19

Problems with Equality Constraints

19.1 INTRODUCTION

In this part, we discuss methods for solving a class of nonlinear constrained optimization problems that can be formulated as:

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & h_i(\boldsymbol{x}) = 0, & i = 1, \ldots, m \\ & g_j(\boldsymbol{x}) \leq 0, & j = 1, \ldots, p, \end{array}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h_i : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$, and $m \leq n$. In vector notation, the problem above can be represented in the following *standard form*:

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \\ & \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}, \end{array}$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, and $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$. As usual, we adopt the following terminology.

Definition 19.1 Any point satisfying the constraints is called a *feasible point*. The set of all feasible points

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$$

is called the *feasible set*. ∎

Optimization problems of the above form are not new to us. Indeed, linear programming problems of the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0, \end{array}$$

which we studied in Part III, are of this type.

As we remarked in Part II, there is no loss of generality by considering only minimization problems. For if we are confronted with a maximization problem, it can be easily transformed into the minimization problem by observing that

$$\text{maximize } f(x) = \text{minimize } -f(x).$$

We illustrate the problems we study in this part by considering the following simple numerical example.

Example 19.1

$$\begin{array}{ll} \text{minimize} & (x_1 - 1)^2 + x_2 - 2 \\ \text{subject to} & x_2 - x_1 = 1, \\ & x_1 + x_2 \leq 2. \end{array}$$

This problem is already in the standard form given earlier, with $f(x_1, x_2) = (x_1 - 1)^2 + x_2 - 2$, $h(x_1, x_2) = x_2 - x_1 - 1$, and $g(x_1, x_2) = x_1 + x_2 - 2$. This problem turns out to be simple enough to be solved graphically (see Figure 19.1). In the figure the set of points that satisfy the constraints (the feasible set) is marked by the heavy solid line. The inverted parabolas represent level sets of the objective function f—the lower the level set, the smaller the objective function value. Therefore, the solution can be obtained by finding the lowest level set that intersects the feasible set. In this case, the minimizer lies on the level set with $f = -1/4$. The minimizer of the objective function is $x^* = [1/2, 3/2]^T$. ∎

In the remainder of this chapter, we discuss constrained optimization problems with only equality constraints. The general constrained optimization problem is discussed in the chapters to follow.

19.2 PROBLEM FORMULATION

The class of optimization problems we analyze in this chapter is

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array}$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $h = [h_1, \ldots, h_m]^T$, and $m \leq n$. We assume that the function h is continuously differentiable, that is, $h \in C^1$.

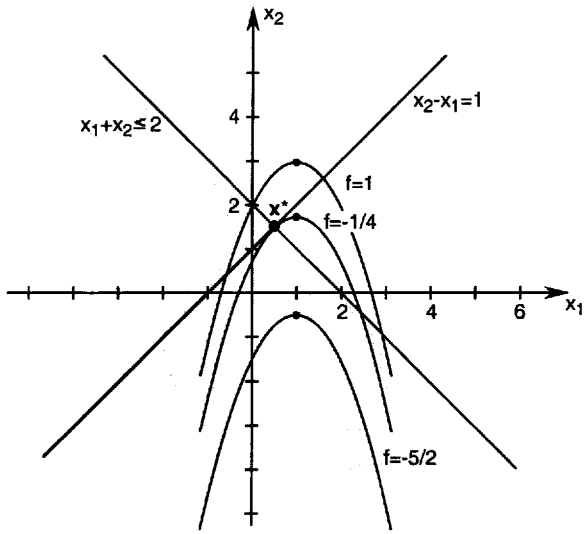


Figure 19.1 Graphical solution to the problem in Example 19.1

We introduce the following definition.

Definition 19.2 A point x^* satisfying the constraints $h_1(x^*) = 0, \ldots, h_m(x^*) = 0$ is said to be a *regular point* of the constraints if the gradient vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent. ∎

Let $Dh(x^*)$ be the Jacobian matrix of $h = [h_1, \ldots, h_m]^T$ at x^*, given by

$$Dh(x^*) = \begin{bmatrix} Dh_1(x^*) \\ \vdots \\ Dh_m(x^*) \end{bmatrix} = \begin{bmatrix} \nabla h_1(x^*)^T \\ \vdots \\ \nabla h_m(x^*)^T \end{bmatrix}.$$

Then, x^* is regular if and only if $\operatorname{rank} Dh(x^*) = m$, that is, the Jacobian matrix is of full rank.

The set of equality constraints $h_1(x) = 0, \ldots, h_m(x) = 0$, $h_i : \mathbb{R}^n \to \mathbb{R}$, describes a surface

$$S = \{x \in \mathbb{R}^n : h_1(x) = 0, \ldots, h_m(x) = 0\}.$$

Assuming the points in S are regular, the dimension of the surface S is $n - m$.

