5.4}$$

Using the properties of polar cones given in Proposition 2.1.2 of Chapter 2, this implies

$$T_X(x^*)^* + V(x^*)^* \subset \big(T_X(x^*) \cap V(x^*)\big)^* \subset T_C(x^*)^*,$$

which combined with Eq. (5.2), yields the desired relation, and concludes the proof.

(b) By definition, the constraint set C admits R-multipliers at x^* if and only if

$$-D(x^*) \subset N_X(x^*) + V(x^*)^*.$$

In view of Eq. (5.1), this implies that the constraint set C admits R-multipliers at x^* if and only if

$$T_C(x^*)^* \subset N_X(x^*) + V(x^*)^*.$$

CHAPTER 4

PSEUDONORMALITY AND CONSTRAINT QUALIFICATIONS

In this chapter, our objective is to identify the structure of the constraint set that guarantees the existence of Lagrange multipliers. We consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C, \end{aligned} \tag{0.1}$$

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

$$C = X \cap \big\{x \mid h_1(x) = 0, \ldots, h_m(x) = 0\big\} \cap \big\{x \mid g_1(x) \leq 0, \ldots, g_r(x) \leq 0\big\}. \tag{0.2}$$

Our purpose is to find and analyze in depth conditions on the above problem that guarantee the existence of Lagrange multipliers. Note that we are interested in conditions that are independent of the cost function f, and are only properties of the constraint set; hence the name *constraint qualifications.* Therefore, if the constraint qualification holds, the Lagrange multiplier rules are valid for the same constraints and any other smooth cost function.

In the case where $X = \Re^n$, a typical approach to asserting the admittance of Lagrange multipliers is to assume structure in the constraint set, which guarantees that the tangent cone $T_C(x^*)$ has the form

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

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

$$V(x^*) = \big\{y \mid \nabla h_i(x^*)'y = 0,\ i = 1, \ldots, m,\ \nabla g_j(x^*)'y \leq 0,\ j \in A(x^*)\big\}. \tag{0.3}$$

In this case we say that x^* is a *quasiregular point* or that *quasiregularity holds at* x^* [other terms used are x^* "satisfies Abadie's constraint qualification" (Abadie [Aba67], Bazaraa, Sherali, and Shetty [BSS93]), or "is a regular point" (Hestenes [Hes75])]. When there is no abstract set constraint, it is well-known (see e.g., Bertsekas [Ber99], p. 332) that for a given smooth f for which x^* is a local minimum, there exist Lagrange multipliers if and only if

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

This result, a direct consequence of Farkas' Lemma, leads to the classical theorem that the constraint set admits Lagrange multipliers at x^* if x^* is a quasiregular point. Therefore, quasiregularity plays a central role in the classical line of development of Lagrange multiplier theory for the case $X = \Re^n$. A common line of analysis is based on establishing various conditions, also known as *constraint qualifications*, which imply quasiregularity, and therefore imply that the constraint set admits Lagrange multipliers. This line of analysis requires fairly complicated proofs to show the relations of constraint qualifications to quasiregularity. Some of the most useful constraint qualifications, for the case when $X = \Re^n$ are the following:

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

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

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

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

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

It is well-known that all of the above constraint qualifications imply the quasiregularity condition $T_C(x^*) = V(x^*)$, and therefore imply that the constraint set admits Lagrange multipliers (see e.g., [BNO02], or [BSS93]; a survey of constraint qualifications is given by

Peterson [Pet73]). These results constitute the classical pathway to Lagrange multipliers for the case where $X = \Re^n$. Figure 4.0.1 summarizes the relationships discussed above for the case $X = \Re^n$.

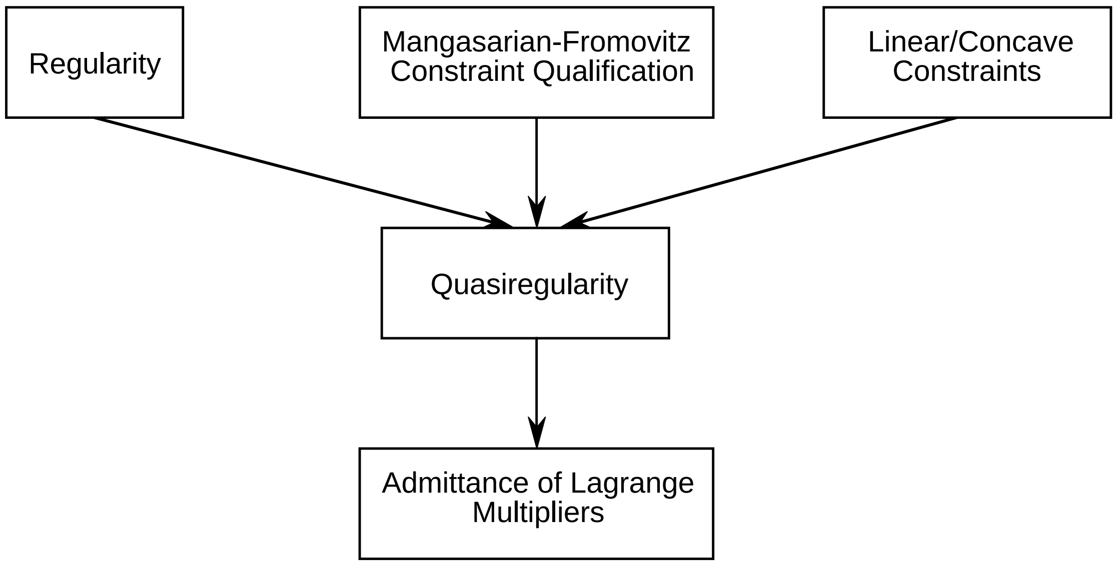


Figure 4.0.1. Characterizations of the constraint set C that imply admittance of Lagrange multipliers in the case where $X = \Re^n$.

Unfortunately, when X is a strict subset of $\Re^n$, the situation changes significantly because there does not appear to be a satisfactory extension of the notion of quasiregularity, which implies admittance of Lagrange multipliers. We will focus on the relation of quasiregularity and the existence of Lagrange multipliers later in this chapter. In the next section, we introduce an alternative notion, and show that it forms the connecting link between major constraint qualifications and the existence of Lagrange multipliers, even for the case when X is a strict subset of $\Re^n$.

4.1. PSEUDONORMALITY

The enhanced Fritz John conditions of Chapter 3 provides Lagrange multiplier-like conditions that hold for the general optimization problem , which includes an abstract set constraint, under no assumption on the constraint set structure. If X is regular, and if we can guarantee that the cost multiplier μ_0^* is positive for some constraint set, then it automatically follows that the constraint set admits Lagrange multipliers. This motivates us to introduce the following general constraint qualification under which the cost multiplier μ_0^* in Prop. 3.2.3 cannot be zero.

Definition 4.1.1: We say that a feasible vector x^* of problem (0.1)-(0.2) is *pseudonormal* if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$, and a sequence $\{x^k\} \subset X$ such that:

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

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

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

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

$$\sum_{i=1}^m \lambda_i h_i(x^k) + \sum_{j=1}^r \mu_j g_j(x^k) > 0, \qquad \forall\ k. \tag{1.1}$$

If x^* is a pseudonormal local minimum, the enhanced Fritz John conditions of Prop. 3.2.3 cannot be satisfied with $\mu_0^* = 0$, so that μ_0^* can be taken equal to 1. Then, if X is regular at x^*, the vector $(\lambda^*, \mu^*) = (\lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*)$ obtained from the enhanced Fritz John conditions is an informative Lagrange multiplier.

4.1.1. Relation to Major Constraint Qualifications

We now focus on various constraint qualifications, which will be shown in this section to imply pseudonormality of a feasible vector x^* and hence also existence of informative Lagrange multipliers (assuming also regularity of X at x^*).

The next constraint qualification applies to the case where X is a strict subset of $\Re^n$. A weaker version of this constraint qualification, for the case where X is a closed convex set and none of the equality constraints is linear, was shown in [Ber99]. We refer to it as the *generalized Mangasarian-Fromovitz constraint qualification* (MFCQ for short), since it reduces to CQ2 when $X = \Re^n$ and none of the equality constraints is linear.

MFCQ:

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

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

are linear.

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

$$-\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*) \tag{1.2}$$

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

(c) The subspace

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

has a nonempty intersection with either the interior of $N_X(x^*)^*$, or, in the case where X is convex, with the relative interior of $N_X(x^*)^*$.

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

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

The following is an example where the above constraint qualification holds. Later in this section, we will show that this constraint qualification implies pseudonormality, and therefore guarantees existence of Lagrange multipliers.

Example 4.1.1:

Let the constraint set be specified by

$$C = \left\{ x \in \Re^2 \mid h_1(x) = 0,\ h_2(x) = 0,\ g(x) \le 0 \right\},$$

where

$$h_1(x) = x_1, \quad h_2(x) = -x_1, \quad g(x) = x_1^3 - x_2,$$

(see Fig. 4.1.2). Consider the feasible vector $x^* = (0,0)$. The vector $[0,1]'$ satisfies condition (d) of MFCQ. Hence, although none of CQ1-CQ3 holds at x^*, MFCQ holds at x^*.

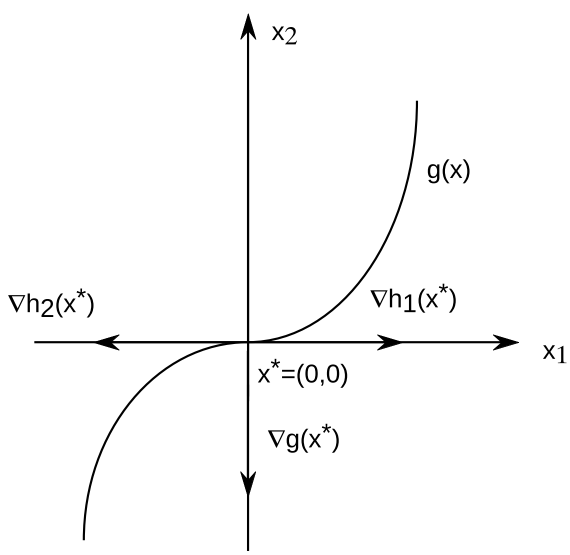


Figure 4.1.2. Constraints of Example 4.1.1.

MFCQ has several special cases for constraint sets that have different representations in terms of equalities and inequalities. For instance, if we assume that all the equality constraints are nonlinear, we get the following special case of MFCQ.

MFCQa:

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

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

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

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

When there are no inequality constraints and no linear equality constraints, the following constraint qualification guarantees that assumption (b) of MFCQ holds.

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

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

contains a point in the interior of $N_X(x^*)^*$.

To see why this condition implies assumption (b) of MFCQ, assume the contrary, i.e., there exists a nonzero vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

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

The vector z cannot be equal to 0, since this would contradict the linear independence assumption of the $\nabla h_i(x^*)$. By the definition of the polar cone, the preceding relation implies that

$$z'y \geq 0, \qquad \forall\, y \in N_X(x^*)^*.$$

Let $\overline{y}$ be the vector of hypothesis in MFCQb, i.e., $\overline{y} \in V(x^*) \cap \mathrm{int}\big(N_X(x^*)^*\big)$. Since $\overline{y} \in \mathrm{int}\big(N_X(x^*)^*\big)$, it follows that, for some sufficiently small positive α, the vector $\overline{y} - \alpha z \in$

$N_X(x^*)^*$. Substituting this vector in the preceding relation, we obtain

$$z'\overline{y} \geq \alpha\|z\|^2 > 0,$$

where the strict inequality follows since $z \neq 0$. But this contradicts the fact that $\overline{y} \in V(x^*)$, which implies that $z'\overline{y} = 0$, thus proving that MFCQb implies assumption (b) of MFCQ. Note that the interior point of $N_X(x^*)^*$ assumption of condition MFCQb cannot be replaced by a relative interior point assumption.

Finally, we state another useful special case of MFCQ, which holds under convexity assumptions.

MFCQc: X is convex, the functions h_i, $i = 1, \ldots, m$ are linear, and the linear manifold

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

contains a point in the relative interior of X. Furthermore, the functions g_j are convex and there exists a feasible vector $\overline{x}$ satisfying

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

The convexity assumptions in MFCQc can be used to establish the corresponding assumptions (c) and (d) of MFCQ. In particular, if X is convex, we have from Proposition 2.2.9 that

$$\mathrm{cl}\big(F_X(x^*)\big) = T_X(x^*) = N_X(x^*)^*,$$

which, using properties of relative interior, implies that

$$\mathrm{ri}\big(F_X(x^*)\big) = \mathrm{ri}\big(N_X(x^*)^*\big). \tag{1.3}$$

Let $\tilde{x}$ be the vector of hypothesis in condition MFCQc, i.e., $\tilde{x} \in L \cap \mathrm{ri}(X)$. Using the Taylor's theorem for affine constraint functions h_i, we see that

$$0 = h_i(\tilde{x}) = h_i(x^*) + \nabla h_i(x^*)'(\tilde{x} - x^*) = \nabla h_i(x^*)'(\tilde{x} - x^*), \quad \forall\, i = 1, \ldots, m,$$

which implies that the vector $\tilde{x} - x^* \in V_L(x^*)$. Since $\tilde{x} \in \mathrm{ri}(X)$ and X is convex, the vector $\tilde{x} - x^*$ belongs to the set $\mathrm{ri}\big(F_X(x^*)\big)$, which in view of relation (1.3) implies that

$\tilde{x} - x^* \in \mathrm{ri}\big(N_X(x^*)^*\big)$, hence showing that condition (c) of MFCQ holds. Similarly, the feasible vector $\overline{x}$ given in MFCQc could be used in conjunction with the linearity of equality constraints and the convexity of inequality constraints to construct the vector y that satisfies the properties of condition (d) of MFCQ.

In the case where there are no equality constraints, MFCQc is a classical constraint qualification, introduced by Slater [Sla50] and known as *Slater's condition*.

The following constraint qualification is simpler to state than the preceding ones, and it is immediately seen to imply pseudonormality. It is the constraint qualification introduced by Rockafellar [Roc93], [RoW98], who used McShanes's line of proof to derive Fritz John conditions in their classical form where the CS condition replaces the CV condition, for problems that involve an abstract set constraint.

RCQ: The set

$$W = \big\{(\lambda, \mu) \mid \lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r \text{ satisfy conditions (i) and (ii) of the definition of pseudonormality}\big\} \tag{1.4}$$

consists of just the vector 0.

It can be shown that the set W of Eq. (1.4) is the recession cone of the set of R-multipliers, provided that the set of R-multipliers is a nonempty convex set (so that we can talk about its recession cone; note that the set of R-multipliers is closed, cf. Chapter 3). To see this, let (λ^*, μ^*) be any R-multiplier. For 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 (λ, μ) is a direction of recession. Conversely, if (λ, μ) is a direction of recession, then for all R-multipliers (λ^*, μ^*), we have for all $\alpha > 0$,

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

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 RCQ holds. In view of Prop. 3.2.3, 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, RCQ is equivalent to the set of R-multipliers being nonempty and compact.*

We will next show that if X is regular at x^*, then RCQ is equivalent to MFCQa. This was shown by Rockafellar and Wets in the case where $X = \Re^n$ (see page 226 of [RoW98]). [1] We generalize it to the case where X is regular in the following proposition.

Proposition 4.1.1: If X is regular at x^*, the constraint qualifications MFCQa and RCQ are equivalent.

Proof: We first show that MFCQa implies RCQ. Assume MFCQa holds:

(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 $d \in N_X(x^*)^* = T_X(x^*)$ (since X is regular at x^*) such that

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

To arrive at a contradiction, assume that RCQ does not hold, i.e., there are scalars $\lambda_1, \ldots, \lambda_m$, $\mu_1, \ldots, \mu_r$, not all of them equal to zero, such that

(i)

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

[1] In fact, it is well known that, for $X = \Re^n$, MFCQa is equivalent to nonemptiness and compactness of the set of Lagrange multipliers, this is a result of Gauvin [Gau77].

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

In view of our assumption that X is regular at x^*, condition (i) can be written as

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

or equivalently,

$$\left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\, y \in T_X(x^*). \tag{1.5}$$

Since not all the λ_i and μ_j are equal to 0, we conclude that $\mu_j > 0$ for at least one $j \in A(x^*)$; otherwise condition (a) of MFCQa would be violated. Since $\mu_j^* \geq 0$ for all j, with $\mu_j^* = 0$ for $j \notin A(x^*)$ and $\mu_j^* > 0$ for at least one j, we obtain

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

where $d \in T_X(x^*)$ is the vector in condition (b) of MFCQa. But this contradicts Eq. (1.5), showing that RCQ holds.

Conversely, assume that RCQ holds. It can be seen that this implies condition (a) of MFCQa. We next show that condition (b) of MFCQa holds. Let H denote the subspace spanned by the vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$, and let G denote the cone generated by the vectors $\nabla g_j(x^*)$, $j \in A(x^*)$. Then, the orthogonal complement of H is given by

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

whereas the polar of G is given by

$$G^* = \big\{y \mid \nabla g_j(x^*)' y \leq 0,\ \forall\, j \in A(x^*)\big\}.$$

The interior of G^* is the set

$$\text{int}(G^*) = \big\{y \mid \nabla g_j(x^*)' y < 0,\ \forall\, j \in A(x^*)\big\}.$$

Assume, to arrive at a contradiction, that condition (b) of MFCQa does not hold. This implies that

$$N_X(x^*)^* \cap \Big(H^\perp \cap \text{int}\big(G^*\big)\Big) = \emptyset.$$

Since X is regular at x^*, the preceding is equivalent to

$$T_X(x^*) \cap \Big(H^\perp \cap \text{int}\big(G^*\big)\Big) = \emptyset.$$

The regularity of X at x^* implies that $T_X(x^*)$ is convex. Similarly, since the interior of a convex set is convex and the intersection of two convex sets is convex, it follows that the set $H^\perp \cap \text{int}\big(G^*\big)$ is convex. By the Separating Hyperplane Theorem, there exists some vector $a \neq 0$ such that

$$a'x \leq a'y, \qquad \forall\, x \in T_X(x^*),\ \forall\, y \in \Big(H^\perp \cap \text{int}\big(G^*\big)\Big),$$

or equivalently,

$$a'(x - y) \leq 0, \qquad \forall\, x \in T_X(x^*),\ \forall\, y \in \big(H^\perp \cap G^*\big),$$

which implies that

$$a \in \big(T_X(x^*) - (H^\perp \cap G^*)\big)^*.$$

Using the properties of cones given in Proposition 2.1.2 of Chapter 2, we have

$$\begin{aligned}
\big(T_X(x^*) - (H^\perp \cap G^*)\big)^* &= T_X(x^*)^* \cap -(H^\perp \cap G^*)^* \\
&= T_X(x^*)^* \cap -\big(\text{cl}(H + G)\big) \\
&= T_X(x^*)^* \cap -(H + G) \\
&= N_X(x^*) \cap -(H + G),
\end{aligned}$$

where the second equality follows since $H^\perp$ and G^* are closed and convex, and the third equality follows since H and G are both polyhedral cones. Combining the preceding relations, it follows that there exists a nonzero vector a that belongs to the set

$$N_X(x^*) \cap -(H + G).$$

But this contradicts RCQ, thus completing our proof. **Q.E.D.**

Clearly RCQ implies pseudonormality, since the vectors in W are not required to satisfy condition (iii) of the definition of pseudonormality. However, CQ3 and MFCQ do not preclude unboundedness of the set of Lagrange multipliers and hence do not imply RCQ. Thus RCQ is not as effective in unifying various constraint qualifications as pseudonormality, which is implied by all these constraint qualifications, as shown in the following proposition.

Proposition 4.1.2: A feasible point x^* of problem (0.1)-(0.2) is pseudonormal if any one of the constraint qualifications CQ1-CQ3, MFCQ, and RCQ is satisfied.

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

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

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

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

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

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

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

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

$$\begin{aligned}\sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x) &\le \sum_{i=1}^{m} \lambda_i h_i(x^*) + \sum_{j=1}^{r} \mu_j g_j(x^*) \\ &\quad + \left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \right)' (x - x^*) \\ &= 0,\end{aligned} \tag{1.6}$$

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. (1.6).

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

$$- \sum_{i=\overline{m}+1}^{m} \lambda_i a_i \in N_X(x^*). \tag{1.7}$$

Consider first the case where X is necessarily convex and there is an interior point $\overline{y}$ of $N_X(x^*)^*$ that satisfies $a_i' \overline{y} = 0$ for all $i = \overline{m}+1, \ldots, m$. Let S be an open sphere centered at the origin such that $\overline{y} + d \in N_X(x^*)^*$ for all $d \in S$. We have from Eq. (1.7),

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

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

Consider now the alternative case where X is convex and there is a relative interior point $\overline{y}$ of $N_X(x^*)^*$ that satisfies $a_i' \overline{y} = 0$ for all $i = \overline{m} + 1, \ldots, m$. Then, we have

$$\sum_{i=\overline{m}+1}^{m} \lambda_i a_i' \overline{y} = 0,$$

while from Eq. (1.7), we have

$$\sum_{i=\overline{m}+1}^{m} \lambda_i a_i' y \geq 0, \qquad \forall\, y \in N_X(x^*)^*.$$

The convexity of X implies that $X - \{x^*\} \subset T(x^*) = N_X(x^*)^*$ and that $N_X(x^*)^*$ is convex (cf. Proposition 2.2.19 of Chapter 2). Since the linear function $\sum_{i=\overline{m}+1}^{m} \lambda_i a_i' y$ attains a minimum over $N_X(x^*)^*$ at the relative interior point $\overline{y}$, it follows that this linear function is constant over $N_X(x^*)^*$. Thus, we have $\sum_{i=\overline{m}+1}^{m} \lambda_i a_i' y = 0$ for all $y \in N_X(x^*)^*$, and hence [since $X - \{x^*\} \subset N_X(x^*)^*$ and $\lambda_i a_i' x^* = 0$ for all i]

$$\sum_{i=\overline{m}+1}^{m} \lambda_i a_i' x = 0, \qquad \forall\, x \in X.$$

This contradicts condition (iii), which requires that for some $x \in X$ we have $\sum_{i=\overline{m}+1}^{m} \lambda_i a_i' x > 0$. This completes the proof that at least one of the $\lambda_1, \ldots, \lambda_{\overline{m}}$ and μ_j, $j \in A(x^*)$ must be nonzero.

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

$$-\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*). \tag{1.8}$$

By what has been proved above, the multipliers $\lambda_1, \ldots, \lambda_{\overline{m}}$ of the nonlinear constraints cannot be all zero, so Eq. (1.8) contradicts assumption (b) of MFCQ.

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

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

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

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

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

Note that the constraint qualifications MFCQ and RCQ guarantee pseudonormality, as per the preceding proposition, but do not guarantee that the constraint set admits Lagrange multipliers at a point x^*, unless X is regular at x^*. As an illustration, in Example 3.4.5, x^* satisfies MFCQ and RCQ, and is therefore pseudonormal. However, as we have seen, in this example, the constraint set does not admit Lagrange multipliers, although there do exist R-multipliers for every smooth cost function f, consistently with the pseudonormality of x^*.

4.1.2. Quasinormality

A general constraint qualification, called *quasinormality*, was introduced for the special case where $X = \Re^n$ by Hestenes in [Hes75]. Hestenes also showed that quasinormality implies quasiregularity (see also Bertsekas [Ber99], Proposition 3.3.17). Since it is simple to show that the major classical constraint qualifications imply quasinormality (see e.g. Bertsekas [Ber99]), this provides an alternative line of proof that these constraint qualifications imply quasiregularity for the case $X = \Re^n$. In this section, we investigate the extension of quasinormality to the case where $X \neq \Re^n$. We subsequently compare this notion with pseudonormality in this section, and also with an extension of the notion of quasiregularity in the next section.

Definition 4.1.2: We say that a feasible vector x^* of problem (1.1)-(1.2) is *quasinormal* if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$, and a sequence $\{x^k\} \subset X$ such that:

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

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

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

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

If x^* is a quasinormal local minimum, the enhanced Fritz John conditions of Prop. 3.2.3 cannot be satisfied with $\mu_0^* = 0$, so that μ_0^* can be taken equal to 1. Then, if X is regular at x^*, the vector $(\lambda^*, \mu^*) = (\lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*)$ obtained from the enhanced Fritz John conditions is an informative Lagrange multiplier. It can be seen that *pseudonormality implies quasinormality.* The following example shows that the reverse is not true.

