Chapter 7 BASIC PROPERTIES OF SOLUTIONS AND ALGORITHMS

In this chapter we consider optimization problems of the form

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \Omega, \end{array} \tag{1}$$

where f is a real-valued function and Ω, the feasible set, is a subset of E^n. Throughout most of the chapter attention is restricted to the case where $\Omega = E^n$, corresponding to the completely unconstrained case, but sometimes we consider cases where Ω is some particularly simple subset of E^n.

The first and third sections of the chapter characterize the first- and second-order conditions that must hold at a solution point of (1). These conditions are simply extensions to E^n of the well-known derivative conditions for a function of a single variable that hold at a maximum or a minimum point. The fourth and fifth sections of the chapter introduce the important classes of convex and concave functions that provide zeroth-order conditions as well as a natural formulation for a global theory of optimization and provide geometric interpretations of the derivative conditions derived in the first two sections.

The final sections of the chapter are devoted to basic convergence characteristics of algorithms. Although this material is not exclusively applicable to optimization problems but applies to general iterative algorithms for solving other problems as well, it can be regarded as a fundamental prerequisite for a modern treatment of optimization techniques. Two essential questions are addressed concerning iterative algorithms. The first question, which is qualitative in nature, is whether a given algorithm in some sense yields, at least in the limit, a solution to the original problem. This question is treated in Section 7.6, and conditions sufficient to guarantee appropriate convergence are established. The second question, the more quantitative one, is related to how fast the algorithm converges to a solution. This question is defined more precisely in Section 7.7. Several special types of convergence, which arise frequently in the development of algorithms for optimization, are explored.

7.1 FIRST-ORDER NECESSARY CONDITIONS

Perhaps the first question that arises in the study of the minimization problem (1) is whether a solution exists. The main result that can be used to address this issue is the theorem of Weierstras, which states that if f is continuous and Ω is compact, a solution exists (see Appendix A.6). This is a valuable result that should be kept in mind throughout our development; however, our primary concern is with characterizing solution points and devising effective methods for finding them.

In an investigation of the general problem (1) we distinguish two kinds of solution points: *local minimum points*, and *global minimum points*.

Definition. A point $\mathbf{x}^* \in \Omega$ is said to be a *relative minimum point* or a *local minimum point* of f over Ω if there is an $\varepsilon > 0$ such that $f(\mathbf{x}) \geqslant f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega$ within a distance ε of $\mathbf{x}^*$ (that is, $\mathbf{x} \in \Omega$ and $|\mathbf{x}-\mathbf{x}^*| < \varepsilon$). If $f(\mathbf{x}) > f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega$, $\mathbf{x} \neq \mathbf{x}^*$, within a distance ε of $\mathbf{x}^*$, then $\mathbf{x}^*$ is said to be a *strict relative minimum point* of f over Ω.

Definition. A point $\mathbf{x}^* \in \Omega$ is said to be a *global minimum point* of f over Ω if $f(\mathbf{x}) \geqslant f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega$. If $f(\mathbf{x}) > f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega$, $\mathbf{x} \neq \mathbf{x}^*$, then $\mathbf{x}^*$ is said to be a *strict global minimum point* of f over Ω.

In formulating and attacking problem (1) we are, by definition, explicitly asking for a global minimum point of f over the set Ω. Practical reality, however, both from the theoretical and computational viewpoint, dictates that we must in many circumstances be content with a relative minimum point. In deriving necessary conditions based on the differential calculus, for instance, or when searching for the minimum point by a convergent stepwise procedure, comparisons of the values of nearby points is all that is possible and attention focuses on relative minimum points. Global conditions and global solutions can, as a rule, only be found if the problem possesses certain convexity properties that essentially guarantee that any relative minimum is a global minimum. Thus, in formulating and attacking problem (1) we shall, by the dictates of practicality, usually consider, implicitly, that we are asking for a relative minimum point. If appropriate conditions hold, this will also be a global minimum point.

Feasible Directions

To derive necessary conditions satisfied by a relative minimum point $\mathbf{x}^*$, the basic idea is to consider movement away from the point in some given direction. Along any given direction the objective function can be regarded as a function of a single variable, the parameter defining movement in this direction, and hence the ordinary calculus of a single variable is applicable. Thus given $\mathbf{x} \in \Omega$ we are motivated to say that a vector $\mathbf{d}$ is a *feasible direction at* $\mathbf{x}$ if there is an $\bar{\alpha} > 0$ such that $\mathbf{x} + \alpha\mathbf{d} \in \Omega$ for all α, $0 \leqslant \alpha \leqslant \bar{\alpha}$. With this simple concept we can state some simple conditions satisfied by relative minimum points.

Proposition 1 (First-order necessary conditions). *Let Ω be a subset of E^n and let $f \in C^1$ be a function on Ω. If $\mathbf{x}^*$ is a relative minimum point of f over Ω, then for any $\mathbf{d} \in E^n$ that is a feasible direction at $\mathbf{x}^*$, we have $\nabla f(\mathbf{x}^*)\mathbf{d} \geqslant 0$.*

Proof. For any α, $0 \leqslant \alpha \leqslant \bar{\alpha}$, the point $\mathbf{x}(\alpha) = \mathbf{x}^* + \alpha\mathbf{d} \in \Omega$. For $0 \leqslant \alpha \leqslant \bar{\alpha}$ define the function $g(\alpha) = f(\mathbf{x}(\alpha))$. Then g has a relative minimum at $\alpha = 0$. A typical g is shown in Fig. 7.1. By the ordinary calculus we have

$$g(\alpha) - g(0) = g'(0)\alpha + o(\alpha), \tag{2}$$

where $o(\alpha)$ denotes terms that go to zero faster than α (see Appendix A). If $g'(0) < 0$ then, for sufficiently small values of $\alpha > 0$, the right side of (2) will be negative, and hence $g(\alpha) - g(0) < 0$, which contradicts the minimal nature of $g(0)$. Thus $g'(0) = \nabla f(\mathbf{x}^*)\mathbf{d} \geqslant 0$. ▮

A very important special case is where $\mathbf{x}^*$ is in the interior of Ω (as would be the case if $\Omega = E^n$). In this case there are feasible directions emanating in every direction from $\mathbf{x}^*$, and hence $\nabla f(\mathbf{x}^*)\mathbf{d} \geqslant 0$ for all $\mathbf{d} \in E^n$. This implies $\nabla f(\mathbf{x}^*) = \mathbf{0}$. We state this important result as a corollary.

Corollary. (Unconstrained case). *Let Ω be a subset of E^n, and let $f \in C^1$ be a function' on Ω. If $\mathbf{x}^*$ is a relative minimum point of f over Ω and if $\mathbf{x}^*$ is an interior point of Ω, then $\nabla f(\mathbf{x}^*) = \mathbf{0}$.*

The necessary conditions in the pure unconstrained case lead to n equations (one for each component of ∇f) in n unknowns (the components of $\mathbf{x}^*$), which in many cases can be solved to determine the solution. In practice, however, as demonstrated in the following chapters, an optimization problem is solved directly without explicitly attempting to solve the equations arising from the necessary conditions. Nevertheless, these conditions form a foundation for the theory.

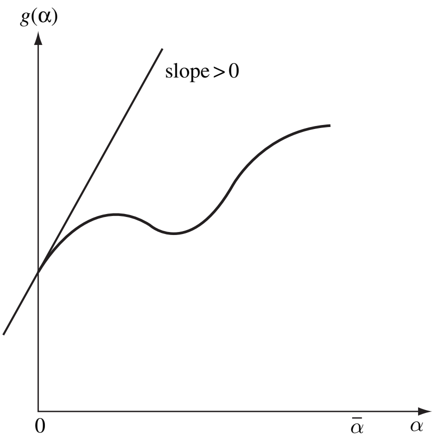

Fig. 7.1 Construction for proof

Example 1. Consider the problem

$$\text{minimize } f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2 - 3x_2.$$

There are no constraints, so $\Omega = E^2$. Setting the partial derivatives of f equal to zero yields the two equations

$$\begin{aligned} 2x_1 - x_2 &= 0 \\ -x_1 + 2x_2 &= 3. \end{aligned}$$

These have the unique solution $x_1 = 1$, $x_2 = 2$, which is a global minimum point of f.

Example 2. Consider the problem

$$\begin{aligned} &\text{minimize} && f(x_1, x_2) = x_1^2 - x_1 + x_2 + x_1x_2 \\ &\text{subject to} && x_1 \geqslant 0, \qquad x_2 \geqslant 0. \end{aligned}$$

This problem has a global minimum at $x_1 = \frac{1}{2}$, $x_2 = 0$. At this point

$$\frac{\partial f}{\partial x_1} = 2x_1 - 1 + x_2 = 0$$

$$\frac{\partial f}{\partial x_2} = 1 + x_1 = \tfrac{3}{2}.$$

Thus, the partial derivatives do not both vanish at the solution, but since any feasible direction must have an x_2 component greater than or equal to zero, we have $\nabla f(\mathbf{x}^*)\mathbf{d} \geqslant 0$ for all $\mathbf{d} \in E^2$ such that $\mathbf{d}$ is a feasible direction at the point (1/2, 0).

7.2 EXAMPLES OF UNCONSTRAINED PROBLEMS

Unconstrained optimization problems occur in a variety of contexts, but most frequently when the problem formulation is simple. More complex formulations often involve explicit functional constraints. However, many problems with constraints are frequently converted to unconstrained problems by using the constraints to establish relations among variables, thereby reducing the effective number of variables. We present a few examples here that should begin to indicate the wide scope to which the theory applies.

Example 1 (Production). A common problem in economic theory is the determination of the best way to combine various inputs in order to produce a certain commodity. There is a known production function $f(x_1, x_2, \ldots, x_n)$ that gives the amount of the commodity produced as a function of the amounts x_i of the inputs, $i = 1, 2, \ldots, n$. The unit price of the produced commodity is q, and the unit prices of the inputs are $p_1, p_2, \ldots, p_n$. The producer wishing to maximize profit must solve the problem

$$\text{maximize} \quad qf(x_1, x_2, \ldots, x_n) - p_1x_1 - p_2x_2 \ldots - p_nx_n.$$

The first-order necessary conditions are that the partial derivatives with respect to the x_i's each vanish. This leads directly to the n equations

$$q\frac{\partial f}{\partial x_i}(x_1,\ x_2, \ldots,\ x_n) = p_i, \qquad i = 1,\ 2, \ldots,\ n.$$

These equations can be interpreted as stating that, at the solution, the marginal value due to a small increase in the ith input must be equal to the price p_i.

Example 2 (Approximation). A common use of optimization is for the purpose of function approximation. Suppose, for example, that through an experiment the value of a function g is observed at m points, $x_1, x_2, \ldots, x_m$. Thus, values $g(x_1), g(x_2), \ldots, g(x_m)$ are known. We wish to approximate the function by a polynomial

$$h(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$$

of degree n (or less), where $n < m$. Corresponding to any choice of the approximating polynomial, there will be a set of errors $\varepsilon_k = g(x_k) - h(x_k)$. We define the best approximation as the polynomial that minimizes the sum of the squares of these errors; that is, minimizes

$$\sum_{k=1}^{m} (\varepsilon_k)^2.$$

This in turn means that we minimize

$$f(\mathbf{a}) = \sum_{k=1}^{m} [g(x_k) - (a_n x_k^n + a_{n-1}x_k^{n-1} + \ldots + a_0)]^2$$

with respect to $\mathbf{a} = (a_0, a_1, \ldots, a_n)$ to find the best coefficients. This is a quadratic expression in the coefficients $\mathbf{a}$. To find a compact representation for this objective we define $q_{ij} = \sum_{k=1}^{m} (x_k)^{i+j}$, $b_j = \sum_{k=1}^{m} g(x_k)(x_k)^j$ and $c = \sum_{k=1}^{m} g(x_k)^2$. Then after a bit of algebra it can be shown that

$$f(\mathbf{a}) = \mathbf{a}^T\mathbf{Q}\mathbf{a} - 2\mathbf{b}^T\mathbf{a} + c$$

where $\mathbf{Q} = [q_{ij}]$, $\mathbf{b} = (b_1, b_2, \ldots, b_{n+1})$.

The first-order necessary conditions state that the gradient of f must vanish. This leads directly to the system of $n+1$ equations

$$\mathbf{Q}\mathbf{a} = \mathbf{b}.$$

These can be solved to determine $\mathbf{a}$.

Example 3 (Selection problem). It is often necessary to select an assortment of factors to meet a given set of requirements. An example is the problem faced by an electric utility when selecting its power-generating facilities. The level of power that the company must supply varies by time of the day, by day of the week, and by season. Its power-generating requirements are summarized by a curve, $h(x)$, as shown in Fig. 7.2(a), which shows the total hours in a year that a power level of at least x is required for each x. For convenience the curve is normalized so that the upper limit is unity.

The power company may meet these requirements by installing generating equipment, such as (1) nuclear or (2) coal-fired, or by purchasing power from a central energy grid. Associated with type i $(i = 1, 2)$ of generating equipment is a yearly unit capital cost b_i and a unit operating cost c_i. The unit price of power purchased from the grid is c_3.

Nuclear plants have a high capital cost and low operating cost, so they are used to supply a base load. Coal-fired plants are used for the intermediate level, and power is purchased directly only for peak demand periods. The requirements are satisfied as shown in Fig. 7.2(b), where x_1 and x_2 denote the capacities of the nuclear and coal-fired plants, respectively. (For example, the nuclear power plant can be visualized as consisting of x_1/Δ small generators of capacity Δ, where Δ is small. The first such generator is on for about $h(\Delta)$ hours, supplying $\Delta h(\Delta)$ units of energy; the next supplies $\Delta h(2\Delta)$ units, and so forth. The total energy supplied by the nuclear plant is thus the area shown.)

The total cost is

$$f(x_1,\ x_2) = b_1x_1 + b_2x_2 + c_1 \int_0^{x_1} h(x)\ dx$$
$$+ c_2 \int_{x_1}^{x_1+x_2} h(x)\ dx + c_3 \int_{x_1+x_2}^{1} h(x)\ dx,$$

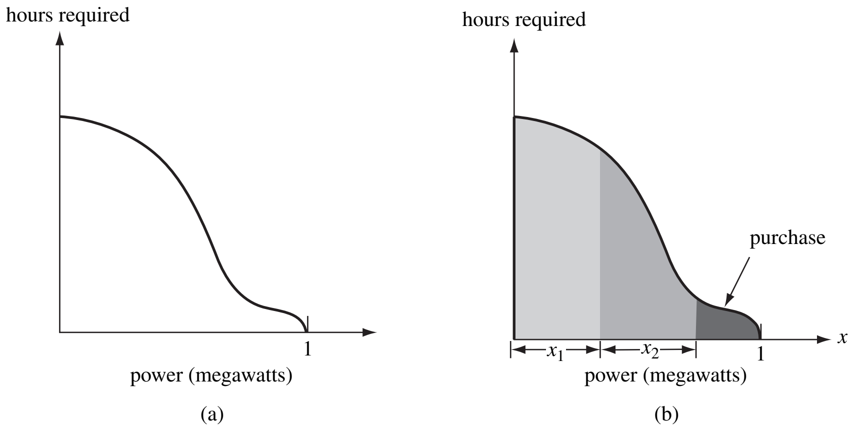

Fig. 7.2 Power requirements curve

and the company wishes to minimize this over the set defined by

$$x_1 \geqslant 0, \qquad x_2 \geqslant 0, \qquad x_1 + x_2 \leqslant 1.$$

Assuming that the solution is interior to the constraints, by setting the partial derivatives equal to zero, we obtain the two equations

$$b_1 + (c_1 - c_2)h(x_1) + (c_2 - c_3)h(x_1 + x_2) = 0$$
$$b_2 + (c_2 - c_3)h(x_1 + x_2) = 0,$$

which represent the necessary conditions.

If $x_1 = 0$, then the general necessary condition theorem shows that the first equality could relax to $\geqslant 0$. Likewise, if $x_2 = 0$, then the second equality could relax to $\geqslant 0$. The case $x_1 + x_2 = 1$ requires a bit more analysis (see Exercise 2).

Example 4 (Control). Dynamic problems, where the variables correspond to actions taken at a sequence of time instants, can often be formulated as unconstrained optimization problems. As an example suppose that the position of a large object is controlled by a series of corrective control forces. The error in position (the distance from the desired position) is governed by the equation

$$x_{k+1} = x_k + u_k,$$

where x_k is the error at time instant k, and u_k is the effective force applied at time u_k (after being normalized to account for the mass of the object and the duration of the force). The value of x_0 is given. The sequence $u_0, u_1, \ldots, u_n$ should be selected so as to minimize the objective

$$J = \sum_{k=0}^{n} \{x_k^2 + u_k^2\}.$$

This represents a compromise between a desire to have x_k equal to zero and recognition that control action u_k is costly.

The problem can be converted to an unconstrained problem by eliminating the x_k variables, $k = 1, 2, \ldots, n$, from the objective. It is readily seen that

$$x_k = x_0 + u_0 + u_1 + \cdots + u_{k-1}.$$

The objective can therefore be rewritten as

$$J = \sum_{k=0}^{n} \{(x_0 + u_0 + \cdots + u_{k-1})^2 + u_k^2\}.$$

This is a quadratic function in the unknowns u_k. It has the same general structure as that of Example 2 and it can be treated in a similar way.

7.3 SECOND-ORDER CONDITIONS

The proof of Proposition 1 in Section 7.1 is based on making a first-order approximation to the function f in the neighborhood of the relative minimum point. Additional conditions can be obtained by considering higher-order approximations. The second-order conditions, which are defined in terms of the Hessian matrix $\nabla^2 f$ of second partial derivatives of f (see Appendix A), are of extreme theoretical importance and dominate much of the analysis presented in later chapters.