Example 19.2 Let $n = 3$ and $m = 1$ (i.e., we are operating in $\mathbb{R}^3$). Assuming that all points in S are regular, the set S is a two-dimensional surface. For example, let

$$h_1(x) = x_2 - x_3^2 = 0.$$

Note that $\nabla h_1(x) = [0, 1, -2x_3]^T$, and hence for any $x \in \mathbb{R}^3$, $\nabla h_1(x) \neq \mathbf{0}$. In this case,

$$\dim S = \dim\{x : h_1(x) = 0\} = n - m = 2.$$

See Figure 19.2 for a graphical illustration. ∎

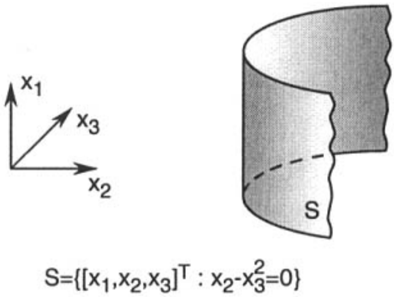


Figure 19.2 A two-dimensional surface in $\mathbb{R}^3$

Example 19.3 Let $n = 3$ and $m = 2$. Assuming regularity, the feasible set S is a one-dimensional object (i.e., a curve in $\mathbb{R}^3$). For example, let

$$\begin{aligned} h_1(\boldsymbol{x}) &= x_1, \\ h_2(\boldsymbol{x}) &= x_2 - x_3^2. \end{aligned}$$

In this case, $\nabla h_1(\boldsymbol{x}) = [1, 0, 0]^T$, and $\nabla h_2(\boldsymbol{x}) = [0, 1, -2x_3]^T$. Hence, the vectors $\nabla h_1(\boldsymbol{x})$ and $\nabla h_2(\boldsymbol{x})$ are linearly independent in $\mathbb{R}^3$. Thus,

$$\dim S = \dim\{\boldsymbol{x} : h_1(\boldsymbol{x}) = 0,\ h_2(\boldsymbol{x}) = 0\} = n - m = 1.$$

See Figure 19.3 for a graphical illustration. ∎

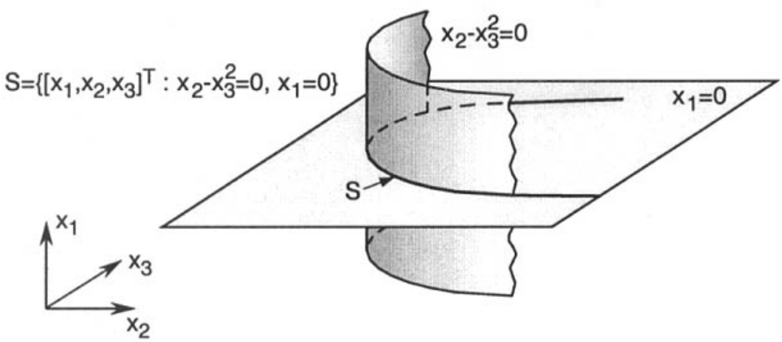


Figure 19.3 A one-dimensional surface in $\mathbb{R}^3$

19.3 TANGENT AND NORMAL SPACES

In this section, we discuss the notion of a tangent space and normal space at a point on a surface. We begin by defining a "curve" on a surface S.

Definition 19.3 A *curve* C on a surface S is a set of points $\{\boldsymbol{x}(t) \in S : t \in (a, b)\}$, continuously parameterized by $t \in (a, b)$; that is, $\boldsymbol{x} : (a, b) \to S$ is a continuous function. ∎

A graphical illustration of the definition of a curve is given in Figure 19.4. The definition of a curve implies that all the points on the curve satisfy the equation describing the surface. The curve C passes through a point $\boldsymbol{x}^*$ if there exists $t^* \in (a, b)$ such that $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$.

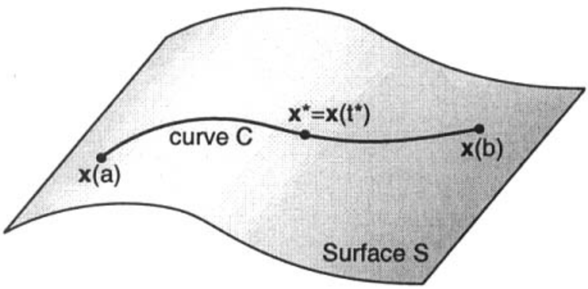


Figure 19.4 A curve on a surface

Intuitively, we can think of a curve $C = \{\boldsymbol{x}(t) : t \in (a, b)\}$ as the path traversed by a point $\boldsymbol{x}$ traveling on the surface S. The position of the point at time t is given by $\boldsymbol{x}(t)$.

Definition 19.4 The curve $C = \{\boldsymbol{x}(t) : t \in (a, b)\}$ is *differentiable* if

$$\dot{\boldsymbol{x}}(t) = \frac{d\boldsymbol{x}}{dt}(t) = \begin{bmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix}$$

exists for all $t \in (a, b)$.

The curve $C = \{\boldsymbol{x}(t) : t \in (a, b)\}$ is *twice differentiable* if

$$\ddot{\boldsymbol{x}}(t) = \frac{d^2\boldsymbol{x}}{dt^2}(t) = \begin{bmatrix} \ddot{x}_1(t) \\ \vdots \\ \ddot{x}_n(t) \end{bmatrix}$$

exists for all $t \in (a, b)$. ∎

Note that both $\dot{\boldsymbol{x}}(t)$ and $\ddot{\boldsymbol{x}}(t)$ are n-dimensional vectors. We can think of $\dot{\boldsymbol{x}}(t)$ and $\ddot{\boldsymbol{x}}(t)$ as the "velocity" and "acceleration," respectively, of a point traversing the curve C with position $\boldsymbol{x}(t)$ at time t. The vector $\dot{\boldsymbol{x}}(t)$ points in the direction of the instantaneous motion of $\boldsymbol{x}(t)$. Therefore, the vector $\dot{\boldsymbol{x}}(t^*)$ is *tangent* to the curve C at $\boldsymbol{x}^*$ (see Figure 19.5).