Example 4.1.2:

Let the constraint set be specified by

$$C = \{x \in X \mid g_1(x) \le 0,\ g_2(x) \le 0,\ g_3(x) \le 0\},$$

where $X = \Re^2$ and

$$\begin{aligned} g_1(x) &= x_1^2 + (x_2 - 1)^2 - 1, \\ g_2(x) &= \big(x_1 - \cos(\pi/6)\big)^2 + \big(x_2 + \sin(\pi/6)\big)^2 - 1, \\ g_3(x) &= \big(x_1 + \cos(\pi/6)\big)^2 + \big(x_2 + \sin(\pi/6)\big)^2 - 1. \end{aligned}$$

(see Fig. 4.1.3). Consider the feasible vector $x^* = (0,0)$. Because there is no x that simultaneously violates all three constraints, quasinormality is satisfied. However, a straightforward calculation shows that we have

$$\nabla g_1(x^*) + \nabla g_2(x^*) + \nabla g_3(x^*) = 0,$$

while

$$g_1(x) + g_2(x) + g_3(x) = 3(x_1^2 + x_2^2) > 0, \qquad \forall\, x \neq x^*,$$

so by using $\mu = (1,1,1)$, the conditions for pseudonormality of x^* are violated. Thus, even when $X = \Re^n$, quasinormality does not imply pseudonormality.

In the next proposition, we show that under the assumption that $N_X(x^*)$ is convex (which is true in particular if X is regular at x^*), quasinormality is in fact equivalent to a slightly weaker version of pseudonormality.

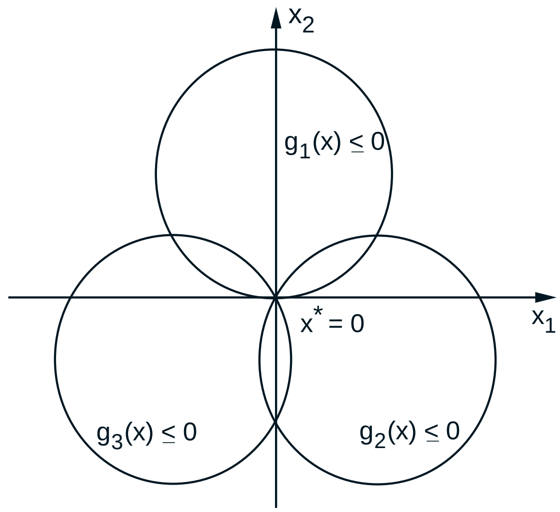


Figure 4.1.3. Constraints of Example 4.1.2.

Proposition 4.1.3 Let x^* be a feasible vector of problem (1.1)-(1.2), and assume that the normal cone $N_X(x^*)$ is convex. Then x^* is quasinormal if and only if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$ satisfying conditions (i)-(iii) of the definition of quasinormality together with the following condition:

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

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

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

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

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

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

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

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

$$-\left(a_1 + \sum_{j=2}^{r} \mu_j a_j\right) \in N_X(x^*). \tag{1.9}$$

If $-a_1 \in N_X(x^*)$, the choice of scalars $\overline{\mu}_1 = 1$ and $\overline{\mu}_j = 0$ for all $j = 2, \ldots, r$, satisfies conditions (i)-(iv) in the definition of quasinormality, hence x^* is not quasinormal and we are done. Assume that $-a_1 \notin N_X(x^*)$. The assumptions of Lemma 3.3.2 are satisfied, so it follows that there exist scalars $\overline{\mu}_2, \ldots, \overline{\mu}_r$, not all 0, such that

$$-\left(a_1 + \sum_{j=2}^{r} \overline{\mu}_j a_j\right) \in N_X(x^*), \tag{1.10}$$

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

$$\nabla g_j(x^*)'\overline{d} > 0, \qquad \forall\ j = 2, \ldots, r \text{ with } \overline{\mu}_j > 0, \tag{1.11}$$

while by Eq. (1.10), the $\overline{\mu}_j$ satisfy

$$-\left(\nabla g_1(x^*) + \sum_{j=2}^{r} \overline{\mu}_j \nabla g_j(x^*)\right) \in N_X(x^*). \tag{1.12}$$

Next, we show that the scalars $\overline{\mu}_1 = 1$ and $\overline{\mu}_2, \ldots, \overline{\mu}_r$ satisfy condition (iv) in the definition of quasinormality, completing the proof. We use Proposition 2.2.18 of Chapter 2 to argue that for the vector $\overline{d} \in N_X(x^*)^*$ and the sequence x^k constructed above, there is a sequence $d^k \in T_X(x^k)$ such that $d^k \to \overline{d}$. Since $x^k \to x^*$ and $d^k \to \overline{d}$, by Eq. (1.11), we obtain for all sufficiently large k,

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

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

$$x_\nu^k \to x^k, \qquad \frac{x_\nu^k - x^k}{\|x_\nu^k - x^k\|} \to \frac{d^k}{\|d^k\|}, \qquad \text{as } \nu \to \infty. \tag{1.13}$$

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

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

where the inequality above follows from Eq. (1.13) and the assumption that $g_j(x^k) \geq 0$, for all j and x^k. It follows that for ν and k sufficiently large, there exists $x_\nu^k \in X$ arbitrarily close to x^k such that $g_j(x_\nu^k) > 0$, for all $j = 2, \ldots, r$ with $\overline{\mu}_j > 0$. Since $g_1(x^k) > 0$ and g_1 is a continuous function, we have that $g_1(\tilde{x}) > 0$ for all $\tilde{x}$ in some neighborhood V_k of x^k. Since $x^k \to x^*$ and $x_\nu^k \to x^k$ for each k, by choosing ν and k sufficiently large, we get $g_j(x_\nu^k) > 0$ for $j = 1$ and each $j = 2, \ldots, r$ with $\overline{\mu}_j > 0$. This together with Eq. (1.12), violates the quasinormality assumption of x^*, which completes the proof. **Q.E.D.**

The following example shows that convexity of $N_X(x^*)$ is an essential assumption for the conclusion of Prop. 4.1.3.

Example 4.1.3:

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

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

(see Fig. 4.1.4). The normal cone $N_X(x^*)$ consists of the three rays shown in Fig. 4.1.4, and is not convex. Let there be two inequality constraints with

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

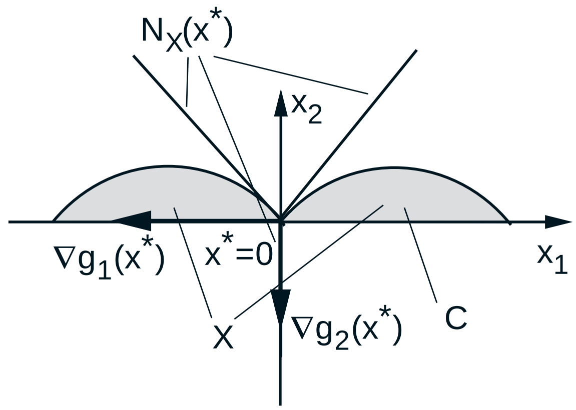


Figure 4.1.4. Constraints of Example 4.1.3.

In order to have $-\sum_j \mu_j \nabla g_j(x^*) \in N_X(x^*)$, we must have $\mu_1 > 0$ and $\mu_2 > 0$. There is no $x \in X$ such that $g_2(x) > 0$, so x^* is quasinormal. However, for $-2 \leq x_1 \leq 0$ and $x_2 = 0$, we have $x \in X$, $g_1(x) > 0$, and $g_2(x) = 0$. Hence x^* does not satisfy the weak form of pseudonormality given in Prop. 4.1.3.

4.1.3. Quasiregularity

We will now provide an extension of the notion of quasiregularity, which also applies to the case where X is a strict subset of $\Re^n$. We will then develop the connection of this notion with pseudonormality and quasinormality, and we explain the reasons why quasiregularity is not a satisfactory vehicle for unification of Lagrange multiplier theory when $X \neq \Re^n$.

We recall that for the case where $X = \Re^n$, a point x in the constraint set

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

is said to be a quasiregular point of C if

$$T_C(x) = V(x), \tag{1.14}$$

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

$$V(x) = \big\{y \mid \nabla h_i(x)'y = 0,\ i = 1, \ldots, m, \nabla g_j(x)'y \leq 0,\ j \in A(x)\big\}, \tag{1.15}$$

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

A classical approach to showing existence of Lagrange multipliers for the case where $X = \Re^n$ is to argue that at a local minimum x^* of f over C, we have $\nabla f(x^*)'y \geq 0$ for all $y \in T_C(x^*)$. Thus, if x^* is quasiregular, we have

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

By Farkas' Lemma, it follows that we either have that $\nabla f(x^*) = 0$ or else there exists a nonzero Lagrange multiplier vector (λ^*, μ^*).

For the case where $X \neq \Re^n$, we say that a feasible vector x^* of problem (0.1)-(0.2) is *quasiregular* if

$$T_C(x^*) = V(x^*) \cap T_X(x^*).$$

Our first aim is to show that under a regularity assumption on X, quasinormality implies quasiregularity. Moreover, since pseudonormality implies quasinormality, it follows that under the given assumption, pseudonormality also implies quasiregularity. This shows that any constraint qualification that implies pseudonormality imply quasiregularity.

We first prove the following result that relates to the properties of the quasinormality condition.

Proposition 4.1.4: If a vector $x^* \in C$ is quasinormal, then all feasible vectors in a neighborhood of x^* are quasinormal.

Proof: We assume for simplicity that all the constraints of problem (0.1)-(0.2) are inequalities; equality constraints can be handled by conversion to inequality constraints. Assume that the claim is not true. Then we can find a sequence $\{x^k\} \subset C$ such that $x^k \neq x^*$ for all k, $x^k \to x^*$ and x^k is not quasinormal for all k. This implies, for each k, the existence of scalars $\xi_1^k, \ldots, \xi_r^k$, and a sequence $\{x_l^k\} \subset X$ such that:

(a)

$$-\left(\sum_{j=1}^{r} \xi_j^k \nabla g_j(x^k)\right) \in N_X(x^k), \tag{1.16}$$

(b) $\xi_j^k \geq 0$, for all $j = 1, \ldots, r$, and $\xi_1^k, \ldots, \xi_r^k$ are not all equal to 0.

(c) $\lim_{l\to\infty} x_l^k = x_k$, and for all l, $\xi_j^k g_j(x_l^k) > 0$ for all j with $\xi_j^k > 0$.

For each k denote,

$$\delta^k = \sqrt{\sum_{j=1}^{r} \left(\xi_j^k\right)^2},$$

$$\mu_j^k = \frac{\xi_j^k}{\delta^k}, \qquad j = 1, \ldots, r, \ \forall\ k.$$

Since $\delta^k \neq 0$ and $N_X(x^k)$ is a cone, conditions (a)-(c) for the scalars $\xi_1^k, \ldots, \xi_r^k$ yield the following set of conditions that hold for each k for the scalars $\mu_1^k, \ldots, \mu_r^k$:

(i)

$$-\left(\sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k)\right) \in N_X(x^k), \tag{1.17}$$

(ii) $\mu_j^k \geq 0$, for all $j = 1, \ldots, r$, and $\mu_1^k, \ldots, \mu_r^k$ are not all equal to 0.

(iii) There exists a sequence $\{x_l^k\} \subset X$ such that $\lim_{l\to\infty} x_l^k = x^k$, and for all l, $\mu_j^k g_j(x_l^k) > 0$ for all j with $\mu_j^k > 0$.

Since by construction we have

$$\sum_{j=1}^{r} \left(\mu_j^k\right)^2 = 1, \tag{1.18}$$

the sequence $\{\mu_1^k, \ldots, \mu_r^k\}$ is bounded and must contain a subsequence that converges to some nonzero limit $\{\mu_1^*, \ldots, \mu_r^*\}$. Assume without loss of generality that $\{\mu_1^k, \ldots, \mu_r^k\}$ converges to $\{\mu_1^*, \ldots, \mu_r^*\}$. Taking the limit in Eq. (1.17), and using the closedness of the normal cone, we see that this limit must satisfy

$$-\left(\sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*). \tag{1.19}$$

Moreover, from condition (ii) and Eq. (1.18), it follows that $\mu_j^* \geq 0$, for all $j = 1, \ldots, r$, and $\mu_1^*, \ldots, \mu_r^*$ are not all equal to 0. Finally, let

$$J = \{j \mid \mu_j^* > 0\}.$$

Then, there exists some k_0 such that for all $k \geq k_0$, we must have $\mu_j^k > 0$ for all $j \in J$. From condition (iii), it follows that for each $k \geq k_0$, there exists a sequence $\{x_l^k\} \subset X$ with

$$\lim_{l \to \infty} x_l^k = x^k, \qquad g_j(x_l^k) > 0, \quad \forall\, l,\ \forall\, j \in J.$$

For each $k \geq k_0$, choose an index l_k such that $l_1 < \ldots < l_{k-1} < l_k$ and

$$\lim_{k \to \infty} x_{l_k}^k = x^*.$$

Consider the sequence $\{y^k\}$ defined by

$$y^k = x_{l_{k_0+k-1}}^{k_0+k-1}, \qquad k = 1, 2, \ldots.$$

It follows from the preceding relations that $\{y^k\} \subset X$ and

$$\lim_{k \to \infty} y^k = x^*, \qquad g_j(y^k) > 0, \quad \forall\, k,\ \forall\, j \in J.$$

The existence of scalars $\mu_1^*, \ldots, \mu_r^*$ that satisfy Eq. (1.19) and the sequence $\{y^k\}$ that satisfies the preceding relation violates the quasinormality of x^*, thus completing the proof. **Q.E.D.**

We next use Proposition 2.3.21 of Chapter 2, i.e., gradient characterization of vectors in the polar of the tangent cone, to obtain a specific representation of a vector that belongs to $T_C(\overline{x})^*$ for some $\overline{x} \in C$ under a quasinormality condition. This result will be central in showing the relation of quasinormality to quasiregularity.

Proposition 4.1.5: If $\overline{x}$ is a quasinormal vector of C, then any $y \in T_C(\overline{x})^*$ can be represented as

$$y = z + \sum_{j=1}^{r} \overline{\mu}_j \nabla g_j(\overline{x}),$$

where $z \in N_X(\overline{x})$, $\overline{\mu}_j \geq 0$, for all $j = 1, \ldots, r$. Furthermore, there exists a sequence $\{x^k\} \subset X$ that converges to $\overline{x}$ and is such that $\overline{\mu}_j g_j(x^k) > 0$ for all k and all j with $\overline{\mu}_j > 0$.

Proof: We assume for simplicity that all the constraints are inequalities. Let y be a vector that belongs to $T_C(\overline{x})^*$. By Prop. 2.3.21, there exists a smooth function F that achieves a strict global minimum over C at $\overline{x}$ with $-\nabla F(\overline{x}) = y$. We use a quadratic penalty function approach. For each $k = 1, 2, \ldots$, choose an $\epsilon > 0$ and consider the "penalized" problem

$$\begin{aligned} &\text{minimize} \;\; F^k(x) \\ &\text{subject to} \;\; x \in X \cap S, \end{aligned}$$

where

$$F^k(x) = F(x) + \frac{k}{2} \sum_{j=1}^{r} \big(g_j^+(x)\big)^2,$$

and $S = \{x \mid ||x - \overline{x}|| \leq \epsilon\}$. Since $X \cap S$ is compact, by Weierstrass' theorem, there exists an optimal solution x^k for the above problem. We have for all k

$$F(x^k) + \frac{k}{2} \sum_{j=1}^{r} \big(g_j^+(x^k)\big)^2 = F^k(x^k) \leq F^k(\overline{x}) = F(\overline{x}) \tag{1.20}$$

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

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

otherwise the left-hand side of Eq. (1.20) would become unbounded from above as $k \to \infty$. Therefore, every limit point $\tilde{x}$ of $\{x^k\}$ is feasible, i.e., $\tilde{x} \in C$. Furthermore, Eq. (1.20) yields $F(x^k) \leq F(\overline{x})$ for all k, so by taking the limit along the relevant subsequence as $k \to \infty$, we obtain

$$F(\tilde{x}) \leq F(\overline{x}).$$

Since $\tilde{x}$ is feasible, we have $F(\overline{x}) < F(\tilde{x})$ (since F achieves a strict global minimum over C at $\overline{x}$), unless $\tilde{x} = \overline{x}$, which when combined with the preceding inequality yields $\tilde{x} = \overline{x}$. Thus the sequence $\{x^k\}$ converges to $\overline{x}$, 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, $\nabla F^k(x^k)'y \geq 0$ for all $y \in T_X(x^k)$, or equivalently $-\nabla F^k(x^k) \in T_X(x^k)^*$, which is written as

$$-\left(\nabla F(x^k) + \sum_{j=1}^{r} \zeta_j^k \nabla g_j(x^k)\right) \in T_X(x^k)^*, \tag{1.21}$$

where

$$\zeta_j^k = k g_j^+(x^k). \tag{1.22}$$

Denote,

$$\delta^k = \sqrt{1 + \sum_{j=1}^{r} (\zeta_j^k)^2}, \tag{1.23}$$

$$\mu_0^k = \frac{1}{\delta^k}, \qquad \mu_j^k = \frac{\zeta_j^k}{\delta^k}, \quad j = 1, \ldots, r. \tag{1.24}$$

Then by dividing Eq. (1.21) with δ^k, we get

$$-\left(\mu_0^k \nabla F(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k)\right) \in T_X(x^k)^*. \tag{1.25}$$

Since by construction the sequence $\{\mu_0^k, \mu_1^k, \ldots, \mu_r^k\}$ is bounded, it must contain a subsequence that converges to some nonzero limit $\{\overline{\mu}_0, \overline{\mu}_1, \ldots, \overline{\mu}_r\}$. From Eq. (1.25) and the defining property of the normal cone $N_X(\overline{x})$ [$x^k \to \overline{x}$, $z^k \to \overline{z}$, and $z^k \in T_X(x^k)^*$ for all k, imply that $\overline{z} \in N_X(\overline{x})$], we see that $\overline{\mu}_0$ and the $\overline{\mu}_j$ must satisfy

$$-\left(\overline{\mu}_0 \nabla F(\overline{x}) + \sum_{j=1}^{r} \overline{\mu}_j \nabla g_j(\overline{x})\right) \in N_X(\overline{x}). \tag{1.26}$$

Furthermore, from Eq. (1.24), we have $g_j(x^k) > 0$ for all j such that $\overline{\mu}_j > 0$ and k sufficiently large. By using the quasinormality of $\overline{x}$, it follows that we cannot have $\overline{\mu}_0 = 0$, and by appropriately normalizing, we can take $\overline{\mu}_0 = 1$ and obtain

$$-\left(\nabla F(\overline{x}) + \sum_{j=1}^{r} \overline{\mu}_j \nabla g_j(\overline{x})\right) \in N_X(\overline{x}).$$

Since $-\nabla F(\overline{x}) = y$, we see that

$$y = z + \sum_{j=1}^{r} \overline{\mu}_j \nabla g_j(\overline{x}),$$

where $z \in N_X(\overline{x})$, and the scalars $\overline{\mu}_1, \ldots, \overline{\mu}_r$ and the sequence $\{x^k\}$ satisfy the desired properties, thus completing the proof. **Q.E.D.**

Next, we prove the main result of this section, namely that under a regularity assumption on X, quasinormality implies quasiregularity.

Proposition 4.1.6: If x^* is a quasinormal vector of C and X is regular at x^*, then x^* is quasiregular.

Proof: We assume for simplicity that all the constraints are inequalities. We must show that $T_C(x^*) = T_X(x^*) \cap V(x^*)$, and to this end, we first show that $T_C(x^*) \subset T_X(x^*) \cap V(x^*)$.

Indeed, since $C \subset X$, using the definition of the tangent cone, we have

$$T_C(x^*) \subset T_X(x^*). \tag{1.27}$$

To show that $T_C(x^*) \subset V(x^*)$, let y be a nonzero tangent of C at x^*. Then there exist sequences $\{\xi^k\}$ and $\{x^k\} \subset C$ such that $x^k \neq x^*$ for all k,

$$\xi^k \to 0, \qquad x^k \to x^*,$$

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

$$\nabla g_j(\tilde{x}^k)' y^k \leq 0, \qquad y^k = y + \xi^k \|y\|.$$

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, $T_C(x^*) \subset V(x^*)$. Together with Eq. (1.27), this shows that

$$T_C(x^*) \subset T_X(x^*) \cap V(x^*). \tag{1.28}$$

To show the reverse inclusion $T_X(x^*) \cap V(x^*) \subset T_C(x^*)$, we first show that

$$N_C(x^*) \subset T_X(x^*)^* + V(x^*)^*.$$

Let y^* be a vector that belongs to $N_C(x^*)$. By the definition of the normal cone, this implies the existence of a sequence $\{x^k\} \subset C$ that converges to x^* and a sequence $\{y^k\}$ that converges to y^*, with $y^k \in T_C(x^k)^*$ for all k. In view of the assumption that x^* is quasinormal, it follows from Prop. 4.1.4 that for all sufficiently large k, x^k is quasinormal. Then, by Prop. 4.1.5, for each sufficiently large k, there exists a vector $z^k \in N_X(x^k)$ and nonnegative scalars $\mu_1^k, \ldots, \mu_r^k$ such that

$$y^k = z^k + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k). \tag{1.29}$$

Furthermore, there exists a sequence $\{x_l^k\} \subset X$ such that

$$\lim_{l \to \infty} x_l^k = x^k,$$

and for all l, $\mu_j^k g_j(x_l^k) > 0$ for all j with $\mu_j^k > 0$.

We will show that the sequence $\{\mu_1^k, \ldots, \mu_r^k\}$ is bounded. Suppose, to arrive at a contradiction, that this sequence is unbounded, and assume without loss of generality, that for each k, at least one of the μ_j^k is nonzero. For each k, denote

$$\delta^k = \frac{1}{\sum_{j=1}^{r} (\mu_j^k)^2},$$

and

$$\xi_j^k = \delta^k \mu_j^k, \qquad \forall\ j = 1, \ldots, r.$$

It follows that $\delta^k > 0$ for all k and $\delta^k \to 0$ as $k \to \infty$. Then, by multiplying Eq. (1.29) by δ^k, we obtain

$$\delta^k y^k = \delta^k z^k + \sum_{j=1}^{r} \xi_j^k \nabla g_j(x^k),$$

or equivalently, since $z^k \in N_X(x^k)$ and $\delta^k > 0$, we have

$$\delta^k z^k = \left(\delta^k y^k - \sum_{j=1}^{r} \xi_j^k \nabla g_j(x^k) \right) \in N_X(x^k).$$

Note that by construction, the sequence $\{\xi_1^k, \ldots, \xi_r^k\}$ is bounded, and therefore has a nonzero limit point $\{\xi_1^*, \ldots, \xi_r^*\}$. Taking the limit in the preceding relation along the relevant subsequence and using the facts $\delta^k \to 0$, $y^k \to y^*$, and $x^k \to x^*$ together with the closedness of the normal cone $N_X(x^*)$ (cf. Prop. 4.1.4), we see that $\delta^k z^k$ converges to some vector z^* in $N_X(x^*)$, where

$$z^* = -\left(\sum_{j=1}^{r} \xi_j^* \nabla g_j(x^*)\right).$$

Furthermore, by defining an index l_k for each k such that $l_1 < \cdots < l_{k-1} < l_k$ and

$$\lim_{k\to\infty} x_{l_k}^k = x^*,$$

we see that for all j with $\xi_j^* > 0$, we have $g_j(x_{l_k}^k) > 0$ for all sufficiently large k. The existence of such scalars $\xi_1^*, \ldots, \xi_r^*$ violates the quasinormality of the vector x^*, thus showing that the sequence $\{\mu_1^k, \ldots, \mu_r^k\}$ is bounded.