Proposition 1 (Second-order necessary conditions). *Let Ω be a subset of E^n and let $f \in C^2$ be a function on Ω. If $\mathbf{x}^*$ is a relative minimum point of f over Ω, then for any $\mathbf{d} \in E^n$ that is a feasible direction at $\mathbf{x}^*$ we have*

i) $\nabla f(\mathbf{x}^*)\mathbf{d} \geqslant 0$ (3)

ii) *if* $\nabla f(\mathbf{x}^*)\mathbf{d} = 0,$ *then* $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geqslant 0.$ (4)

Proof. The first condition is just Proposition 1, and the second applies only if $\nabla f(\mathbf{x}^*)\mathbf{d} = 0$. In this case, introducing $\mathbf{x}(\alpha) = \mathbf{x}^* + \alpha\mathbf{d}$ and $g(\alpha) = f(\mathbf{x}(\alpha))$ as before, we have, in view of $g'(0) = 0$,

$$g(\alpha) - g(0) = \tfrac{1}{2} g''(0)\alpha^2 + o(\alpha^2).$$

If $g''(0) < 0$ the right side of the above equation is negative for sufficiently small α which contradicts the relative minimum nature of $g(0)$. Thus

$$g''(0) = \mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geqslant 0. \blacksquare$$

Example 1. For the same problem as Example 2 of Section 7.1, we have for $\mathbf{d} = (d_1, d_2)$

$$\nabla f(\mathbf{x}^*)\mathbf{d} = \tfrac{3}{2} d_2.$$

Thus condition (ii) of Proposition 1 applies only if $d_2 = 0$. In that case we have $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} = 2d_1^2 \geqslant 0$, so condition (ii) is satisfied.

Again of special interest is the case where the minimizing point is an interior point of Ω, as, for example, in the case of completely unconstrained problems. We then obtain the following classical result.

Proposition 2 (Second-order necessary conditions—unconstrained case). *Let $\mathbf{x}^*$ be an interior point of the set Ω, and suppose $\mathbf{x}^*$ is a relative minimum point over Ω of the function $f \in C^2$. Then*

i) $\nabla f(\mathbf{x}^*) = \mathbf{0}$ (5)

ii) *for all* $\mathbf{d}$, $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geqslant 0.$ (6)

For notational simplicity we often denote $\nabla^2 f(\mathbf{x})$, the $n \times n$ matrix of the second partial derivatives of f, the Hessian of f, by the alternative notation $\mathbf{F}(\mathbf{x})$. Condition (ii) is equivalent to stating that the matrix $\mathbf{F}(\mathbf{x}^*)$ is positive semidefinite. As we shall see, the matrix $\mathbf{F}(\mathbf{x}^*)$, which arises here quite naturally in a discussion of necessary conditions, plays a fundamental role in the analysis of iterative methods for solving unconstrained optimization problems. The structure of this matrix is the primary determinant of the rate of convergence of algorithms designed to minimize the function f.

Example 2. Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x_1,\ x_2) = x_1^3 - x_1^2 x_2 + 2x_2^2 \\ \text{subject to} & x_1 \geqslant 0, \qquad x_2 \geqslant 0. \end{array}$$

If we assume that the solution is in the interior of the feasible set, that is, if $x_1 > 0$, $x_2 > 0$, then the first-order necessary conditions are

$$3x_1^2 - 2x_1 x_2 = 0, \qquad -x_1^2 + 4x_2 = 0.$$

There is a solution to these at $x_1 = x_2 = 0$ which is a boundary point, but there is also a solution at $x_1 = 6$, $x_2 = 9$. We note that for x_1 fixed at $x_1 = 6$, the objective attains a relative minimum with respect to x_2 at $x_2 = 9$. Conversely, with x_2 fixed at $x_2 = 9$, the objective attains a relative minimum with respect to x_1 at $x_1 = 6$. Despite this fact, the point $x_1 = 6$, $x_2 = 9$ is not a relative minimum point, because the Hessian matrix is

$$\mathbf{F} = \begin{bmatrix} 6x_1 - 2x_2 & -2x_1 \\ -2x_1 & 4 \end{bmatrix},$$

which, evaluated at the proposed solution $x_1 = 6$, $x_2 = 9$, is

$$\mathbf{F} = \begin{bmatrix} 18 & -12 \\ -12 & 4 \end{bmatrix}.$$

This matrix is not positive semidefinite, since its determinant is negative. Thus the proposed solution is not a relative minimum point.

Sufficient Conditions for a Relative Minimum

By slightly strengthening the second condition of Proposition 2 above, we obtain a set of conditions that imply that the point $\mathbf{x}^*$ is a relative minimum. We give here the conditions that apply only to unconstrained problems, or to problems where the minimum point is interior to the feasible region, since the corresponding conditions for problems where the minimum is achieved on a boundary point of the feasible set are a good deal more difficult and of marginal practical or theoretical value. A more general result, applicable to problems with functional constraints, is given in Chapter 11.

Proposition 3 (Second-order sufficient conditions—unconstrained case). *Let $f \in C^2$ be a function defined on a region in which the point $\mathbf{x}^*$ is an interior point. Suppose in addition that*

i) $\nabla f(\mathbf{x}^*) = \mathbf{0}$ (7)

ii) $\mathbf{F}(\mathbf{x}^*)$ *is positive definite.* (8)

Then $\mathbf{x}^*$ is a strict relative minimum point of f.

Proof. Since $\mathbf{F}(\mathbf{x}^*)$ is positive definite, there is an $a > 0$ such that for all $\mathbf{d}$, $\mathbf{d}^T\mathbf{F}(\mathbf{x}^*)\mathbf{d} \geqslant a|\mathbf{d}|^2$. Thus by the Taylor's Theorem (with remainder)

$$\begin{aligned} f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) &= \tfrac{1}{2}\mathbf{d}^T\mathbf{F}(\mathbf{x}^*)\mathbf{d} + o(|\mathbf{d}|^2) \\ &\geqslant (a/2)|\mathbf{d}|^2 + o(|\mathbf{d}|^2) \end{aligned}$$

For small $|\mathbf{d}|$ the first term on the right dominates the second, implying that both sides are positive for small $\mathbf{d}$. ▮

7.4 CONVEX AND CONCAVE FUNCTIONS

In order to develop a theory directed toward characterizing global, rather than local, minimum points, it is necessary to introduce some sort of convexity assumptions. This results not only in a more potent, although more restrictive, theory but also provides an interesting geometric interpretation of the second-order sufficiency result derived above.

Definition. A function f defined on a convex set Ω is said to be *convex* if, for every $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and every α, $0 \leqslant \alpha \leqslant 1$, there holds

$$f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leqslant \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2).$$

If, for every α, $0 < \alpha < 1$, and $\mathbf{x}_1 \neq \mathbf{x}_2$, there holds

$$f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) < \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2),$$

then f is said to be *strictly convex*.

Several examples of convex or nonconvex functions are shown in Fig. 7.3. Geometrically, a function is convex if the line joining two points on its graph lies nowhere below the graph, as shown in Fig. 7.3(a), or, thinking of a function in two dimensions, it is convex if its graph is bowl shaped.

Next we turn to the definition of a concave function.

Definition. A function g defined on a convex set Ω is said to be *concave* if the function $f = -g$ is convex. The function g is *strictly concave* if $-g$ is strictly convex.

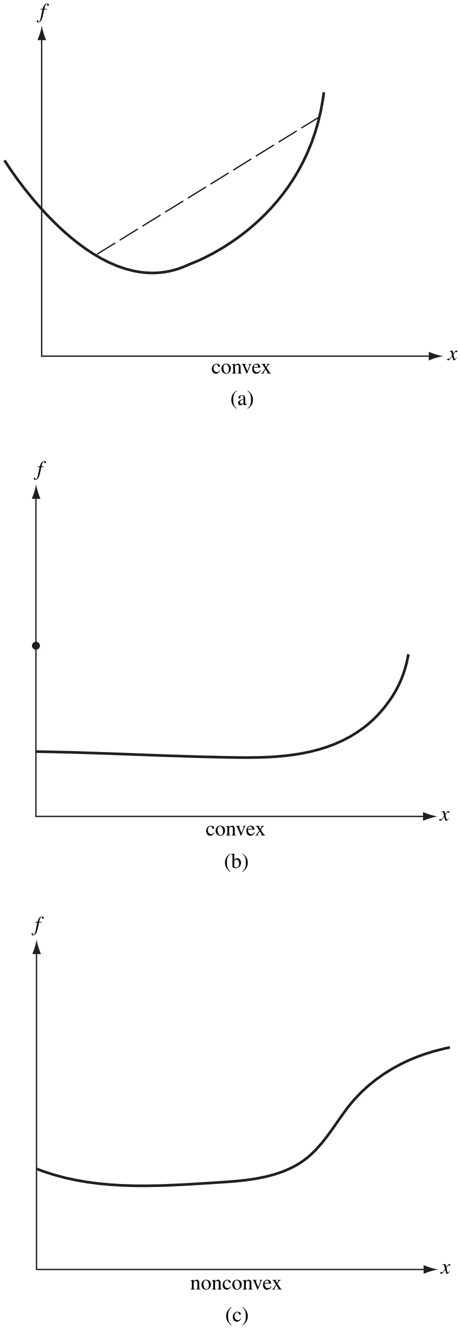

Fig. 7.3 Convex and nonconvex functions

Combinations of Convex Functions

We show that convex functions can be combined to yield new convex functions and that convex functions when used as constraints yield convex constraint sets.

Proposition 1. *Let f_1 and f_2 be convex functions on the convex set Ω. Then the function $f_1 + f_2$ is convex on Ω.*

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$, and $0 < \alpha < 1$. Then

$$f_1(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) + f_2(\alpha\mathbf{x}_1) + (1-\alpha)\mathbf{x}_2)$$
$$\leqslant \alpha[f_1(\mathbf{x}_1) + f_2(\mathbf{x}_1)] + (1-\alpha)[f_1(\mathbf{x}_2) + f_2(\mathbf{x}_2)]. \blacksquare$$

Proposition 2. *Let f be a convex function over the convex set Ω. Then the function af is convex for any $a \geqslant 0$.*

Proof. Immediate. ▮

Note that through repeated application of the above two propositions it follows that a positive combination $a_1 f_1 + a_2 f_2 + \ldots + a_m f_m$ of convex functions is again convex.

Finally, we consider sets defined by convex inequality constraints.

Proposition 3. *Let f be a convex function on a convex set Ω. The set $\Gamma_c = \{\mathbf{x} : \mathbf{x} \in \Omega, f(\mathbf{x}) \leqslant c\}$ is convex for every real number c.*

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in \Gamma_c$. Then $f(\mathbf{x}_1) \leqslant c$, $f(\mathbf{x}_2) \leqslant c$ and for $0 < \alpha < 1$,

$$f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leqslant \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \leqslant c.$$

Thus $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in \Gamma_c$. ▮

We note that, since the intersection of convex sets is also convex, the set of points simultaneously satisfying

$$f_1(\mathbf{x}) \leqslant c_1, \quad f_2(\mathbf{x}) \leqslant c_2, \ldots, f_m(\mathbf{x}) \leqslant c_m,$$

where each f_i is a convex function, defines a convex set. This is important in mathematical programming, since the constraint set is often defined this way.

Properties of Differentiable Convex Functions

If a function f is differentiable, then there are alternative characterizations of convexity.

Proposition 4. *Let $f \in C^1$. Then f is convex over a convex set Ω if and only if*

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}) \tag{9}$$

for all $\mathbf{x}, \mathbf{y} \in \boldsymbol{\Omega}$.

Proof. First suppose f is convex. Then for all α, $0 \leqslant \alpha \leqslant 1$,

$$f(\alpha\mathbf{y} + (1-\alpha)\mathbf{x}) \leqslant \alpha f(\mathbf{y}) + (1-\alpha)f(\mathbf{x}).$$

Thus for $0 < \alpha \leqslant 1$

$$\frac{f(\mathbf{x} + \alpha(\mathbf{y}-\mathbf{x})) - f(\mathbf{x})}{\alpha} \leqslant f(\mathbf{y}) - f(\mathbf{x}).$$

Letting $\alpha \to 0$ we obtain

$$\boldsymbol{\nabla} f(\mathbf{x})(\mathbf{y}-\mathbf{x}) \leqslant f(\mathbf{y}) - f(\mathbf{x}).$$

This proves the "only if" part.

Now assume

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \boldsymbol{\nabla} f(\mathbf{x})(\mathbf{y}-\mathbf{x})$$

for all $\mathbf{x}$, $\mathbf{y} \in \Omega$. Fix $\mathbf{x}_1$, $\mathbf{x}_2 \in \Omega$ and α, $0 \leqslant \alpha \leqslant 1$. Setting $\mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ and alternatively $\mathbf{y} = \mathbf{x}_1$ or $\mathbf{y} = \mathbf{x}_2$, we have

$$f(\mathbf{x}_1) \geqslant f(\mathbf{x}) + \boldsymbol{\nabla} f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x}) \tag{10}$$

$$f(\mathbf{x}_2) \geqslant f(\mathbf{x}) + \boldsymbol{\nabla} f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}). \tag{11}$$

Multiplying (10) by α and (11) by $(1-\alpha)$ and adding, we obtain

$$\alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \geqslant f(\mathbf{x}) + \boldsymbol{\nabla} f(\mathbf{x})[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 - \mathbf{x}].$$

But substituting $\mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$, we obtain

$$\alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \geqslant f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2). \blacksquare$$

The statement of the above proposition is illustrated in Fig. 7.4. It can be regarded as a sort of dual characterization of the original definition illustrated in Fig. 7.3. The original definition essentially states that linear interpolation between two points overestimates the function, while the above proposition states that linear approximation based on the local derivative underestimates the function.

For twice continuously differentiable functions, there is another characterization of convexity.

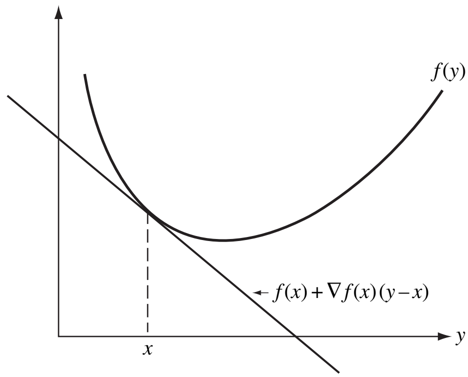

Fig. 7.4 Illustration of Proposition 4

Proposition 5. *Let $f \in C^2$. Then f is convex over a convex set Ω containing an interior point if and only if the Hessian matrix $\mathbf{F}$ of f is positive semidefinite throughout Ω.*

Proof. By Taylor's theorem we have

$$f(\mathbf{y}) = f(\mathbf{x}) = \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}) + \tfrac{1}{2}(\mathbf{y}-\mathbf{x})^T\mathbf{F}(\mathbf{x}+\alpha(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x}) \tag{12}$$

for some α, $0 \leqslant \alpha \leqslant 1$. Clearly, if the Hessian is everywhere positive semidefinite, we have

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}), \tag{13}$$

which in view of Proposition 4 implies that f is convex.

Now suppose the Hessian is not positive semidefinite at some point $\mathbf{x} \in \Omega$. By continuity of the Hessian it can be assumed, without loss of generality, that $\mathbf{x}$ is an interior point of Ω. There is a $\mathbf{y} \in \Omega$ such that $(\mathbf{y}-\mathbf{x})^T\mathbf{F}(\mathbf{x})(\mathbf{y}-\mathbf{x}) < 0$. Again by the continuity of the Hessian, $\mathbf{y}$ may be selected so that for all α, $0 \leqslant \alpha \leqslant 1$,

$$(\mathbf{y}-\mathbf{x})^T\mathbf{F}(\mathbf{x}+\alpha(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x}) < 0.$$

This in view of (12) implies that (13) does not hold; which in view of Proposition 4 implies that f is not convex. ▌

The Hessian matrix is the generalization to E^n of the concept of the curvature of a function, and correspondingly, positive definiteness of the Hessian is the generalization of positive curvature. Convex functions have positive (or at least nonnegative) curvature in every direction. Motivated by these observations, we sometimes refer to a function as being *locally convex* if its Hessian matrix is positive semidefinite in a small region, and *locally strictly convex* if the Hessian is positive definite in the region. In these terms we see that the second-order sufficiency result of the last section requires that the function be locally strictly convex at the point $\mathbf{x}^*$. Thus, even the local theory, derived solely in terms of the elementary calculus, is actually intimately related to convexity—at least locally. For this reason we can view the two theories, local and global, not as disjoint parallel developments but as complementary and interactive. Results that are based on convexity apply even to nonconvex problems in a region near the solution, and conversely, local results apply to a global minimum point.

7.5 MINIMIZATION AND MAXIMIZATION OF CONVEX FUNCTIONS

We turn now to the three classic results concerning minimization or maximization of convex functions.

Theorem 1. *Let f be a convex function defined on the convex set Ω. Then the set Γ where f achieves its minimum is convex, and any relative minimum of f is a global minimum.*

Proof. If f has no relative minima the theorem is valid by default. Assume now that c_0 is the minimum of f. Then clearly $\Gamma = \{\mathbf{x} : f(\mathbf{x}) \leqslant c_0, \mathbf{x} \in \Omega\}$ and this is convex by Proposition 3 of the last section.

Suppose now that $\mathbf{x}^* \in \Omega$ is a relative minimum point of f, but that there is another point $\mathbf{y} \in \Omega$ with $f(\mathbf{y}) < f(\mathbf{x}^*)$. On the line $\alpha\mathbf{y} + (1-\alpha)\mathbf{x}^*$, $0 < \alpha < 1$ we have

$$f(\alpha\mathbf{y} + (1-\alpha)\mathbf{x}^*) \leqslant \alpha f(\mathbf{y}) + (1-\alpha) f(\mathbf{x}^*) < f(\mathbf{x}^*),$$

contradicting the fact that $\mathbf{x}^*$ is a relative minimum point. ▮

We might paraphrase the above theorem as saying that for convex functions, all minimum points are located together (in a convex set) and all relative minima are global minima. The next theorem says that if f is continuously differentiable and convex, then satisfaction of the first-order necessary conditions are both necessary and sufficient for a point to be a global minimizing point.