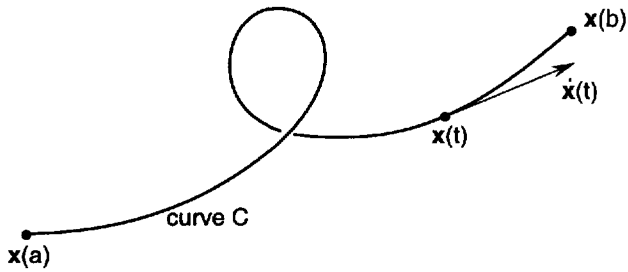


Figure 19.5 Geometric interpretation of the differentiability of a curve

We are now ready to introduce the notions of a tangent space. For this, recall the set

$$S = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\},$$

where $\boldsymbol{h} \in \mathcal{C}^1$. We think of S as a surface in $\mathbb{R}^n$.

Definition 19.5 The *tangent space* at a point $\boldsymbol{x}^*$ on the surface $S = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\}$ is the set

$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} : D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \mathbf{0}\}.$$

∎

Note that the tangent space $T(\boldsymbol{x}^*)$ is the nullspace of the matrix $D\boldsymbol{h}(\boldsymbol{x}^*)$, that is,

$$T(\boldsymbol{x}^*) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*)).$$

The tangent space is therefore a subspace of $\mathbb{R}^n$.

Assuming $\boldsymbol{x}^*$ is regular, the dimension of the tangent space is $n - m$, where m is the number of equality constraints $h_i(\boldsymbol{x}^*) = 0$. Note that the tangent space passes through the origin. However, it is often convenient to picture the tangent space as a plane that passes through the point $\boldsymbol{x}^*$. For this, we define the *tangent plane* at $\boldsymbol{x}^*$ to be the set

$$TP(\boldsymbol{x}^*) = T(\boldsymbol{x}^*) + \boldsymbol{x}^* = \{\boldsymbol{x} + \boldsymbol{x}^* : \boldsymbol{x} \in T(\boldsymbol{x}^*)\}.$$

Figure 19.6 illustrates the notion of a tangent plane. Figure 19.7 illustrates the relationship between the tangent plane and the tangent space.

Example 19.4 Let

$$S = \{\boldsymbol{x} \in \mathbb{R}^3 : h_1(\boldsymbol{x}) = x_1 = 0,\ h_2(\boldsymbol{x}) = x_1 - x_2 = 0\}.$$

Then, S is the x_3-axis in $\mathbb{R}^3$ (see Figure 19.8).

We have

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x})^T \\ \nabla h_2(\boldsymbol{x})^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}.$$

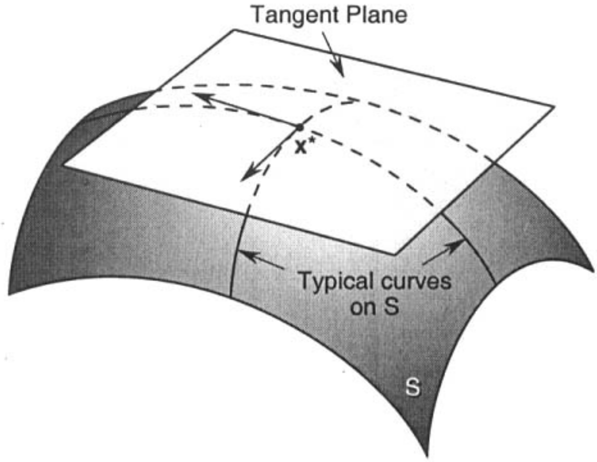


Figure 19.6 The tangent plane to the surface S at the point $\boldsymbol{x}^*$

Because ∇h_1 and ∇h_2 are linearly independent when evaluated at any $\boldsymbol{x} \in S$, all the points of S are regular. The tangent space at an arbitrary point of S is

$$
\begin{aligned}
T(\boldsymbol{x}) &= \{\boldsymbol{y} : \nabla h_1(\boldsymbol{x})^T \boldsymbol{y} = 0,\ \nabla h_2(\boldsymbol{x})^T \boldsymbol{y} = 0\} \\
&= \left\{ \boldsymbol{y} : \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \mathbf{0} \right\} \\
&= \{[0, 0, \alpha]^T : \alpha \in \mathbb{R}\} \\
&= \text{the } x_3\text{-axis in } \mathbb{R}^3.
\end{aligned}
$$

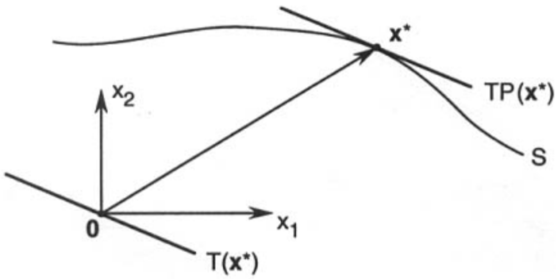
In this example, the tangent space $T(\boldsymbol{x})$ at any point $\boldsymbol{x} \in S$ is a one-dimensional subspace of $\mathbb{R}^3$. ∎

Intuitively, we would expect the definition of the tangent space at a point on a surface to be the collection of all "tangent vectors" to the surface at that point. We have seen that the derivative of a curve on a surface at a point is a tangent vector to the curve, and hence to the surface. The above intuition agrees with our definition whenever $\boldsymbol{x}^*$ is regular, as stated in the theorem below.