Let $\{\mu_1^*, \ldots, \mu_r^*\}$ be a limit point of the sequence $\{\mu_1^k, \ldots, \mu_r^k\}$, and assume without loss of generality that $\{\mu_1^k, \ldots, \mu_r^k\}$ converges to $\{\mu_1^*, \ldots, \mu_r^*\}$. Taking the limit as $k \to \infty$ in Eq. (1.29), we see that z^k converges to some z^*, where

$$z^* = y^* - \left(\sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right). \tag{1.30}$$

By closedness of the normal cone $N_X(x^*)$ and in view of the assumption that X is regular at x^*, so that $N_X(x^*) = T_X(x^*)^*$, we have that $z^* \in T_X(x^*)^*$. Furthermore, by defining an index l_k for each k such that $l_1 < \cdots < l_{k-1} < l_k$ and

$$\lim_{k\to\infty} x_{l_k}^k = x^*,$$

we see that for all j with $\mu_j^* > 0$, we have $g_j(x_{i_k}^k) > 0$ for all sufficiently large k, showing that $g_j(x^*) = 0$. Hence, it follows that $\mu_j^* = 0$ for all $j \notin A(x^*)$, and using Eq. (1.30), we can write y^* as

$$y^* = z^* + \left(\sum_{j\in A(x^*)} \mu_j^* \nabla g_j(x^*)\right).$$

By Farkas' Lemma, $V(x^*)^*$ is the cone generated by $\nabla g_j(x^*),\ j \in A(x^*)$. Hence, it follows that $y^* \in T_X(x^*)^* + V(x^*)^*$, and we conclude that

$$N_C(x^*) \subset T_X(x^*)^* + V(x^*)^*. \tag{1.31}$$

Finally, using the properties relating to cones and their polars and the fact that $T_X(x^*)$ is convex (which follows by the regularity of X at x^*, cf. Proposition 2.2.19 of Chapter 2), we obtain

$$\big(T_X(x^*)^* + V(x^*)^*\big)^* = T_X(x^*) \cap V(x^*) \ \subset \ N_C(x^*)^*. \tag{1.32}$$

Using the relation $N_C(x^*)^* \subset T_C(x^*)$ (cf. Proposition 2.2.18 of Chapter 2), this shows that $T_X(x^*) \cap V(x^*) \subset T_C(x^*)$, which together with Eq. (1.28) concludes the proof. **Q.E.D.**

Note that in the preceding proof, we showed

$$T_C(x^*) \subset T_X(x^*) \cap V(x^*),$$

which implies that

$$T_X(x^*)^* + V(x^*)^* \subset (T_X(x^*) \cap V(x^*))^* \subset T_C(x^*)^*. \tag{1.33}$$

We also proved that if X is regular at x^* and x^* is quasinormal, we have

$$N_C(x^*) \subset T_X(x^*)^* + V(x^*)^*,$$

[cf. Eq. (1.31)]. Combining the preceding two relations with the relation $T_C(x^*)^* \subset N_C(x^*)$, we obtain

$$T_C(x^*)^* = N_C(x^*),$$

thus showing that quasinormality of x^* together with regularity of X at x^* implies that C is regular at x^*.

Note that contrary to the case where $X = \Re^n$, quasiregularity is not sufficient to guarantee the existence of a Lagrange multiplier. What is happening here is that the constraint set admits Lagrange multipliers at x^* if and only if

$$T_C(x^*)^* = T_X(x^*)^* + V(x^*)^*, \tag{1.34}$$

(cf. Chapter 3). Note that this condition is equivalent to the following two conditions[1] :

(a) $V(x^*) \cap \mathrm{cl}\Big(\mathrm{conv}\big(T_X(x^*)\big)\Big) = \mathrm{cl}\Big(\mathrm{conv}\big(T_C(x^*)\big)\Big)$,

(b) $V(x^*)^* + T_X(x^*)^*$ is a closed set.

Quasiregularity is a weaker condition, even under the assumption that X is regular, since the vector sum $V(x^*)^* + T_X(x^*)^*$ need not be closed even if both of these cones themselves are closed, as shown in the following example.

Example 4.1.4:

Consider the constraint set $C \subset \Re^3$ specified by,

$$C = \big\{x \in X \mid h(x) = 0\big\},$$

where

$$X = \big\{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 \leq x_3^2,\ x_3 \leq 0\big\},$$

and

$$h(x) = x_2 + x_3.$$

Let x^* denote the origin. Since X is closed and convex, we have that $X = T_X(x^*)$, and that X is regular at x^*. The cone of first order feasible variations, $V(x^*)$, is given by

$$V(x^*) = \big\{(x_1, x_2, x_3) \mid x_2 + x_3 = 0\big\}.$$

It can be seen that the set $V(x^*)^* + T_X(x^*)^*$ is not closed, implying that C does not admit Lagrange multipliers. On the other hand, we have

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

i.e., x^* is quasiregular.

Hence, quasiregularity is not powerful enough to assert the existence of Lagrange multipliers for the general case $X \neq \Re^n$, unless additional assumptions are imposed. It is

[1] Note that this is exactly the same condition given by Guignard [Gui69] as a sufficient condition for the constraint set to admit Lagrange multipliers at x^*.

effective only for special cases, for instance, when $T_X(x^*)$ is a convex and polyhedral cone, in which case the vector sum $V(x^*)^* + T_X(x^*)^*$ is closed, and quasiregularity implies the admittance of Lagrange multipliers. Thus the importance of quasiregularity, the classical pathway to Lagrange multipliers when $X = \Re^n$, diminishes when $X \neq \Re^n$. By contrast, pseudonormality provides satisfactory unification of the theory.

4.2. EXACT PENALTY FUNCTIONS

In this section, we relate the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C, \end{aligned} \tag{2.1}$$

where

$$C = X \cap \{x \mid h_1(x) = 0, \ldots, h_m(x) = 0\} \cap \{x \mid g_1(x) \leq 0, \ldots, g_r(x) \leq 0\}, \tag{2.2}$$

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*, and determines the extent to which the penalized problem approximates the original constrained problem. As c takes higher values, the penalty approximation becomes increasingly accurate. In fact, it can be shown that the optimal solution of the original constrained problem can be obtained by solving a sequence of problems, where we minimize F_{c^k} over X, for some sequence $\{c^k\}$ that goes to infinity (see [Ber99]). However, although the penalized problem is less constrained, it becomes seriously ill-conditioned as c goes to infinity. It turns out that, under some conditions on the constraint set, the optimal solution of the original constrained problem can be obtained by a single optimization problem, which minimizes the penalized cost function F_c over X, for a finite value of the parameter c.

Definition 4.2.3: 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.

Note that, like admittance of Lagrange multipliers, admittance of an exact penalty is a *property of the constraint set* C, and does not depend on the cost function f of problem (2.1).

Traditionally exact penalty functions have been viewed as useful computational devices and they have not been fully integrated within the theory of constraint qualifications. There has been research on finding conditions that guarantee that the constraint set admits an exact penalty for optimization problems that do not have an abstract set constraint. In fact, it was shown by Pietrzykowski [Pie69] that the constraint set admits an exact penalty if CQ1 holds. Similarly, the fact that CQ2 implies admittance of an exact penalty was studied by Zangwill [Zan67], Han and Mangasarian [HaM79], and Bazaraa and Goode [BaG82]. In this work, we will clarify the relations of exact penalty functions, Lagrange multipliers, and constraint qualifications. We show that pseudonormality is the key property that places all these notions in one big picture. In the process we prove in a unified way that the constraint set admits an exact penalty for a much larger variety of constraint qualifications than has been known up to now.

Note that, 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 4.2.5:

Consider the 2-dimensional constraint set specified by

$$h_1(x) = \frac{x_2}{x_1^2 + 1} = 0, \qquad x \in X = \Re^2.$$

The feasible points are of the form $x = (x_1, 0)$ with $x_1 \in \Re$, and at each of them the gradient $\nabla h_1(x^*)$ 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

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

(take $x_1 = x_2$ as $x_2 \to -\infty$). Thus, the penalty function is not exact for any $c > 0$. It follows that 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 next 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 MFCQ to obtain a necessary condition for a local minimum of the exact penalty function.

Proposition 4.2.7: Let x^* be a local minimum of

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

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

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

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

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

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

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

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

$$\begin{aligned} &\text{minimize} \;\; f(x) + c\left(\sum_{i=1}^{m} w_i + \sum_{j=1}^{r} v_j\right) \\ &\text{subject to} \;\; x \in X, \;\; h_i(x) \le w_i, \; -h_i(x) \le w_i, \; i = 1, \ldots, m, \\ &\qquad\qquad g_j(x) \le v_j, \; 0 \le v_j, \; j = 1, \ldots, r, \end{aligned}$$

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 MFCQ is satisfied. Thus, by Prop. 4.1.2, this local minimum is pseudonormal, and hence there exist multipliers satisfying the enhanced Fritz John conditions (Prop. 3.2.3) 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 4.2.8: If x^* is a feasible vector of problem (2.1)-(2.2), which is pseudonormal, the constraint set admits an exact penalty at x^*.

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

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

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

$$f(x^*) < f(x), \qquad \forall\, x \in C \text{ with } x \neq x^* \text{ and } \|x - x^*\| \leq \epsilon. \tag{2.3}$$

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 (2.2), i.e.,

$$\sum_{i=1}^{m} |h_i(x^k)| + \sum_{j=1}^{r} g_j^+(x^k) > 0. \tag{2.4}$$

We have

$$F_k(x^k) = f(x^k) + k\left(\sum_{i=1}^{m} |h_i(x^k)| + \sum_{j=1}^{r} g_j^+(x^k)\right) \leq f(x^*), \tag{2.5}$$

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. (2.3) and (2.5) 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. 4.2.7):

$$-\left(\frac{1}{k}\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k)\right) \in N_X(x^k), \tag{2.6}$$

where

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

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

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

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

In view of Eq. (2.4), we can find a subsequence $\{\lambda^k, \mu^k\}_{k\in\mathcal{K}}$ such that for some equality constraint index i we have $|\lambda_i^k| = 1$ and $h_i(x^k) \neq 0$ for all $k \in \mathcal{K}$ or for some inequality constraint index j we have $\mu_j^k = 1$ and $g_j(x^k) > 0$ for all $k \in \mathcal{K}$. Let (λ, μ) be a limit point of this subsequence. We then have $(\lambda, \mu) \neq (0, 0)$, $\mu \geq 0$. Using the closure of the mapping $x \mapsto N_X(x)$, Eq. (2.6) yields

$$-\left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right) \in N_X(x^*). \tag{2.7}$$

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

$$\sum_{i=1}^{m} \lambda_i h_i(x^k) + \sum_{j=1}^{r} \mu_j g_j(x^k) > 0. \tag{2.8}$$

Thus, Eqs. (2.7) and (2.8) violate the hypothesis that x^* is pseudonormal. **Q.E.D.**

The following example shows that the converse of Prop. 4.2.8 does not hold. In particular, the admittance of an exact penalty function at a point x^* does not imply pseudonormality.

Example 4.2.6:

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

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

(see Fig. 4.2.5). The only feasible solution is $x^* = (0,0)$ and the constraint gradients are given by

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

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

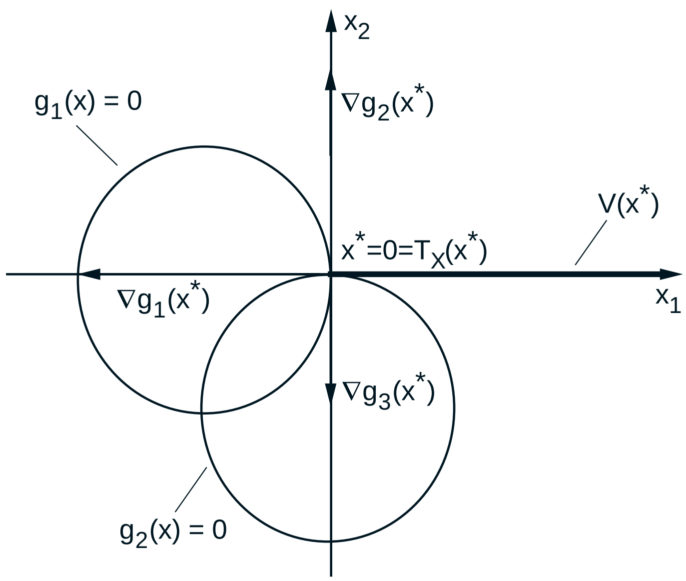


Figure 4.2.5. Constraints of Example 4.2.6. The only feasible point is $x^* = (0,0)$. The tangent cone $T(x^*)$ and the cone of first order feasible variations $V(x^*)$ are also illustrated in the figure.

However, it can be seen that the directional derivative of the function $P(x) = \sum_{j=1}^{3} g_j^+(x)$ at x^* is positive in all directions. This implies that we can choose a sufficiently large penalty parameter c, so that x^* is a local 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 4.2.9: Let x^* be a feasible vector of problem (2.1)-(2.2), and let X be regular at x^*. If the constraint set admits an exact penalty at x^*, it admits Lagrange multipliers at x^*.

Proof: Suppose that a given smooth function $f(x)$ has a local minimum at x^*. Then the function $f(x)+\|x-x^*\|^2$ has a strict local minimum at x^*. Since C admits an exact penalty at x^*, there exist λ_i^* and μ_j^* satisfying the conditions of Prop. 4.2.7. (The term $\|x-x^*\|^2$ in the cost function is inconsequential, since its gradient at x^* is 0.) In view of the regularity of X at x^*, the λ_i^* and μ_j^* are Lagrange multipliers. **Q.E.D.**

Note that because Prop. 4.2.7 does not require regularity of X, the proof of Prop. 4.2.9 can be used to establish that *admittance of an exact penalty implies the admittance of R-multipliers*, as defined in Section 5.3. On the other hand, Example 3.4.5 shows that the regularity assumption on X in Prop. 4.2.9 cannot be dispensed with. Indeed, in that example, x^* is pseudonormal, the constraint set admits an exact penalty at x^* (consistently with Prop. 4.2.8), but it does not admit Lagrange multipliers.

The relations shown thus far are summarized in Fig. 4.2.6, 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 additional examples and counterexamples).

4.3. USING THE EXTENDED REPRESENTATION

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

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

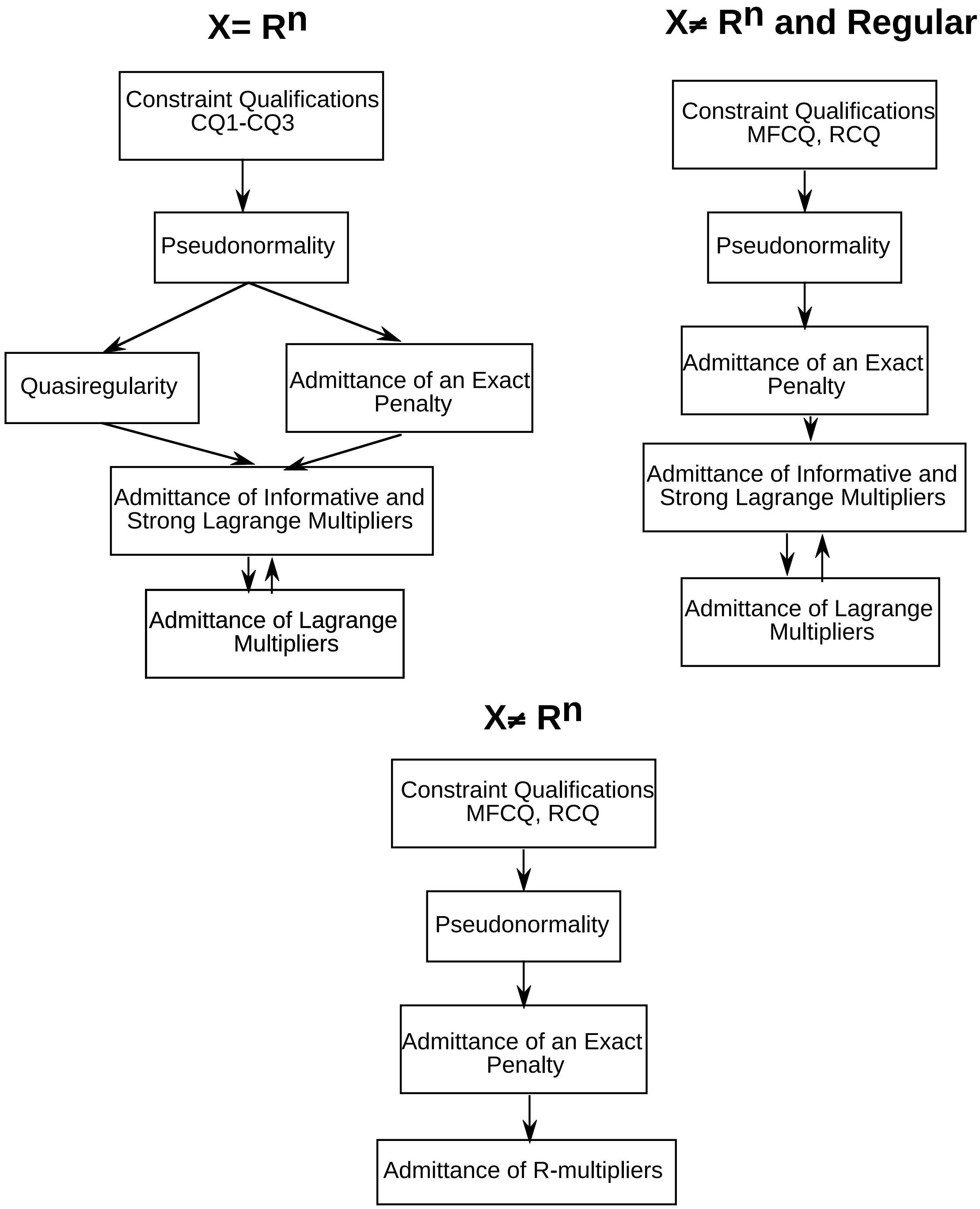


Figure 4.2.6. Relations between various conditions, which when satisfied at a local minimum x^*, guarantee the admittance of an exact penalty and corresponding multipliers. In the case where X is regular, the tangent and normal cones are convex. Hence, by Prop. 3.3.6, the admittance of Lagrange multipliers implies the admittance of an informative Lagrange multiplier, while by Prop. 4.2.5, pseudonormality implies the admittance of an exact penalty.

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

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

We call this the *extended representation* of C, to contrast it with the representation (2.2), which we call the *original representation*. Issues relating to exact penalty functions and Lagrange multipliers can be investigated for the extended representation and results can be carried over to the original representation by using the following proposition.

Proposition 4.3.10:

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

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

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

$$\nabla f(x^*) + \sum_{i=1}^{\overline{m}} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{\overline{r}} \mu_j^* \nabla g_j(x^*) = 0, \tag{3.1}$$

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

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

where

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

For $y \in T_X(x^*)$, we have $\nabla h_i(x^*)'y = 0$ for all $i = m+1, \ldots, \overline{m}$, and $\nabla g_j(x^*)'y \leq 0$ for all $j = r+1, \ldots, \overline{r}$ with $j \in \overline{A}(x^*)$. Hence Eq. (3.1) implies that

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

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

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

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

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

Prop. 4.3.10 can be used in the case when all the constraints are linear and X is a polyhedron. Here, the constraint set need not satisfy pseudonormality (as shown in the following example). However, by Prop. 4.1.2, it satisfies pseudonormality in the extended representation, so using Prop. 4.3.10, it admits Lagrange multipliers and an exact penalty at any feasible point in the original representation.

Example 4.3.7:

Let

$$C = \{x \in X \mid a'x \leq 0,\ b'x \leq 0\},$$

where $a = (1, -1)$, $b = (-1, -1)$, and $X = \{x \in \Re^2 \mid\ a'x \geq 0,\ b'x \geq 0\}$. The constraint set is depicted in Fig. 4.3.7.

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

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

while in every neighborhood N of x^* there is an $x \in X \cap N$ such that $a'x > 0$ and $b'x > 0$ simultaneously. Hence x^* is not pseudonormal. This constraint set, however, admits Lagrange multipliers at $x^* = (0, 0)$ with respect to its extended representation (cf. Prop. 4.3.10), and hence it admits Lagrange multipliers at $x^* = (0, 0)$ with respect to the original representation.

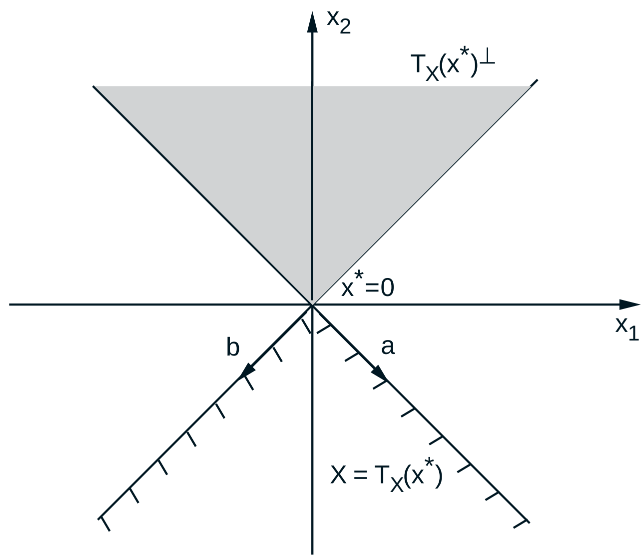


Figure 4.3.7. Constraints of Example 4.3.7. The only feasible point is $x^* = (0,0)$. The tangent cone $T_X(x^*)$ and its polar $T_X(x^*)^*$ are shown in the figure.

Note that part (a) of Prop. 4.3.10 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 Proposition 3.3.6 applies.

Example 4.3.8:

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

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

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. Since $X = \Re^2$ in this representation, the constraint set also admits informative Lagrange multipliers, as shown by Proposition 3.3.6.

Now let the same constraint set be specified by the two linear constraint functions $a_1'x \le 0$ and $a_2'x \le 0$ together with the abstract constraint set

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

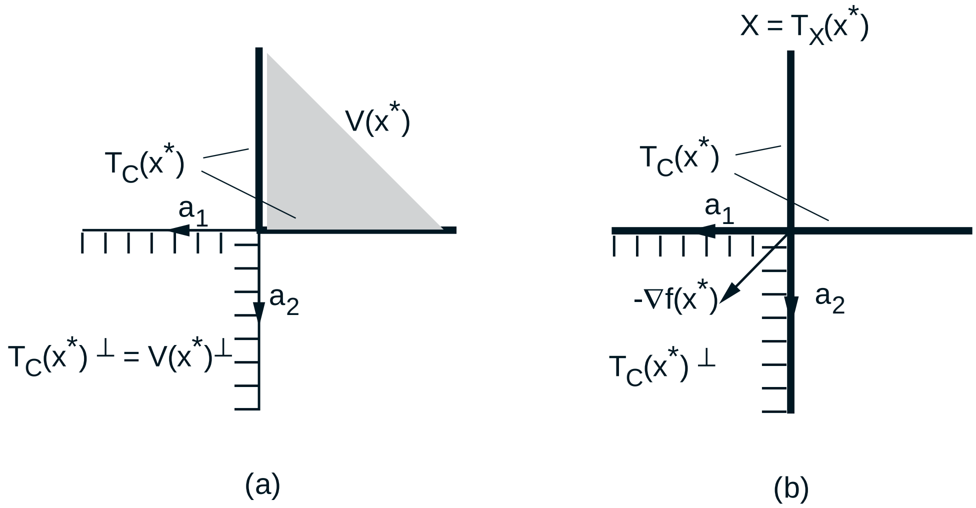


Figure 4.3.8. Constraints and relevant cones for different representations of the problem in Example 4.3.8.

Here $T_X(x^*) = X$ and $T_X(x^*)^* = \{0\}$. The normal cone $N_X(x^*)$ consists of the coordinate axes. Since $N_X(x^*) \neq T_X(x^*)^*$, X is not regular at x^*. Furthermore, $T_X(x^*)$ is not convex, so Prop. 3.3.6(a) cannot be used to guarantee the admittance of an informative Lagrange multiplier. For any f for which x^* is a local minimum, we must have $-\nabla f(x^*) \in T_C(x^*)^*$ (see Fig. 4.3.8). The candidate multipliers are determined from the requirement that

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

which uniquely determines μ_1 and μ_2. If $\nabla f(x^*)$ lies in the interior of the positive orthant, we need to have $\mu_1 > 0$ and $\mu_2 > 0$. However, there exists no $x \in X$ that violates both constraints $a_1'x \leq 0$ and $a_2'x \leq 0$, so the multipliers do not qualify as informative. Thus, the constraint set does not admit informative Lagrange multipliers in the original representation.

CHAPTER 5

MULTIPLIERS AND CONVEX PROGRAMMING

In this chapter, our objective is to extend the theory of the preceding chapters to problems in which continuity/differentiability assumptions are replaced by convexity assumptions. For this purpose, we adopt a different approach based on tools from convex analysis, such as hyperplanes, convex set support/separation arguments, and saddle point theory. Because of the geometric character of the analysis, the results and their proofs admit insightful visualization. Moreover, since this line of analysis does not depend on using gradients at a specific local or global minimum, it allows us to analyze the global problem structure. Thus, it becomes possible to develop a similar theory for optimization problems without guaranteeing the existence of an optimal solution. This development motivates us to define an extended notion of pseudonormality, which is a property of the constraint set, as opposed to being tied to a specific feasible vector of the constraint set. Through the notion of pseudonormality, this development provides an alternative pathway to obtain strong duality results of convex programming. Pseudonormality also admits an insightful geometric visualization under convexity assumptions.