Theorem 2. *Let $f \in C^1$ be convex on the convex set Ω. If there is a point $\mathbf{x}^* \in \Omega$ such that, for all $\mathbf{y} \in \Omega$, $\nabla f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geqslant 0$, then $\mathbf{x}^*$ is a global minimum point of f over Ω.*

Proof. We note parenthetically that since $\mathbf{y} - \mathbf{x}^*$ is a feasible direction at $\mathbf{x}^*$, the given condition is equivalent to the first-order necessary condition stated in Section 7.1. The proof of the proposition is immediate, since by Proposition 4 of the last section

$$f(\mathbf{y}) \geqslant f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geqslant f(\mathbf{x}^*). \ ▮$$

Next we turn to the question of maximizing a convex function over a convex set. There is, however, no analog of Theorem 1 for maximization; indeed, the tendency is for the occurrence of numerous nonglobal relative maximum points. Nevertheless, it is possible to prove one important result. It is not used in subsequent chapters, but it is useful for some areas of optimization.

Theorem 3. *Let f be a convex function defined on the bounded, closed convex set Ω. If f has a maximum over Ω it is achieved at an extreme point of Ω.*

Proof. Suppose f achieves a global maximum at $\mathbf{x}^* \in \Omega$. We show first that this maximum is achieved at some boundary point of Ω. If $\mathbf{x}^*$ is itself a boundary point, then there is nothing to prove, so assume $\mathbf{x}^*$ is not a boundary point. Let L be any line passing through the point $\mathbf{x}^*$. The intersection of this line with Ω is an interval of the line L having end points $\mathbf{y}_1$, $\mathbf{y}_2$ which are boundary points of Ω, and we have $\mathbf{x}^* = \alpha\mathbf{y}_1 + (1-\alpha)\mathbf{y}_2$ for some α, $0 < \alpha < 1$. By convexity of f

$$f(\mathbf{x}^*) \leqslant \alpha f(\mathbf{y}_1) + (1-\alpha)f(\mathbf{y}_2) \leqslant \max\{f(\mathbf{y}_1),\ f(\mathbf{y}_2)\}.$$

Thus either $f(\mathbf{y}_1)$ or $f(\mathbf{y}_2)$ must be at least as great as $f(\mathbf{x}^*)$. Since $\mathbf{x}^*$ is a maximum point, so is either $\mathbf{y}_1$ or $\mathbf{y}_2$.

We have shown that the maximum, if achieved, must be achieved at a boundary point of Ω. If this boundary point, $\mathbf{x}^*$, is an extreme point of Ω there is nothing more to prove. If it is not an extreme point, consider the intersection of Ω with a supporting hyperplane H at $\mathbf{x}^*$. This intersection, T_1, is of dimension $n-1$ or less and the global maximum of f over T_1 is equal to $f(\mathbf{x}^*)$ and must be achieved at a boundary point $\mathbf{x}_1$ of T_1. If this boundary point is an extreme point of T_1, it is also an extreme point of Ω by Lemma 1, Section B.4, and hence the theorem is proved. If $\mathbf{x}_1$ is not an extreme point of T_1, we form T_2, the intersection of T_1 with a hyperplane in E^{n-1} supporting T_1 at $\mathbf{x}_1$. This process can continue at most a total of n times when a set T_n of dimension zero, consisting of a single point, is obtained. This single point is an extreme point of T_n and also, by repeated application of Lemma 1, Section B.4, an extreme point of Ω. ▌

7.6 ZERO-ORDER CONDITIONS

We have considered the problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \Omega \end{aligned} \tag{14}$$

to be unconstrained because there are no functional constraints of the form $g(\mathbf{x}) \leqslant b$ or $h(\mathbf{x}) = c$. However, the problem is of course constrained by the set Ω. This constraint influences the first- and second-order necessary and sufficient conditions through the relation between feasible directions and derivatives of the function f. Nevertheless, there is a way to treat this constraint without reference to derivatives. The resulting conditions are then of zero order. These necessary conditions require that the problem be convex is a certain way, while the sufficient conditions require no assumptions at all. The simplest assumptions for the necessary conditions are that Ω is a convex set and that f is a convex function on all of E^n.

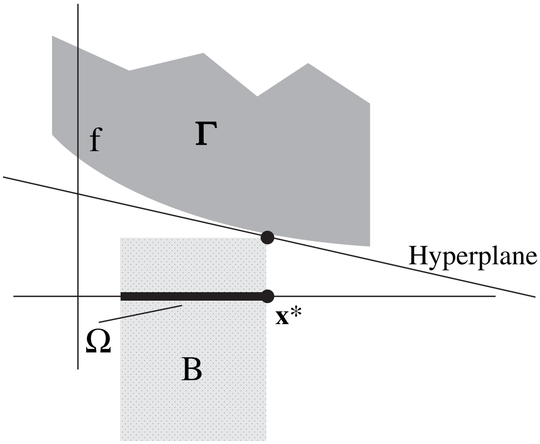

Fig. 7.5 The epigraph, the tubular region, and the hyperplane

To derive the necessary conditions under these assumptions consider the set $\Gamma \subset E^{n+1} = \{(r, \mathbf{x}) : r \geqslant f(\mathbf{x}), \mathbf{x} \in E^n\}$. In a figure of the graph of f, the set Γ is the region above the graph, shown in the upper part of Fig. 7.5. This set is called the *epigraph* of f. It is easy to verify that the set Γ is convex if f is a convex function.

Suppose that $\mathbf{x}^* \in \Omega$ is the minimizing point with value $f^* = f(\mathbf{x}^*)$. We construct a tubular region with cross section Ω and extending vertically from $-\infty$ up to f^*, shown as B in the upper part of Fig. 7.5. This is also a convex set, and it overlaps the set Γ only at the boundary point $(f^*, \mathbf{b}^*)$ above $\mathbf{x}^*$ (or possibly many boundary points if f is flat near $\mathbf{x}^*$).

According to the separating hyperplane theorem (Appendix B), there is a hyperplane separating these two sets. This hyperplane can be represented by a nonzero vector of the form $(s, \ \boldsymbol{\lambda}) \in E^{n+1}$ with s a scalar and $\boldsymbol{\lambda} \in E^n$, and a separation constant c. The separation conditions are

$$sr + \boldsymbol{\lambda}^T\mathbf{x} \geq c \quad \text{for all } \mathbf{x} \in E^n \text{ and } r \geq f(\mathbf{x}) \tag{15}$$

$$sr + \boldsymbol{\lambda}^T\mathbf{x} \leq c \quad \text{for all } \mathbf{x} \in \Omega \text{ and } r \leq f^*. \tag{16}$$

It follows that $s \neq 0$; for otherwise $\boldsymbol{\lambda} \neq \mathbf{0}$ and then (15) would be violated for some $\mathbf{x} \in E^n$. It also follows that $s \geqslant 0$ since otherwise (16) would be violated by very negative values of r. Hence, together we find $s > 0$ and by appropriate scaling we may take $s = 1$.

It is easy to see that the above conditions can be expressed alternatively as two optimization problems, as stated in the following proposition.

Proposition 1 (Zero-order necessary conditions). *If $\mathbf{x}^*$ solves (14) under the stated convexity conditions, then there is a nonzero vector $\boldsymbol{\lambda} \in E^n$ such that $\mathbf{x}^*$ is a solution to the two problems:*

$$\begin{aligned} &\textit{minimize} \quad f(\mathbf{x}) + \boldsymbol{\lambda}^T\mathbf{x} \\ &\textit{subject to} \quad \mathbf{x} \in E^n \end{aligned} \tag{17}$$

and

$$\begin{aligned} &\textit{maximize} \quad \boldsymbol{\lambda}^T\mathbf{x} \\ &\textit{subject to} \quad \mathbf{x} \in \Omega. \end{aligned} \tag{18}$$

Proof. Problem (17) follows from (15) (with $s = 1$) and the fact that $f(\mathbf{x}) \leq r$ for $r \geq f(\mathbf{x})$. The value c is attained from above at $(f^*, \mathbf{x}^*)$. Likewise (18) follows from (16) and the fact that $\mathbf{x}^*$ and the appropriate r attain c from below. ∎

Notice that problem (17) is completely unconstrained, since $\mathbf{x}$ may range over all of E^n. The second problem (18) is constrained by Ω but has a linear objective function.

It is clear from Fig. 7.5 that the slope of the hyperplane is equal to the slope of the function f when f is continuously differentiable at the solution $\mathbf{x}^*$.

If the optimal solution $\mathbf{x}^*$ is in the interior of Ω, then the second problem (18) implies that $\boldsymbol{\lambda} = \mathbf{0}$, for otherwise there would be a direction of movement from $\mathbf{x}^*$ that increases the product $\boldsymbol{\lambda}^T\mathbf{x}$ above $\boldsymbol{\lambda}^T\mathbf{x}^*$. The hyperplane is horizontal in that case. The zeroth-order conditions provide no new information in this situation. However, when the solution is on a boundary point of Ω the conditions give very useful information.

Example 1 (Minimization over an interval). Consider a continuously differentiable function f of a single variable $x \in E^1$ defined on the unit interval [0,1] which plays the role of Ω here. The first problem (17) implies $f'(x^*) = -\lambda$. If the solution is at the left end of the interval (at $x = 0$) then the second problem (18) implies that $\lambda \leq 0$ which means that $f'(x^*) \geq 0$. The reverse holds if x^* is at the right end. These together are identical to the first-order conditions of section 7.1.

Example 2 As a generalization of the above example, let $f \in C^1$ on E^n, and let f have a minimum with respect to Ω at $\mathbf{x}^*$. Let $\mathbf{d} \in E^n$ be a feasible direction at x^*. Then it follows again from (17) that $\nabla f(\mathbf{x}^*)\mathbf{d} \geq 0$.

Sufficient Conditions. The conditions of Proposition 1 are sufficient for $\mathbf{x}^*$ to be a minimum even without the convexity assumptions.

Proposition 2 (Zero-order sufficiency conditions). *If there is a* $\boldsymbol{\lambda}$ *such that* $\mathbf{x}^* \in \Omega$ *solves the problems (17) and (18), then* $\mathbf{x}^*$ *solves (14).*

Proof. Suppose $\mathbf{x}_1$ is any other point in Ω. Then from (17)

$$f(\mathbf{x}_1) + \boldsymbol{\lambda}^T\mathbf{x}_1 \geqslant f(\mathbf{x}^*) + \boldsymbol{\lambda}^T\mathbf{x}^*.$$

This can be rewritten as

$$f(\mathbf{x}_1) - f(\mathbf{x}^*) \geqslant \boldsymbol{\lambda}^T\mathbf{x}^* - \boldsymbol{\lambda}^T\mathbf{x}_1.$$

7.10 EXERCISES

1. To approximate a function g over the interval [0, 1] by a polynomial p of degree n (or less), we minimize the criterion

$$f(\mathbf{a}) = \int_0^1 [g(x) - p(x)]^2 \, dx,$$

where $p(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0$. Find the equations satisfied by the optimal coefficients $\mathbf{a} = (a_0, a_1, \ldots, a_n)$.

2. In Example 3 of Section 7.2 show that if the solution has $x_1 > 0$, $x_1 + x_2 = 1$, then it is necessary that

$$b_1 - b_2 + (c_1 - c_2)h(x_1) = 0$$
$$b_2 + (c_2 - c_3)h(x_1 + x_2) \leqslant 0.$$

Hint: One way is to reformulate the problem in terms of the variables x_1 and $y = x_1 + x_2$.

3. a) Using the first-order necessary conditions, find a minimum point of the function

$$f(x,\ y,\ z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9.$$

   b) Verify that the point is a relative minimum point by verifying that the second-order sufficiency conditions hold.
   c) Prove that the point is a global minimum point.

4. In this exercise and the next we develop a method for determining whether a given symmetric matrix is positive definite. Given an $n \times n$ matrix $\mathbf{A}$ let $\mathbf{A}_k$ denote the principal submatrix made up of the first k rows and columns. Show (by induction) that if the first $n-1$ principal submatrices are nonsingular, then there is a unique lower triangular matrix $\mathbf{L}$ with unit diagonal and a unique upper triangular matrix $\mathbf{U}$ such that $\mathbf{A} = \mathbf{LU}$. (See Appendix C.)

5. A symmetric matrix is positive definite if and only if the determinant of each of its principal submatrices is positive. Using this fact and the considerations of Exercise 4, show that an $n \times n$ symmetric matrix $\mathbf{A}$ is positive definite if and only if it has an **LU** decomposition (without interchange of rows) and the diagonal elements of $\mathbf{U}$ are all positive.

6. Using Exercise 5 show that an $n \times n$ matrix $\mathbf{A}$ is symmetric and positive definite if and only if it can be written as $\mathbf{A} = \mathbf{G}\mathbf{G}^T$ where $\mathbf{G}$ is a lower triangular matrix with positive diagonal elements. This representation is known as the *Cholesky factorization* of $\mathbf{A}$.

7. Let f_i, $i \in I$ be a collection of convex functions defined on a convex set Ω. Show that the function f defined by $f(\mathbf{x}) = \sup_{i \in I} f_i(\mathbf{x})$ is convex on the region where it is finite.

8. Let γ be a monotone nondecreasing function of a single variable (that is, $\gamma(r) \leqslant \gamma(r')$ for $r' > r$) which is also convex; and let f be a convex function defined on a convex set Ω. Show that the function $\gamma(f)$ defined by $\gamma(f)(\mathbf{x}) = \gamma[f(\mathbf{x})]$ is convex on Ω.

9. Let f be twice continuously differentiable on a region $\Omega \subset E^n$. Show that a sufficient condition for a point $\mathbf{x}^*$ in the interior of Ω to be a relative minimum point of f is that $\nabla f(\mathbf{x}^*) = \mathbf{0}$ and that f be locally convex at $\mathbf{x}^*$.

10. Define the point-to-set mapping on E^n by

$$\mathbf{A}(\mathbf{x}) = \{\mathbf{y} : \mathbf{y}^T\mathbf{x} \leqslant b\},$$

where b is a fixed constant. Is $\mathbf{A}$ closed?

11. Prove the two corollaries in Section 7.6 on the closedness of composite mappings.

12. Show that if $\mathbf{A}$ is a continuous point-to-point mapping, the Global Convergence Theorem is valid even without assumption (i). Compare with Example 2, Section 7.7.

13. Let $\{r_k\}_{k=0}^{\infty}$ and $\{c_k\}_{k=0}^{\infty}$ be sequences of real numbers. Suppose $r_k \to 0$ average linearly and that there are constants $c > 0$ and C such that $c \leqslant c_k \leqslant C$ for all k. Show that $c_k r_k \to 0$ average linearly.

14. Prove a proposition, similar to the one in Section 7.8, showing that the order of convergence is insensitive to the error function.

15. Show that if $r_k \to r^*$ (step-wise) linearly with convergence ratio β, then $r_k \to r^*$ (average) linearly with average convergence ratio no greater than β.

REFERENCES

7.1–7.5 For alternative discussions of the material in these sections, see Hadley [H2], Fiacco and McCormick [F4], Zangwill [Z2] and Luenberger [L8].

7.6 Although the general concepts of this section are well known, the formulation as zero-order conditions appears to be new.

7.7 The idea of using a descent function (usually the objective itself) in order to guarantee convergence of minimization algorithms is an old one that runs through most literature on optimization, and has long been used to establish global convergence. Formulation of the general Global Convergence Theorem, which captures the essence of many previously diverse arguments, and the idea of representing an algorithm as a point-to-set mapping are both due to Zangwill [Z2].

7.8 Most of the definitions given in this section have been standard for quite some time. A thorough discussion which contributes substantially to the unification of these concepts is contained in Ortega and Rheinboldt [O7].

Chapter 11 CONSTRAINED MINIMIZATION CONDITIONS

We turn now, in this final part of the book, to the study of minimization problems having constraints. We begin by studying in this chapter the necessary and sufficient conditions satisfied at solution points. These conditions, aside from their intrinsic value in characterizing solutions, define Lagrange multipliers and a certain Hessian matrix which, taken together, form the foundation for both the development and analysis of algorithms presented in subsequent chapters.

The general method used in this chapter to derive necessary and sufficient conditions is a straightforward extension of that used in Chapter 7 for unconstrained problems. In the case of equality constraints, the feasible region is a curved surface embedded in E^n. Differential conditions satisfied at an optimal point are derived by considering the value of the objective function along curves on this surface passing through the optimal point. Thus the arguments run almost identically to those for the unconstrained case; families of curves on the constraint surface replacing the earlier artifice of considering feasible directions. There is also a theory of zero-order conditions that is presented in the final section of the chapter.

11.1 CONSTRAINTS

We deal with general nonlinear programming problems of the form

$$\begin{array}{lll}
\text{minimize} & f(\mathbf{x}) & \\
\text{subject to} & h_1(\mathbf{x}) = 0 & g_1(\mathbf{x}) \leqslant 0 \\
 & h_2(\mathbf{x}) = 0 & g_2(\mathbf{x}) \leqslant 0 \\
 & \vdots & \vdots \\
 & h_m(\mathbf{x}) = 0 & g_p(\mathbf{x}) \leqslant 0 \\
 & \mathbf{x}, \in \Omega \subset E^n, &
\end{array} \tag{1}$$

where $m \leqslant n$ and the functions f, $h_i, i = 1, 2, \ldots, m$ and $g_j, j = 1, 2, \ldots, p$ are continuous, and usually assumed to possess continuous second partial derivatives. For notational simplicity, we introduce the vector-valued functions $\mathbf{h} = (h_1, h_2, \ldots, h_m)$ and $\mathbf{g} = (g_1, g_2, \ldots, g_P)$ and rewrite (1) as

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{h}(\mathbf{x}) = \mathbf{0},\ \mathbf{g}(\mathbf{x}) \leqslant \mathbf{0} \\ & \mathbf{x} \in \Omega. \end{array} \tag{2}$$

The constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leqslant \mathbf{0}$ are referred to as *functional constraints*, while the constraint $\mathbf{x} \in \Omega$ is a *set constraint*. As before we continue to de-emphasize the set constraint, assuming in most cases that either Ω is the whole space E^n or that the solution to (2) is in the interior of Ω. A point $\mathbf{x} \in \Omega$ that satisfies all the functional constraints is said to be *feasible*.