Theorem 19.1 *Suppose $\boldsymbol{x}^* \in S$ is a regular point, and $T(\boldsymbol{x}^*)$ is the tangent space at $\boldsymbol{x}^*$. Then, $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ if and only if there exists a differentiable curve in S passing through $\boldsymbol{x}^*$ with derivative $\boldsymbol{y}$ at $\boldsymbol{x}^*$.* □

Proof. $\Leftarrow$: Suppose there exists a curve $\{\boldsymbol{x}(t) : t \in (a, b)\}$ in S such that $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$ and $\dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$ for some $t^* \in (a, b)$. Then,

$$\boldsymbol{h}(\boldsymbol{x}(t)) = \mathbf{0}$$



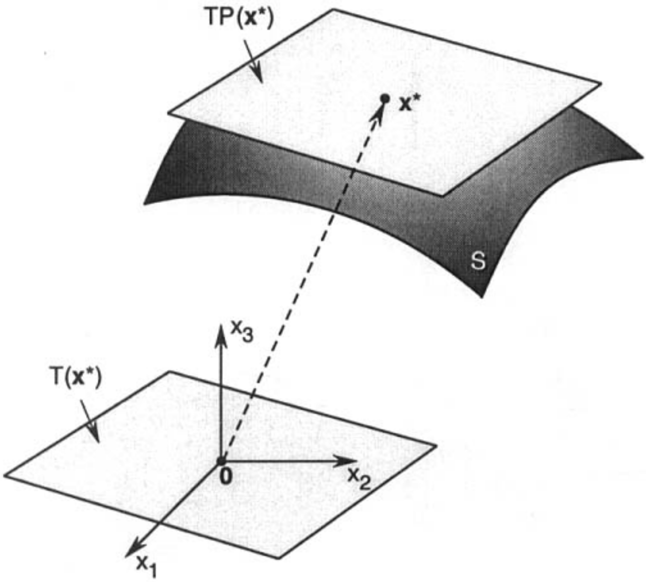


Figure 19.7 Tangent spaces and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$

for all $t \in (a, b)$. If we differentiate the function $\boldsymbol{h}(\boldsymbol{x}(t))$ with respect to t using the chain rule, we obtain

$$\frac{d}{dt}\boldsymbol{h}(\boldsymbol{x}(t)) = D\boldsymbol{h}(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t) = \boldsymbol{0}$$

for all $t \in (a, b)$. Therefore, at t^*, we get

$$D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0},$$

and hence $\boldsymbol{y} \in T(\boldsymbol{x}^*)$.

$\Rightarrow$: To prove this, we need to use the implicit function theorem. We refer the reader to [64, p. 298]. ∎

We now introduce the notion of a normal space.

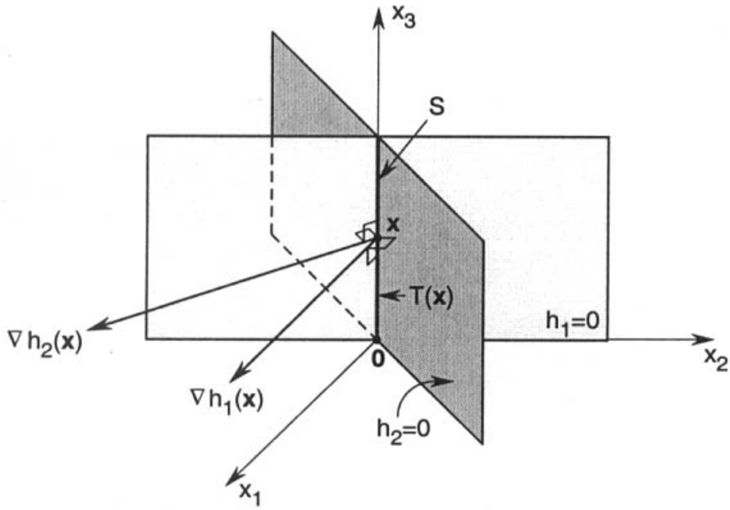


Figure 19.8 The surface $S = \{x \in \mathbb{R}^3 : x_1 = 0,\ x_1 - x_2 = 0\}$

Definition 19.6 The *normal space* $N(x^*)$ at a point x^* on the surface $S = \{x \in \mathbb{R}^n : h(x) = 0\}$ is the set

$$N(x^*) = \{x \in \mathbb{R}^n : x = Dh(x^*)^T z,\ z \in \mathbb{R}^m\}.$$

■

We can express the normal space $N(x^*)$ as

$$N(x^*) = \mathcal{R}\left(Dh(x^*)^T\right),$$

that is, the range of the matrix $Dh(x^*)^T$. Note that the normal space $N(x^*)$ is the subspace of $\mathbb{R}^n$ spanned by the vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$, that is,

$$\begin{aligned} N(x^*) &= \operatorname{span}[\nabla h_1(x^*), \ldots, \nabla h_m(x^*)] \\ &= \{x \in \mathbb{R}^n : x = z_1 \nabla h_1(x^*) + \cdots + z_m \nabla h_m(x^*),\ z_1, \ldots, z_m \in \mathbb{R}\}. \end{aligned}$$

Note that the normal space contains the zero vector. Assuming x^* is regular, the dimension of the normal space $N(x^*)$ is m. As in the case of the tangent space, it is often convenient to picture the normal space $N(x^*)$ as passing through the point x^* (rather than through the origin of $\mathbb{R}^n$). For this, we define the *normal plane* at x^* as the set

$$NP(x^*) = N(x^*) + x^* = \{x + x^* \in \mathbb{R}^n : x \in N(x^*)\}.$$

Figure 19.9 illustrates the normal space and plane in $\mathbb{R}^3$ (i.e., $n = 3$ and $m = 1$).