We first present a straightforward extension of the theory of the preceding chapters to convex programming problems by using subgradients, instead of gradients, for convex possibly nondifferentiable functions. For this purpose, we use generic optimality conditions given in Chapter 2 for minimizing a convex function over a constraint set. However, using subgradients requires more stringent assumptions than necessary on the cost and the constraint functions. Therefore, in Section 5.2, we use a different line of analysis based on convexity and saddle point theory to derive optimality conditions for convex problems with optimal solutions.

We next introduce a new notion of a multiplier vector, called *geometric*, that is not tied to a specific local or global minimum and does not require differentiability or even continuity of the cost and constraint functions. We show that, under convexity assumptions, these multipliers are related to Lagrange multipliers defined in Chapter 3. Then we focus on problems that do not necessarily have optimal solutions and develop enhanced Fritz John conditions for these problems. We consider special geometric multipliers that carry sensitivity information regarding constraints of the problem (similar to 'informative Lagrange multipliers'), and investigate the conditions required for their existence. Finally, we derive Fritz John optimality conditions for a dual optimization problem. Based on these conditions, we define a special dual optimal solution that carries sensitivity information and show its existence under general assumptions.

5.1. EXTENSIONS OF THE DIFFERENTIABLE CASE

We consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g(x) \le 0, \end{aligned} \tag{1.1}$$

where $g(x) = \big(g_1(x), \ldots, g_r(x)\big)$, under the following assumption:

Assumption 5.1.1: The set X is nonempty and closed, and the functions f and g_j are real-valued and convex over $\Re^n$.

For simplicity, we assume no equality constraints. The extension of the following analysis to cover equality constraints is straightforward.

The theory of the preceding chapters can be generalized by substituting the gradients of convex but nondifferentiable functions with subgradients. In particular, we use the necessary optimality condition given in Chapter 2 for the problem of minimizing a convex function $F(x)$ over X: if x^* is a local minimum and the tangent cone $T_X(x^*)$ is convex, then

$$0 \in \partial F(x^*) + T_X(x^*)^*. \tag{1.2}$$

We have the following proposition.

Proposition 5.1.1: Let x^* be a local minimum of problem (1.1). Let Assumption 5.1.1 hold, and assume that the tangent cone $T_X(x^*)$ is convex. Then, there exist scalars μ_0^*, and $\mu_1^*, \ldots, \mu_r^*$, satisfying the following conditions:

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

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

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

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

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

$$g_j^+(x^k) = o\big(\min_{j \in J} g_j(x)\big), \quad \forall\, j \notin J.$$

Proof: The proof is similar to the proof of Prop. 3.2.3, in that we use the condition $0 \in \partial F^k(x^k) + T_X(x^k)^*$ in place of $-\nabla F^k(x^k) \in T_X(x^k)^*$, together with the closedness of $N_X(x^*)$. **Q.E.D.**

The straightforward extensions for the definitions of Lagrange multiplier and pseudonormality are as follows.

Definition 5.1.1: Consider problem (1.1), and let x^* be a local minimum. A vector μ^* is called a *Lagrange multiplier vector corresponding to* f *and* x^* if

$$0 \in \partial f(x^*) + \sum_{j=1}^r \mu_j^* \partial g_j(x^*) + T_X(x^*)^*, \tag{1.3}$$

$$\mu^* \geq 0, \qquad \mu^{*\prime} g(x^*) = 0. \tag{1.4}$$

Definition 5.1.2: Consider problem (1.1) under Assumption 5.1.1. A feasible vector x^* is said to be *pseudonormal* if there do not exist any scalars $\mu_1, \ldots, \mu_r$, and any sequence $\{x^k\} \subset X$ such that:

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

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

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

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

If a local minimum x^* is pseudonormal and the tangent cone $T_X(x^*)$ is convex, by Prop. 5.1.1, there exists a Lagrange multiplier vector, which also satisfies the extra condition (iv) of that proposition, hence qualifies as an 'informative' Lagrange multiplier.

The theory of Chapter 4 can be extended to relate constraint qualifications to pseudonormality. In particular, it can be seen that a feasible vector x^* is pseudonormal under any of the following conditions:

(1) *Linearity criterion*: X is a polyhedron and the constraint functions g_j are affine.

(2) *Slater criterion*: X is closed and convex, the functions $g_j : \Re^n \mapsto \Re$ are convex over $\Re^n$,[1] and there exists a feasible vector $\bar{x}$ such that

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

Thus, under either any one of these criteria, a Lagrange multiplier vector, satisfying the extra CV condition (iv) of Prop. 5.1.1, is guaranteed to exist.

[1] The assumptions that X is closed and the functions g_j are convex over $\Re^n$ can be relaxed using the proposition given in the next section.

We now consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g(x) \le 0, \end{aligned} \tag{2.1}$$

where $g(x) = \big(g_1(x), \ldots, g_r(x)\big)$, under weaker convexity assumptions. [The extension of the analysis to the case where there are affine equality constraints is straightforward: we replace each equality constraint with two affine inequality constraints.] In particular, we assume the following:

Assumption 5.2.2: The set X is nonempty and convex, and the functions $f : X \mapsto \Re$ and $g_j : X \mapsto \Re$ are closed and convex.

Since in this section we do not require convexity of the cost and constraint functions over the entire space $\Re^n$, the line of analysis using subgradients breaks down. In the next proposition, we use a different line of proof that does not rely on gradients and subgradients, and is based instead on the saddle point theory.

Proposition 5.2.2: (Enhanced Fritz John Conditions for Convex Problems with an Optimal Solution) Consider problem (2.1) under Assumption 5.2.2, and let x^* be a global minimum. Then there exists a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, satisfying the following conditions:

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

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

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

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

$$f(x^k) < f(x^*), \qquad \mu_j^* g_j(x^k) > 0, \quad \forall\, j \in J, \tag{2.2}$$

$$g_j^+(x^k) = o\left(\min_{j \in J} g_j(x^k)\right), \qquad \forall\, j \notin J. \tag{2.3}$$

Proof: For each positive integer k and m, we consider the function

$$L_{k,m}(x, \xi) = f(x) + \frac{1}{k^3}\|x - x^*\|^2 + \xi' g(x) - \frac{\|\xi\|^2}{2m},$$

For each k, we consider the set

$$X^k = X \cap \big\{x \mid \|x - x^*\| \le k)\big\}.$$

Since the functions f and the g_j are closed and convex over X, they are closed and convex over X^k, which implies that, for each $\xi \ge 0$, $L_{k,m}(\cdot, \xi)$ is closed and convex over X^k. Similarly, for each $x \in \Re^n$, $L_{k,m}(x, \cdot)$ is closed, convex, and coercive in ξ. Since X^k is also bounded, we use the saddle point theorem given in Chapter 2 to assert that $L_{k,m}$ has a saddle point over $x \in X^k$ and $\xi \ge 0$, which we denote by $(x^{k,m}, \xi^{k,m})$.

Since $(x^{k,m}, \xi^{k,m})$ is a saddle point of $L_{k,m}$ over $x \in X^k$ and $\xi \ge 0$, the infimum of $L_{k,m}(x, \xi^{k,m})$ over $x \in X^k$ is attained at $x^{k,m}$, implying that

$$\begin{aligned}
f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m} - x^*\|^2 + \xi^{k,m\prime} g(x^{k,m}) &= \inf_{x \in X^k}\left\{f(x) + \frac{1}{k^3}\|x - x^*\|^2 + \xi^{k,m\prime} g(x)\right\} \\
&\le \inf_{x \in X^k,\ g(x) \le 0}\left\{f(x) + \frac{1}{k^3}\|x - x^*\|^2 + \xi^{k,m\prime} g(x)\right\} \\
&\le \inf_{x \in X^k,\ g(x) \le 0}\left\{f(x) + \frac{1}{k^3}\|x - x^*\|^2\right\} \\
&= f(x^*).
\end{aligned} \tag{2.4}$$

Hence, we have

$$\begin{aligned} L_{k,m}(x^{k,m},\xi^{k,m}) &= f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m}-x^*\|^2 + \xi^{k,m'}g(x^{k,m}) - \frac{1}{2m}\|\xi^{k,m}\|^2 \\ &\le f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m}-x^*\|^2 + \xi^{k,m'}g(x^{k,m}) \\ &\le f(x^*). \end{aligned} \tag{2.5}$$

Since $L_{k,m}$ is quadratic in ξ, the supremum of $L_{k,m}(x^{k,m},\xi)$ over $\xi \ge 0$ is attained at

$$\xi_j^{k,m} = kg_j^+(x^{k,m}), \qquad j = 1,\dots,r. \tag{2.6}$$

This implies that

$$\begin{aligned} L_{k,m}(x^{k,m},\xi^{k,m}) &= f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m}-x^*\|^2 + \frac{m}{2}\|g^+(x^{k,m})\|^2 \\ &\ge f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m}-x^*\|^2 \\ &\ge f(x^{k,m}). \end{aligned} \tag{2.7}$$

From Eqs. (2.5) and (2.7), we see that the sequence $\{x^{k,m}\}$ belongs to the set $\big\{x \in X^k \mid f(x) \le f(x^*)\big\}$, which is compact. Hence, $\{x^{k,m}\}$ has a limit point (as $m \to \infty$), denoted by $\overline{x}^k$, which belongs to $\big\{x \in X^k \mid f(x) \le f(x^*)\big\}$. By passing to a subsequence if necessary, we can assume without loss of generality that $\{x^{k,m}\}$ converges to $\overline{x}^k$ as $m \to \infty$. For each k, the sequence $f(x^{k,m})$ is bounded from below by $\inf_{x\in X^k} f(x)$, which is finite by Weierstrass' Theorem since f is closed and coercive over X^k. Also, for each k, $L_{k,m}(x^k,\xi^k)$ is bounded from above by $f(x^*)$ [cf. Eq. (2.5)], therefore, Eq. (2.7) implies that

$$\limsup_{m\to\infty} g(x^{k,m}) \le 0, \tag{2.8}$$

from which, by using the lower semicontinuity of g_j, we obtain $g(\overline{x}^k) \le 0$, implying that $\overline{x}^k$ is a feasible point of problem (2.1), so that $f(\overline{x}^k) \ge f(x^*)$. Using Eqs. (2.5) and (2.7) together with the lower semicontinuity of f, we also have

$$f(\overline{x}^k) \le \liminf_{m\to\infty} f(x^{k,m}) \le \limsup_{m\to\infty} f(x^{k,m}) \le f(x^*),$$

hence showing that for each k,

$$\lim_{m\to\infty} f(x^{k,m}) = f(x^*). \tag{2.9}$$

Together with Eq. (2.7), this also implies that for each k,

$$\lim_{m\to\infty} x^{k,m} = x^*.$$

Combining the preceding relations with Eqs. (2.5) and (2.7), for each k, we obtain

$$\lim_{m\to\infty} \big(f(x^{k,m}) - f(x^*) + \xi^{k,m\prime} g(x^{k,m})\big) = 0. \tag{2.10}$$

Denote

$$\delta^{k,m} = \sqrt{1 + \sum_{j=1}^{r} (\xi_j^{k,m})^2},$$

$$\mu_0^{k,m} = \frac{1}{\delta^{k,m}}, \qquad \mu_j^{k,m} = \frac{\xi_j^{k,m}}{\delta^{k,m}},\ j = 1, \ldots, r. \tag{2.11}$$

Since $\delta^{k,m}$ is bounded from below, Eq. (2.10) yields

$$\lim_{k\to\infty} \mu_0^{k,m}\big(f(x^{k,m}) - f(x^*)\big) + \sum_{j=1}^{r} \mu_j^{k,m} g_j(x^{k,m}) = 0. \tag{2.12}$$

Dividing both sides of the first relation in Eq. (2.4) by $\delta^{k,m}$, we get

$$\mu_0^{k,m} f(x^{k,m}) + \frac{1}{k^3\delta^{k,m}} \|x^{k,m} - x^*\|^2 + \sum_{j=1}^{r} \mu_j^{k,m} g_j(x^{k,m}) \le \mu_0^{k,m} f(x) + \sum_{j=1}^{r} \mu_j^{k,m} g_j(x) + \frac{1}{k\delta^{k,m}},$$

$$\forall\, x \in X^k.$$

Since the sequence $\{\mu_0^{k,m}, \mu_1^{k,m}, \ldots, \mu_r^{k,m}\}$ is bounded, it has a limit point (as $k \to \infty$ and $m \to \infty$), denoted by $\{\mu_0^*, \mu_1^*, \ldots, \mu_r^*\}$. Taking the limit along the relevant subsequences in the preceding relation together with Eq. (2.12) yields

$$\mu_0^* f(x^*) \le \mu_0^* f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x), \qquad \forall\, x \in X,$$

which implies that

$$\begin{aligned}
\mu_0^* f^* &\le \inf_{x\in X} \Big\{\mu_0^* f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x)\Big\} \\
&\le \inf_{x\in X,\ g_j(x)\le 0,} \Big\{\mu_0^* f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x)\Big\} \\
&\le \inf_{x\in X,\ g_j(x)\le 0} \mu_0^* f(x) \\
&= \mu_0^* f^*.
\end{aligned}$$

Thus, equality holds throughout, and we have

$$\mu_0^* f(x^*) = \inf_{x \in X} \bigl\{ \mu_0^* f(x) + \sum_{j=1}^r \mu_j^* g_j(x) \bigr\},$$

showing that $\mu_0^*, \ldots, \mu_r^*$ satisfy conditions (i), (ii), and (iii) of the proposition.

For each k, choose an index m_k such that

$$0 < \|x^{k,m_k} - x^*\| \le \frac{1}{k}, \quad |f(x^{k,m_k}) - f(x^*)| \le \frac{1}{k}, \quad \|g^+(x^{k,m_k})\| \le \frac{1}{k},$$

$$|\mu_j^{k,m_k} - \mu_j^*| \le \frac{1}{k}, \quad j = 1, \ldots, r.$$

Dividing both sides of Eq. (2.6) by δ^{k,m_k}, and using Eq. (2.11), we obtain

$$\lim_{m \to \infty} \frac{m_k g_j^+(x^{k,m_k})}{\delta^{k,m_k}} = \mu_j^*, \qquad j = 1, \ldots, r.$$

From Eq. (2.7), we also have

$$f(x^{k,m_k}) < f(x^*),$$

for all k sufficiently large (the case where $x^{k,m_k} = x^*$ for infinitely many k is excluded by the assumption that the set J is nonempty). Hence the sequence $\{x^{k,m_k}\}$ satisfies condition (iv) of the proposition as well, concluding the proof. **Q.E.D.**

We next provide a geometric interpretation for the proof of the preceding proposition. Consider the function $L_{k,m}$, introduced in the proof of the proposition,

$$L_{k,m}(x, \xi) = f(x) + \frac{1}{k^3}\|x - x^*\|^2 + \xi' g(x) - \frac{\|\xi\|^2}{2m}.$$

Note that the term $(1/k^3)\|x - x^*\|^2$ is introduced to ensure that x^* is a strict local minimum of the function $f(x) + (1/k^3)\|x - x^*\|^2$. In the following discussion, let us assume without loss of generality that f is strictly convex, so that this term can be ommitted from the definition of $L_{k,m}$.

For any nonnegative vector $u \in \Re^r$, we consider the following perturbed version of the original problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g(x) \le u, \\ & && x \in X^k = X \cap \bigl\{ x \bigm| \|x - x^*\| \le k \bigr\}, \end{aligned} \tag{2.13}$$

where u is a nonnegative vector in $\Re^r$. In this problem, the abstract set constraint is replaced by the bounded set X^k, which has the property that any $x \in X$ belongs to X^k for all k sufficiently large. We denote the optimal value of problem (2.13) by $p^k(u)$. For each k and m, the saddle point of the function $L_{k,m}(x, \xi)$, denoted by $(x^{k,m}, \xi^{k,m})$, can be characterized in terms of $p^k(u)$ as follows.

Because of the quadratic nature of $L_{k,m}$, the maximization of $L_{k,m}(x, \xi)$ over $\xi \geq 0$ for any fixed $x \in X^k$ yields

$$\xi_j = m g_j^+(x), \qquad j = 1, \ldots, r, \tag{2.14}$$

so that we have

$$\begin{aligned} L_{k,m}(x^{k,m}, \xi^{k,m}) &= \inf_{x \in X^k} \sup_{\xi \geq 0} \left\{ f(x) + \xi' g(x) - \frac{\|\xi\|^2}{2m} \right\} \\ &= \inf_{x \in X^k} \left\{ f(x) + \frac{m}{2} \|g^+(x)\|^2 \right\}. \end{aligned}$$

This minimization can be written as

$$\begin{aligned} L_{k,m}(x^{k,m}, \xi^{k,m}) &= \inf_{u \in \Re^r} \inf_{x \in X^k,\ g(x) \leq u} \left\{ f(x) + \frac{m}{2} \|g^+(x)\|^2 \right\} \\ &= \inf_{u \in \Re^r} \left\{ p^k(u) + \frac{m}{2} \|u^+\|^2 \right\}. \end{aligned} \tag{2.15}$$

The vector $u^{k,m} = g(x^{k,m})$ attains the infimum in the preceding relation. This minimization can be visualized geometrically as in Fig. 5.2.1. The point of contact of the functions $p^k(u)$ and $L_{k,m}(x^{k,m}, \xi^{k,m}) - m/2\|u^+\|^2$ corresponds to the vector $u^{k,m}$ that attains the infimum in Eq. (2.15).

We can also interpret $\xi^{k,m}$ in terms of the function p^k. In particular, the infimum of $L_{k,m}(x, \xi^{k,m})$ over $x \in X^k$ is attained at $x^{k,m}$, implying that

$$\begin{aligned} f(x^{k,m}) + \xi^{k,m\prime} g(x^{k,m}) &= \inf_{x \in X^k} \bigl\{ f(x) + \xi^{k,m\prime} g(x) \bigr\} \\ &= \inf_{u \in \Re^r} \bigl\{ p^k(u) + \xi^{k,m\prime} u \bigr\}. \end{aligned}$$

Replacing $u^{k,m} = g(x^{k,m})$ in the preceding, and using the fact that $x^{k,m}$ is feasible for problem (2.13) with $u = u^{k,m}$, we obtain

$$p^k(u^{k,m}) \leq f(x^{k,m}) = \inf_{u \in \Re^r} \bigl\{ p^k(u) + \xi^{k,m\prime} (u - u^{k,m}) \bigr\}.$$

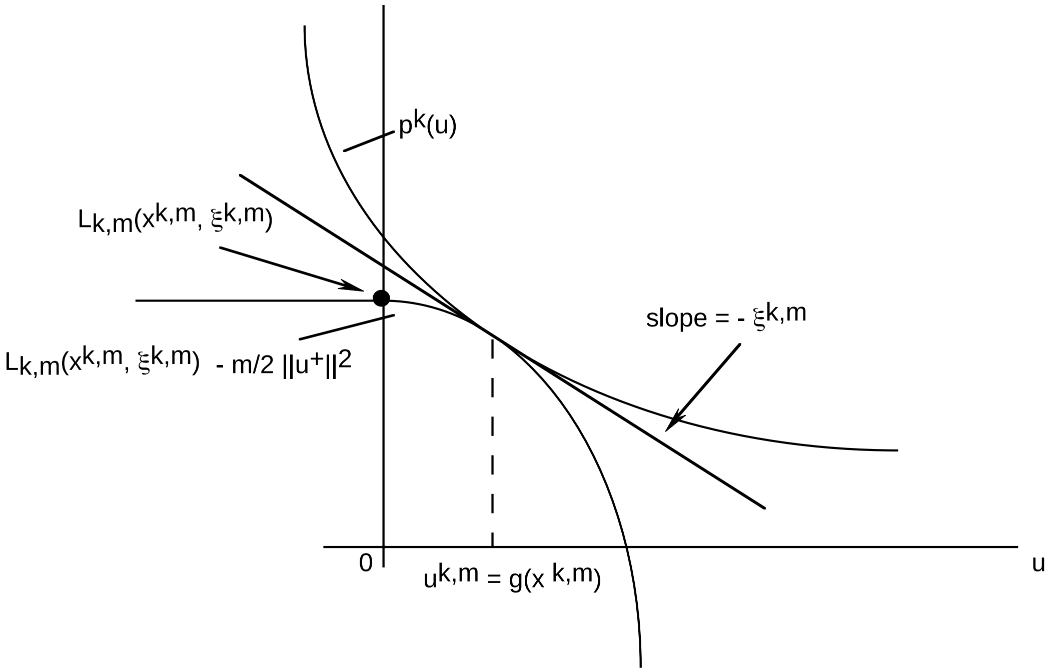


Figure 5.2.1. Illustration of the saddle points of the function $L_{k,m}(x, \xi)$ over $x \in X^k$ and $\xi \geq 0$ in terms of the function $p^k(u)$, which is the optimal value of problem (2.13) as a function of u.

Thus, we see that

$$p^k(u^{k,m}) \leq p^k(u) + \xi^{k,m\prime}(u - u^{k,m}), \qquad \forall\, u \in \Re^r,$$

which, by the definition of the subgradient of a function, implies that

$$-\xi^{k,m} \in \partial p^k(u^{k,m}),$$

(cf. Fig. 5.2.1). It can be seen from this interpretation that, the limit of $L_{k,m}(x^{k,m}, \xi^{k,m})$ as $m \to \infty$ is equal to $p^k(0)$, which is equal to $f(x^*)$ for each k. The limit of the normalized sequence

$$\left\{ \frac{(1, \xi^{k,m})}{\sqrt{1 + \|\xi^{k,m}\|^2}} \right\}$$

as $k \to \infty$ and $m \to \infty$ yields the Fritz John multipliers that satisfy conditions (i)-(iii) of the proposition, and the sequence $\{x^{k,m}\}$ is used to construct the sequence that satisfies condition (iv) of the proposition.

The next example demonstrates the saddle points of the function $L_{k,m}(x, \xi)$ and the effect of using a bounded approximation of the abstract set constraint [cf. problem (2.13)] on the original problem.

Example 5.2.1:

Consider the two-dimensional problem

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x_1 \leq 0, \qquad x \in X = \{x \mid x \geq 0\}, \end{aligned}$$

where

$$f(x) = e^{-\sqrt{x_1 x_2}}, \qquad \forall\, x \in X.$$

It can be seen that f is convex and closed. Since for feasibility, we must have $x_1 = 0$, we see that $f(x^*) = 1$. Consider the following perturbed version of the problem

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x_1 \leq u, \qquad x \in X = \{x \mid x \geq 0\}. \end{aligned}$$

The optimal value of this problem, which we denote by $p(u)$, is given by

$$p(u) = \inf_{x \in X} \sup_{\mu \geq 0} \big\{ e^{-\sqrt{x_1 x_2}} + \mu(x_1 - u) \big\} = \begin{cases} \infty & \text{if } u < 0, \\ 1 & \text{if } u = 0, \\ 0 & \text{if } u > 0, \end{cases}$$

(cf. Fig. 5.2.2). Thus, even though $p(0)$ is finite, p is not lower semicontinuous at 0. We also consider the following problem,

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x_1 \leq u, \qquad x \in X^k = \{x \mid x \geq 0,\; \|x\| \leq k\}, \end{aligned}$$

where the abstract set constraint is approximated by a compact set X^k around $x^* = (0, 0)$. The approximation is parameterized by the nonnegative scalar k and becomes increasingly accurate as $k \to \infty$. The optimal value of this problem, denoted by $p^k(u)$, is given by

$$p^k(u) = \inf_{x \in X^k} \sup_{\mu \geq 0} \big\{ e^{-\sqrt{x_1 x_2}} + \mu(x_1 - u) \big\} = \begin{cases} \infty & \text{if } u < 0, \\ e^{-(u^2 k^2 - u^4)^{\frac{1}{4}}} & \text{if } u \geq 0. \end{cases}$$

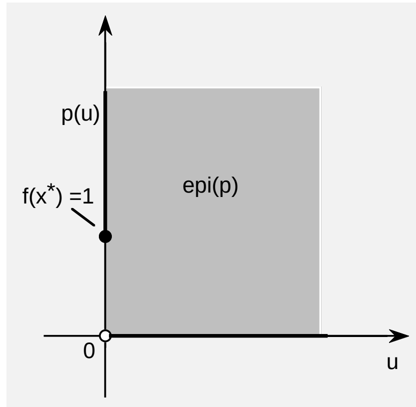


Figure 5.2.2. The function p for Example 5.2.1:

$$p(u) = \begin{cases} \infty & \text{if } u < 0, \\ 1 & \text{if } u = 0, \\ 0 & \text{if } u > 0. \end{cases}$$

Here p is not lower semicontinuous at 0.