A fundamental concept that provides a great deal of insight as well as simplifying the required theoretical development is that of an *active constraint*. An inequality constraint $g_i(\mathbf{x}) \leqslant 0$ is said to be *active* at a feasible point $\mathbf{x}$ if $g_i(\mathbf{x}) = 0$ and *inactive* at $\mathbf{x}$ if $g_i(\mathbf{x}) < 0$. By convention we refer to any equality constraint $h_i(\mathbf{x}) = 0$ as *active* at any feasible point. The constraints active at a feasible point $\mathbf{x}$ restrict the domain of feasibility in neighborhoods of $\mathbf{x}$, while the other, inactive constraints, have no influence in neighborhoods of $\mathbf{x}$. Therefore, in studying the properties of a local minimum point, it is clear that attention can be restricted to the active constraints. This is illustrated in Fig. 11.1 where local properties satisfied by the solution $\mathbf{x}^*$ obviously do not depend on the inactive constraints g_2 and g_3.

It is clear that, if it were known *a priori* which constraints were active at the solution to (1), the solution would be a local minimum point of the problem defined by ignoring the inactive constraints and treating all active constraints as equality constraints. Hence, with respect to local (or relative) solutions, the problem could be regarded as having equality constraints only. This observation suggests that the majority of insight and theory applicable to (1) can be derived by consideration of equality constraints alone, later making additions to account for the selection of the

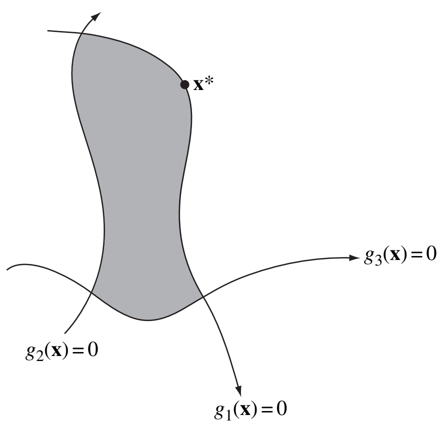

Fig. 11.1 Example of inactive constraints

active constraints. This is indeed so. Therefore, in the early portion of this chapter we consider problems having only equality constraints, thereby both economizing on notation and isolating the primary ideas associated with constrained problems. We then extend these results to the more general situation.

11.2 TANGENT PLANE

A set of equality constraints on E^n

$$\begin{aligned} h_1(\mathbf{x}) &= 0 \\ h_2(\mathbf{x}) &= 0 \\ &\vdots \\ h_m(\mathbf{x}) &= 0 \end{aligned} \tag{3}$$

defines a subset of E^n which is best viewed as a hypersurface. If the constraints are everywhere regular, in a sense to be described below, this hypersurface is of dimension $n-m$. If, as we assume in this section, the functions h_i, $i = 1, 2, \ldots, m$ belong to C^1, the surface defined by them is said to be *smooth*.

Associated with a point on a smooth surface is the *tangent plane* at that point, a term which in two or three dimensions has an obvious meaning. To formalize the general notion, we begin by defining curves on a surface. A *curve* on a surface S is a family of points $\mathbf{x}(t) \in S$ continuously parameterized by t for $a \leqslant t \leqslant b$. The curve is *differentiable* if $\dot{\mathbf{x}} \equiv (\mathrm{d}/\mathrm{d}t)\mathbf{x}(t)$ exists, and is *twice differentiable* if $\ddot{\mathbf{x}}(t)$ exists. A curve $\mathbf{x}(t)$ is said to pass through the point $\mathbf{x}^*$ if $\mathbf{x}^* = \mathbf{x}(t^*)$ for some t^*, $a \leqslant t^* \leqslant b$. The derivative of the curve at $\mathbf{x}^*$ is, of course, defined as $\dot{\mathbf{x}}(t^*)$. It is itself a vector in E^n.

Now consider all differentiable curves on S passing through a point $\mathbf{x}^*$. The *tangent plane* at $\mathbf{x}^*$ is defined as the collection of the derivatives at $\mathbf{x}^*$ of all these differentiable curves. The tangent plane is a subspace of E^n.

For surfaces defined through a set of constraint relations such as (3), the problem of obtaining an explicit representation for the tangent plane is a fundamental problem that we now address. Ideally, we would like to express this tangent plane in terms of derivatives of functions h_i that define the surface. We introduce the subspace

$$M = \{\mathbf{y} : \nabla\mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$$

and investigate under what conditions M is equal to the tangent plane at $\mathbf{x}^*$. The key concept for this purpose is that of a *regular point*. Figure 11.2 shows some examples where for visual clarity the tangent planes (which are sub-spaces) are translated to the point $\mathbf{x}^*$.

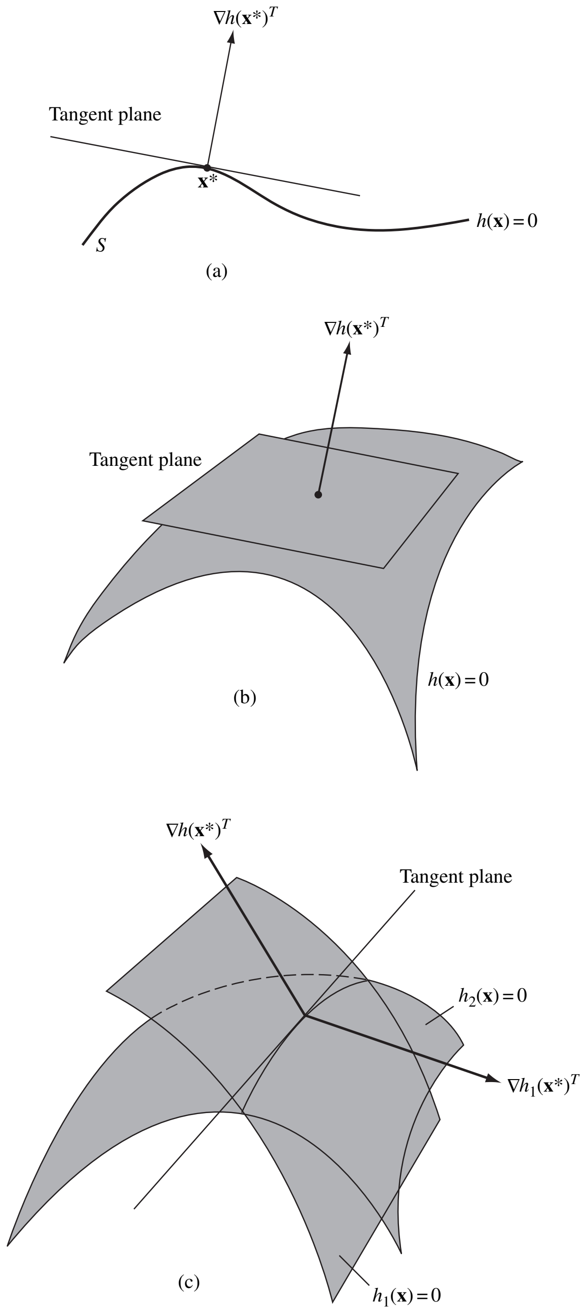

Fig. 11.2 Examples of tangent planes (translated to $\mathbf{x}^*$)

Definition. A point $\mathbf{x}^*$ satisfying the constraint $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ is said to be a *regular point* of the constraint if the gradient vectors $\nabla h_1(\mathbf{x}^*), \nabla h_2(\mathbf{x}^*), \ldots, \nabla h_m(\mathbf{x}^*)$ are linearly independent.

Note that if $\mathbf{h}$ is affine, $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$, regularity is equivalent to $\mathbf{A}$ having rank equal to m, and this condition is independent of $\mathbf{x}$.

In general, at regular points it is possible to characterize the tangent plane in terms of the gradients of the constraint functions.

Theorem. *At a regular point $\mathbf{x}^*$ of the surface S defined by $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ the tangent plane is equal to*

$$M = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}.$$

Proof. Let T be the tangent plane at $\mathbf{x}^*$. It is clear that $T \subset M$ whether $\mathbf{x}^*$ is regular or not, for any curve $\mathbf{x}(t)$ passing through $\mathbf{x}^*$ at $t = t^*$ having derivative $\dot{\mathbf{x}}(t^*)$ such that $\nabla \mathbf{h}(\mathbf{x}^*)\dot{\mathbf{x}}(t^*) \neq \mathbf{0}$ would not lie on S.

To prove that $M \subset T$ we must show that if $\mathbf{y} \in M$ then there is a curve on S passing through $\mathbf{x}^*$ with derivative $\mathbf{y}$. To construct such a curve we consider the equations

$$\mathbf{h}(\mathbf{x}^* + t\mathbf{y} + \nabla \mathbf{h}(\mathbf{x}^*)^T \mathbf{u}(t)) = \mathbf{0}, \tag{4}$$

where for fixed t we consider $\mathbf{u}(t) \in E^m$ to be the unknown. This is a nonlinear system of m equations and m unknowns, parameterized continuously, by t. At $t = 0$ there is a solution $\mathbf{u}(0) = \mathbf{0}$. The Jacobian matrix of the system with respect to $\mathbf{u}$ at $t = 0$ is the $m \times m$ matrix

$$\nabla \mathbf{h}(\mathbf{x}^*) \nabla \mathbf{h}(\mathbf{x}^*)^T,$$

which is nonsingular, since $\nabla \mathbf{h}(\mathbf{x}^*)$ is of full rank if $\mathbf{x}^*$ is a regular point. Thus, by the Implicit Function Theorem (see Appendix A) there is a continuously differentiable solution $\mathbf{u}(t)$ in some region $-a \leqslant t \leqslant a$.

The curve $\mathbf{x}(t) = \mathbf{x}^* + t\mathbf{y} + \nabla \mathbf{h}(\mathbf{x}^*)^T \mathbf{u}(t)$ is thus, by construction, a curve on S. By differentiating the system (4) with respect to t at $t = 0$ we obtain

$$\mathbf{0} = \frac{d}{dt}\mathbf{h}(\mathbf{x}(t))\bigg]_{t=0} = \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} + \nabla \mathbf{h}(\mathbf{x}^*) \nabla \mathbf{h}(\mathbf{x}^*)^T \dot{\mathbf{u}}(0).$$

By definition of $\mathbf{y}$ we have $\nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}$ and thus, again since $\nabla \mathbf{h}(\mathbf{x}^*) \nabla \mathbf{h}(\mathbf{x}^*)^T$ is nonsingular, we conclude that $\dot{\mathbf{x}}(0) = \mathbf{0}$. Therefore

$$\dot{\mathbf{x}}(0) = \mathbf{y} + \nabla \mathbf{h}(\mathbf{x}^*)^T \dot{\mathbf{x}}(0) = \mathbf{y},$$

and the constructed curve has derivative $\mathbf{y}$ at $\mathbf{x}^*$. ∎

It is important to recognize that the condition of being a regular point is not a condition on the constraint surface itself but on its representation in terms of an $\mathbf{h}$. The tangent plane is defined independently of the representation, while M is not.

Example. In E^2 let $h(x_1, x_2) = x_1$. Then $h(\mathbf{x}) = 0$ yields the x_2 axis, and every point on that axis is regular. If instead we put $h(x_1, x_2) = x_1^2$, again S is the x_2 axis but now no point on the axis is regular. Indeed in this case $M = E^2$, while the tangent plane is the x_2 axis.

11.3 FIRST-ORDER NECESSARY CONDITIONS (EQUALITY CONSTRAINTS)

The derivation of necessary and sufficient conditions for a point to be a local minimum point subject to equality constraints is fairly simple now that the representation of the tangent plane is known. We begin by deriving the first-order necessary conditions.

Lemma. *Let* $\mathbf{x}^*$ *be a regular point of the constraints* $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ *and a local extremum point (a minimum or maximum) of* f *subject to these constraints. Then all* $\mathbf{y} \in E^n$ *satisfying*

$$\nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0} \tag{5}$$

must also satisfy

$$\nabla f(\mathbf{x}^*)\mathbf{y} = 0. \tag{6}$$

Proof. Let $\mathbf{y}$ be any vector in the tangent plane at $\mathbf{x}^*$ and let $\mathbf{x}(t)$ be any smooth curve on the constraint surface passing through $\mathbf{x}^*$ with derivative $\mathbf{y}$ at $\mathbf{x}^*$; that is, $\mathbf{x}(0) = \mathbf{x}^*$, $\dot{\mathbf{x}}(0) = \mathbf{y}$, and $\mathbf{h}(\mathbf{x}(\mathbf{t})) = \mathbf{0}$ for $-a \leqslant t \leqslant a$ for some $a > 0$.

Since $\mathbf{x}^*$ is a regular point, the tangent plane is identical with the set of $\mathbf{y}$'s satisfying $\nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}$. Then, since $\mathbf{x}^*$ is a constrained local extremum point of f, we have

$$\left.\frac{d}{dt} f(\mathbf{x}(t))\right]_{t=0} = 0,$$

or equivalently,

$$\nabla f(\mathbf{x}^*)\mathbf{y} = 0. \quad \blacksquare$$

The above Lemma says that $\nabla f(\mathbf{x}^*)$ is orthogonal to the tangent plane. Next we conclude that this implies that $\nabla f(\mathbf{x}^*)$ is a linear combination of the gradients of $\mathbf{h}$ at $\mathbf{x}^*$, a relation that leads to the introduction of Lagrange multipliers.

Theorem. *Let* $\mathbf{x}^*$ *be a local extremum point of* f *subject to the constraints* $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. *Assume further that* $\mathbf{x}^*$ *is a regular point of these constraints. Then there is a* $\boldsymbol{\lambda} \in E^m$ *such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}. \tag{7}$$

Proof. From the Lemma we may conclude that the value of the linear program

$$\begin{aligned} &\text{maximize} && \nabla f(\mathbf{x}^*)\mathbf{y} \\ &\text{subject to} && \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0} \end{aligned}$$

is zero. Thus, by the Duality Theorem of linear programming (Section 4.2) the dual problem is feasible. Specifically, there is $\boldsymbol{\lambda} \in E^m$ such that $\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$. ▮

It should be noted that the first-order necessary conditions

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

together with the constraints

$$\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

give a total of $n+m$ (generally nonlinear) equations in the $n+m$ variables comprising $\mathbf{x}^*$, $\boldsymbol{\lambda}$. Thus the necessary conditions are a complete set since, at least locally, they determine a unique solution.

It is convenient to introduce the *Lagrangian* associated with the constrained problem, defined as

$$l(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}). \tag{8}$$

The necessary conditions can then be expressed in the form

$$\nabla_{\mathbf{x}} l(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0} \tag{9}$$

$$\nabla_{\boldsymbol{\lambda}} l(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}, \tag{10}$$

the second of these being simply a restatement of the constraints.

11.4 EXAMPLES

We digress briefly from our mathematical development to consider some examples of constrained optimization problems. We present five simple examples that can be treated explicitly in a short space and then briefly discuss a broader range of applications.

Example 1. Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_1x_2 + x_2x_3 + x_1x_3 \\ \text{subject to} \quad & x_1 + x_2 + x_3 = 3. \end{aligned}$$

The necessary conditions become

$$\begin{aligned} x_2 + x_3 + \lambda &= 0 \\ x_1 \qquad + x_3 + \lambda &= 0 \\ x_1 + x_2 \qquad + \lambda &= 0. \end{aligned}$$

These three equations together with the one constraint equation give four equations that can be solved for the four unknowns x_1, x_2, x_3, λ. Solution yields $x_1 = x_2 = x_3 = 1$, $\lambda = -2$.

Example 2 (Maximum volume). Let us consider an example of the type that is now standard in textbooks and which has a structure similar to that of the example above. We seek to construct a cardboard box of maximum volume, given a fixed area of cardboard.

Denoting the dimensions of the box by x, y, z, the problem can be expressed as

$$\begin{aligned} \text{maximize} \quad & xyz \\ \text{subject to} \quad & (xy + yz + xz) = \frac{c}{2}, \end{aligned} \tag{11}$$

where $c > 0$ is the given area of cardboard. Introducing a Lagrange multiplier, the first-order necessary conditions are easily found to be

$$\begin{aligned} yz + \lambda(y+z) &= 0 \\ xz + \lambda(x+z) &= 0 \\ xy + \lambda(x+y) &= 0 \end{aligned} \tag{12}$$

together with the constraint. Before solving these, let us note that the sum of these equations is $(xy + yz + xz) + 2\lambda(x + y + z) = 0$. Using the constraint this becomes $c/2 + 2\lambda(x + y + z) = 0$. From this it is clear that $\lambda \neq 0$. Now we can show that x, y, and z are nonzero. This follows because $x = 0$ implies $z = 0$ from the second equation and $y = 0$ from the third equation. In a similar way, it is seen that if either x, y, or z are zero, all must be zero, which is impossible.

To solve the equations, multiply the first by x and the second by y, and then subtract the two to obtain

$$\lambda(x - y)z = 0.$$

Operate similarly on the second and third to obtain

$$\lambda(y - z)x = 0.$$

Since no variables can be zero, it follows that $x = y = z = \sqrt{c/6}$ is the unique solution to the necessary conditions. The box must be a cube.

Example 3 (Entropy). Optimization problems often describe natural phenomena. An example is the characterization of naturally occurring probability distributions as maximum entropy distributions.

As a specific example consider a discrete probability density corresponding to a measured value taking one of n values $x_1, x_2, \ldots, x_n$. The probability associated with x_i is p_i. The p_i's satisfy $p_i \geqslant 0$ and $\sum_{i=1}^{n} p_i = 1$.

The *entropy* of such a density is

$$\varepsilon = -\sum_{i=1}^{n} p_i \log(p_i).$$

The *mean value* of the density is $\sum_{i=1}^{n} x_i p_i$.

If the value of mean is known to be m (by the physical situation), the maximum entropy argument suggests that the density should be taken as that which solves the following problem:

$$\begin{aligned} &\text{maximize} - \sum_{i=1}^{n} p_i \log(p_i) \\ &\text{subject to } \sum_{i=1}^{n} p_i = 1 \\ &\qquad\qquad \sum_{i=1}^{n} x_i p_i = m \\ &\qquad\qquad p_i \geqslant 0, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{13}$$

We begin by ignoring the nonnegativity constraints, believing that they may be inactive. Introducing two Lagrange multipliers, λ and μ, the Lagrangian is

$$l = \sum_{i=1}^{n} \{-p_i \log p_i + \lambda p_i + \mu x_i p_i\} - \lambda - \mu m.$$

The necessary conditions are immediately found to be

$$-\log p_i - 1 + \lambda + \mu x_i = 0, \quad i = 1, 2, \ldots, n.$$

This leads to

$$p_i = \exp\{(\lambda - 1) + \mu x_i\}, \quad i = 1, 2, \ldots, n. \tag{14}$$

We note that $p_i > 0$, so the nonnegativity constraints are indeed inactive. The result (14) is known as an exponential density. The Lagrange multipliers λ and μ are parameters that must be selected so that the two equality constraints are satisfied.