We now show that the tangent space and normal space are orthogonal complements of each other (see Section 3.3).

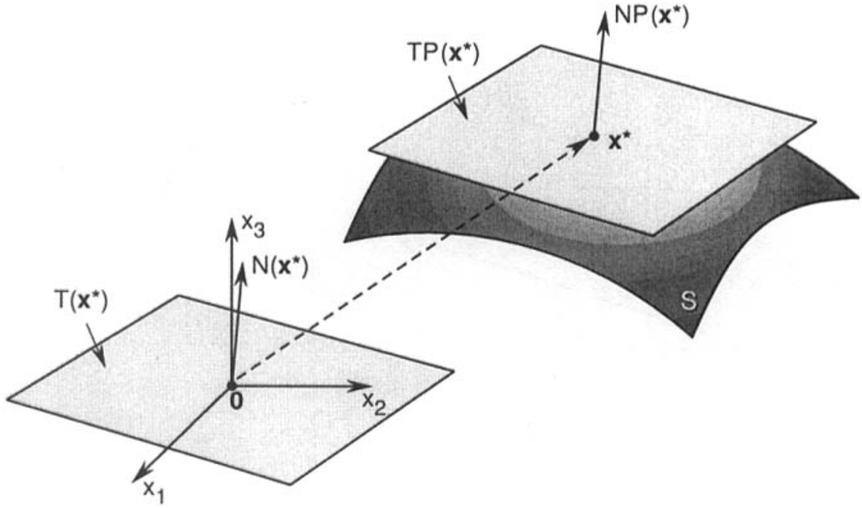


Figure 19.9 Normal space in $\mathbb{R}^3$

Lemma 19.1 *We have* $T(\boldsymbol{x}^*) = N(\boldsymbol{x}^*)^\perp$ *and* $T(\boldsymbol{x}^*)^\perp = N(\boldsymbol{x}^*)$. □

Proof. By definition of $T(\boldsymbol{x}^*)$, we may write

$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{x}^T\boldsymbol{y} = 0 \text{ for all } \boldsymbol{x} \in N(\boldsymbol{x}^*)\}.$$

Hence, by definition of $N(\boldsymbol{x}^*)$, we have $T(\boldsymbol{x}^*) = N(\boldsymbol{x}^*)^\perp$. By Exercise 3.6, we also have $T(\boldsymbol{x}^*)^\perp = N(\boldsymbol{x}^*)$. ∎

By the above lemma, we can write $\mathbb{R}^n$ as the direct sum decomposition (see Section 3.3):

$$\mathbb{R}^n = N(\boldsymbol{x}^*) \oplus T(\boldsymbol{x}^*),$$

that is, given any vector $\boldsymbol{v} \in \mathbb{R}^n$, there are unique vectors $\boldsymbol{w} \in N(\boldsymbol{x}^*)$ and $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ such that

$$\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{y}.$$

19.4 LAGRANGE CONDITION

In this section, we present a first-order necessary condition for extremum problems with constraints. The result is the well-known *Lagrange's theorem.* To better understand the idea underlying this theorem, we first consider functions of two variables and only one equality constraint. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be the constraint function. Recall that at each point $\boldsymbol{x}$ of the domain, the gradient vector $\nabla h(\boldsymbol{x})$ is orthogonal to the level set that passes through that point. Indeed, let us choose a point $\boldsymbol{x}^* = [x_1^*, x_2^*]^T$ such that $h(\boldsymbol{x}^*) = 0$, and assume $\nabla h(\boldsymbol{x}^*) \neq \boldsymbol{0}$. The level set through the point x^* is the set $\{x : h(x) = 0\}$. We then parameterize this level set in a neighborhood of x^* by a curve $\{x(t)\}$, that is, a continuously differentiable vector function $x : \mathbb{R} \to \mathbb{R}^2$ such that

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad t \in (a,b), \quad x^* = x(t^*), \quad \dot{x}(t^*) \neq \mathbf{0}, \quad t^* \in (a,b).$$

We can now show that $\nabla h(x^*)$ is orthogonal to $\dot{x}(t^*)$. Indeed, because h is constant on the curve $\{x(t) : t \in (a,b)\}$, we have that for all $t \in (a,b)$,

$$h(x(t)) = 0.$$

Hence, for all $t \in (a,b)$,

$$\frac{d}{dt}h(x(t)) = 0.$$

Applying the chain rule, we get

$$\frac{d}{dt}h(x(t)) = \nabla h(x(t))^T \dot{x}(t) = 0.$$

Therefore, $\nabla h(x^*)$ is orthogonal to $\dot{x}(t^*)$.