Note that p^k is lower semicontinuous at 0. Hence, the compactification has the effect of regularizing the function $p(u)$ around $u = 0$ by using the approximation $p^k(u)$, which is lower semicontinuous at 0.

Figure 5.2.3 illustrates the function $p^k(u)$ and the quadratic term $-m/2\|u^+\|^2$ for different values of k and m. For each fixed k, it can be seen that $L_{k,m}(x^{k,m}, \xi^{k,m}) \to f(x^*)$, $g(x^{k,m}) = x_1^{k,m} \to 0$ and $f(x^{k,m}) \to f(x^*)$ as $m \to \infty$.

5.3 GEOMETRIC MULTIPLIERS AND DUALITY

We consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g_j(x) \le 0, \ j = 1, \ldots, r, \end{aligned} \tag{3.1}$$

where $f : \Re^n \mapsto \Re$, $g_j : \Re^n \mapsto \Re$, $j = 1, \ldots, r$, are given functions, and X is a nonempty subset of $\Re^n$, and we use the notation

$$g(x) = \big(g_1(x), \ldots, g_r(x)\big),$$

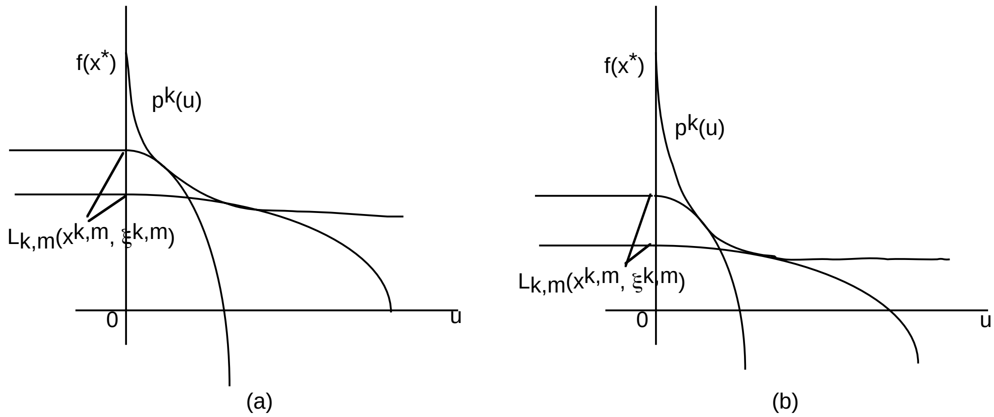


Figure 5.2.3. Illustration of the function $p^k(u)$ and the quadratic term $-m/2\|u^+\|^2$ for different values of k and m. The figure in (a) corresponds to the case where $k = 2$, and $m = 1,\ 10$, whereas the figure in (b) corresponds to the case where $k = 50$, and $m = 1,\ 10$.

for the constraint functions. We denote the optimal value of problem (3.1) by f^*, and assume throughout this section that $-\infty < f^* < \infty$, i.e., the problem is feasible and the cost function is bounded from below over the constraint set. Again, for clarity and simplicity of presentation, we only consider inequality constraints and note that the extension to problems with equality constraints is straightforward.

We have the following notion of a multiplier vector, that is not tied to a specific optimal solution, and does not require any differentiability assumptions on the cost and constraint functions.

Definition 5.3.3: We say that *there exists a geometric multiplier vector* (or simply a *geometric multiplier*) for problem (3.1) if there exists a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*) \geq 0$ that satisfies

$$f^* = \inf_{x \in X} L(x, \mu^*), \tag{3.2}$$

where $L(x, \mu^*) = f(x) + {\mu^*}' g(x)$ denotes the Lagrangian function.

To get the intuition behind the preceding definition, assume that problem (3.1) has an optimal solution x^*, and the functions f and the g_j are smooth. Then, by using the necessary optimality condition given in Chapter 2 and the definition of a geometric multiplier, we obtain

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

$$\mu_j^* g_j(x^*) = 0, \qquad \forall\, j = 1, \ldots, r.$$

Hence, the geometric multiplier is the vector that renders the Lagrangian function stationary and satisfies the complementary slackness condition, hence is the Lagrange multiplier for problem (3.1) under these assumptions.

The geometric multiplier can be visualized using hyperplanes in the constraint-cost space. In particular, it can be seen that geometric multipliers correspond to slopes of nonvertical hyperplanes that support the set of constraint-cost pairs as x ranges over the set X, denoted by S,

$$S = \Big\{ \big(g(x), f(x)\big) \;\big|\; x \in X \Big\},$$

(cf. Fig. 5.3.4).

However, it may not always be possible to find a vector μ^* that satisfies Eq. (3.2). Figure 5.3.5 shows some examples where there exist no geometric multipliers. Therefore, a natural question is to find conditions under which problem (3.1) has at least one geometric multiplier. One of our goals in this chapter is to develop an approach that addresses this question under convexity assumptions, through using Fritz John-type optimality conditions.

5.3.1. Relation between Geometric and Lagrange Multipliers

As indicated in the previous section, there is a strong connection between geometric and Lagrange multipliers for problems with a convex structure. Consider the following assumption:

Assumption 5.3.3: The set X is nonempty closed and convex, and the functions f and g_j are real-valued and convex over $\Re^n$.

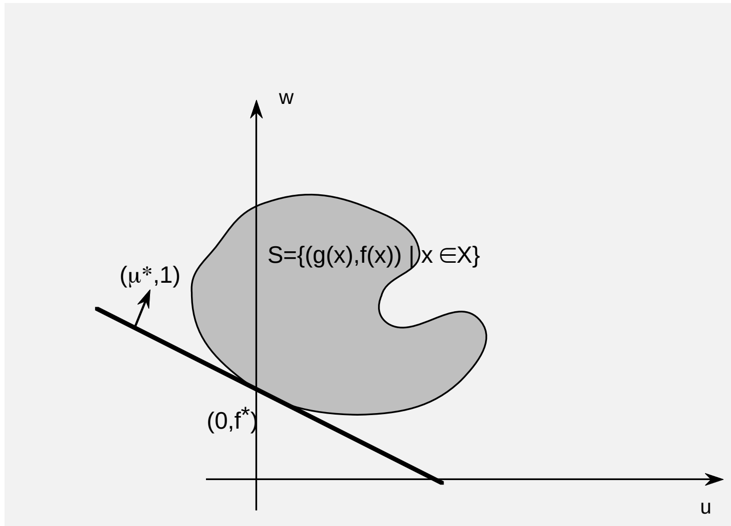


Figure 5.3.4. Geometrical interpretation of a geometric multiplier in terms of hyperplanes supporting the set

$$S = \Big\{ \big(g(x), f(x) \big) \ \Big|\ x \in X \Big\}.$$

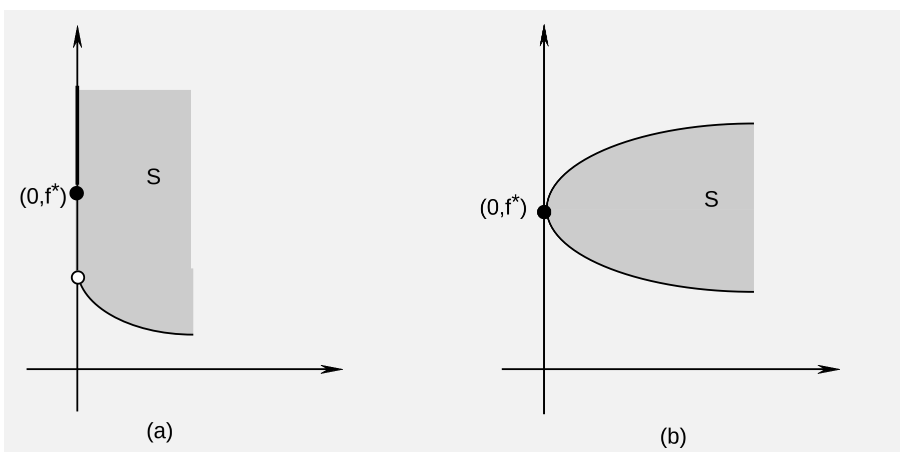


Figure 5.3.5. Examples where there are no geometric multipliers.

Before presenting the connection between geometric and Lagrange multipliers, we recall the definition of Lagrange multipliers, given in Section 5.1.

Definition 5.3.4: Consider problem (3.1), and let x^* be a global minimum. A vector μ^* is called a *Lagrange multiplier vector* if

$$0 \in \partial f(x^*) + \sum_{j=1}^{r} \mu_j^* \partial g_j(x^*) + T_X(x^*)^*, \tag{3.3}$$

$$\mu^* \geq 0, \qquad \mu^{*\prime} g(x^*) = 0. \tag{3.4}$$

Proposition 5.3.3: Consider problem (3.1) under Assumption 5.3.3. Assume further that problem (3.1) has at least one optimal solution x^*. Then, the set of Lagrange multipliers associated with x^* is the same as the set of geometric multipliers.

Proof: If μ^* is a geometric multiplier, we have $\mu^* \geq 0$. By definition (3.2), we have

$$f(x^*) \leq f(x) + {\mu^*}' g(x), \qquad \forall\, x \in X,$$

which, since $x^* \in X$ and $g(x^*) \leq 0$, implies that

$$f(x^*) \leq f(x^*) + {\mu^*}' g(x^*) \leq f(x^*),$$

form which we obtain $\mu^{*\prime} g(x^*) = 0$, i.e., CS condition holds. The preceding also implies that x^* minimizes $L(x, \mu^*)$ over X, so by using the necessary conditions of Chapter 2, we get

$$0 \in \partial f(x^*) + \sum_{j=1}^{r} \mu_j^* \partial g_j(x^*) + T_X(x^*)^*.$$

Thus all the conditions of Definition 5.3.4 are satisfied and μ^* is a Lagrange multiplier.

Conversely, if μ^* is a Lagrange multiplier, Definition 5.3.4 and the convexity assumptions imply that x^* minimizes $L(x, \mu^*)$ over X, so using also the CS condition $[\mu^{*\prime} g(x^*) = 0]$, we have

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

Since $\mu^* \geq 0$ as well, it follows that μ^* is a geometric multiplier. **Q.E.D.**

Prop. 5.3.3 implies that for a convex problem that has multiple optimal solutions, all the optimal solutions have the same set of Lagrange multipliers, which is the same as the set of geometric multipliers. However, even for convex problems, the notions of geometric and Lagrange multipliers are different. In particular, there may exist geometric multipliers, but no optimal solution and hence no Lagrange multipliers. As an example, consider the one-dimensional convex problem of minimizing e^{-x} subject to the single inequality constraint $x \geq 0$; it has the optimal value $f^* = 0$ and the geometric multiplier $\mu^* = 0$, but it has no optimal solution, and therefore no Lagrange multipliers.

Note that if problem (3.1) has at least one optimal solution that is pseudonormal, then a Lagrange multiplier is guaranteed to exist by the theory of Chapter 4. Under Assumption 5.3.3, this Lagrange multiplier is geometric by Prop. 5.3.3. Recall from Section 5.1 two criteria that relate to pseudonormality in the convex case and guarantee the existence of at least one Lagrange multiplier.

(a) *Linearity criterion*: X is a polyhedron, and the functions g_j are affine.

(b) *Slater criterion*: X is convex, the functions $g_j : X \mapsto \Re$ are convex, and there exists a feasible vector $\bar{x}$ such that

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

Thus using Prop. 5.3.3, we obtain the following proposition.

Proposition 5.3.4: Consider problem (3.1) under Assumption 5.3.3. Assume further that problem (3.1) has at least one optimal solution x^*. Then under either the linearity criterion or the Slater criterion there exists at least one geometric multiplier.

In Section 5.4, we will derive conditions on the constraint set that guarantee existence of a geometric multiplier, without requiring that the problem has an optimal solution.

5.3.2. Dual Optimization Problem

In this section, we show that conditions related to existence of geometric multipliers give information about a 'dual optimization problem'. We consider problem (3.1) and introduce the related optimization problem

$$\begin{aligned} &\text{maximize } q(\mu) \\ &\text{subject to } \ \mu \geq 0, \end{aligned} \tag{3.5}$$

where

$$q(\mu) = \inf_{x \ inX} \big\{ f(x) + \mu' g(x) \big\}.$$

We call problem (3.5) the *dual problem* and denote its optimal value by q^*. It is well-known that regardless of the structure of the original problem (3.1), the dual problem has nice convexity properties, given in the following proposition. (For the proof, see [BNO02].)

Proposition 5.3.5: The function q is concave and upper semicontinuous over $\Re^r$.

The optimal values of the dual problem and problem (3.1) satisfy the following relation:

Proposition 5.3.6: (Weak Duality Theorem) We always have

$$q^* \leq f^*.$$

If $q^* = f^*$, we say that *there is no duality gap*, or that *strong duality holds*. The next proposition shows that the existence of a geometric multiplier is related to the no duality gap condition. (For the proof, see [Ber99].)

Proposition 5.3.7: Assume that $-\infty < f^* < \infty$. There exists a geometric multiplier if and only if there is no duality gap and the dual problem has an optimal solution.

5.4 OPTIMALITY CONDITIONS FOR PROBLEMS WITH NO OPTIMAL SOLUTION

In this section, we consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g(x) \leq 0, \end{aligned} \tag{4.1}$$

where $g(x) = \big(g_1(x), \ldots, g_r(x)\big)$, under various convexity assumptions, and we focus on Fritz John-type of optimality conditions. For simplicity, we assume no equality constraints. The following analysis extends to the case where we have affine equality constraints, by replacing each equality constraint by two affine (and hence convex) inequality constraints. We denote the optimal value of problem (4.1) by f^*, and we assume that $-\infty < f^* < \infty$, i.e., the problem is feasible and the cost function is bounded from below over the constraint set.

We have already derived in Section 5.2 Fritz John conditions in the case where there exists an optimal solution x^*. These conditions were shown in their enhanced form, which includes the CV condition and relates to the notion of pseudonormality.

Our goal in this section is to derive Fritz John optimality conditions without being tied to a specific optimal solution. In fact we allow problem (4.1) to have no optimal solution at all. The next proposition presents Fritz John conditions in their classical form, i.e., the corresponding Fritz John multipliers satisfy the CS condition as opposed to CV-type conditions. (We include the proof of this proposition here for the sake of completeness of the analysis.)

Proposition 5.4.8: (Fritz John Conditions in Classical Form) Consider problem (4.1), and assume that X is convex, the functions f and g_j are convex over X, and $-\infty < f^* < \infty$. Then there exists a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, satisfying the following conditions:

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

(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.

Proof: Consider the subset of $\Re^{r+1}$ given by

$$M = \big\{ (u_1, \ldots, u_r, w) \mid \text{there exists } x \in X \text{ such that}$$
$$g_j(x) \leq u_j,\ j = 1, \ldots, r,\ f(x) \leq w \big\},$$

(cf. Fig. 5.4.6). We first show that M is convex. To this end, we consider vectors $(u, w) \in M$ and $(\tilde{u}, \tilde{w}) \in M$, and we show that their convex combinations lie in M. The definition of M implies that for some $x \in X$ and $\tilde{x} \in X$, we have

$$f(x) \leq w, \qquad g_j(x) \leq u_j, \qquad j = 1, \ldots, r,$$

$$f(\tilde{x}) \leq \tilde{w}, \qquad g_j(\tilde{x}) \leq \tilde{u}_j, \qquad j = 1, \ldots, r.$$

For any $\alpha \in [0, 1]$, we multiply these relations with α and $1 - \alpha$, respectively, and add. By using the convexity of f and g_j, we obtain

$$f\big(\alpha x + (1-\alpha)\tilde{x}\big) \leq \alpha f(x) + (1-\alpha) f(\tilde{x}) \leq \alpha w + (1-\alpha)\tilde{w},$$

$$g_j\big(\alpha x + (1-\alpha)\tilde{x}\big) \leq \alpha g_j(x) + (1-\alpha) g_j(\tilde{x}) \leq \alpha u_j + (1-\alpha)\tilde{u}_j, \quad j = 1, \ldots, r.$$

In view of the convexity of X, we have $\alpha x + (1-\alpha)\tilde{x} \in X$, so these equations imply that the convex combination of (u, w) and $(\tilde{u}, \tilde{w})$, i.e., $\big(\alpha u + (1-\alpha)\tilde{u}, \alpha w + (1-\alpha)\tilde{w}\big)$, belongs to M, proving the convexity of M.

We next note that $(0, f^*)$ is not an interior point of M; otherwise, for some $\epsilon > 0$, the point $(0, f^* - \epsilon)$ would belong to M, contradicting the definition of f^* as the optimal

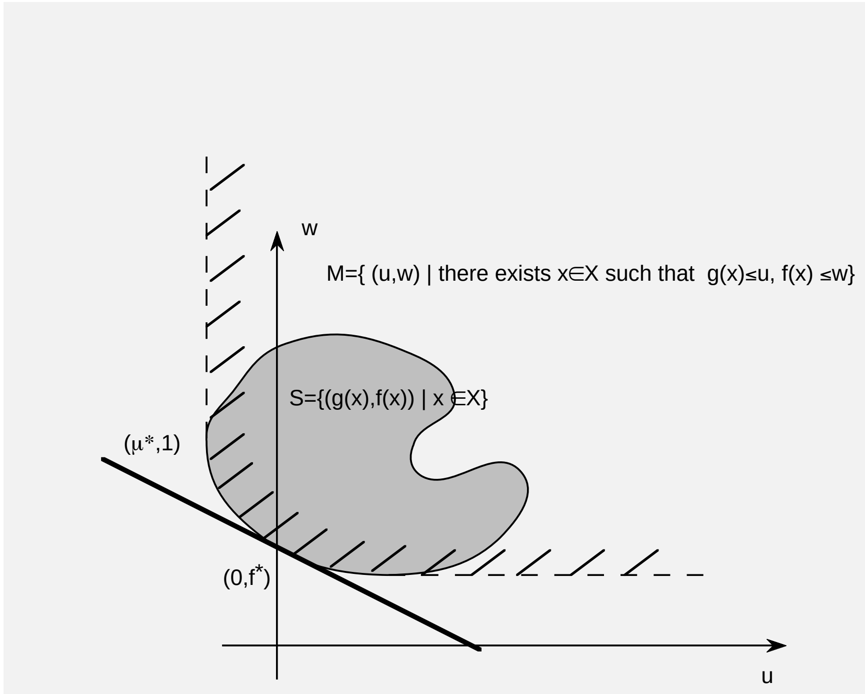


Figure 5.4.6. Illustration of the set

$$S = \Big\{ \big(g(x), f(x) \big) \mid x \in X \Big\}$$

and the set

$$M = \Big\{ (u_1, \ldots, u_r, w) \mid \text{there exists } x \in X \text{ such that} \\ g_j(x) \le u_j,\ j = 1, \ldots, r,\ f(x) \le w \Big\},$$

used in the proof of Prop. 5.4.8. The idea of the proof is to show that M is convex and that $(0, f^*)$ is not an interior point of M. A hyperplane passing through $(0, f^*)$ and supporting M is used to define the Fritz John multipliers.

value of problem (4.1). Therefore, by the supporting hyperplane theorem (cf. Section 2.1), there exists a hyperplane passing through $(0, f^*)$ and containing M in one of the two corresponding closed halfspaces. In particular, there exists a vector $(\mu^*, \mu_0^*) \neq (0, 0)$ such that

$$\mu_0^* f^* \le \mu_0^* w + \mu^{*\prime} u, \qquad \forall\ (u, w) \in M. \tag{4.2}$$

This equation implies that

$$\mu_0^* \ge 0, \qquad \mu_j^* \ge 0, \qquad \forall\ j = 1, \ldots, r,$$

since for each $(u, w) \in M$, we have that $(u, w + \gamma) \in M$ and $(u_1, \ldots, u_j + \gamma, \ldots, u_r, w) \in M$ for all $\gamma > 0$ and $j = 1, \ldots, r$.

Finally, since for all $x \in X$, we have $\big(g(x), f(x)\big) \in M$, Eq. (4.2) implies that

$$\mu_0^* f^* \leq \mu_0^* f(x) + \mu^{*\prime} g(x), \qquad \forall\, x \in X.$$

Taking the infimum over all $x \in X$, it follows that

$$\begin{aligned}
\mu_0^* f^* &\leq \inf_{x \in X} \big\{\mu_0^* f(x) + \mu^{*\prime} g(x)\big\} \\
&\leq \inf_{x \in X,\ g(x) \leq 0} \big\{\mu_0^* f(x) + \mu^{*\prime} g(x)\big\} \\
&\leq \inf_{x \in X,\ g(x) \leq 0} \mu_0^* f(x) \\
&= \mu_0^* f^*
\end{aligned}$$

Hence, equality holds throughout above, proving the desired result. **Q.E.D.**

Note that, in the above proposition, if it can be guaranteed that the scalar μ_0^* can be taken to be nonzero (and without loss of generality equal to 1), then the remaining scalars $\mu_1^*, \ldots, \mu_r^*$ constitute a geometric multiplier.

We next prove a stronger Fritz John theorem for problems with *linear constraints*, i.e., the constraint functions g_j are affine. Our ultimate goal is to use the preceding proposition (Fritz John conditions in their classical form) and the next proposition (Fritz John conditions for linear constraints) to show the existence of geometric multipliers under various constraint qualifications without requiring that the problem has an optimal solution. Using Prop. 5.3.7, this also provides conditions under which there is no duality gap and the dual problem has an optimal solution. Hence, this line of analysis provides an alternative pathway to obtain the strong duality results of convex programming.

Proposition 5.4.9: (Fritz John Conditions for Linear Constraints) Consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g_j(x) \le 0, \ j = 1, \ldots, r, \end{aligned} \tag{4.3}$$

where X is a nonempty convex set, the function $f : X \mapsto \Re$ is convex and the functions $g_j : X \mapsto \Re$ are affine, and $-\infty < f^* < \infty$. Then there exists a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, satisfying the following conditions:

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

(ii) $\mu_j^* \ge 0$ 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 vector $\overline{x} \in X$ such that

$$\mu^{*\prime} g(\overline{x}) > 0.$$

Proof: If $f^* = \inf_{x \in X} f(x)$, then set $\mu_0^* = 1$, and $\mu^* = 0$, and we are done. Hence, assume that there exists some $\overline{x} \in X$ with $f(\overline{x}) < f^*$. Consider the convex sets defined by

$$C_1 = \big\{ (x, w) \mid \text{there is a vector } x \in X, \ f(x) < w \big\},$$

$$C_2 = \big\{ (x, f^*) \mid g(x) \le 0 \big\}.$$

The set C_2 is polyhedral. Also C_1 and C_2 are disjoint. To see this, note that if $(x, f^*) \in C_1 \cap C_2$, then by the definition of C_2, this would imply that $x \in X$, $g(x) \le 0$, and $f(x) < f^*$, contradicting the fact that f^* is the optimal value of problem (4.3).

Hence, by the polyhedral proper separation theorem (cf. Section 2.1), there exists a hyperplane that separates C_1 and C_2 and does not contain C_1, i.e., a vector (ξ, μ_0^*) such that

$$\mu_0^* f^* + \xi' z \le \mu_0^* w + \xi' x, \qquad \forall\ x \in X \text{ with } f(x) < w, \ g(z) \le 0, \tag{4.4}$$

$$\inf_{(x,w) \in C_1} \{ \mu_0^* w + \xi' x \} < \sup_{(x,w) \in C_1} \{ \mu_0^* w + \xi' x \}.$$

The preceding two relations also imply that

$$\mu_0^* f^* + \sup_{g(z)\le 0} \xi' z < \sup_{(x,w)\in C_1} \{\mu_0^* w + \xi' x\}. \tag{4.5}$$

Note that Eq. (4.4) imply that $\mu_0^* \ge 0$, otherwise it would be possible to increase w to ∞ and break the inequality.