Example 4 (Hanging chain). A chain is suspended from two thin hooks that are 16 feet apart on a horizontal line as shown in Fig. 11.3. The chain itself consists of 20 links of stiff steel. Each link is one foot in length (measured inside). We wish to formulate the problem to determine the equilibrium shape of the chain.

The solution can be found by minimizing the potential energy of the chain. Let us number the links consecutively from 1 to 20 starting with the left end. We let link i span an x distance of x_i and a y distance of y_i. Then $x_i^2 + y_i^2 = 1$. The potential energy of a link is its weight times its vertical height (from some reference). The potential energy of the chain is the sum of the potential energies of each link. We may take the top of the chain as reference and assume that the mass of each link is concentrated at its center. Assuming unit weight, the potential energy is then

$$\frac{1}{2}y_1 + \left(y_1 + \frac{1}{2}y_2\right) + \left(y_1 + y_2 + \frac{1}{2}y_3\right) + \cdots$$
$$+ \left(y_1 + y_2 + \cdots + y_{n-1} + \frac{1}{2}y_n\right) = \sum_{i=1}^{n}\left(n - i + \frac{1}{2}\right)y_i,$$

where $n = 20$ in our example.

The chain is subject to two constraints: The total y displacement is zero, and the total x displacement is 16. Thus the equilibrium shape is the solution of

$$\begin{aligned} &\text{minimize} \sum_{i=1}^{n}\left(n - i + \frac{1}{2}\right)y_i \\ &\text{subject to } \sum_{i=1}^{n} y_i = 0 \\ &\qquad\qquad \sum_{i=1}^{n}\sqrt{1 - y_i^2} = 16. \end{aligned} \tag{15}$$

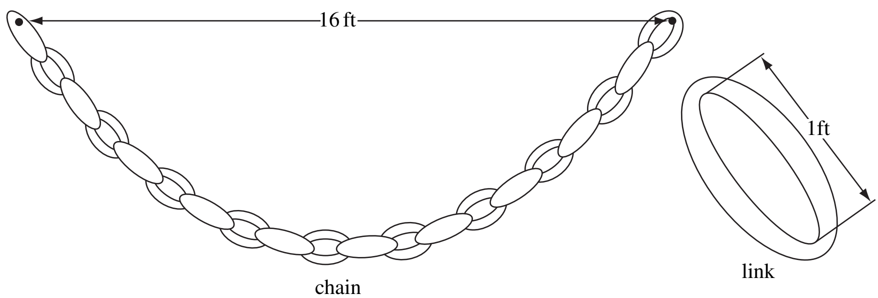

Fig. 11.3 A hanging chain

The first-order necessary conditions are

$$\left(n-i+\frac{1}{2}\right)+\lambda-\frac{\mu y_i}{\sqrt{1-y_i^2}}=0 \tag{16}$$

for $i=1,2,\ldots,n$. This leads directly to

$$y_i=-\frac{n-i+\frac{1}{2}+\lambda}{\sqrt{\mu^2+(n-i+\frac{1}{2}+\lambda)^2}}. \tag{17}$$

As in Example 2 the solution is determined once the Lagrange multipliers are known. They must be selected so that the solution satisfies the two constraints.

It is useful to point out that problems of this type may have local minimum points. The reader can examine this by considering a short chain of, say, four links and v and w configurations.

Example 5 (Portfolio design). Suppose there are n securities indexed by $i=1,2,\ldots,n$. Each security i is characterized by its random rate of return r_i which has mean value $\overline{r}_i$. Its covariances with the rates of return of other securtities are σ_{ij}, for $j=1,2,\ldots,n$. The portfolio problem is to allocate total available wealth among these n securities, allocating a fraction w_i of wealth to the security i.

The overall rate of return of a portfolio is $r=\sum_{i=1}^n w_i r_i$. This has mean value $\overline{r}=\sum_{i=1}^n w_i\overline{r}_i$ and variance $\sigma^2=\sum_{i,j=1}^n w_i\sigma_{ij}w_j$.

Markowitz introduced the concept of devising *efficient* portfolios which for a given expected rate of return $\overline{r}$ have minimum possible variance. Such a portfolio is the solution to the problem

$$\begin{aligned}\min_{w_i,w_2,\ldots,w_n}\ &\sum\nolimits_{i,j=1}^n w_i\sigma_{ij}w_j\\ \text{subject to}\ &\sum\nolimits_{i=1}^n w_i\overline{r}_i=\overline{r}\\ &\sum\nolimits_{i=1}^n w_i=1.\end{aligned}$$

The second constraint forces the sum of the weights to equal one. There may be the further restriction that each $w_i \geq 0$ which would imply that the securities must not be shorted (that is, sold short).

Introducing Lagrange multipliers λ and μ for the two constraints leads easily to the $n+2$ linear equations

$$\sum_{j=1}^n \sigma_{ij}w_j+\lambda\overline{r}_i+\mu=0 \quad \text{for } i=1,2,\ldots,n$$

$$\sum_{i=1}^n w_i\overline{r}_i=\overline{r}$$

$$\sum_{i=1}^n w_i=1$$

in the $n+2$ unknowns (the w_i's, λ and μ).

Large-Scale Applications

The problems that serve as the primary motivation for the methods described in this part of the book are actually somewhat different in character than the problems represented by the above examples, which by necessity are quite simple. Larger, more complex, nonlinear programming problems arise frequently in modern applied analysis in a wide variety of disciplines. Indeed, within the past few decades nonlinear programming has advanced from a relatively young and primarily analytic subject to a substantial general tool for problem solving.

Large nonlinear programming problems arise in problems of mechanical structures, such as determining optimal configurations for bridges, trusses, and so forth. Some mechanical designs and configurations that in the past were found by solving differential equations are now often found by solving suitable optimization problems. An example that is somewhat similar to the hanging chain problem is the determination of the shape of a stiff cable suspended between two points and supporting a load.

A wide assortment, of large-scale optimization problems arise in a similar way as methods for solving partial differential equations. In situations where the underlying continuous variables are defined over a two- or three-dimensional region, the continuous region is replaced by a grid consisting of perhaps several thousand discrete points. The corresponding discrete approximation to the partial differential equation is then solved indirectly by formulating an equivalent optimization problem. This approach is used in studies of plasticity, in heat equations, in the flow of fluids, in atomic physics, and indeed in almost all branches of physical science.

Problems of optimal control lead to large-scale nonlinear programming problems. In these problems a dynamic system, often described by an ordinary differential equation, relates control variables to a trajectory of the system state. This differential equation, or a discretized version of it, defines one set of constraints. The problem is to select the control variables so that the resulting trajectory satisfies various additional constraints and minimizes some criterion. An early example of such a problem that was solved numerically was the determination of the trajectory of a rocket to the moon that required the minimum fuel consumption.

There are many examples of nonlinear programming in industrial operations and business decision making. Many of these are nonlinear versions of the kinds of examples that were discussed in the linear programming part of the book. Nonlinearities can arise in production functions, cost curves, and, in fact, in almost all facets of problem formulation.

Portfolio analysis, in the context of both stock market investment and evaluation of a complex project within a firm, is an area where nonlinear programming is becoming increasingly useful. These problems can easily have thousands of variables.

In many areas of model building and analysis, optimization formulations are increasingly replacing the direct formulation of systems of equations. Thus large economic forecasting models often determine equilibrium prices by minimizing an objective termed *consumer surplus*. Physical models are often formulated as minimization of energy. Decision problems are formulated as maximizing expected utility. Data analysis procedures are based on minimizing an average error or maximizing a probability. As the methodology for solution of nonlinear programming improves, one can expect that this trend will continue.

11.5 SECOND-ORDER CONDITIONS

By an argument analogous to that used for the unconstrained case, we can also derive the corresponding second-order conditions for constrained problems. Throughout this section it is assumed that $f, \mathbf{h} \in C^2$.

Second-Order Necessary Conditions. *Suppose that $\mathbf{x}^*$ is a local minimum of f subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and that $\mathbf{x}^*$ is a regular point of these constraints. Then there is a $\boldsymbol{\lambda} \in E^m$ such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}. \tag{18}$$

If we denote by M the tangent plane $M = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$, then the matrix

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{H}(\mathbf{x}^*) \tag{19}$$

is positive semidefinite on M, that is, $\mathbf{y}^T \mathbf{L}(\mathbf{x}^*)\mathbf{y} \geqslant \mathbf{0}$ for all $\mathbf{y} \in M$.

Proof. From elementary calculus it is clear that for every twice differentiable curve on the constraint surface S through $\mathbf{x}^*$ (with $\mathbf{x}(0) = \mathbf{x}^*$) we have

$$\left.\frac{d^2}{dt^2} f(\mathbf{x}(t))\right]_{t=0} \geqslant 0. \tag{20}$$

By definition

$$\left.\frac{d^2}{dt^2} f(\mathbf{x}(t))\right]_{t=0} = \dot{\mathbf{x}}(0)^T \mathbf{F}(\mathbf{x}^*)\dot{\mathbf{x}}(0) + \nabla f(\mathbf{x}^*)\ddot{\mathbf{x}}(0). \tag{21}$$

Furthermore, differentiating the relation $\boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}(t)) = 0$ twice, we obtain

$$\dot{\mathbf{x}}(0)^T \boldsymbol{\lambda}^T \mathbf{H}(\mathbf{x}^*)\dot{\mathbf{x}}(0) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*)\ddot{\mathbf{x}}(0) = 0. \tag{22}$$

Adding (22) to (21), while taking account of (20), yields the result

$$\left.\frac{d^2}{dt^2} f(\mathbf{x}(t))\right]_{t=0} = \dot{\mathbf{x}}(0)^T \mathbf{L}(\mathbf{x}^*)\dot{\mathbf{x}}(0) \geqslant 0.$$

Since $\dot{\mathbf{x}}(0)$ is arbitrary in M, we immediately have the stated conclusion. ▮

The above theorem is our first encounter with the matrix $\mathbf{L} = \mathbf{F} + \boldsymbol{\lambda}^T \mathbf{H}$ which is the matrix of second partial derivatives, with respect to $\mathbf{x}$, of the Lagrangian l.

(See Appendix A, Section A.6, for a discussion of the notation $\boldsymbol{\lambda}^T\mathbf{H}$ used here.) This matrix is the backbone of the theory of algorithms for constrained problems, and it is encountered often in subsequent chapters.

We next state the corresponding set of sufficient conditions.

Second-Order Sufficiency Conditions. *Suppose there is a point* $\mathbf{x}^*$ *satisfying* $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$, *and a* $\boldsymbol{\lambda} \in E^m$ *such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}. \tag{23}$$

Suppose also that the matrix $\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \boldsymbol{\lambda}^T\mathbf{H}(\mathbf{x}^*)$ *is positive definite on* $M = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$, *that is, for* $\mathbf{y} \in M$, $\mathbf{y} \neq \mathbf{0}$ *there holds* $\mathbf{y}^T\mathbf{L}(\mathbf{x}^*)\mathbf{y} > \mathbf{0}$. *Then* $\mathbf{x}^*$ *is a strict local minimum of* f *subject to* $\mathbf{h}(\mathbf{x}) = \mathbf{0}$.

Proof. If $\mathbf{x}^*$ is not a strict relative minimum point, there exists a sequence of feasible points $\{\mathbf{y}_k\}$ converging to $\mathbf{x}^*$ such that for each k, $f(\mathbf{y}_k) \leqslant f(\mathbf{x}^*)$. Write each $\mathbf{y}_k$ in the form $\mathbf{y}_k = \mathbf{x}^* + \delta_k\mathbf{s}_k$ where $\mathbf{s}_k \in \mathbf{E}^n$, $|\mathbf{s}_k| = 1$, and $\delta_k > 0$ for each k. Clearly, $\delta_k \to 0$ and the sequence $\{\mathbf{s}_k\}$, being bounded, must have a convergent subsequence converging to some $\mathbf{s}^*$. For convenience of notation, we assume that the sequence $\{\mathbf{s}_k\}$ is itself convergent to $\mathbf{s}^*$. We also have $\mathbf{h}(\mathbf{y}_k) - \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$, and dividing by δ_k and letting $k \to \infty$ we see that $\nabla \mathbf{h}(\mathbf{x}^*)\mathbf{s}^* = \mathbf{0}$.

Now by Taylor's theorem, we have for each j

$$0 = h_j(\mathbf{y}_k) = h_j(\mathbf{x}^*) + \delta_k \nabla h_j(\mathbf{x}^*)\mathbf{s}_k + \frac{\delta_k^2}{2}\mathbf{s}_k^T \nabla^2 h_j(\boldsymbol{\eta}_j)\mathbf{s}_k \tag{24}$$

and

$$0 \geqslant f(\mathbf{y}_k) - f(\mathbf{x}^*) = \delta_k \nabla f(\mathbf{x}^*)\mathbf{s}_k + \frac{\delta_k^2}{2}\mathbf{s}_k^T \nabla^2 f(\boldsymbol{\eta}_0)\mathbf{s}_k, \tag{25}$$

where each $\boldsymbol{\eta}_j$ is a point on the line segment joining $\mathbf{x}^*$ and $\mathbf{y}_k$. Multiplying (24) by $\boldsymbol{\lambda}_j$ and adding these to (25) we obtain, on accounting for (23),

$$0 \geqslant \frac{\delta_k^2}{2}\mathbf{s}_k^T \left\{ \nabla^2 f(\boldsymbol{\eta}_0) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i \nabla^2 h_i(\boldsymbol{\eta}_i) \right\} \mathbf{s}_k,$$

which yields a contradiction as $k \to \infty$. ▮

Example 1. Consider the problem

$$\begin{array}{ll} \text{maximize} & x_1x_2 + x_2x_3 + x_1x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 3. \end{array}$$

In Example 1 of Section 11.4 it was found that $x_1 = x_2 = x_3 = 1$, $\lambda = -2$ satisfy the first-order conditions. The matrix $\mathbf{F} + \boldsymbol{\lambda}^T\mathbf{H}$ becomes in this case

$$L = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix},$$

which itself is neither positive nor negative definite. On the subspace $M = \{\mathbf{y} : y_1 + y_2 + y_3 = 0\}$, however, we note that

$$\begin{aligned} \mathbf{y}^T\mathbf{L}\mathbf{y} &= y_1(y_2 + y_3) + y_2(y_1 + y_3) + y_3(y_1 + y_2) \\ &= -(y_1^2 + y_2^2 + y_3^2), \end{aligned}$$

and thus $\mathbf{L}$ is negative definite on M. Therefore, the solution we found is at least a local maximum.

11.6 EIGENVALUES IN TANGENT SUBSPACE

In the last section it was shown that the matrix $\mathbf{L}$ restricted to the subspace M that is tangent to the constraint surface plays a role in second-order conditions entirely analogous to that of the Hessian of the objective function in the unconstrained case. It is perhaps not surprising, in view of this, that the structure of $\mathbf{L}$ restricted to M also determines rates of convergence of algorithms designed for constrained problems in the same way that the structure of the Hessian of the objective function does for unconstrained algorithms. Indeed, we shall see that the eigenvalues of $\mathbf{L}$ restricted to M determine the natural rates of convergence for algorithms designed for constrained problems. It is important, therefore, to understand what these restricted eigenvalues represent. We first determine geometrically what we mean by the restriction of $\mathbf{L}$ to M which we denote by $\mathbf{L}_M$. Next we define the eigenvalues of the operator $\mathbf{L}_M$. Finally we indicate how these various quantities can be computed.

Given any vector $\mathbf{y} \in M$, the vector $\mathbf{L}\mathbf{y}$ is in E^n but not necessarily in M. We project $\mathbf{L}\mathbf{y}$ orthogonally back onto M, as shown in Fig. 11.4, and the result is said to be the restriction of $\mathbf{L}$ to M operating on $\mathbf{y}$. In this way we obtain a linear transformation from M to M. The transformation is determined somewhat implicitly, however, since we do not have an explicit matrix representation.

A vector $\mathbf{y} \in M$ is an *eigenvector* of $\mathbf{L}_M$ if there is a real number λ such that $\mathbf{L}_M\mathbf{y} = \lambda\mathbf{y}$; the corresponding λ is an *eigenvalue* of $\mathbf{L}_M$. This coincides with the standard definition. In terms of $\mathbf{L}$ we see that $\mathbf{y}$ is an eigenvector of $\mathbf{L}_M$ if $\mathbf{L}\mathbf{y}$ can be written as the sum of $\lambda\mathbf{y}$ and a vector orthogonal to M. See Fig. 11.5.

To obtain a matrix representation for $\mathbf{L}_M$ it is necessary to introduce a basis in the subspace M. For simplicity it is best to introduce an orthonormal basis, say $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{n-m}$. Define the matrix $\mathbf{E}$ to be the $n \times (n - m)$ matrix whose columns

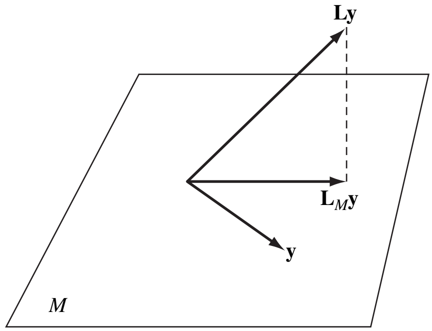

Fig. 11.4 Definition of $\mathbf{L}_M$

consist of the vectors $\mathbf{e}_i$. Then any vector $\mathbf{y}$ in M can be written as $\mathbf{y} = \mathbf{Ez}$ for some $\mathbf{z} \in E^{n-m}$ and, of course, $\mathbf{LEz}$ represents the action of $\mathbf{L}$ on such a vector. To project this result back into M and express the result in terms of the basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{n-m}$, we merely multiply by $\mathbf{E}^T$. Thus $\mathbf{E}^T\mathbf{LEz}$ is the vector whose components give the representation in terms of the basis; and, correspondingly, the $(n-m) \times (n-m)$ matrix $\mathbf{E}^T\mathbf{LE}$ is the matrix representation of $\mathbf{L}$ restricted to M.