Now suppose that x^* is a minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ on the set $\{x : h(x) = 0\}$. We claim that $\nabla f(x^*)$ is orthogonal to $\dot{x}(t^*)$. To see this, it is enough to observe that the composite function of t given by

$$\phi(t) = f(x(t))$$

achieves a minimum at t^*. Consequently, the first-order necessary condition for the unconstrained extremum problem implies

$$\frac{d\phi}{dt}(t^*) = 0.$$

Applying the chain rule yields

$$0 = \frac{d}{dt}\phi(t^*) = \nabla f(x(t^*))^T \dot{x}(t^*) = \nabla f(x^*)^T \dot{x}(t^*).$$

Thus, $\nabla f(x^*)$ is orthogonal to $\dot{x}(t^*)$. The fact that $\dot{x}(t^*)$ is tangent to the curve $\{x(t)\}$ at x^* means that $\nabla f(x^*)$ is orthogonal to the curve at x^* (see Figure 19.10).

Recall that $\nabla h(x^*)$ is also orthogonal to $\dot{x}(t^*)$. Therefore, the vectors $\nabla h(x^*)$ and $\nabla f(x^*)$ are "parallel", that is, $\nabla f(x^*)$ is a scalar multiple of $\nabla h(x^*)$. The above observations allow us now to formulate *Lagrange's theorem* for functions of two variables with one constraint.

Theorem 19.2 ***Lagrange's Theorem for*** $n = 2$, $m = 1$. *Let the point x^* be a minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ subject to the constraint $h(x) = 0$, $h : \mathbb{R}^2 \to \mathbb{R}$. Then, $\nabla f(x^*)$ and $\nabla h(x^*)$ are parallel. That is, if $\nabla h(x^*) \neq \mathbf{0}$, then there exists a scalar λ^* such that*

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

□

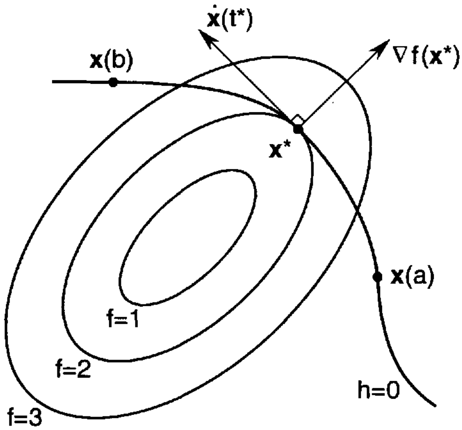


Figure 19.10 The gradient $\nabla f(\boldsymbol{x}^*)$ is orthogonal to the curve $\{\boldsymbol{x}(t)\}$ at the point $\boldsymbol{x}^*$ that is a minimizer of f on the curve

In the above theorem, we refer to λ^* as the *Lagrange multiplier.* Note that the theorem also holds for maximizers. Figure 19.11 gives an illustration of Lagrange's theorem for the case where $\boldsymbol{x}^*$ is a maximizer of f over the set $\{\boldsymbol{x} : h(\boldsymbol{x}) = 0\}$.

Lagrange's theorem provides a first-order necessary condition for a point to be a local minimizer. We call this condition the *Lagrange condition*, which consists of two equations:

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \lambda^* \nabla h(\boldsymbol{x}^*) &= \mathbf{0} \\ h(\boldsymbol{x}) &= 0. \end{aligned}$$

Note that the Lagrange condition is only necessary but not sufficient. In Figure 19.12, we illustrate a variety of points where the Lagrange condition is satisfied, including a case where the point is not an extremizer (neither a maximizer nor a minimizer).

We now generalize Lagrange's theorem for the case when $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m, m \leq n$.

Theorem 19.3 ***Lagrange's Theorem.*** *Let $\boldsymbol{x}^*$ be a local minimizer (or maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$, subject to $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$. Assume that $\boldsymbol{x}^*$ is a regular point. Then, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that*

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}^T.$$

□

Proof. We need to prove that

$$\nabla f(\boldsymbol{x}^*) = -D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^*$$

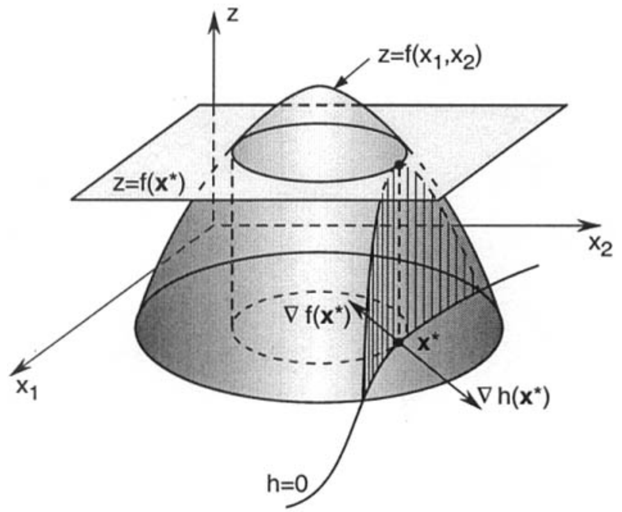


Figure 19.11 Illustration of Lagrange's theorem for $n = 2$, $m = 1$

for some $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, that is, $\nabla f(\boldsymbol{x}^*) \in \mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*)^T) = N(\boldsymbol{x}^*)$. But, by Lemma 19.1, $N(\boldsymbol{x}^*) = T(\boldsymbol{x}^*)^\perp$. Therefore, it remains to show that $\nabla f(\boldsymbol{x}^*) \in T(\boldsymbol{x}^*)^\perp$.