Next, we focus on the linear program

$$\text{maximize } \xi' z$$
$$\text{subject to } \; g(z) \le 0,$$

where $g(z) = Az - b$. By Eq. (4.4), this linear program is bounded (since the set C_1 is nonempty), and therefore has an optimal solution, which we denote by z^*. The dual of this program is

$$\text{maximize } \; -b'\mu$$
$$\text{subject to } \; \xi = A'\mu, \quad \mu \ge 0.$$

By linear programming duality, it follows that this problem has a dual optimal solution $\mu^* \ge 0$, and satisfies

$$\xi' z^* = {\mu^*}' b, \qquad A'\mu^* = \xi, \tag{4.6}$$

which implies that

$${\mu^*}' A z^* = {\mu^*}' b.$$

From Eq. (4.4), we have

$$\mu_0^* f^* + \sup_{g(z)\le 0} \xi' z \le \mu_0^* w + \xi' x, \qquad \forall\, x \in X \text{ with } f(x) < w,$$

which by using the preceding relations imply that

$$\mu_0^* f^* \le \mu_0^* w + {\mu^*}'(Ax - b), \qquad \forall\, x \in X \text{ with } f(x) < w.$$

In particular, we have

$$\mu_0^* f^* \le \mu_0^* f(x) + {\mu^*}'(Ax - b). \qquad \forall\, x \in X,$$

from which we get

$$\begin{aligned}
\mu_0^* f^* &\leq \inf_{x\in X}\left\{\mu_0^* f(x) + \sum_{j=1}^r \mu_j^* g_j(x)\right\} \\
&\leq \inf_{x\in X,\ g(x)\leq 0}\left\{\mu_0^* f(x) + \sum_{j=1}^r \mu_j^* g_j(x)\right\} \\
&\leq \inf_{x\in X,\ g(x)\leq 0} \mu_0^* f(x) \\
&= \mu_0^* f^*
\end{aligned}$$

Hence, equality holds throughout above, which proves condition (i) of the proposition.

Substituting the relations in (4.6) in Eq. (4.5) yields

$$\mu_0^* f^* < \sup_{x\in X, f(x)<w} \big\{\mu_0^* w + {\mu^*}' g(x)\big\}.$$

If $\mu_0^* = 0$, we obtain

$$0 < \sup_{x\in X,\ f(x)<w} {\mu^*}' g(x),$$

thus showing condition (iv). Assume that $\mu_0^* > 0$. We claim that

$$\mu_0^* f^* < \sup_{x\in X,\ f(x)<w\leq f^*} \big\{\mu_0^* w + {\mu^*}' g(x)\big\}. \tag{4.7}$$

To show this, assume to arrive at a contradiction that

$$\mu_0^* f^* \geq \sup_{x\in X,\ f(x)<w\leq f^*} \big\{\mu_0^* w + {\mu^*}' g(x)\big\}. \tag{4.8}$$

From Eq. (4.4), we have

$$\mu_0^* f^* + \sup_{g(z)\leq 0} \xi' z \leq \mu_0^* w + \xi' x, \qquad \forall\ x\in X,\ f(x)<w,$$

Substituting the relations in (4.6) in the preceding relation, we obtain

$$\mu_0^* f^* \leq \mu_0^* w + {\mu^*}' g(x), \qquad \forall\ x\in X,\ f(x)<w.$$

In particular, this implies that

$$\mu_0^* f^* \leq \mu_0^* w + {\mu^*}' g(x), \qquad \forall\ x\in X,\ f(x)<w\leq f^*.$$

Combining with Eq. (4.8), it follows that

$$\mu_0^* w + \mu^{*\prime} g(x) = \mu_0^* f^*, \qquad \forall\, x \in X,\ \forall\, w \text{ such that } f(x) < w \le f^*. \tag{4.9}$$

By assumption, there exists some $\overline{x} \in X$ with $f(\overline{x}) < f^*$. Let $\epsilon = f^* - f(\overline{x}) > 0$, and $w = f(\overline{x}) + \epsilon/4$. Then, from Eq. (4.9), we have

$$\mu_0^* f^* = \mu_0^* \big(f(\overline{x}) + \epsilon/4\big) + \mu^{*\prime} g(\overline{x}).$$

Since $f(\overline{x}) + \epsilon/2 \le f^*$, we have, combining Eqs. (4.9) and the preceding relation, that

$$\mu_0^* f^* = \mu_0^* f^* + \mu_0 \epsilon/4,$$

which is a contradiction, showing that Eq. (4.8) holds. Hence, there exists some $\overline{x} \in C$, and $\overline{w}$ with $f(\overline{x}) < \overline{w} \le f^*$ such that

$$\mu_0^* f^* < \mu_0^* \overline{w} + \mu^{*\prime} g(\overline{x}),$$

which implies that

$$0 \le \mu_0^{*\prime} \big(f^* - \overline{w}\big) < \mu^{*\prime} g(\overline{x}),$$

thus showing condition (iv), and concluding the proof. **Q.E.D.**

Note that the proof of the preceding proposition relies on a special type of separation result for polyhedral sets. We will see in the next section that these separation arguments form the basis for bringing out the structure in the constraint set that guarantees the existence of geometric multipliers.

5.4.1 Existence of Geometric Multipliers

In this section, we use the Fritz John Theorems of Propositions 5.4.8 and 5.4.9 to assert the existence of geometric multipliers under some conditions. This development parallels our analysis of the relation of the constraint qualifications and the existence of Lagrange multipliers of Chapter 4. However, here we do not require that the problem has an optimal solution.

Before, we present these conditions, we show the following result, related to extended representations of the constraint set, which is analogous to Prop. 4.3.10 of Chapter 4. In particular, consider problem (4.1), and assume that the set X is partially described in terms of inequality constraints:

$$X = \big\{x \in \tilde{X} \mid g_j(x) \le 0,\ j = r+1, \ldots, \overline{r}\big\}.$$

Then the constraint set can alternatively be described as:

$$\big\{x \in \tilde{X} \mid g_j(x) \le 0,\ j = 1, \ldots, r, r+1, \ldots, \overline{r}\big\}.$$

We call this the *extended representation* of the constraint set (cf. Section 4.3), whereas we call the representation given in problem (4.1), the *original representation.* We have the following result, which relates issues about existence of geometric multipliers between the two representations.

Proposition 5.4.10: If there exist geometric multipliers in the extended representation, there exist geometric multipliers in the original representation.

Proof: The hypothesis implies that there exist nonnegative scalars $\mu_1^*, \ldots, \mu_r^*, \mu_{r+1}^*, \ldots, \mu_{\overline{r}}^*$ such that

$$f^* = \inf_{x \in \tilde{X}} \left\{ f(x) + \sum_{j=1}^{\overline{r}} \mu_j^* g_j(x) \right\}.$$

Since $X \subset \tilde{X}$, this implies that

$$f^* \le f(x) + \sum_{j=1}^{\overline{r}} \mu_j^* g_j(x), \qquad \forall\, x \in X.$$

For any $x \in X$, we have $g_j(x) \le 0$, for all $j = r+1, \ldots, \overline{r}$, so that $\mu_j^* g_j(x) \le 0$, for all $j = r+1, \ldots, \overline{r}$. Therefore, it follows from the preceding relation that

$$f^* \le f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x), \qquad \forall\, x \in X.$$

Taking the infimum over all $x \in X$, it follows that

$$\begin{aligned} f^* &\leq \inf_{x \in X} \left\{ f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \right\} \\ &\leq \inf_{x \in X,\ g_j(x) \leq 0} \left\{ f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \right\} \\ &\leq \inf_{x \in X,\ g_j(x) \leq 0} f(x) \\ &= f^* \end{aligned}$$

Hence, equality holds throughout above, showing that $\mu_1^* \dots, \mu_r^*$ constitute a geometric multiplier for the original representation. **Q.E.D.**

We will use this result when we are examining a problem with affine constraint functions and a polyhedral set constraint.

5.4.1.1 Convex Constraints

We consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g_j(x) \leq 0,\ j = 1, \dots, r, \end{aligned} \tag{4.10}$$

under the following assumption:

Assumption 5.4.4: (Slater Condition) The optimal value f^* of problem (4.10) is finite, the set X is convex, and the functions $f : X \mapsto \Re$ and $g_j : X \mapsto \Re$ are convex. Furthermore, there exists a vector $\overline{x} \in X$ such that $g_j(\overline{x}) < 0$ for all $j = 1, \dots, r$.

We have the following proposition.

Proposition 5.4.11: (Strong Duality Theorem - Convex Constraints) Let Assumption 5.4.4 hold for problem (4.10). Then, there exists a geometric multiplier.

Proof: Under the given assumptions, it follows from Prop. 5.4.8 that there exist nonnegative scalars $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$, not all of which are zero, such that

$$\mu_0^* f^* = \inf_{x \in X} \big\{ \mu_0^* f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \big\}.$$

We will show that under Slater Condition, μ_0^* can not be zero. Assume to arrive at a contradiction that μ_0^* is equal to zero. then it follows from the preceding relation that

$$0 \leq \sum_{j=1}^{r} \mu_j^* g_j(x), \qquad \forall\, x \in X. \tag{4.11}$$

By assumption, $\mu_j^* \geq 0$ for all $j = 1, \ldots, r$, and at least one of them is nonzero. Therefore, in view of the assumption that there exists some $\overline{x} \in X$ such that $g_j(\overline{x}) < 0$, for all j, we obtain

$$\sum_{j=1}^{r} \mu_j^* g_j(\overline{x}) < 0,$$

thus contradicting Eq. (4.11), and showing that $\mu_0^* > 0$. Without loss of generality, we can assume that $\mu_0^* = 1$, and the remaining $\mu_1^*, \ldots, \mu_r^*$ constitute a geometric multiplier. **Q.E.D.**

5.4.1.2 Linear Constraints

We consider the linearly-constrained problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \qquad a_j' x - b_j \leq 0, \; j = 1, \ldots, r. \end{aligned} \tag{4.12}$$

Without loss of generality, we assume that there are no equality constraints (each equality constraint can be converted into two inequality constraints). We have the following assumption:

Assumption 5.4.5: (Linear Constraints) The optimal value f^* of problem (4.12) is finite, and the following hold:

(1) The set X is the intersection of a polyhedral set P and a convex set C.

(2) The cost function $f : \Re^n \mapsto \Re$ is convex over C.

(3) There exists a feasible solution of the problem that belongs to the relative interior of C.

Proposition 5.4.12: (Strong Duality Theorem - Linear Constraints) Let Assumption 5.4.5 hold for problem (4.12). Then, there exists a geometric multiplier.

Proof: Let $X = P \cap C$, where P is a polyhedral set expressed in terms of linear inequalities as

$$P = \{x \mid a_j' x - b_j \le 0,\ j = r+1, \ldots, p\},$$

for some integer $p > r$. We consider the following extended representation of the constraint set,

$$\{x \in C \mid a_j' x - b_j \le 0,\ j = 1, \ldots, p\}.$$

By Prop. 5.4.9, there exist nonnegative scalars $\mu_0^*, \mu_1^*, \ldots, \mu_p^*$, not all of which are zero, such that

$$\mu_0^* f^* = \inf_{x \in C} \left\{ \mu_0^* f(x) + \sum_{j=1}^{p} \mu_j^* (a_j' x - b_j) \right\}. \tag{4.13}$$

Furthermore, if the index set $J = \{j \neq 0 \mid \mu_j^* > 0\}$ is nonempty, there exists a vector $\overline{x} \in C$ such that

$$\sum_{j=1}^{p} \mu_j^* (a_j' \overline{x} - b_j) > 0. \tag{4.14}$$

We will show that under Assumption 5.4.5, μ_0^* can not be zero. Assume to arrive at a contradiction that μ_0^* is equal to zero. Then it follows from Eq. (4.13) that

$$0 \leq \sum_{j=1}^{p} \mu_j^*(a_j'x - b_j), \qquad \forall\, x \in C. \tag{4.15}$$

By assumption, there exists some $\tilde{x} \in \mathrm{ri}(C)$, that satisfies $a_j'\tilde{x} - b_j \leq 0$, for all $j = 1, \ldots, p$. Combining with the preceding relation, we obtain

$$0 \leq \sum_{j=1}^{p} \mu_j^*(a_j'\tilde{x} - b_j) \leq 0.$$

Hence, the function $\sum_{j=1}^{p} \mu_j^*(a_j'x - b_j)$ attains its minimum over $x \in C$ at some relative interior point of C, implying that

$$\sum_{j=1}^{p} \mu_j^*(a_j'x - b_j) = 0, \qquad \forall\, x \in C.$$

But this contradicts Eq. (4.14), showing that $\mu_0^* > 0$, and therefore, the scalars $\mu_1^*, \ldots, \mu_p^*$ constitute a geometric multiplier for the extended representation of problem (4.12). By Prop. 5.4.10, this implies that there exists a geometric multiplier in the original representation of problem (4.12) as well. **Q.E.D.**

5.4.2. Enhanced Primal Fritz John Conditions

The Fritz John conditions of Propositions 5.4.8 and 5.4.9 are weaker than the ones that we have encountered in the preceding sections in that they do not include conditions analogous to the CV condition, which formed the basis for the notion of pseudonormality and our analysis of Chapter 4. A natural form of this condition would assert the existence of a sequence $\{x^k\} \subset X$ such that

$$\lim_{k\to\infty} f(x^k) = f^*, \qquad \limsup_{k\to\infty} g(x^k) \leq 0, \tag{4.16}$$

and for all k,

$$f(x^k) < f^*, \qquad g_j(x^k) > 0, \quad \forall\, j \text{ with } \mu_j^* > 0, \tag{4.17}$$

$$g_j^+(x^k) = o\left(\min_{j\in J} g_j^+(x^k)\right), \qquad \forall\, j \text{ with } \mu_j^* = 0, \tag{4.18}$$

(assuming that $\mu^* \neq 0$). Unfortunately, such a condition does not hold in the absence of additional assumptions, as can be seen in the following example.

Example 5.4.2

Consider the one-dimensional problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g(x) = x \le 0, \ x \in X = \{x \mid x \ge 0\}, \end{aligned}$$

where

$$f(x) = \begin{cases} -1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x < 0. \end{cases}$$

Then f is convex over X and the assumptions of Prop. 5.4.8 are satisfied. Indeed the Fritz John multipliers that satisfy conditions (i)-(iii) of Prop. 5.4.8 must have the form $\mu_0^* = 0$ and $\mu^* > 0$ (cf. Fig. 5.4.7). However, here we have $f^* = 0$, and for all x with $g(x) > 0$, we have $x > 0$ and $f(x) = -1$. Thus, there is no sequence $\{x^k\} \subset X$ satisfying conditions (4.16) and (4.17) simultaneously.

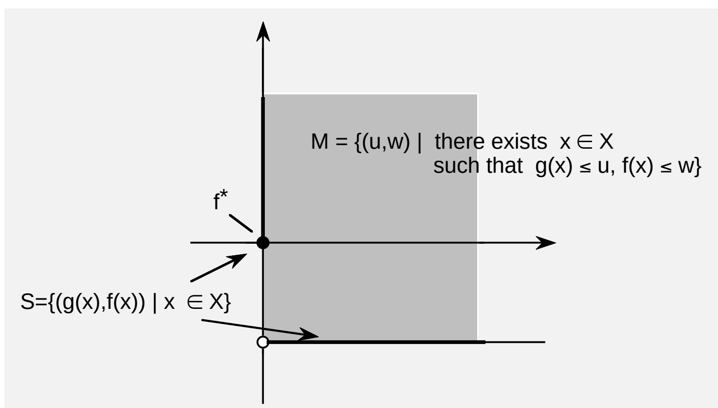


Figure 5.4.7. Illustration of the set

$$S = \Big\{ \big(g(x), f(x)\big) \mid x \in X \Big\}$$

and the set

$$M = \Big\{ (u, w) \mid \text{there exists } x \in X \text{ such that } g(x) \le u, \ f(x) \le w \Big\},$$

for Example 5.4.2.

The following proposition imposes slightly stronger assumptions in order to derive an enhanced set of Fritz John conditions. In particular we assume the following:

Assumption 5.4.6: The set X is nonempty and convex, and the functions f and g_j, viewed as functions from X to $\Re$, are closed and convex. Furthermore, $-\infty < f^* < \infty$.

We have the following proposition.

Proposition 5.4.13: (Enhanced Fritz John Conditions) Consider problem (4.1) under Assumption 5.4.6. Then there exists a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, satisfying the following conditions:

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

(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 $\{x^k\} \subset X$ such that

$$\lim_{k \to \infty} f(x^k) = f^*, \qquad \limsup_{k \to \infty} g(x^k) \leq 0, \tag{4.19}$$

and for all k,

$$\mu_j^* g_j(x^k) > 0, \quad \forall\, j \in J, \tag{4.20}$$

$$g_j^+(x^k) = o\left(\min_{j \in J} g_j^+(x^k) \right), \qquad \forall\, j \notin J. \tag{4.21}$$

Proof: If $f(x) \geq f^*$ for all $x \in X$, then we set $\mu_0^* = 1$ and $\mu^* = 0$, and we are done. We will thus assume that there exists some $\overline{x} \in X$ such that $f(\overline{x}) < f^*$. Consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \ \ g(x) \leq 0, \quad x \in X^k, \end{aligned} \tag{4.22}$$

where

$$X^k = X \cap \left\{ x \;\middle|\; \|x\| \leq k \right\}.$$

We assume without loss of generality that for all $k \geq 1$, the constraint set is nonempty. Since the constraint set of this problem, $\{x \in X \mid g(x) \leq 0\} \cap \{x \mid \|x\| \leq k\}$ is compact, and f is lower semicontinuous over X, this problem has an optimal solution, which we denote by $\overline{x}^k$. Since this is a more constrained problem than the original, we have $f^* \leq f(\overline{x}^k)$ and $f(\overline{x}^k) \downarrow f^*$ as $k \to \infty$. Let $\gamma^k = f(\overline{x}^k) - f^*$. Note that if $\gamma^k = 0$ for some k, then it follows that $\overline{x}^k$ is an optimal solution for problem (4.1), and the result follows by the enhanced Fritz John conditions for convex problems with an optimal solution (cf. Prop. 5.2.2). Therefore, we assume that $\gamma^k > 0$ for all k.

For positive integers k and positive scalars m, we consider the function

$$L_{k,m}(x,\xi) = f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \overline{x}^k\|^2 + \xi' g(x) - \frac{\|\xi\|^2}{2m},$$

and we note that $L_{k,m}$ is convex in x, and concave and coercive in ξ. Since f and g_j are closed and convex, they are closed, convex, and coercive when restricted to X^k. Hence, we can use the saddle point theorem to assert that $L_{k,m}$ has a saddle point over $x \in X^k$ and $\xi \geq 0$, which we denote by $(x^{k,m}, \xi^{k,m})$.

The infimum of $L_{k,m}(x, \xi^{k,m})$ over $x \in X^k$ is attained at $x^{k,m}$, implying that

$$\begin{aligned} f(x^{k,m}) &+ \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x^{k,m}) \\ &= \inf_{x \in X^k} \left\{ f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x) \right\} \\ &\leq \inf_{x \in X^k,\, g(x) \leq 0} \left\{ f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x) \right\} \\ &\leq \inf_{x \in X^k,\, g(x) \leq 0} \left\{ f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \overline{x}^k\|^2 \right\} \\ &= f(\overline{x}^k). \end{aligned} \tag{4.23}$$

Hence, we have

$$\begin{aligned} L_{k,m}(x^{k,m}, \xi^{k,m}) &= f(x^{k,m}) + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x^{k,m}) - \frac{1}{2m}\|\xi^{k,m}\|^2 \\ &\leq f(x^{k,m}) + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x^{k,m}) \\ &\leq f(\overline{x}^k). \end{aligned} \tag{4.24}$$

Since $L_{k,m}$ is quadratic in ξ, the supremum of $L_{k,m}(x^{k,m}, \xi)$ over $\xi \geq 0$ is attained at

$$\xi_j^{k,m} = m g_j^+(x^{k,m}), \qquad j = 1, \ldots, r, \tag{4.25}$$

This implies that

$$\begin{aligned} L_{k,m}(x^{k,m}, \xi^{k,m}) &= f(x^{k,m}) + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \frac{m}{2}\|g^+(x^{k,m})\|^2 \\ &\geq f(x^{k,m}). \end{aligned} \tag{4.26}$$

For any sequence $m \to \infty$, consider the corresponding sequence $\{x^{k,m}\}$. From Eqs. (4.24) and (4.26), we see that the sequence $\{x^{k,m}\}$ belongs to the set $\big\{x \in X^k \mid f(x) \leq f(\overline{x}^k)\big\}$, which is compact, since f is closed. Hence, $\{x^{k,m}\}$ has a limit point, denoted by $\hat{x}^k$, which belongs to $\big\{x \in X^k \mid f(x) \leq f(\overline{x}^k)\big\}$. By passing to a subsequence if necessary, we can assume without loss of generality that $\{x^{k,m}\}$ converges to $\hat{x}^k$. For each k, the sequence $\big\{f(x^{k,m})\big\}$ is bounded from below by $\inf_{x\in X^k} f(x)$, which is finite by Weierstrass' Theorem since f is closed and coercive when restricted to X^k. Also, for each k, $L_{k,m}(x^{k,m}, \xi^{k,m})$ is bounded from above by $f(\overline{x}^k)$ [cf. Eq. (4.24)], so Eq. (4.26) implies that

$$\limsup_{m\to\infty} g(x^{k,m}) \leq 0.$$

Therefore, by using the lower semicontinuity of the g_j, we obtain $g(\hat{x}^k) \leq 0$, implying that $\hat{x}^k$ is a feasible point of problem (4.22), so that $f(\hat{x}^k) \geq f(\overline{x}^k)$. Using Eqs. (4.24) and (4.26) together with the closedness of f, we also have

$$f(\hat{x}^k) \leq \liminf_{m\to\infty} f(x^{k,m}) \leq \limsup_{m\to\infty} f(x^{k,m}) \leq f(\overline{x}^k),$$

thereby showing that for each k,

$$\lim_{m\to\infty} f(x^{k,m}) = f(\overline{x}^k) = f^* + \gamma^k. \tag{4.27}$$

Let $\gamma = f^* - f(\overline{x})$. For sufficiently large k, we have $\overline{x} \in X^k$ and $\gamma^k < \gamma$. Consider the vector

$$z^k = \left(1 - \frac{2\gamma^k}{\gamma^k + \gamma}\right)\overline{x}^k + \frac{2\gamma^k}{\gamma^k + \gamma}\overline{x},$$

which belongs to X^k for sufficiently large k [by the convexity of X^k and the fact that $(2\gamma^k/\gamma^k + \gamma) < 1$]. By the convexity of f, we have

$$\begin{aligned} f(z^k) &\leq \left(1 - \frac{2\gamma^k}{\gamma^k + \gamma}\right) f(\overline{x}^k) + \frac{2\gamma^k}{\gamma^k + \gamma} f(\overline{x}) \\ &= \left(1 - \frac{2\gamma^k}{\gamma^k + \gamma}\right)(f^* + \gamma^k) + \frac{2\gamma^k}{\gamma^k + \gamma}(f^* - \gamma) \\ &= f^* - \gamma^k. \end{aligned} \tag{4.28}$$

Similarly, by the convexity of the g_j, we have

$$\begin{aligned} g_j(z^k) &\leq \left(1 - \frac{2\gamma^k}{\gamma^k + \gamma}\right) g_j(\overline{x}^k) + \frac{2\gamma^k}{\gamma^k + \gamma} g_j(\overline{x}) \\ &\leq \frac{2\gamma^k}{\gamma^k + \gamma} g_j(\overline{x}) \\ &\leq o(\sqrt{\gamma^k}). \end{aligned} \tag{4.29}$$

Using Eqs. (4.24) and (4.25), we obtain

$$\begin{aligned} f(x^{k,m}) &\leq f(x^{k,m}) + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \frac{m}{2}\|g^+(x^{k,m})\|^2 \\ &\leq f(x) + (\gamma^k)^2 + \frac{m}{2}\|g^+(x)\|^2, \qquad \forall\, x \in X^k. \end{aligned}$$

Substituting $x = z^k$ in the preceding relation, and using Eqs. (4.28) and (4.29), we see that for large k,

$$f(x^{k,m}) \leq f^* - \gamma^k + (\gamma^k)^2 + m o(\gamma^k).$$

Since $\gamma^k \to 0$, this implies that for sufficiently large k and for all scalars $m \leq 1/\sqrt{\gamma^k}$, we have

$$f(x^{k,m}) \leq f^* - \frac{\gamma^k}{2}. \tag{4.30}$$

We next show that for sufficiently large k, there exists some scalar $m_k \geq 1/\sqrt{\gamma^k}$ such that

$$f(x^{k,m_k}) = f^* - \frac{\gamma^k}{2}. \tag{4.31}$$

For this purpose, we first show that $L_{k,m}(x^{k,m}, \xi^{k,m})$ changes continuously with m, i.e, for all $\overline{m} > 0$, we have $L_{k,m}(x^{k,m}, \xi^{k,m}) \to L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}})$ as $m \to \overline{m}$. [By this we mean, for every sequence $\{m^t\}$ that converges to $\overline{m}$, we have that the corresponding sequence $L_{k,m^t}(x^{k,m^t}, \xi^{k,m^t})$ converges to $L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}})$.] Denote

$$\overline{f}(x^{k,m}) = f(x^{k,m}) + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2.$$

Note that for all $m \geq \overline{m}$, we have

$$\begin{aligned} L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}}) &= \overline{f}(x^{k,\overline{m}}) + \frac{\overline{m}}{2}\|g^+(x^{k,\overline{m}})\|^2 \\ &\leq \overline{f}(x^{k,m}) + \frac{\overline{m}}{2}\|g^+(x^{k,m})\|^2 \\ &\leq \overline{f}(x^{k,m}) + \frac{m}{2}\|g^+(x^{k,m})\|^2 \\ &\leq \overline{f}(x^{k,\overline{m}}) + \frac{m}{2}\|g^+(x^{k,\overline{m}})\|^2, \end{aligned}$$

showing that $L_{k,m}(x^{k,m}, \xi^{k,m}) \to L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}})$ as $m \downarrow \overline{m}$. Similarly, we have for all $m \leq \overline{m}$,

$$\begin{aligned}
\overline{f}(x^{k,\overline{m}}) + \frac{m}{2}\|g^+(x^{k,\overline{m}})\|^2 &\leq \overline{f}(x^{k,\overline{m}}) + \frac{\overline{m}}{2}\|g^+(x^{k,\overline{m}})\|^2 \\
&\leq \overline{f}(x^{k,m}) + \frac{\overline{m}}{2}\|g^+(x^{k,m})\|^2 \\
&= \overline{f}(x^{k,m}) + \frac{m}{2}\|g^+(x^{k,m})\|^2 + \frac{\overline{m}-m}{2}\|g^+(x^{k,m})\|^2 \\
&\leq \overline{f}(x^{k,\overline{m}}) + \frac{m}{2}\|g^+(x^{k,\overline{m}})\|^2 + \frac{\overline{m}-m}{2}\|g^+(x^{k,m})\|^2.
\end{aligned}$$

For each k, the sequence $g_j(x^{k,m})$ is bounded from below by $\inf_{x \in X^k} g_j(x)$, which is finite by Weierstrass' Theorem since g_j is closed and coercive when restricted to X^k. Therefore, we have from the preceding that $L_{k,m}(x^{k,m}, \xi^{k,m}) \to L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}})$ as $m \uparrow \overline{m}$, which shows that $L_{k,m}(x^{k,m}, \xi^{k,m})$ changes continuously with m.