The eigenvalues of $\mathbf{L}$ restricted to M can be found by determining the eigenvalues of $\mathbf{E}^T\mathbf{LE}$. These eigenvalues are independent of the particular orthonormal basis $\mathbf{E}$.

Example 1. In the last section we considered

$$L = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

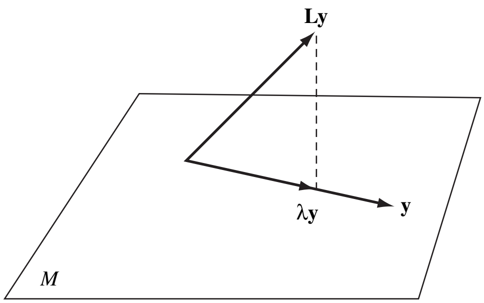

Fig. 11.5 Eigenvector of $\mathbf{L}_M$

restricted to $M = \{\mathbf{y} : y_1 + y_2 + y_3 = 0\}$. To obtain an explicit matrix representation on M let us introduce the orthonormal basis:

$$\mathbf{e}_1 = \frac{1}{\sqrt{2}}(1, 0, -1)$$
$$\mathbf{e}_2 = \frac{1}{\sqrt{6}}(1, -2, 1).$$

This gives, upon expansion,

$$\mathbf{E}^T\mathbf{L}\mathbf{E} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix},$$

and hence $\mathbf{L}$ restricted to M acts like the negative of the identity.

Example 2. Let us consider the problem

$$\begin{aligned} &\text{extremize} \quad x_1 + x_2^2 + x_2 x_3 + 2x_3^2 \\ &\text{subject to} \quad \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) = 1. \end{aligned}$$

The first-order necessary conditions are

$$\begin{aligned} 1 + \qquad \lambda x_1 &= 0 \\ 2x_2 + x_3 + \lambda x_2 &= 0 \\ x_2 + 4x_3 + \lambda x_3 &= 0. \end{aligned}$$

One solution to this set is easily seen to be $x_1 = 1$, $x_2 = 0$, $x_3 = 0$, $\lambda = -1$. Let us examine the second-order conditions at this solution point. The Lagrangian matrix there is

$$\mathbf{L} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 3 \end{bmatrix},$$

and the corresponding subspace M is

$$M = \{\mathbf{y} : y_1 = 0\}.$$

In this case M is the subspace spanned by the second two basis vectors in E^3 and hence the restriction of $\mathbf{L}$ to M can be found by taking the corresponding submatrix of $\mathbf{L}$. Thus, in this case,

$$\mathbf{E}^T\mathbf{L}\mathbf{E} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}.$$

The characteristic polynomial of this matrix is

$$\det\begin{bmatrix} 1-\lambda & 1 \\ 1 & 3-\lambda \end{bmatrix} = (1-\lambda)(3-\lambda)-1 = \lambda^2-4\lambda+2.$$

The eigenvalues of $\mathbf{L}_M$ are thus $\lambda = 2\pm\sqrt{2}$, and $\mathbf{L}_M$ is positive definite.

Since the $\mathbf{L}_M$ matrix is positive definite, we conclude that the point found is a relative minimum point. This example illustrates that, in general, the restriction of $\mathbf{L}$ to M can be thought of as a submatrix of $\mathbf{L}$, although it can be read directly from the original matrix only if the subspace M is spanned by a subset of the original basis vectors.

Bordered Hessians

The above approach for determining the eigenvalues of $\mathbf{L}$ projected onto M is quite direct and relatively simple. There is another approach, however, that is useful in some theoretical arguments and convenient for simple applications. It is based on constructing matrices and determinants of order $n+m$ rather than $n-m$, so dimension is increased.

Let us first characterize all vectors orthogonal to M. M itself is the set of all $\mathbf{x}$ satisfying $\nabla\mathbf{h}\mathbf{x}=\mathbf{0}$. A vector $\mathbf{z}$ is orthogonal to M if $\mathbf{z}^T\mathbf{x}=0$ for all $\mathbf{x}\in M$. It is not hard to show that $\mathbf{z}$ is orthogonal to M if and only if $\mathbf{z}=\nabla\mathbf{h}^T\mathbf{w}$ for some $\mathbf{w}\in E^m$. The proof that this is sufficient follows from the calculation $\mathbf{z}^T\mathbf{x}=\mathbf{w}^T\nabla\mathbf{h}\mathbf{x}=0$. The proof of necessity follows from the Duality Theorem of Linear Programming (see Exercise 6).

Now we may explicitly characterize an eigenvector of $\mathbf{L}_M$. The vector $\mathbf{x}$ is such an eigenvector if it satisfies these two conditions: (1) $\mathbf{x}$ belongs to M, and (2) $\mathbf{L}\mathbf{x}=\lambda\mathbf{x}+\mathbf{z}$, where $\mathbf{z}$ is orthogonal to M. These conditions are equivalent, in view of the characterization of $\mathbf{z}$, to

$$\begin{aligned} \nabla\mathbf{h}\mathbf{x} &= 0 \\ \mathbf{L}\mathbf{x} &= \lambda\mathbf{x}+\nabla\mathbf{h}^T\mathbf{w}. \end{aligned}$$

This can be regarded as a homogeneous system of $n+m$ linear equations in the unknowns $\mathbf{w},\mathbf{x}$. It possesses a nonzero solution if and only if the determinant of the coefficient matrix is zero. Denoting this determinant $p(\lambda)$, we have

$$\det\begin{bmatrix} \mathbf{0} & \nabla\mathbf{h} \\ -\nabla\mathbf{h}^T & \mathbf{L}-\lambda\mathbf{I} \end{bmatrix} \equiv p(\lambda) = 0 \tag{26}$$

as the condition. The function $p(\lambda)$ is a polynomial in λ of degree $n-m$. It is, as we have derived, the characteristic polynomial of $\mathbf{L}_M$.

Example 3. Approaching Example 2 in this way we have

$$p(\lambda) \equiv \det \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & -(1+\lambda) & 0 & 0 \\ 0 & 0 & (1-\lambda) & 1 \\ 0 & 0 & 1 & (3-\lambda) \end{bmatrix}.$$

This determinant can be evaluated by using Laplace's expansion down the first column. The result is

$$p(\lambda) = (1-\lambda)(3-\lambda) - 1,$$

which is identical to that found earlier.

The above treatment leads one to suspect that it might be possible to extend other tests for positive definiteness over the whole space to similar tests in the constrained case by working in $n+m$ dimensions. We present (but do not derive) the following classic criterion, which is of this type. It is expressed in terms of the *bordered Hessian* matrix

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \nabla \mathbf{h} \\ \nabla \mathbf{h}^T & \mathbf{L} \end{bmatrix}. \tag{27}$$

(Note that by convention the minus sign in front of $\nabla \mathbf{h}^T$ is deleted to make $\mathbf{B}$ symmetric; this only introduces sign changes in the conclusions.)

Bordered Hessian Test. *The matrix* $\mathbf{L}$ *is positive definite on the subspace* $M = \{\mathbf{x} : \nabla \mathbf{h}\mathbf{x} = \mathbf{0}\}$ *if and only if the last* $n-m$ *principal minors of* $\mathbf{B}$ *all have sign* $(-1)^m$.

For the above example we form

$$\mathbf{B} = \det \left[\begin{array}{ccc:c} 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \hdashline 0 & 0 & 1 & 3 \end{array} \right]$$

and check the last two principal minors—the one indicated by the dashed lines and the whole determinant. These are -1, -2, which both have sign $(-1)^1$, and hence the criterion is satisfied.

11.7 SENSITIVITY

The Lagrange multipliers associated with a constrained minimization problem have an interpretation as prices, similar to the prices associated with constraints in linear programming. In the nonlinear case the multipliers are associated with the particular solution point and correspond to incremental or marginal prices, that is, prices associated with small variations in the constraint requirements.

Suppose the problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{aligned} \tag{28}$$

has a solution at the point $\mathbf{x}^*$ which is a regular point of the constraints. Let $\boldsymbol{\lambda}$ be the corresponding Lagrange multiplier vector. Now consider the family of problems

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{c}, \end{aligned} \tag{29}$$

where $\mathbf{c} \in E^m$. For a sufficiently small range of $\mathbf{c}$ near the zero vector, the problem will have a solution point $\mathbf{x}(\mathbf{c})$ near $\mathbf{x}(\mathbf{0}) \equiv \mathbf{x}^*$. For each of these solutions there is a corresponding value $f(\mathbf{x}(\mathbf{c}))$, and this value can be regarded as a function of $\mathbf{c}$, the right-hand side of the constraints. The components of the gradient of this function can be interpreted as the incremental rate of change in value per unit change in the constraint requirements. Thus, they are the incremental prices of the constraint requirements measured in units of the objective. We show below how these prices are related to the Lagrange multipliers of the problem having $\mathbf{c} = \mathbf{0}$.

Sensitivity Theorem*. Let f, $\mathbf{h} \in C^2$ and consider the family of problems*

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{h}(\mathbf{x}) = \mathbf{c}. \end{aligned} \tag{29}$$

Suppose for $\mathbf{c} = \mathbf{0}$ there is a local solution $\mathbf{x}^*$ that is a regular point and that, together with its associated Lagrange multiplier vector $\boldsymbol{\lambda}$, satisfies the second-order sufficiency conditions for a strict local minimum. Then for every $\mathbf{c} \in E^m$ in a region containing $\mathbf{0}$ there is an $\mathbf{x}(\mathbf{c})$, depending continuously on $\mathbf{c}$, such that $\mathbf{x}(\mathbf{0}) = \mathbf{x}^*$ and such that $\mathbf{x}(\mathbf{c})$ is a local minimum of (29). Furthermore,

$$\nabla_{\mathbf{c}} f(\mathbf{x}(\mathbf{c}))\Big]_{\mathbf{c}=\mathbf{0}} = -\boldsymbol{\lambda}^T.$$

Proof. Consider the system of equations

$$\nabla f(\mathbf{x}) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}) = \mathbf{0} \tag{30}$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{c}. \tag{31}$$

By hypothesis, there is a solution $\mathbf{x}^*$, $\boldsymbol{\lambda}$ to this system when $\mathbf{c} = \mathbf{0}$. The Jacobian matrix of the system at this solution is

$$\begin{bmatrix} \mathbf{L}(\mathbf{x}^*) & \nabla \mathbf{h}(\mathbf{x}^*)^T \\ \nabla \mathbf{h}(\mathbf{x}^*) & \mathbf{0} \end{bmatrix}.$$

Because by assumption $\mathbf{x}^*$ is a regular point and $\mathbf{L}(\mathbf{x}^*)$ is positive definite on M, it follows that this matrix is nonsingular (see Exercise 11). Thus, by the Implicit Function Theorem, there is a solution $\mathbf{x}(\mathbf{c}), \boldsymbol{\lambda}(\mathbf{c})$ to the system which is in fact continuously differentiable.

By the chain rule we have

$$\nabla_{\mathbf{c}} f(\mathbf{x}(\mathbf{c}))\Big]_{\mathbf{c}=\mathbf{0}} = \nabla_{\mathbf{x}} f(\mathbf{x}^*)\nabla_{\mathbf{c}}\mathbf{x}(\mathbf{0}).$$

and

$$\nabla_{\mathbf{c}} \mathbf{h}(\mathbf{x}(\mathbf{c}))\Big]_{\mathbf{c}=\mathbf{0}} = \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*)\nabla_{\mathbf{c}}\mathbf{x}(\mathbf{0}).$$

In view of (31), the second of these is equal to the identity $\mathbf{I}$ on E^m, while this, in view of (30), implies that the first can be written

$$\nabla_{\mathbf{c}} f(\mathbf{x}(\mathbf{c}))\Big]_{\mathbf{c}=\mathbf{0}} = -\boldsymbol{\lambda}^T. \quad \blacksquare$$

11.8 INEQUALITY CONSTRAINTS

We consider now problems of the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{g}(\mathbf{x}) \leqslant \mathbf{0}. \end{aligned} \tag{32}$$

We assume that f and $\mathbf{h}$ are as before and that $\mathbf{g}$ is a p-dimensional function. Initially, we assume $f, \mathbf{h}, \mathbf{g} \in C^1$.

There are a number of distinct theories concerning this problem, based on various regularity conditions or constraint qualifications, which are directed toward obtaining definitive general statements of necessary and sufficient conditions. One can by no means pretend that all such results can be obtained as minor extensions of the theory for problems having equality constraints only. To date, however, these alternative results concerning necessary conditions have been of isolated theoretical interest only—for they have not had an influence on the development of algorithms, and have not contributed to the theory of algorithms. Their use has been limited to small-scale programming problems of two or three variables. We therefore choose to emphasize the simplicity of incorporating inequalities rather than the possible complexities, not only for ease of presentation and insight, but also because it is this viewpoint that forms the basis for work beyond that of obtaining necessary conditions.

First-Order Necessary Conditions

With the following generalization of our previous definition it is possible to parallel the development of necessary conditions for equality constraints.

Definition. Let $\mathbf{x}^*$ be a point satisfying the constraints

$$\mathbf{h}(\mathbf{x}^*) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}^*) \leqslant \mathbf{0}, \tag{33}$$

and let J be the set of indices j for which $g_j(\mathbf{x}^*) = 0$. Then $\mathbf{x}^*$ is said to be a *regular point* of the constraints (33) if the gradient vectors $\nabla h_i(\mathbf{x}^*)$, $\nabla g_j(\mathbf{x}^*)$, $1 \leqslant i \leqslant m$, $j \in J$ are linearly independent.

We note that, following the definition of active constraints given in Section 11.1, a point $\mathbf{x}^*$ is a regular point if the gradients of the active constraints are linearly independent. Or, equivalently, $\mathbf{x}^*$ is regular for the constraints if it is regular in the sense of the earlier definition for equality constraints applied to the active constraints.

Karush–Kuhn–Tucker Conditions*. Let* $\mathbf{x}^*$ *be a relative minimum point for the problem*

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}) \leqslant \mathbf{0}, \end{aligned} \tag{34}$$

and suppose $\mathbf{x}^*$ *is a regular point for the constraints. Then there is a vector* $\boldsymbol{\lambda} \in E^m$ *and a vector* $\boldsymbol{\mu} \in E^p$ *with* $\boldsymbol{\mu} \geqslant \mathbf{0}$ *such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0} \tag{35}$$

$$\boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) = 0. \tag{36}$$

Proof. We note first, since $\boldsymbol{\mu} \geqslant \mathbf{0}$ and $\mathbf{g}(\mathbf{x}^*) \leqslant \mathbf{0}$, (36) is equivalent to the statement that a component of $\boldsymbol{\mu}$ may be nonzero only if the corresponding constraint is active. This a *complementary slackness* condition, stating that $\mathbf{g}(\mathbf{x}^*)_i < 0$ implies $\mu_i = 0$ and $\mu_i > 0$ implies $\mathbf{g}(\mathbf{x}^*)_i = 0$.

Since $\mathbf{x}^*$ is a relative minimum point over the constraint set, it is also a relative minimum over the subset of that set defined by setting the active constraints to zero. Thus, for the resulting equality constrained problem defined in a neighborhood of $\mathbf{x}^*$, there are Lagrange multipliers. Therefore, we conclude that (35) holds with $\mu_j = 0$ if $g_j(\mathbf{x}^*) \neq 0$ (and hence (36) also holds).

It remains to be shown that $\boldsymbol{\mu} \geqslant \mathbf{0}$. Suppose $\mu_k < 0$ for some $k \in \boldsymbol{J}$. Let S and M be the surface and tangent plane, respectively, defined by all other active constraints at $\mathbf{x}^*$. By the regularity assumption, there is a $\mathbf{y}$ such that $\mathbf{y} \in M$ and

$\nabla g_k(\mathbf{x}^*)\mathbf{y} < 0$. Let $\mathbf{x}(t)$ be a curve on S passing through $\mathbf{x}^*$ (at $t = 0$) with $\dot{\mathbf{x}}(0) = \mathbf{y}$. Then for small $t \geqslant 0$, $\mathbf{x}(t)$ is feasible, and

$$\left.\frac{df}{dt}(\mathbf{x}(t))\right]_{t=0} = \nabla f(\mathbf{x}^*)\mathbf{y} < \mathbf{0}$$

by (35), which contradicts the minimality of $\mathbf{x}^*$. ▌

Example. Consider the problem

$$\begin{aligned} \text{minimize} \quad & 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 \leqslant 5 \\ & 3x_1 + x_2 \leqslant 6. \end{aligned}$$

The first-order necessary conditions, in addition to the constraints, are

$$\begin{gathered} 4x_1 + 2x_2 - 10 + 2\mu_1 x_1 + 3\mu_2 = 0 \\ 2x_1 + 2x_2 - 10 + 2\mu_1 x_2 + \mu_2 = 0 \\ \mu_1 \geqslant 0, \quad \mu_2 \geqslant 0 \\ \mu_1(x_1^2 + x_2^2 - 5) = 0 \\ \mu_2(3x_1 + x_2 - 6) = 0. \end{gathered}$$

To find a solution we define various combinations of active constraints and check the signs of the resulting Lagrange multipliers. In this problem we can try setting none, one, or two constraints active. Assuming the first constraint is active and the second is inactive yields the equations

$$\begin{gathered} 4x_1 + 2x_2 - 10 + 2\mu_1 x_1 = 0 \\ 2x_1 + 2x_2 - 10 + 2\mu_1 x_2 = 0 \\ x_1^2 + x_2^2 = 5, \end{gathered}$$

which has the solution

$$x_1 = 1, \quad x_2 = 2, \quad \mu_1 = 1.$$

This yields $3x_1 + x_2 = 5$ and hence the second constraint is satisfied. Thus, since $\mu_1 > 0$, we conclude that this solution satisfies the first-order necessary conditions.

Second-Order Conditions

The second-order conditions, both necessary and sufficient, for problems with inequality constraints, are derived essentially by consideration only of the equality constrained problem that is implied by the active constraints. The appropriate tangent plane for these problems is the plane tangent to the active constraints.