We proceed as follows. Suppose

$$\boldsymbol{y} \in T(\boldsymbol{x}^*).$$

Then, by Theorem 19.1, there exists a differentiable curve $\{\boldsymbol{x}(t) : t \in (a, b)\}$ such that for all $t \in (a, b)$,

$$\boldsymbol{h}(\boldsymbol{x}(t)) = \boldsymbol{0},$$

and there exists $t^* \in (a, b)$ satisfying

$$\boldsymbol{x}(t^*) = \boldsymbol{x}^*, \quad \dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}.$$

Consider now the composite function $\phi(t) = f(\boldsymbol{x}(t))$. Note that t^* is a local minimizer of this function. By the first-order necessary condition for unconstrained local minimizers (see Theorem 6.1),

$$\frac{d\phi}{dt}(t^*) = 0.$$

Applying the chain rule yields

$$\frac{d\phi}{dt}(t^*) = Df(\boldsymbol{x}^*)\dot{\boldsymbol{x}}(t^*) = Df(\boldsymbol{x}^*)\boldsymbol{y} = \nabla f(\boldsymbol{x}^*)^T\boldsymbol{y} = 0.$$

So all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ satisfy

$$\nabla f(\boldsymbol{x}^*)^T\boldsymbol{y} = 0,$$

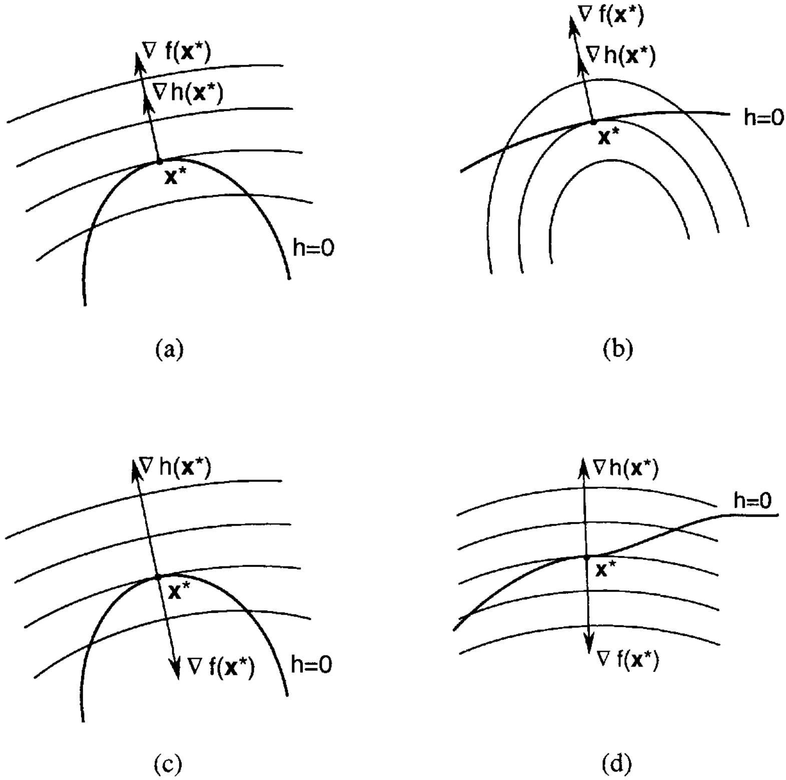


Figure 19.12 Four examples where the Lagrange condition is satisfied: (a) maximizer, (b) minimizer, (c) minimizer, (d) not an extremizer (adapted from [87])

that is

$$\nabla f(x^*) \in T(x^*)^{\perp}.$$

This completes the proof. ∎

Lagrange's theorem states that if x^* is an extremizer, then the gradient of the objective function f can be expressed as a linear combination of the gradients of the constraints. We refer to the vector λ^* in the above theorem as the *Lagrange multiplier vector*, and its components the *Lagrange multipliers*.

Observe that x^* cannot be an extremizer if

$$\nabla f(x^*) \notin N(x^*).$$

This situation is illustrated in Figure 19.13

It is convenient to introduce the so-called *Lagrangian* function $l : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, given by

$$l(x, \lambda) \triangleq f(x) + \lambda^T h(x).$$

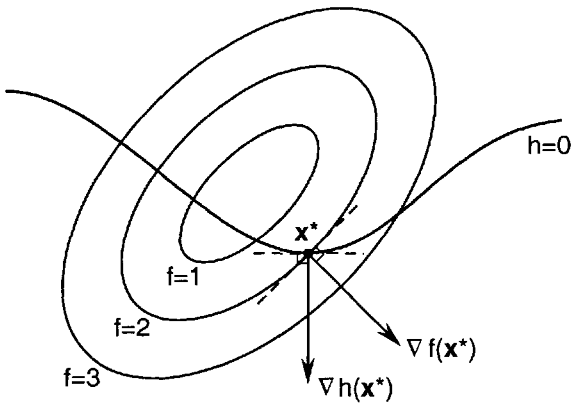


Figure 19.13 An example where the Lagrange condition does not hold

The Lagrange condition for a local minimizer $\boldsymbol{x}^*$ can be represented using the Lagrangian function as

$$Dl(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}^T$$

for some $\boldsymbol{\lambda}^*$, where the derivative operation D is with respect to the entire argument $[\boldsymbol{x}^T, \boldsymbol{\lambda}^T]^T$. In other words, the necessary condition in Lagrange's theorem is equivalent to the first-order necessary condition for unconstrained optimization applied to the Lagrangian function.