Next, we show that $x^{k,m} \to x^{k,\overline{m}}$ as $m \to \overline{m}$. Since, for each k, the sequence $x^{k,m}$ belongs to a compact set, it has a limit point as $m \to \overline{m}$. Let $\hat{x}$ be a limit point of $x^{k,m}$. Using the continuity of $L_{k,m}$ and the closedness of f and the g_j, we obtain

$$\begin{aligned}
L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}}) &= \liminf_{m \to \overline{m}} \overline{f}(x^{k,m}) + \frac{m}{2}\|g^+(x^{k,m})\|^2 \\
&\geq \overline{f}(\hat{x}) + \frac{\overline{m}}{2}\|g^+(\hat{x})\|^2 \\
&\geq \inf_{x \in X^k} \left\{ \overline{f}(x) + \frac{\overline{m}}{2}\|g^+(x)\|^2 \right\} \\
&= L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}}).
\end{aligned}$$

This shows that $\hat{x}$ attains the infimum of $\overline{f}(x) + \frac{\overline{m}}{2}\|g^+(x)\|^2$ over $x \in X^k$. Since this function is strictly convex, it has a unique optimal solution, showing that $\hat{x} = x^{k,\overline{m}}$.

Finally, we show that $f(x^{k,m}) \to f(x^{k,\overline{m}})$ as $m \to \overline{m}$. Assume that $f(x^{k,\overline{m}}) < \limsup_{m \to \overline{m}} f(x^{k,m})$. Using the continuity of $L_{k,m}$ and the fact that $x^{k,m} \to x^{k,\overline{m}}$ as $m \to \overline{m}$, we have

$$\begin{aligned}
\overline{f}(x^{k,\overline{m}}) + \liminf_{m \to \overline{m}} \|g^+(x^{k,m})\|^2 &< \limsup_{m \to \overline{m}} L_{k,m}(x^{k,m}, \xi^{k,m}) \\
&= L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}}) \\
&= \overline{f}(x^{k,\overline{m}}) + \|g^+(x^{k,\overline{m}})\|^2.
\end{aligned}$$

But this contradicts the lower semicontinuity of the g_j, hence showing that $f(x^{k,\overline{m}}) \geq \limsup_{m \to \overline{m}} f(x^{k,m})$, which together with the lower semicontinuity of f yields the desired result.

Eqs. (4.30), (4.27), and the continuity of $f(x^{k,m})$ in m imply the existence of some scalar $m_k \geq 1/\sqrt{\gamma^k}$ such that

$$f(x^{k,m_k}) = f^* - \frac{\gamma^k}{2}. \tag{4.32}$$

Combining the preceding relation with Eqs. (4.24) and (4.26) (for $m = m_k$), together with the facts that $f(\overline{x}^k) \to f^*$ and $\gamma^k \to 0$ as $k \to \infty$, we obtain

$$\lim_{k\to\infty} \left(f(x^{k,m_k}) - f^* + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m_k} - \overline{x}^k\|^2 + \xi^{k,m_k'} g(x^{k,m_k}) \right) = 0. \tag{4.33}$$

Denote

$$\delta^k = \sqrt{1 + \sum_{j=1}^{r} (\xi_j^{k,m_k})^2},$$

$$\mu_0^k = \frac{1}{\delta^k}, \qquad \mu_j^k = \frac{\xi_j^{k,m_k}}{\delta^k}, \quad j = 1, \ldots, r. \tag{4.34}$$

Since δ^k is bounded from below, Eq. (4.33) yields

$$\lim_{k\to\infty} \left(\mu_0^k f(x^{k,m_k}) - \mu_0^k f^* + \frac{(\gamma^k)^2}{4k^2\delta^k}\|x^{k,m_k} - \overline{x}^k\|^2 + \sum_{j=1}^{r} \mu_j^k g_j(x^{k,m_k}) \right) = 0. \tag{4.35}$$

Substituting $m = m_k$ in Eq. (4.23) and dividing both sides of the first relation by δ^k, we get

$$\begin{aligned} \mu_0^k f(x^{k,m_k}) + \frac{(\gamma^k)^2}{4k^2\delta^k}\|x^{k,m_k} - \overline{x}^k\|^2 + \sum_{j=1}^{r} \mu_j^k g_j(x^{k,m_k}) \\ \leq \mu_0^k f(x) + \sum_{j=1}^{r} \mu_j^k g_j(x) + \frac{(\gamma^k)^2}{\delta^k}, \qquad \forall\, x \in X^k. \end{aligned}$$

Since the sequence $\{\mu_0^k, \mu_1^k, \ldots, \mu_r^k\}$ is bounded, it has a limit point, denoted by $\{\mu_0^*, \mu_1^*, \ldots, \mu_r^*\}$. Taking the limit along the relevant subsequences in the preceding relation together with Eq. (4.35) yields

$$\mu_0^* f^* \leq \mu_0^* f(x) + \mu^{*\prime} g(x), \qquad \forall\, x \in X,$$

which implies that

$$\begin{aligned} \mu_0^* f^* &\leq \inf_{x\in X} \{\mu_0^* f(x) + \mu^{*\prime} g(x)\} \\ &\leq \inf_{x\in X,\, g(x)\leq 0} \{\mu_0^* f(x) + \mu^{*\prime} g(x)\} \\ &\leq \inf_{x\in X,\, g(x)\leq 0} \mu_0^* f(x) \\ &= \mu_0^* f^*. \end{aligned}$$

Thus we have

$$\mu_0^* f^* = \inf_{x \in X} \{\mu_0^* f(x) + \mu^{*\prime} g(x)\},$$

so that $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$ satisfy conditions (i), (ii), and (iii) of the proposition.

Finally, dividing both sides of Eq. (4.25) by δ^k, and using Eq. (4.34) and the fact that $\mu_j^k \to \mu_j^*$, as $k \to \infty$, we obtain

$$\lim_{k \to \infty} \frac{m_k g_j^+(x^{k,m_k})}{\delta^k} = \mu_j^*, \qquad j = 1, \ldots, r.$$

Since, we also have from Eq. (4.32) that

$$f(x^{k,m_k}) < f^*, \qquad \lim_{k \to \infty} f(x^{k,m_k}) = f^*,$$

it follows that the sequence $\{x^{k,m_k}\}$ satisfies condition (iv) of the proposition, thereby completing the proof. **Q.E.D.**

The preceding proposition motivates a definition of pseudonormality that is not tied to a specific optimal primal solution.

Definition 5.4.5: Consider problem (4.1), under Assumption 5.4.6. The constraint set of problem (4.1) is said to be *pseudonormal* if there do not exist a vector $\mu = (\mu_1, \ldots, \mu_r) \geq 0$, and a sequence $\{x^k\} \subset X$ such that:

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

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

Figure 5.4.8 provides a geometric interpretation of pseudonormality. As an example, it is easily seen with the aid of Fig. 5.4.8 that if f is convex over $\Re^n$, the functions g_j are affine, and $X = \Re^n$, then the constraint set is pseudonormal.

In view of Prop. 5.4.13, if problem (P) has a closed and convex cost function f and a pseudonormal constraint set, there exists a geometric multiplier and there is no duality gap. This geometric multiplier satisfies in addition the special condition (iv) of Prop. 5.4.13.

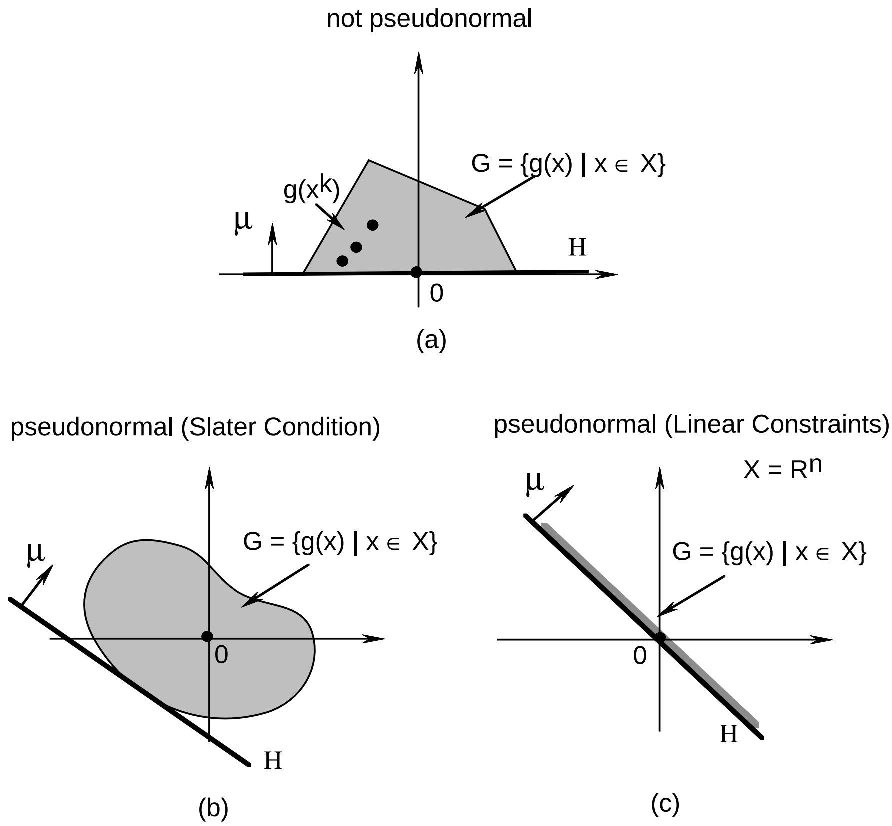


Figure 5.4.8. Geometric interpretation of pseudonormality, assuming for simplicity that there are no equality constraints. Consider the set

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

For feasibility, G should intersect the nonpositive orthant $\{z \mid z \leq 0\}$. The first condition in the definition of pseudonormality means that there is a hyperplane with normal μ, which simultaneously supports G and passes through 0 [note that, as illustrated in figure (b), this cannot happen if G intersects the interior of the nonpositive orthant; cf. the Slater criterion]. The second and third conditions in the definition of pseudonormality mean that the negative orthant can be approached by a sequence $\big\{g(x^k)\big\} \subset G \cap \mathrm{int}(\overline{H})$, where $\overline{H}$ is the positive halfspace defined by the hyperplane,

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

[cf. figure (a)]. Pseudonormality means that there is no $\mu \geq 0$ and $\{x^k\} \subset X$ satisfying both of these conditions.

5.4.3. Enhanced Dual Fritz John Conditions

The Fritz John multipliers of Props. 5.4.8-5.4.13 define a hyperplane with normal (μ^*, μ_0^*) that supports the set of constraint-cost pairs (i.e., the set M of Fig. 5.4.6) at $(0, f^*)$. On the other hand, it is possible to construct a hyperplane that supports the set M at the point $(0, q^*)$, where q^* is the optimal dual value

$$q^* = \sup_{\mu \geq 0} q(\mu) = \sup_{\mu \geq 0} \inf_{x \in X} \big\{ f(x) + \mu' g(x) \big\},$$

while asserting the existence of a sequence that satisfies a condition analogous to the CV condition. The next proposition addresses this question.

Proposition 5.4.14: (Enhanced Dual Fritz John Conditions) Consider problem (4.1) under assumption 5.4.6. We also assume that

$$q^* > -\infty.$$

Then there exists a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, satisfying the following conditions:

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

(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 $\{x^k\} \subset X$ such that

$$\lim_{k\to\infty} f(x^k) = q^*, \qquad \limsup_{k\to\infty} g(x^k) \leq 0, \tag{4.36}$$

and for all k,

$$f(x^k) < q^*, \qquad \mu_j^* g_j(x^k) > 0, \quad \forall\, j \in J, \tag{4.37}$$

$$g_j^+(x^k) = o\left(\min_{j\in J} g_j(x^k)\right), \qquad \forall\, j \notin J. \tag{4.38}$$

Proof: First we prove the following lemma.

Lemma 5.4.1: Consider problem (4.1), and assume that X is convex, the functions f and the g_j are convex over X, and $-\infty < f^* < \infty$. For each $\delta > 0$, let

$$r^\delta = \inf_{\substack{x\in X \\ g(x)\leq \delta e}} f(x), \tag{4.39}$$

where $e \in \Re^r$ is a vector, whose components are all equal to 1. Then $r^\delta \leq q^*$ for all $\delta > 0$ and

$$q^* = \lim_{\delta\downarrow 0} r^\delta.$$

Proof: Since f^* is finite, there exists some $\overline{x} \in X$ such that $g(\overline{x}) \leq 0$. Hence, for each $\delta > 0$ such that $r^\delta > -\infty$, the Slater condition is satisfied for problem (4.39), and therefore, by Prop. 5.4.11, this problem has a geometric multiplier, i.e., there exists a nonnegative vector μ^δ such that

$$\begin{aligned} r^\delta &= \inf_{x\in X}\Big\{f(x) + \mu^{\delta\prime}\big(g(x) - \delta e\big)\Big\} \\ &\leq \inf_{x\in X}\big\{f(x) + \mu^{\delta\prime} g(x)\big\} \\ &= q(\mu^\delta) \\ &\leq q^*. \end{aligned}$$

For each $\delta > 0$ such that $r^\delta = -\infty$, we also have $r^\delta \le q^*$, so that

$$r^\delta \le q^*, \qquad \forall\ \delta > 0.$$

Taking the limit as $\delta \downarrow 0$, we obtain

$$\lim_{\delta \downarrow 0} r^\delta \le q^*.$$

To show the converse, for each $\delta > 0$, choose $x^\delta \in X$ such that $g_j(x^\delta) \le \delta$ for all j and $f(x^\delta) \le r^\delta + \delta$. Then, for any $\mu \ge 0$,

$$q(\mu) = \inf_{x \in X} \big\{ f(x) + \mu' g(x) \big\} \le f(x^\delta) + \mu' g(x^\delta) \le r^\delta + \delta + \delta \sum_{j=1}^{r} \mu_j.$$

Taking the limit as $\delta \downarrow 0$, we obtain

$$q(\mu) \le \lim_{\delta \downarrow 0} r^\delta,$$

so that $q^* \le \lim_{\delta \downarrow 0} r^\delta$. **Q.E.D.**

We now return to the proof of the proposition. Consider the problem

$$\begin{aligned} &\text{minimize} \ \ f(x) \\ &\text{subject to} \ \ x \in X,\ g(x) \le \frac{1}{k^4} e. \end{aligned}$$

By the previous lemma, for each k, the optimal value of this problem is less than or equal to q^*. For each k, let $\tilde{x}^k \in X$ be a vector that satisfies

$$f(\tilde{x}^k) \le q^* + \frac{1}{k^2}, \qquad g(\tilde{x}^k) \le \frac{1}{k^4} e.$$

Consider also the problem

$$\begin{aligned} &\text{minimize} \ \ f(x) \\ &\text{subject to} \ \ g(x) \le \frac{1}{k^4} e, \\ &\qquad\qquad x \in \tilde{X}^k = X \cap \left\{ x \ \Big|\ \|x\| \le k\big(\max_{1 \le i \le k} \|\tilde{x}^i\| + 1 \big) \right\}. \end{aligned} \tag{4.40}$$

Since $\{x \mid g(x) \leq (1/k^4)e\} \cap X^k$ is the intersection of the closed set $\{x \in X \mid g(x) \leq (1/k^4)e\}$ and the compact set $\left\{x \;\middle|\; \|x\| \leq \|\tilde{x}^k\| + k\right\}$, we see that the constraint set of the preceding problem is compact. Since f is closed, and therefore lower semicontinuous over X^k, by Weierstrass' Theorem, the preceding minimization problem has an optimal solution, which we denote by $\overline{x}^k$. Note that since $\tilde{x}^k$ belongs to the feasible set of this problem, we have

$$f(\overline{x}^k) \leq f(\tilde{x}^k) \leq q^* + \frac{1}{k^2}. \tag{4.41}$$

We consider the function

$$L_k(x, \xi) = f(x) + \xi' g(x) - \frac{\|\xi\|^2}{2k},$$

and we note that L_k is convex in x, and concave and coercive in ξ. For each k, we consider the saddle points of L_k over x in

$$X^k = \tilde{X}^k \cap \{x \mid g(x) \leq ke\} \tag{4.42}$$

and $\xi \geq 0$. Note that $X^k = \left\{x \;\middle|\; \|x\| \leq k\left(\max_{1 \leq i \leq k} \|\tilde{x}^i\| + 1\right)\right\} \cap \{x \in X \mid g_j(x) \leq k,\ j = 1, \ldots, r\}$, from which using the closedness of the g_j, we see that X^k is compact. Therefore, we can use the saddle point theorem to assert that L_k has a saddle point over $x \in X^k$ and $\xi \geq 0$, denoted by (x^k, ξ^k).

Since L_k is quadratic in ξ, the supremum of $L_k(x^k, \xi)$ over $\xi \geq 0$ is attained at

$$\xi_j^k = k g_j^+(x^k), \qquad j = 1, \ldots, r. \tag{4.43}$$

Similarly, the infimum of $L_k(x, \xi^k)$ over $x \in X^k$ is attained at x^k, implying that

$$\begin{aligned}
f(x^k) + \xi^{k\prime} g(x^k) &= \inf_{x \in X^k} \{f(x) + \xi^{k\prime} g(x)\} \\
&= \inf_{x \in X^k} \{f(x) + k g^+(x^k)' g(x)\} \\
&\leq \inf_{x \in X^k,\ g(x) \leq \frac{1}{k^4}} \{f(x) + k g^+(x^k)' g(x)\} \\
&\leq \inf_{x \in X^k,\ g(x) \leq \frac{1}{k^4}} f(x) + \frac{r}{k^2} \\
&= f(\overline{x}^k) + \frac{r}{k^2} \\
&\leq q^* + \frac{r+1}{k^2},
\end{aligned} \tag{4.44}$$

where the last inequality follows from Eq. (4.41).

We also have

$$\begin{aligned} L_k(x^k, \xi^k) &= \sup_{\xi \geq 0} \inf_{x \in X^k} L_k(x, \xi) \\ &\geq \sup_{\xi \geq 0} \inf_{x \in X} L_k(x, \xi) \\ &= \sup_{\xi \geq 0} \left\{ \inf_{x \in X} \big(f(x) + \xi' g(x) \big) - \frac{\|\xi\|^2}{2k} \right\} \\ &= \sup_{\xi \geq 0} \left\{ q(\xi) - \frac{\|\xi\|^2}{2k} \right\} \\ &\geq q(\lambda^k) - \frac{\|\lambda^k\|^2}{2k}, \end{aligned} \tag{4.45}$$

for each k, where $\{\lambda^k\}$ is a nonnegative sequence such that

$$q(\lambda^k) \to q^*, \qquad \frac{\|\lambda^k\|^2}{2k} \to 0 \tag{4.46}$$

as $k \to \infty$.

Combining Eqs. (4.45) and (4.44), we obtain

$$\begin{aligned} q(\lambda^k) - \frac{\|\lambda^k\|^2}{2k} &\leq L_k(x^k, \xi^k) = f(x^k) + \xi^{k\prime} g(x^k) - \frac{1}{2k} \|\xi^k\|^2 \\ &\leq f(x^k) + \xi^{k\prime} g(x^k) \\ &\leq q^* + \frac{r+1}{k^2}. \end{aligned} \tag{4.47}$$

Taking the limit in the preceding relation, and using Eq. (4.46), we have

$$\lim_{k \to \infty} \big(f(x^k) - q^* + \xi^{k\prime} g(x^k) \big) = 0. \tag{4.48}$$

Denote

$$\delta^k = \sqrt{1 + \sum_{j=1}^{r} (\xi_j^k)^2},$$

$$\mu_0^k = \frac{1}{\delta^k}, \qquad \mu_j^k = \frac{\xi_j^k}{\delta^k}, \ j = 1, \ldots, r. \tag{4.49}$$

Since δ^k is bounded from below, Eq. (4.48) yields

$$\lim_{k \to \infty} \mu_0^k \big(f(x^k) - q^* \big) + \sum_{j=1}^{r} \mu_j^k g_j(x^k) = 0. \tag{4.50}$$

Dividing both sides of the first relation in Eq. (4.44) by δ^k, we get

$$\mu_0^k f(x^k) + \sum_{j=1}^r \mu_j^k g_j(x^k) \le \mu_0^k f(x) + \sum_{j=1}^r \mu_j^k g_j(x), \qquad \forall\, x \in X^k.$$

Since the sequence $\{\mu_0^k, \mu_1^k, \ldots, \mu_r^k\}$ is bounded, it has a limit point, denoted by $\{\mu_0^*, \mu_1^*, \ldots, \mu_r^*\}$. Taking the limit along the relevant subsequence in the preceding relation together with Eq. (4.50) yields

$$\mu_0^* q^* \le \inf_{x \in X} \big\{\mu_0^* f(x) + \sum_{j=1}^r \mu_j^* g_j(x)\big\}.$$

If $\mu_0^* > 0$, we obtain from the preceding relation

$$q^* \le \inf_{x \in X} \left\{ f(x) + \sum_{j=1}^r \frac{\mu_j^*}{\mu_0^*} g_j(x) \right\} = q\left(\frac{\mu^*}{\mu_0^*}\right) \le q^*.$$

Similarly, if $\mu_0^* = 0$, it can be seen that $0 = \inf_{x \in X} {\mu^*}' g(x)$. Hence, in both cases, we have

$$\mu_0^* q^* = \inf_{x \in X} \left\{ \mu_0^* f(x) + \sum_{j=1}^r \mu_j^* g_j(x) \right\},$$

thus showing that $\mu_0^*, \ldots, \mu_r^*$ satisfy conditions (i)-(iii) of the proposition.