Second-Order Necessary Conditions. *Suppose the functions $f, \mathbf{g}, \mathbf{h} \in C^2$ and that $\mathbf{x}^*$ is a regular point of the constraints (33). If $\mathbf{x}^*$ is a relative minimum point for problem (32), then there is a $\boldsymbol{\lambda} \in E^m$, $\boldsymbol{\mu} \in E^p$, $\boldsymbol{\mu} \geqslant \mathbf{0}$ such that (35) and (36) hold and such that*

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{H}(\mathbf{x}^*) + \boldsymbol{\mu}^T \mathbf{G}(\mathbf{x}^*) \tag{37}$$

is positive semidefinite on the tangent subspace of the active constraints at $\mathbf{x}^*$.

Proof. If $\mathbf{x}^*$ is a relative minimum point over the constraints (33), it is also a relative minimum point for the problem with the active constraints taken as equality constraints. ▌

Just as in the theory of unconstrained minimization, it is possible to formulate a converse to the Second-Order Necessary Condition Theorem and thereby obtain a Second-Order Sufficiency Condition Theorem. By analogy with the unconstrained situation, one can guess that the required hypothesis is that $\mathbf{L}(\mathbf{x}^*)$ be positive definite on the tangent plane M. This is indeed sufficient in most situations. However, if there are *degenerate inequality constraints* (that is, active inequality constraints having zero as associated Lagrange multiplier), we must require $\mathbf{L}(\mathbf{x}^*)$ to be positive definite on a subspace that is larger than M.

Second-Order Sufficiency Conditions. *Let $f, \mathbf{g}, \mathbf{h} \in C^2$. Sufficient conditions that a point $\mathbf{x}^*$ satisfying (33) be a strict relative minimum point of problem (32) is that there exist $\boldsymbol{\lambda} \in E^m$, $\boldsymbol{\mu} \in E^p$, such that*

$$\boldsymbol{\mu} \geqslant 0 \tag{38}$$

$$\boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) = 0 \tag{39}$$

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^{T_1} \nabla \mathbf{g}(\mathbf{x}^*) = 0, \tag{40}$$

and the Hessian matrix

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{H}(\mathbf{x}^*) + \boldsymbol{\mu}^T \mathbf{G}(\mathbf{x}^*) \tag{41}$$

is positive definite on the subspace

$$M' = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = 0, \nabla g_j(\mathbf{x}^*)\mathbf{y} = 0 \textit{ for all } j \in J\},$$

where

$$J = \{j : g_j(\mathbf{x}^*) = 0, \mu_j > 0\}.$$

Proof. As in the proof of the corresponding theorem for equality constraints in Section 11.5, assume that $\mathbf{x}^*$ is not a strict relative minimum point; let $\{\mathbf{y}_k\}$ be a sequence of feasible points converging to $\mathbf{x}^*$ such that $f(\mathbf{y}_k) \leqslant f(\mathbf{x}^*)$, and write each $\mathbf{y}_k$ in the form $\mathbf{y}_k = \mathbf{x}^* + \delta_k \mathbf{s}_k$ with $|\mathbf{s}_k| = 1$, $\delta_k > 0$. We may assume that $\delta_k \to 0$ and $\mathbf{s}_k \to \mathbf{s}^*$. We have $0 \geqslant \nabla f(\mathbf{x}^*)\mathbf{s}^*$, and for each $i = 1, \ldots, m$ we have

$$\nabla h_i(\mathbf{x}^*)\mathbf{s}^* = 0.$$

Also for each active constraint g_j we have $g_j(\mathbf{y}_k) - g_j(\mathbf{x}^*) \leqslant 0$, and hence

$$\nabla g_j(\mathbf{x}^*)\mathbf{s}^* \leqslant 0.$$

If $\nabla g_j(\mathbf{x}^*)\mathbf{s}^* = 0$ for all $j \in J$, then the proof goes through just as in Section 11.5. If $\nabla g_j(\mathbf{x}^*)\mathbf{s}^* < 0$ for at least one $j \in J$, then

$$0 \geqslant \nabla f(\mathbf{x}^*)\mathbf{s}^* = -\boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{s}^* - \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}^*)\mathbf{s}^* > 0,$$

which is a contradiction. ▌

We note in particular that if all active inequality constraints have strictly positive corresponding Lagrange multipliers (no degenerate inequalities), then the set J includes all of the active inequalities. In this case the sufficient condition is that the Lagrangian be positive definite on M, the tangent plane of active constraints.

Sensitivity

The sensitivity result for problems with inequalities is a simple restatement of the result for equalities. In this case, a nondegeneracy assumption is introduced so that the small variations produced in Lagrange multipliers when the constraints are varied will not violate the positivity requirement.

Sensitivity Theorem. *Let $f, \mathbf{g}, \mathbf{h} \in C^2$ and consider the family of problems*

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{h}(\mathbf{x}) = \mathbf{c} \\ & \mathbf{g}(\mathbf{x}) \leqslant \mathbf{d}. \end{array} \tag{42}$$

Suppose that for $\mathbf{c} = \mathbf{0}$, $\mathbf{d} = \mathbf{0}$, there is a local solution $\mathbf{x}^*$ that is a regular point and that, together with the associated Lagrange multipliers, $\boldsymbol{\lambda}, \boldsymbol{\mu} \geqslant \mathbf{0}$, satisfies the second-order sufficiency conditions for a strict local minimum. Assume further that no active inequality constraint is degenerate. Then for every $(\mathbf{c}, \mathbf{d}) \in E^{m+p}$ in a region containing $(\mathbf{0}, \mathbf{0})$ there is a solution $\mathbf{x}(\mathbf{c}, \mathbf{d})$, depending continuously on $(\mathbf{c}, \mathbf{d})$, such that $\mathbf{x}(\mathbf{0}, \mathbf{0}) = \mathbf{x}^*$, and such that $\mathbf{x}(\mathbf{c}, \mathbf{d})$ is a relative minimum point of (42). Furthermore,

$$\nabla_{\mathbf{c}} f(\mathbf{x}(\mathbf{c}, \mathbf{d}))\Big]_{\mathbf{0},\mathbf{0}} = -\boldsymbol{\lambda}^T \tag{43}$$

$$\nabla_{\mathbf{d}} f(\mathbf{x}(\mathbf{c}, \mathbf{d}))\Big]_{\mathbf{0},\mathbf{0}} = -\boldsymbol{\mu}^T. \tag{44}$$

11.9 ZERO-ORDER CONDITIONS AND LAGRANGE MULTIPLIERS

Zero-order conditions for functionally constrained problems express conditions in terms of Lagrange multipliers without the use of derivatives. This theory is not only of great practical value, but it also gives new insight into the meaning of Lagrange multipliers. Rather than regarding the Lagrange multipliers as separate scalars, they are identified as components of a single vector that has a strong geometric interpretation. As before, the basic constrained problem is

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0} \\ &\qquad\qquad \mathbf{x} \in \Omega, \end{aligned} \tag{45}$$

where $\mathbf{x}$ is a vector in E^n, and $\mathbf{h}$ and $\mathbf{g}$ are m-dimensional and p-dimensional functions, respectively.

In purest form, zero-order conditions require that the functions that define the objective and the constraints are convex functions and sets. (See Appendix B).

The vector-valued function $\mathbf{g}$ consisting of p individual component functions $g_1, g_2, \ldots, g_p$ is said to be *convex* if each of the component functions is convex.

The programming problem (45) above is termed a *convex programming problem* if the functions f and $\mathbf{g}$ are convex, the function $\mathbf{h}$ is affine (that is, linear plus a constant), and the set $\Omega \subset E^n$ is convex.

Notice that according to Proposition 3, Section 7.4, the set defined by each of the inequalities $g_j(\mathbf{x}) \le 0$ is convex. This is true also of a set defined by $h_i(\mathbf{x}) = 0$. Since the overall constraint set is the intersection of these and Ω it follows from Proposition 1 of Appendix B that this overall constraint set is itself convex. Hence the problem can be regarded as minimize $f(\mathbf{x})$, $\mathbf{x} \in \Omega_1$ where Ω_1 is a convex subset of Ω.

With this view, one could apply the zero-order conditions of Section 7.6 to the problem with constraint set Ω_1. However, in the case of functional constraints it is common to keep the structure of the constraints explicit instead of folding them into an amorphous set.

Although it is possible to derive the zero-order conditions for (45) all at once, treating both equality and inequality constraints together, it is notationally cumbersome to do so and it may obscure the basic simplicity of the arguments. For this reason, we treat equality constraints first, then inequality constraints, and finally the combination of the two.

The equality problem is

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ &\qquad\qquad \mathbf{x} \in \Omega. \end{aligned} \tag{46}$$

Letting $Y = E^n$, we have $\mathbf{h}(\mathbf{x}) \in Y$ for all $\mathbf{x}$. For this problem we require a regularity condition.

Definition. An affine function $\mathbf{h}$ is *regular* with respect to Ω if the set C in Y defined by $C = \{\mathbf{y} : \mathbf{h}(\mathbf{x}) = \mathbf{y} \text{ for some } \mathbf{x} \in \Omega\}$ contains an open sphere around $\mathbf{0}$; that is, C contains a set of the form $\{\mathbf{y} : |\mathbf{y}| < \varepsilon\}$ for some $\varepsilon > 0$.

This condition means that $\mathbf{h}(\mathbf{x})$ can attain $\mathbf{0}$ and can vary in arbitrary directions from $\mathbf{0}$.

Notice that this condition is similar to the definition of a regular point in the context of first-order conditions. If $\mathbf{h}$ has continuous derivatives at a point $\mathbf{x}^*$ the earlier regularity condition implies that $\nabla\mathbf{h}(\mathbf{x}^*)$ is of full rank and the Implicit Function Theorem (of Appendix A) then guarantees that there is an $\varepsilon > 0$ such that for any $\mathbf{y}$ with $|\mathbf{y} - \mathbf{h}(\mathbf{x}^*)| < \varepsilon$ there is an $\mathbf{x}$ such that $\mathbf{h}(\mathbf{x}) = \mathbf{y}$. In other words, there is an open sphere around $\mathbf{y}^* = \mathbf{h}(\mathbf{x}^*)$ that is attainable. In the present situation we assume this attainability directly, at the point $\mathbf{0} \in \mathbf{Y}$.

Next we introduce the following important construction.

Definition. The *primal function* associated with problem (46) is

$$w(\mathbf{y}) = \inf\{f(\mathbf{x}) : \mathbf{h}(\mathbf{x}) = \mathbf{y}, \mathbf{x} \in \Omega\},$$

defined for all $\mathbf{y} \in C$.

Notice that the primal function is defined by varying the right hand side of the constraint. The original problem (46) corresponds to $\omega(\mathbf{0})$. The primal function is illustrated in Fig. 11.6.

***Proposition* 1.** *Suppose Ω is convex, the function f is convex, and $\mathbf{h}$ is affine. Then the primal function ω is convex.*

Proof. For simplicity of notation we assume that Ω is the entire space X. Then we observe

$$\begin{aligned}
\omega(\alpha\mathbf{y}_1 + (1-\alpha)\mathbf{y}_2) &= \inf\{f(\mathbf{x}) : \mathbf{h}(\mathbf{x}) = \alpha\mathbf{y}_1 + (1-\alpha)\mathbf{y}_2\} \\
&\leq \inf\{f(\mathbf{x}) : \mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, \mathbf{h}(\mathbf{x}_1) = \mathbf{y}_1, \mathbf{h}(\mathbf{x}_2) = \mathbf{y}_2\} \\
&\leq \alpha \inf\{f(\mathbf{x}_1) : \mathbf{h}(\mathbf{x}_1) = \mathbf{y}_1\} + (1-\alpha)\inf\{f\mathbf{x}_2) : \mathbf{h}(\mathbf{x}_2) = \mathbf{y}_2\} \\
&= \alpha\,\omega(\mathbf{y}_1) + (1-\alpha)\omega(\mathbf{y}_2). \quad \blacksquare
\end{aligned}$$

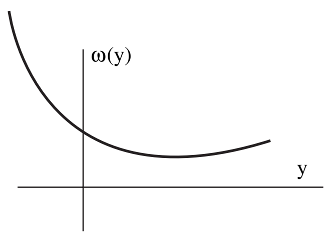

Fig. 11.6 The primal function

We now turn to the derivation of the Lagrange multiplier result for (46).

Proposition 2. *Assume that* $\Omega \subset E^n$ *is convex,* f *is a convex function on* Ω *and* $\mathbf{h}$ *is an m-dimensional affine function on* Ω. *Assume that* $\mathbf{h}$ *is regular with respect to* Ω. *If* $\mathbf{x}^*$ *solves (46), then there is* $\boldsymbol{\lambda} \in E^m$ *such that* $\mathbf{x}^*$ *solves the Lagrangian problem*

$$\begin{aligned} \textit{minimize} \quad & f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) \\ \textit{subject to} \quad & \mathbf{x} \in \Omega. \end{aligned}$$

Proof. Let $f^* = f(\mathbf{x}^*)$. Define the sets A and B in E^{m+1} as

$$\begin{aligned} A &= \{(r, \mathbf{y}) : r \geq \omega(\mathbf{y}),\ \mathbf{y} \in C\} \\ B &= \{(r, \mathbf{y}) : r \leq f^*,\ \mathbf{y} = \mathbf{0}\}. \end{aligned}$$

A is the epigraph of ω (see Section 7.6) and B is the vertical line extending below f^* and aligned with the origin. Both A and B are convex sets. Their only common point is at $(f^*, \mathbf{0})$. See Fig. 11.7

According to the separating hyperplane theorem (Appendix B), there is a hyperplane separating A and B. This hyperplane can be represented by a nonzero vector in E^{m+1} of the form $(s, \boldsymbol{\lambda})$, with $\boldsymbol{\lambda} \in E^m$, and a separation constant c. The separation conditions are

$$\begin{aligned} sr + \boldsymbol{\lambda}^T \mathbf{y} &\geq c \text{ for all } (r, \mathbf{y}) \in A \\ sr + \boldsymbol{\lambda}^T \mathbf{y} &\leq c \text{ for all } (r, \mathbf{y}) \in B. \end{aligned}$$

It follows immediately that $s \geq 0$ for otherwise points $(r, \mathbf{0}) \in B$ with r very negative would violate the second inequality.

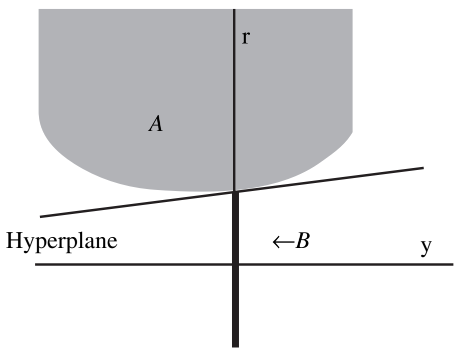

Fig. 11.7 The sets A and B and the separating hyperplane

Geometrically, if $s = 0$ the hyperplane would be vertical. We wish to show that $s \neq 0$, and it is for this purpose that we make use of the regularity condition. Suppose $s = 0$. Then $\boldsymbol{\lambda} \neq \mathbf{0}$ since both s and $\boldsymbol{\lambda}$ cannot be zero. It follows from the second separation inequality that $c = 0$ because the hyperplane must include the point $(f^*, \mathbf{0})$. Now, as $\mathbf{y}$ ranges over a sphere centered at $\mathbf{0} \in C$, the left hand side of the first separation inequality ranges correspondingly over $\boldsymbol{\lambda}^T\mathbf{y}$ which is negative for some $\mathbf{y}$'s. This contradicts the first separation inequality. Thus $s \neq 0$ and thus in fact $s > 0$. Without loss of generality we may, by rescaling if necessary, assume that $s = 1$.

Finally, suppose $\mathbf{x} \in \Omega$. Then $(f(\mathbf{x}), \mathbf{h}(\mathbf{x})) \in A$ and $(f(\mathbf{x}^*), \mathbf{0}) \in B$. Thus, from the separation inequality (with $s = 1$) we have

$$f(\mathbf{x}) + \boldsymbol{\lambda}^T\mathbf{h}(\mathbf{x}) \geq f(\mathbf{x}^*) = f(\mathbf{x}^*) + \boldsymbol{\lambda}^T\mathbf{h}(\mathbf{x}^*).$$

Hence $\mathbf{x}^*$ solves the stated minimization problem. ∎

Example 1 (Best rectangle). Consider the classic problem of finding the rectangle of of maximum area while limiting the perimeter to a length of 4. This can be formulated as

$$\begin{aligned} \text{minimize} \quad & -x_1x_2 \\ \text{subject to} \quad & x_1 + x_2 - 2 = 0 \\ & x_1 \geq 0, \quad x_2 \geq 0. \end{aligned}$$

The regularity condition is met because it is possible to make the right hand side of the functional constraint slightly positive or slightly negative with nonnegative x_1 and x_2. We know the answer to the problem is $x_1 = x_2 = 1$. The Lagrange multiplier is $\lambda = 1$. The Lagrangian problem of Proposition 2 is

$$\begin{aligned} \text{minimize} \quad & -x_1x_2 + 1 \cdot (x_1 + x_2 - 2) \\ \text{subject to} \quad & x_1 \geq 0, \quad x_2 \geq 0. \end{aligned}$$

This can be solved by differentiation to obtain $x_1 = x_2 = 1$.

However the conclusion of the proposition is not satisfied! The value of the Lagrangian at the solution is $V = -1 + 1 + 1 - 2 = -1$. However, at $x_1 = x_2 = 0$ the value of the Lagrangian is $V' = -2$ which is less than V. The Lagrangian is *not* minimized at the solution. The proposition breaks down because the objective function $f(x_1, x_2) = -x_1x_2$ is not convex.