Let $J = \{j \neq 0 \mid \mu_j^* > 0\}$, and assume that J is nonempty. Dividing both sides of Eq. (4.43) by δ^k, and using Eq. (4.49) and the fact that $\mu_j^k \to \mu_j^*$, for all $j = 1, \ldots, r$, we obtain

$$\lim_{k \to \infty} \frac{k g_j^+(x^k)}{\delta^k} = \mu_j^*, \qquad j = 1, \ldots, r.$$

This implies that for all sufficiently large k,

$$g_j(x^k) > 0, \qquad \forall\, j \in J,$$

and

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

Note also that, for all k, we have from Eq. (4.47) that

$$k\big(f(x^k) - q^*\big) + {\xi^k}' k g(x^k) \le \frac{r+1}{k}.$$

Using Eq. (4.43), this yields

$$k\big(f(x^k) - q^*\big) + \sum_{j=1}^{r} (\xi_j^k)^2 \le \frac{r+1}{k}.$$

Dividing by $(\delta^k)^2$ and taking the limit, we get

$$\limsup_{k\to\infty} \frac{k\big(f(x^k) - q^*\big)}{(\delta^k)^2} \le -\sum_{j=1}^{r} (\mu_j^*)^2, \tag{4.51}$$

implying that $f(x^k) < q^*$ for all sufficiently large k, since the index set J is nonempty.

We finally show that $f(x^k) \to q^*$ and $\limsup_{k\to\infty} g(x^k) \le 0$. By Eq. (4.47), we have

$$\limsup_{k\to\infty} \frac{\|\xi^k\|^2}{2k} \le 0,$$

which implies that

$$\lim_{k\to\infty} \frac{\|\xi^k\|^2}{2k} = 0. \tag{4.52}$$

Similarly, combining Eqs. (4.47) and (4.43), we obtain

$$\lim_{k\to\infty} \big(f(x^k) - q^*\big) + \frac{\|\xi^k\|^2}{2k} = 0,$$

which together with Eq. (4.52) shows that $f(x^k) \to q^*$. Moreover, Eqs. (4.52) and (4.43) imply that

$$\lim_{k\to\infty} k \sum_{j=1}^{r} \big(g_j^+(x^k)\big)^2 = 0,$$

showing that $\limsup_{k\to\infty} g(x^k) \le 0$. Therefore, the sequence $\{x^k\}$ satisfies condition (iv) of the proposition, completing the proof. **Q.E.D.**

In the preceding proposition, if we can guarantee that $\mu_0^* > 0$, then there exists a dual optimal solution, which satisfies the special condition (iv) of Prop. 5.4.14.

The proof of this proposition is similar to the proof of Prop. 5.2.2. Essentially, the proof generates saddle points of the function

$$L_k(x, \xi) = f(x) + \xi' g(x) - \frac{\|\xi\|^2}{2k},$$

over x in the compact set X^k [cf. Eq. (4.42)] and $\xi \geq 0$. It can be shown that

$$L_k(x,\xi) = \inf_{u \in \Re^r} \left\{ p^k(u) + \frac{k}{2}\|u^+\|^2 \right\},$$

where $p^k(u)$ is the optimal value of the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g(x) \leq u, \ x \in X^k, \end{aligned} \tag{4.53}$$

(see the discussion following the proof of Prop. 5.2.2). For each k, the value $L_k(x^k,\xi^k)$ can be visualized geometrically as in Fig. 5.2.1.

Note that $p^k(u)$ is the primal function corresponding to the problem where the set constraint X in the original problem is replaced by X^k. Using a compact set approximation of the abstract set constraint X has the effect of approximating the original problem with one that has no duality gap. Hence, the corresponding primal function $p^k(u)$ is lower semicontinuous at 0 and approximates the primal function $p(u)$ with greater accuracy as $k \to \infty$. In this proof, the rate at which X^k approaches X is chosen high enough so that $L_k(x^k,\xi^k)$ converges to q^* as $k \to \infty$ [cf. Eq. (4.47)], and not to f^*, as in the proof of Prop. 5.2.2.

5.4.4. Informative Geometric Multipliers and Dual Optimal Solutions

In this section, we focus on geometric multipliers and dual optimal solutions, which are special in that they satisfy conditions analogous to the CV condition. Consistent with our analysis in Chapter 4, we refer to geometric multipliers that satisfy these conditions as being *informative*, since they provide sensitivity information by indicating the constraints to violate in order to effect a cost reduction.

Definition 5.4.6: A vector $\mu^* \geq 0$ is said to be an *informative geometric multiplier* if the following two conditions hold:

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

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

$$\lim_{k \to \infty} f(x^k) = f^*, \qquad \limsup_{k \to \infty} g(x^k) \leq 0,$$

and for all k,

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

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

In the next proposition, we show the existence of an informative geometric multiplier under very general assumptions.

Proposition 5.4.15: (Existence of Informative Geometric Multipliers) Consider problem (4.1) under Assumption 5.4.6. We assume that the set of geometric multipliers, denoted by

$$M = \left\{\mu \geq 0 \mid f^* = \inf_{x \in X} \{f(x) + \mu' g(x)\}\right\},$$

is nonempty. Then the vector of minimum norm in M is informative.

Proof: Let $\tilde{x}^k$ be a feasible sequence for problem (4.1) such that $f(\tilde{x}^k) \to f^*$ and

$$f(\tilde{x}^k) \leq f^* + \frac{1}{k^2}, \qquad k = 1, 2, \ldots$$

Consider the problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \leq 0, \\ & x \in X^k = X \cap \left\{ x \;\middle|\; \|x\| \leq k \max_{1 \leq i \leq k} \|\tilde{x}^i\| \right\}. \end{aligned}$$

Since $\{x \mid g(x) \leq 0\} \cap X^k$ is the intersection of the closed set $\{x \in X \mid g(x) \leq 0\}$ and the compact set $\left\{x \;\middle|\; \|x\| \leq \|\tilde{x}^k\| + k\right\}$, we see that the constraint set of the preceding problem is compact. Since f is closed, and therefore lower semicontinuous over X^k, by Weierstrass' Theorem, the preceding minimization problem has an optimal solution, which we denote by $\overline{x}^k$. Note that since $\tilde{x}^k$ belongs to the feasible set of this problem, we have

$$f(\overline{x}^k) \leq f(\tilde{x}^k) \leq f^* + \frac{1}{k^2}. \tag{4.54}$$

We consider the function

$$L_k(x, \mu) = kf(x) + \mu' kg(x) - \frac{\|\mu\|^2}{2},$$

and we note that L_k is convex in x, and concave and coercive in μ. For each k, we consider the saddle points of L_k over $x \in X^k$ and $\mu \geq 0$. Since X^k is bounded, we use the saddle point theorem to assert that L_k has a saddle point for each k, denoted by (x^k, μ^k).

Let μ^* be the vector of minimum norm in M. If $\mu^* = 0$, then μ^* is an informative geometric multiplier and we are done, so assume that $\mu^* \neq 0$. For any $\mu \in M$, we have by the definition of M,

$$\inf_{x \in X} \{f(x) + \mu' g(x)\} = f^*,$$

so that

$$\inf_{x \in X^k} L_k(x, \mu) \geq \inf_{x \in X} L_k(x, \mu) = kf^* - \frac{1}{2}\|\mu\|^2.$$

Therefore,

$$\begin{aligned} L_k(x^k, \mu^k) &= \sup_{\mu \geq 0} \inf_{x \in X^k} L_k(x, \mu) \\ &\geq \sup_{\mu \in M} \inf_{x \in X^k} L_k(x, \mu) \\ &\geq \sup_{\mu \in M} \left(kf^* - \frac{1}{2}\|\mu\|^2 \right) \\ &= kf^* - \frac{1}{2}\|\mu^*\|^2, \end{aligned} \tag{4.55}$$

where μ^* denotes the vector of minimum norm in the set M. Since (x^k, μ^k) is a saddle point of L_k over $x \in X^k$ and $\mu \geq 0$, the minimum in the left hand side of

$$\inf_{x \in X^k} L_k(x, \mu^k) = \inf_{x \in X^k} k\big\{f(x) + \mu^{k\prime} g(x)\big\} - \frac{\|\mu^k\|^2}{2},$$

is attained at x^k, implying that

$$\begin{aligned} f(x^k) + \mu^{k\prime} g(x^k) &= \inf_{x \in X^k} \big\{f(x) + \mu^{k\prime} g(x)\big\} \\ &\leq \inf_{x \in X^k,\ g(x) \leq 0} \big\{f(x) + \mu^{k\prime} g(x)\big\} \\ &\leq \inf_{x \in X^k,\ g(x) \leq 0} f(x) \\ &= f(\overline{x}^k) \\ &\leq f^* + \frac{1}{k^2}, \end{aligned} \tag{4.56}$$

where the last inequality follows from Eq. (4.54). Combining Eqs. (4.55) and (4.56), we obtain

$$\begin{aligned} f^* - \frac{1}{2}\|\mu^*\|^2 \leq L_k(x^k, \mu^k) &= k\big\{f(x^k) + \mu^{k\prime} g(x^k)\big\} - \frac{1}{2}\|\mu^k\|^2 \\ &\leq kf^* + \frac{1}{k} - \frac{1}{2}\|\mu^k\|^2. \end{aligned} \tag{4.57}$$

It follows from the preceding relation that $\|\mu^k\|$ remains bounded as $k \to \infty$. Let $\overline{\mu}$ be a limit point of $\{\mu^k\}$. We also have from the preceding relation that

$$\lim_{k \to \infty} \big(f(x_k) - f^*\big) + \mu^{k\prime} g(x^k) = 0.$$

Hence, taking the limit along the relevant subsequence in the first relation in Eq. (4.56) yields

$$f^* \leq \inf_{x \in X} \big\{f(x) + \overline{\mu}' g(x)\big\} = q(\overline{\mu}) \leq f^*,$$

where the last inequality follows from weak duality. Hence $\overline{\mu}$ belongs to set M, and since $\|\overline{\mu}\| \leq \|\mu^*\|$ [which follows by taking the limit in Eq. (4.57)], by using the minimum norm property of μ^*, we conclude that any limit point $\overline{\mu}$ of μ^k must be equal to μ^*. Thus $\mu^k \to \mu^*$, and using Eq. (4.57), we obtain

$$\lim_{k \to \infty} \big(L_k(x^k, \mu^k) - kf^*\big) = -\frac{1}{2}\|\mu^*\|^2. \tag{4.58}$$

Since L_k is quadratic in μ, the supremum of $L_k(x^k, \mu)$ over $\mu \geq 0$ is attained at

$$\mu_j^k = k g_j(x^k)^+, \qquad j = 1, \ldots, r, \tag{4.59}$$

so that

$$L_k(x^k, \mu^k) = \sup_{\mu \geq 0} L_k(x^k, \mu) = k f(x^k) + \frac{1}{2}\|\mu^k\|^2,$$

which combined with Eq. (4.58) yields

$$\lim_{k \to \infty} k\big(f(x^k) - f^*\big) = -\|\mu^*\|^2,$$

implying that $f(x^k) < f^*$ for all sufficiently large k, since $\mu^* \neq 0$. Since, $\mu^k \to \mu^*$, Eq. (4.59) also implies

$$\lim_{k \to \infty} k g_j^+(x^k) = \mu_j^*, \qquad j = 1, \ldots, r.$$

Thus the sequence $\{x^k\}$ fulfills condition (ii) of the definition of an informative geometric multiplier, thereby completing the proof. **Q.E.D.**

When there is a duality gap, there exists no geometric multipliers, even if there is a dual optimal solution. In this case, we are motivated to investigate the existence of a special dual optimal solution, which satisfies condition (iv) of Proposition 5.4.14,

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

$$\lim_{k \to \infty} f(x^k) = q^*, \qquad \limsup_{k \to \infty} g(x^k) \leq 0,$$

and for all k,

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

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

We call such a dual optimal solution *informative*, since it provides information by indicating the constraints to violate to result in a cost reduction by an amount which is strictly greater than the size of the duality gap, i.e., $f^* - q^*$.

We have the following result, which is analogous to the preceding proposition.

Proposition 5.4.16: (Existence of Informative Dual Optimal Solutions) Consider problem (4.1) under Assumption 5.4.6. Assume further that q^* is finite and that there exists a dual optimal solution, i.e., there exists a nonnegative vector μ such that $q(\mu) = q^*$. Then the dual optimal solution with minimum norm is informative.

Proof: Let μ^* be the dual optimal solution of minimum norm. If $\mu^* = 0$, then μ^* is an informative dual optimal solution and we are done, so assume that $\mu^* \neq 0$. Consider the problem

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x \in X, \; g(x) \leq \frac{1}{k^4} e. \end{aligned}$$

By Lemma 5.4.1, for each k, the optimal value of this problem is less than or equal to q^*. For each k, let $\tilde{x}^k \in X$ be a vector that satisfies

$$f(\tilde{x}^k) \leq q^* + \frac{1}{k^2}, \qquad g(\tilde{x}^k) \leq \frac{1}{k^4} e.$$

Consider also the problem

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; g(x) \leq \frac{1}{k^4} e, \\ &\qquad\qquad\;\; x \in \tilde{X}^k = X \cap \left\{ x \;\middle|\; \|x\| \leq k\big(\max_{1 \leq i \leq k} \|\tilde{x}^i\| + 1\big) \right\}. \end{aligned}$$

Since $\{x \mid g(x) \leq 0\} \cap X^k$ is the intersection of the closed set $\{x \in X \mid g(x) \leq 0\}$ and the compact set $\left\{x \;\middle|\; \|x\| \leq \|\tilde{x}^k\| + k\right\}$, we see that the constraint set of the preceding problem is compact. Since f is closed, and therefore lower semicontinuous over X^k, by Weierstrass' Theorem, the preceding minimization problem has an optimal solution, which we denote by $\overline{x}^k$. Note that since $\tilde{x}^k$ belongs to the feasible set of this problem, we have

$$f(\overline{x}^k) \leq f(\tilde{x}^k) \leq q^* + \frac{1}{k^2}. \tag{4.60}$$

We consider the function

$$L_k(x, \mu) = kf(x) + \mu' k g(x) - \frac{\|\mu\|^2}{2},$$

and we note that L_k is convex in x, and concave and coercive in μ. For each k, we consider the saddle points of L_k over x in

$$X^k = \tilde{X}^k \cap \big\{x \mid g(x) \le ke\big\}$$

and $\mu \ge 0$. Since X^k is bounded, L_k has a saddle point for each k, denoted by (x^k, μ^k).

Since L_k is quadratic in μ, the supremum of $L_k(x^k, \mu)$ over $\mu \ge 0$ is attained at

$$\mu_j^k = kg_j^+(x^k), \qquad j = 1, \ldots, r. \tag{4.61}$$

Similarly, the infimum of $L_k(x, \mu^k)$ over $x \in X^k$ is attained at x^k, implying that

$$\begin{aligned} f(x^k) + \mu^{k\prime} g(x^k) &= \inf_{x \in X^k} \big\{f(x) + \mu^{k\prime} g(x)\big\} \\ &= \inf_{x \in X^k} \big\{f(x) + kg_j^+(x^k) g(x)\big\} \\ &\le \inf_{x \in X^k,\ g(x) \le \frac{1}{k^4} e} \big\{f(x) + kg_j^+(x^k) g(x)\big\} \\ &\le \inf_{x \in X^k,\ g(x) \le \frac{1}{k^4} e} f(x) + \frac{r}{k^2} \\ &= f(\overline{x}^k) + \frac{r}{k^2} \\ &\le q^* + \frac{r+1}{k^2}, \end{aligned} \tag{4.62}$$

where the last inequality follows from Eq. (4.60).

We also have

$$\begin{aligned} L_k(x^k, \mu^k) &= \sup_{\mu \ge 0} \inf_{x \in X^k} L_k(x, \mu) \\ &\ge \sup_{\mu \ge 0} \inf_{x \in X} L_k(x, \mu) \\ &= \sup_{\mu \ge 0} \left\{ \inf_{x \in X} \big(kf(x) + \mu' k g(x)\big) - \frac{\|\mu\|^2}{2} \right\} \\ &= \sup_{\mu \ge 0} \left\{ kq(\mu) - \frac{\|\mu\|^2}{2} \right\} \\ &\ge kq(\mu^*) - \frac{\|\mu^*\|^2}{2}, \\ &= kq^* - \frac{\|\mu^*\|^2}{2}, \end{aligned} \tag{4.63}$$

where μ^* is the dual optimal solution with the minimum norm.

Combining Eqs. (4.63) and (4.62), we obtain

$$
\begin{aligned}
kq^* - \frac{1}{2}\|\mu^*\|^2 \le L_k(x^k,\mu^k) &= k\big\{f(x^k) + \mu^{k\prime} g(x^k)\big\} - \frac{1}{2}\|\mu^k\|^2 \\
&\le kq^* + \frac{2}{k} - \frac{1}{2}\|\mu^k\|^2.
\end{aligned}
\tag{4.64}
$$

It follows from the preceding relation that $\|\mu^k\|$ remains bounded as $k \to \infty$. Let $\overline{\mu}$ be a limit point of $\{\mu^k\}$. We also have from the preceding relation that

$$\lim_{k\to\infty} \big(f(x_k) - q^*\big) + \mu^{k\prime} g(x^k) = 0.$$

Hence, taking the limit along the relevant subsequence in the first relation in Eq. (4.62) yields

$$q^* \le \inf_{x\in X} \big\{f(x) + \overline{\mu}' g(x)\big\} = q(\overline{\mu}) \le q^*.$$

Hence $\overline{\mu}$ is a dual optimal solution, and since $\|\overline{\mu}\| \le \|\mu^*\|$ [which follows by taking the limit in Eq. (4.64)], by using the minimum norm property of μ^*, we conclude that any limit point $\overline{\mu}$ of μ^k must be equal to μ^*. Thus $\mu^k \to \mu^*$, and using Eq. (4.64), we obtain

$$\lim_{k\to\infty} \big(L_k(x^k,\mu^k) - kq^*\big) = -\frac{1}{2}\|\mu^*\|^2. \tag{4.65}$$

Using Eq. (4.61), it follows that

$$L_k(x^k,\mu^k) = \sup_{\mu\ge 0} L_k(x^k,\mu) = kf(x^k) + \frac{1}{2}\|\mu^k\|^2,$$

which combined with Eq. (4.65) yields

$$\lim_{k\to\infty} k\big(f(x^k) - q^*\big) = -\|\mu^*\|^2.$$

implying that $f(x^k) < q^*$ for all sufficiently large k, since $\mu^* \ne 0$. Since, $\mu^k \to \mu^*$, Eq. (4.61) also implies

$$\lim_{k\to\infty} kg_j^+(x^k) = \mu_j^*, \qquad j = 1,\ldots,r,$$

thus completing the proof. **Q.E.D.**

CHAPTER 6

CONCLUSIONS

In this thesis, we present a new development of Lagrange multiplier theory that significantly differs from the classical treatments. Our objective is to generalize, unify, and streamline the theory of constraint qualifications, which are conditions on the constraint set that guarantee existence of Lagrange multipliers.

Our analysis is motivated by an enhanced set of necessary optimality conditions of the Fritz John-type, which are stronger than the classical Karush-Kuhn-Tucker conditions (they include extra conditions, which may narrow down the set of candidate optima). They are also more general in that they apply even when there is a possibly nonconvex abstract set constraint, in addition to smooth equality and inequality constraints. For this purpose, we use concepts from nonsmooth analysis to analyze the local structure of the abstract set constraint. We show that the notion of 'regularity of constraint sets' is a crucial property in identifying problems that have satisfactory Lagrange multiplier theory.

A Lagrange multiplier theory should determine the fundamental constraint set structure that guarantees the existence of Lagrange multipliers. Without an abstract set constraint, this structure is identified by the notion of quasiregularity. The classical line of analysis has been either to relate constraint qualifications to quasiregularity, or to show existence of Lagrange multipliers under each constraint qualification separately, using a different and complicated proof. In the presence of an abstract set constraint, quasiregularity fails as a central unification concept, as we have shown in our work. Based on the enhanced Fritz John conditions, we introduce a new general constraint qualification, called pseudonormality. Pseudonormality unifies and expands the major constraint qualifications, and simplifies the proofs of Lagrange multiplier theorems.

Fritz John conditions also motivate us to introduce a taxonomy of different types of Lagrange multipliers. In particular, under mild convexity assumptions, we show that there exists a special Lagrange multiplier, called informative. The nonzero components of informative Lagrange multipliers identify the constraints that need to be violated in order to improve the optimal cost function value.

A notion that is related to pseudonormality, called quasinormality, is given by Hestenes [Hes75] (for the case where $X = \Re^n$). In this thesis, we extend this notion to the case where X is a closed set and we discuss the relation between pseudonormality and quasinormality. We show that pseudonormality is better suited as a unifying vehicle for Lagrange multiplier theory. Quasinormality serves almost the same purpose as pseudonormality when X is regular, but fails to provide the desired theoretical unification when X is not regular. For this reason, it appears that pseudonormality is a theoretically more interesting characteristic than quasinormality.

In this thesis, we also examine the connection between Lagrange multiplier theory and exact penalty functions. In particular, we show that pseudonormality implies the admittance of an exact penalty function. This provides in a unified way a much larger set of constraint qualifications under which we can guarantee that the constraint set admits an exact penalty.

Using a different line of analysis that does not involve gradients or subgradients, we extend the theory we developed regarding Fritz John conditions and pseudonormality to nonsmooth problems under convexity assumptions. Finally, we consider problems that do not necessarily have an optimal solution. We introduce a new notion of a multiplier, called geometric, that is not tied to a specific optimal solution. We develop Fritz John optimality conditions for such problems under different sets of assumptions. In particular, under convexity assumptions, we derive Fritz John conditions, which provides an alternative approach to obtain strong duality results of convex programming. Under additional closedness assumptions, we develop Fritz John conditions that involve conditions analogous to the complementary violation condition. This motivates us to introduce special types of geometric multipliers, called informative (consistent with informative Lagrange multipliers), that carry significant amount of sensitivity information regarding the constraints of the problem. We show that if the set of geometric multipliers is nonempty, then there exists an informative geometric multiplier. We also consider a dual optimization problem associated with the original problem, and we derive Fritz John-type optimality conditions for the dual problem. When there is no duality gap, the set of geometric multipliers and the set of optimal solutions of the dual problem coincide. When there is a duality gap, there exists no geometric multiplier; however the dual problem may still have an optimal solution. Based on dual Fritz John optimality conditions, we introduce special types of dual optimal solutions, called informative (similar to informative geometric multipliers), that carries sensitivity information. We show that an informative dual optimal solution always exists when the dual problem has an optimal solution. An interesting research direction for the future is to use the sensitivity information provided by the informative multipliers that we have defined in this thesis in various computational methods.

REFERENCES

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

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

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

[BNO02] Bertsekas, D. P., with Nedić, A., and Ozdaglar, A. E., 2002. Lectures Notes on Convexity, Duality, and Lagrange Multipliers, M.I.T., Cambridge, MA.

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

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

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

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

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

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

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

[Gal67] Gale, D., 1967. "A Geometric Duality Theorem with Economic Applications," Review of Economic Studies, 34, pp. 19-24.

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

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

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

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

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

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

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

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

[Kar39] Karush, W., 1939. "Minima of Functions of Several Variables with Inequalities as Side Conditions," Master's Thesis, Department of Mathematics, University of Chicago.

[Kuh76] Kuhn, H. W., 1976. "Nonlinear Programming: A Historical View," in Nonlinear Programming, Cottle, R. W., Lemke, C. E., (Eds.) SIAM-AMS Proc., Vol IX, American Math. Soc., Providence, RI, pp. 1-26.

[KuT50] Kuhn, H. W., and Tucker, A. W., 1950. "Nonlinear Programming," in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 481-492.

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

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

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

[Neu28] Neumann, von J., 1928. "Zur Theorie der Gesellschaftsspiele," Math. Ann., 100, pp. 295-320.

[NeM44] Neumann, von J., and Morgenstern, O., 1944. Theory of Games and Economic Behavior, Princeton University Press, Princeton, N. J.

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

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

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

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

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

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

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

[Tak74] Takayama, A., 1974. Mathematical Economics, Hinsdale, Illinois, Dryden Press.

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