Example 2 (Best diagonal). As an alternative problem, consider minimizing the length of the diagonal of a rectangle subject to the perimeter being of length 4. This problem can be formulated as

$$\text{minimize} \quad \frac{1}{2}(x_1^2 + x_2^2)$$

$$\begin{aligned} \text{subject to} \quad & x_1 + x_2 - 2 = 0 \\ & x_1 \geq 0, \quad x_2 \geq 0. \end{aligned}$$

In this case the objective function is convex. The solution is $x_1 = x_2 = 1$ and the Lagrange multiplier is $\lambda = -1$. The Lagrangian problem is

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}(x_1^2 + x_2^2) - 1 \cdot (x_1 + x_2 - 2) \\ \text{subject to} \quad & x_1 \geq 0, \quad x_2 \geq 0. \end{aligned}$$

The value of the Lagrangian at the solution is $V = 1$ which in this case is a minimum as guaranteed by the proposition. (The value at $x_1 = x_2 = 0$ is $V' = 2$.)

Inequality constraints

We outline the parallel results for the inequality constrained problem

$$\begin{aligned} & \text{minimize } f(\mathbf{x}) \\ & \text{subject to } \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \qquad \mathbf{x} \in \Omega, \end{aligned} \tag{47}$$

where $\mathbf{g}$ is a p-dimensional function.

We let $Z = E^p$ and define $D \subset Z$ as $D = \{z \in Z : \mathbf{g}(\mathbf{x}) \leq \mathbf{z}$ for some $\mathbf{x} \in \Omega\}$. The regularity condition (called the *Slater condition*) is that there is a $z_1 \in D$ with $z_1 < 0$.

As before we introduce the primal function.

Definition. The *primal function* associated with problem (47) is

$$w(\mathbf{z}) = \inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \mathbf{z},\ \mathbf{x} \in \Omega\}.$$

The primal function is again defined by varying the right hand side of the constraint function, using the variable $\mathbf{z}$. Now the primal function in monotonically decreasing with $\mathbf{z}$, since an increase in $\mathbf{z}$ enlarges the constraint region.

Proposition 3. *Suppose $\Omega \subset E^n$ is convex and f and $\mathbf{g}$ are convex functions. Then the primal function ω is also convex.*

Proof. The proof parallels that of Proposition 1. One simply substitutes $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ for $\mathbf{h}(\mathbf{x}) = \mathbf{y}$ throughout the series of inequalities. ∎

The zero-order necessary Lagrangian conditions are then given by the proposition below.

Proposition 4. *Assume Ω is a convex subset of E^n and that f and $\mathbf{g}$ are convex functions. Assume also that there is a point $\mathbf{x}_1 \in \Omega$ such that $\mathbf{g}(\mathbf{x}_1) < \mathbf{0}$.*

Then, if $\mathbf{x}^*$ *solves (47), there is a vector* $\boldsymbol{\mu} \in E^p$ *with* $\boldsymbol{\mu} \geq \mathbf{0}$ *such that* $\mathbf{x}^*$ *solves the Lagrangian problem*

$$\begin{array}{ll} \textit{minimize} & f(\mathbf{x}^*) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) \\ \textit{subject to} & \mathbf{x} \in \Omega. \end{array} \tag{48}$$

Furthermore, $\boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) = 0$.

Proof. Here is the proof outline. Let $f^* = f(\mathbf{x}^*)$. In this case define in E^{p+1} the two sets

$$\begin{aligned} A =& \{(r, \mathbf{0}) : r \geq f(\mathbf{x}),\ \mathbf{0} \geq \mathbf{g}(\mathbf{x}),\ \text{for some } \mathbf{x} \in \Omega\} \\ B =& \{(r, \mathbf{0}) \ :\ r \leq f^*,\ \mathbf{0} \leq \mathbf{0}\}. \end{aligned}$$

A is the epigraph of the primal function ω. The set B is the rectangular region at or to the left of the vertical axis and at or lower than f^*. Both A and B are convex. See Fig. 11.8.

The proof is made by constructing a hyperplane separating A and B. The regularity condition guarantees that this hyperplane is not vertical. ∎

The condition $\boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) = 0$ is the complementary slackness condition that is characteristic of necessary conditions for problems with inequality constraints.

Example 4. (Quadratic program). Consider the quadratic program

$$\begin{array}{ll} \text{minimize} & \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x} \\ \text{subject to} & \mathbf{a}^T \mathbf{x} \leq b \\ & \quad \mathbf{x} \geq \mathbf{0}. \end{array}$$

Let $\Omega = \{\mathbf{x} : \mathbf{x} \geq \mathbf{0}\}$ and $g(\mathbf{x}) = \mathbf{a}^T \mathbf{x} - b$. Assume that the $n \times n$ matrix $\mathbf{Q}$ is positive definite, in which case the objective function is convex. Assuming that $b > 0$, the Slater regularity condition is satisfied. Hence there is a Lagrange multiplier $\mu \geq 0$

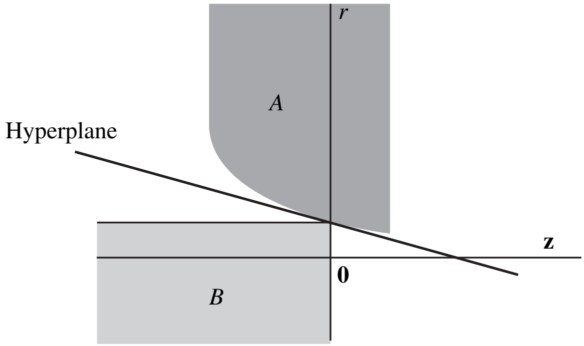

Fig. 11.8 The sets A and B and the separating hyperplane for inequalities

(a scalar in this case) such that the solution $\mathbf{x}^*$ to the quadratic program is also a solution to

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{c}^T\mathbf{x} + \mu(\mathbf{a}^T\mathbf{x} - b) \\ \text{subject to} \quad & \mathbf{x} \geq \mathbf{0}, \end{aligned}$$

and $\mu(\mathbf{a}^T\mathbf{x}^* - b) = 0$.

Mixed constraints

The two previous results can be combined to obtain zero-order conditions for the problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in \Omega. \end{aligned} \tag{49}$$

Zero-order Lagrange Theorem. *Assume that $\Omega \subset E^n$ is a convex set, f and $\mathbf{g}$ are convex functions of dimension 1 and p, respectively, and $\mathbf{h}$ is affine of dimension m. Assume also that $\mathbf{h}$ satisfies the regularity condition with respect to Ω and that there is an $\mathbf{x}_1 \in \Omega$ with $\mathbf{h}(\mathbf{x}_1) = \mathbf{0}$ and $\mathbf{g}(\mathbf{x}_1) < \mathbf{0}$. Suppose $\mathbf{x}^*$ solves (49). Then there are vectors $\boldsymbol{\lambda} \in E^m$ and $\boldsymbol{\mu} \in E^p$ with $\boldsymbol{\mu} \geq \mathbf{0}$ such that $\mathbf{x}^*$ solves the Lagrangian problem*

$$\begin{aligned} \textit{minimize} \quad & f(\mathbf{x}) + \boldsymbol{\lambda}^T\mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T\mathbf{g}(\mathbf{x}) \\ \textit{subject to} \quad & \mathbf{x} \in \Omega. \end{aligned} \tag{50}$$

Furthermore, $\boldsymbol{\mu}^T\mathbf{g}(\mathbf{x}^*) = 0$.

The convexity requirements of this result are satisfied in many practical problems. Indeed convex programming problems are both pervasive and relatively well treated by theory and numerical methods. The corresponding theory also motivates many approaches to general nonlinear programming problems. In fact, it will be apparent that many methods attempt to "convexify" a general nonlinear problem either by changing the formulation of the underlying application or by introducing devices that temporarily relax as the method progresses.

Zero-order sufficient conditions

The sufficiency conditions are very strong and do not require convexity.

Proposition 5. *(Sufficiency Conditions). Suppose f is a real-valued function on a set $\Omega \subset E^n$. Suppose also that $\mathbf{h}$ and $\mathbf{g}$ are, respectively, m-dimensional* *and p-dimensional functions on* Ω. *Finally, suppose there are vectors* $\mathbf{x}^* \in \Omega$, $\boldsymbol{\lambda} \in E^m$, *and* $\boldsymbol{\mu} \in E^p$ *with* $\boldsymbol{\mu} \geq \mathbf{0}$ *such that*

$$f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) \leq f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x})$$

for all $\mathbf{x} \in \Omega$. *Then* $\mathbf{x}^*$ *solves*

$$\begin{aligned} \textit{minimize} \quad & f(\mathbf{x}) \\ \textit{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{h}(\mathbf{x}^*) \\ & \mathbf{g}(\mathbf{x}) \leq \mathbf{g}(\mathbf{x}^*) \\ & \mathbf{x} \in \Omega. \end{aligned}$$

Proof. Suppose there is $\mathbf{x}_1 \in \Omega$ with $f(\mathbf{x}_1) < f(\mathbf{x}^*)$, $\mathbf{h}(\mathbf{x}_1) = \mathbf{h}(\mathbf{x}^*)$, and $\mathbf{g}(\mathbf{x}_1) \leq \mathbf{g}(\mathbf{x}^*)$. From $\boldsymbol{\mu} \geq \mathbf{0}$ it is clear that $\boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}_1) \leq \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*)$. It follows that $f(\mathbf{x}_1) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}_1) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}_1) < f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*)$, which is a contradiction. ∎

This result suggests that Lagrange multiplier values might be guessed and used to define a Lagrangian which is subsequently minimized. This will produce a special value of $\mathbf{x}$ and special values of the right hand sides of the constraints for which this $\mathbf{x}$ is optimal. Indeed, this approach is characteristic of duality methods treated in Chapter 14.

The theory of this section has an inherent geometric simplicity captured clearly by Figs. 11.7 and 11.8. It raises ones's level of understanding of Lagrange multipliers and sets the stage for the theory of convex duality presented in Chapter 14. It is certainly possible to jump ahead and read that now.

11.10 SUMMARY

Given a minimization problem subject to equality constraints in which all functions are smooth, a necessary condition satisfied at a minimum point is that the gradient of the objective function is orthogonal to the tangent plane of the constraint surface. If the point is regular, then the tangent plane has a simple representation in terms of the gradients of the constraint functions, and the above condition can be expressed in terms of Lagrange multipliers.

If the functions have continuous second partial derivatives and Lagrange multipliers exist, then the Hessian of the Lagrangian restricted to the tangent plane plays a role in second-order conditions analogous to that played by the Hessian of the objective function in unconstrained problems. Specifically, the restricted Hessian must be positive semidefinite at a relative minimum point and, conversely, if it is positive definite at a point satisfying the first-order conditions, that point is a strict local minimum point.

Inequalities are treated by determining which of them are active at a solution. An active inequality then acts just like an equality, except that its associated

Lagrange multiplier can never be negative because of the sensitivity interpretation of the multipliers.

The necessary conditions for convex problems can be expressed without derivatives, and these are according termed zero-order conditions. These conditions are highly geometric in character and explicitly treat the Lagrange multiplier as a vector in a space having dimension equal to that of the right-hand-side of the constraints. This Lagrange multiplier vector defines a hyperplane that separates the epigraph of the primal function from a set of unattainable objective and constraint value combinations.

11.11 EXERCISES

1. In E^2 consider the constraints

$$x_1 \geqslant 0$$
$$x_2 \geqslant 0$$
$$x_2 - (x_1 - 1)^2 \leqslant 0.$$

   Show that the point $x_1 = 1$, $x_2 = 0$ is feasible but is not a regular point.

2. Find the rectangle of given perimeter that has greatest area by solving the first-order necessary conditions. Verify that the second-order sufficiency conditions are satisfied.

3. Verify the second-order conditions for the entropy example of Section 11.4.

4. A cardboard box for packing quantities of small foam balls is to be manufactured as shown in Fig. 11.9. The top, bottom, and front faces must be of double weight (i.e., two pieces of cardboard). A problem posed is to find the dimensions of such a box that maximize the volume for a given amount of cardboard, equal to 72 sq. ft.

   a) What are the first-order necessary conditions?
   b) Find x, y, z.
   c) Verify the second-order conditions.

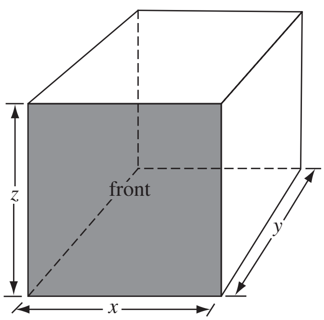

Fig. 11.9 Packing box

5. Define

$$\mathbf{L} = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}, \qquad \mathbf{h} = (1, 1, 0),$$

and let M be the subspace consisting of those points $\mathbf{x} = (x_1, x_2, x_3)$ satisfying $\mathbf{h}^T\mathbf{x} = 0$.

a) Find $\mathbf{L}_M$.
b) Find the eigenvalues of $\mathbf{L}_M$.
c) Find

$$p(\lambda) = \det \begin{bmatrix} 0 & \mathbf{h}^T \\ -\mathbf{h} & \mathbf{L} - \mathbf{I}\lambda \end{bmatrix}.$$

d) Apply the bordered Hessian test.

6. Show that $\mathbf{z}^T\mathbf{x} = 0$ for all $\mathbf{x}$ satisfying $\mathbf{A}\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{z} = \mathbf{A}^T\mathbf{w}$ for some $\mathbf{w}$. (*Hint:* Use the Duality Theorem of Linear Programming.)

7. After a heavy military campaign a certain army requires many new shoes. The quartermaster can order three sizes of shoes. Although he does not know precisely how many of each size are required, he feels that the demand for the three sizes are independent and the demand for each size is uniformly distributed between zero and three thousand pairs. He wishes to allocate his shoe budget of four thousand dollars among the three sizes so as to maximize the expected number of men properly shod. Small shoes cost one dollar per pair, medium shoes cost two dollars per pair, and large shoes cost four dollars per pair. How many pairs of each size should he order?

8. *Optimal control.* A one-dimensional dynamic process is governed by a difference equation

$$x(k+1) = \phi(x(k), u(k), k)$$

with initial condition $x(0) = x_0$. In this equation the value $x(k)$ is called the *state* at step k and $u(k)$ is the *control* at step k. Associated with this system there is an *objective function* of the form

$$J = \sum_{k=0}^{N} \psi(x(k), u(k), k).$$

In addition, there is a *terminal constraint* of the form

$$g(x(N+1)) = 0.$$

The problem is to find the sequence of controls $u(0), u(1), u(2), \ldots, u(N)$ and corresponding state values to minimize the objective function while satisfying the terminal constraint. Assuming all functions have continuous first partial derivatives and that the regularity condition is satisfied, show that associated with an optimal solution there is a sequence $\lambda(k)$, $k = 0, 1, \ldots, N$ and a μ such that

$$\lambda(k-1) = \lambda(k)\phi_x(x(k), u(k), k) + \psi_x(x(k), u(k), k), \qquad k = 1, 2, \ldots, N$$

$$\lambda(N) = \mu g_x(x(N+1))$$

$$\psi_u(x(k),\ u(k), k) + \lambda(k)\phi_u(x(k), u(k), k) = 0, \quad k = 0, 1, 2, \ldots, N.$$

9. Generalize Exercise 9 to include the case where the state $\mathbf{x}(k)$ is an n-dimensional vector and the control $\mathbf{u}(k)$ is an m-dimensional vector at each stage k.

10. An egocentric young man has just inherited a fortune F and is now planning how to spend it so as to maximize his total lifetime enjoyment. He deduces that if $x(k)$ denotes his capital at the beginning of year k, his holdings will be approximately governed by the difference equation

$$x(k+1) = \alpha x(k) - u(k)$$
$$x(0) = F,$$

where $\alpha \geqslant 1$ (with $\alpha - 1$ as the interest rate of investment) and where $u(k)$ is the amount spent in year k. He decides that the enjoyment achieved in year k can be expressed as $\psi(u(k))$ where ψ, his utility function, is a smooth function, and that his total lifetime enjoyment is

$$J = \sum_{k=0}^{N} \psi(u(k))\beta^k,$$

where the term β^k $(0 < \beta < 1)$ reflects the notion that future enjoyment is counted less today. The young man wishes to determine the sequence of expenditures that will maximize his total enjoyment subject to the condition $x(N+1) = 0$.

a) Find the general optimality relationship for this problem.
b) Find the solution for the special case $\psi(u) = u^{1/2}$.

11. Let $\mathbf{A}$ be an $m \times n$ matrix of rank m and let $\mathbf{L}$ be an $n \times n$ matrix that is symmetric and positive definite on the subspace $M = \{\mathbf{x} : \mathbf{Ax} = \mathbf{0}\}$. Show that the $(n+m) \times (n+m)$ matrix

$$\begin{bmatrix} \mathbf{L} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix}$$

is nonsingular.

12. Consider the quadratic program

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Qx} - \mathbf{b}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{Ax} = \mathbf{c}. \end{aligned}$$

Prove that $\mathbf{x}^*$ is a local minimum point if and only if it is a global minimum point. (No convexity is assumed.)

13. Maximize $14x - x^2 + 6y - y^2 + 7$ subject to $x + y \leqslant 2$, $x + 2y \leqslant 3$.

14. In the quadratic program example of Section 11.9, what are more general conditions on $\mathbf{a}$ and b that satisfy the Slater condition?

15. What are the general zero-order Lagrangian conditions for the problem (46) without the regularity condition? [The coefficient of f will be zero, so there is no real condition.]

16. Show that the problem of finding the rectangle of maximum area with a diagonal of unit length can be formulated as an unconstrained convex programming problem using trigonometric functions. [Hint: use variable θ over the range $0 \leq \theta \leq 45$ degrees.]

REFERENCES

11.1–11.5 For a classic treatment of Lagrange multipliers see Hancock [H4]. Also see Fiacco and McCormick [F4], Luenberger [L8], or McCormick [M2].

11.6 The simple formula for the characteristic polynomial of $\mathbf{L}_M$ as an $(n+m)$th-order determinant is apparently due to Luenberger [L17].

11.8 The systematic treatment of inequality constraints was published by Kuhn and Tucker [K11]. Later it was found that the essential elements of the theory were contained in the 1939 unpublished M.Sci Disertation of W. Karush in the Department of Mathematics, University of Chicago. It is common to recognize this contribution by including his name to the conditions for optimality.

11.9 The theory of convex problems and the corresponding Lagrange multiplier theory was developed by Slater [S7]. For presentations similiar to this section, see Hurwicz [H14] and Luenberger [L8].