To see the above, denote the derivative of l with respect to $\boldsymbol{x}$ as $D_x l$, and the derivative of l with respect to $\boldsymbol{\lambda}$ as $D_\lambda l$. Then,

$$Dl(\boldsymbol{x}, \boldsymbol{\lambda}) = [D_x l(\boldsymbol{x}, \boldsymbol{\lambda}), D_\lambda l(\boldsymbol{x}, \boldsymbol{\lambda})].$$

Note that $D_x l(\boldsymbol{x}, \boldsymbol{\lambda}) = Df(\boldsymbol{x}) + \boldsymbol{\lambda}^T D\boldsymbol{h}(\boldsymbol{x})$ and $D_\lambda l(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{h}(\boldsymbol{x})^T$. Therefore, the Lagrange's theorem for a local minimizer $\boldsymbol{x}^*$ can be stated as

$$\begin{aligned} D_x l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= \mathbf{0}^T \\ D_\lambda l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= \mathbf{0}^T \end{aligned}$$

for some $\boldsymbol{\lambda}^*$, which is equivalent to

$$Dl(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}^T.$$

In other words, the Lagrange condition can be expressed as $Dl(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}^T$.

The Lagrange condition is used to find possible extremizers. This entails solving the equations:

$$\begin{aligned} D_x l(\boldsymbol{x}, \boldsymbol{\lambda}) &= \mathbf{0}^T \\ D_\lambda l(\boldsymbol{x}, \boldsymbol{\lambda}) &= \mathbf{0}^T. \end{aligned}$$

The above represents $n + m$ equations in $n + m$ unknowns. Keep in mind that the Lagrange condition is only necessary, but not sufficient; that is, a point x^* satisfying the above equations need not be an extremizer.

Example 19.5 Given a fixed area of cardboard, we wish to construct a closed cardboard box with maximum volume. We can formulate and solve this problem using the Lagrange condition. Denote the dimensions of the box with maximum volume by x_1, x_2, and x_3, and let the given fixed area of cardboard be A. The problem then can be formulated as

$$\begin{array}{ll} \text{maximize} & x_1 x_2 x_3 \\ \text{subject to} & x_1 x_2 + x_2 x_3 + x_3 x_1 = \dfrac{A}{2}. \end{array}$$

We denote $f(x) = -x_1 x_2 x_3$, and $h(x) = x_1 x_2 + x_2 x_3 + x_3 x_1 - A/2$. We have $\nabla f(x) = -[x_2 x_3, x_1 x_3, x_1 x_2]^T$ and $\nabla h(x) = [x_2 + x_3, x_1 + x_3, x_1 + x_2]^T$. Note that all feasible points are regular in this case. By the Lagrange condition, the dimensions of the box with maximum volume satisfies

$$\begin{aligned} x_2 x_3 - \lambda(x_2 + x_3) &= 0 \\ x_1 x_3 - \lambda(x_1 + x_3) &= 0 \\ x_1 x_2 - \lambda(x_1 + x_2) &= 0 \\ x_1 x_2 + x_2 x_3 + x_3 x_1 &= \frac{A}{2}, \end{aligned}$$

where $\lambda \in \mathbb{R}$.

We now solve the above equations. First, we show that that x_1, x_2, x_3, and λ are all nonzero. Suppose $x_1 = 0$. By the constraints, we have $x_2 x_3 = A/2$. However, the second and third equations in the Lagrange condition yield $\lambda x_2 = \lambda x_3 = 0$, which together with the first equation implies $x_2 x_3 = 0$. This contradicts the constraints. A similar argument applies to x_2 and x_3.

Next, suppose $\lambda = 0$. Then, the sum of the three Lagrange equations gives $x_2 x_3 + x_1 x_3 + x_1 x_2 = 0$, which contradicts the constraints.

We now solve for x_1, x_2, and x_3 in the Lagrange equations. First, multiply the first equation by x_1 and the second by x_2, and subtract one from the other. We arrive at $x_3 \lambda(x_1 - x_2) = 0$. Because neither x_3 nor λ can be zero (by part b), we conclude that $x_1 = x_2$. We similarly deduce that $x_2 = x_3$. From the constraint equation, we obtain $x_1 = x_2 = x_3 = \sqrt{A/6}$.

Notice that we have ignored the constraints that x_1, x_2, and x_3 are positive so that we can solve the problem using Lagrange's theorem. However, there is only one solution to the Lagrange equations, and the solution is positive. Therefore, if a solution exists for the problem with positivity constraints on the variables x_1, x_2, and x_3, then this solution must necessarily be equal to above solution obtained by ignoring the positivity constraints. ∎

Next we provide an example with a quadratic objective function and a quadratic constraint.

Example 19.6 Consider the problem of extremizing the objective function

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$

on the ellipse

$$\{[x_1, x_2]^T : h(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1 = 0\}.$$

We have

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= [2x_1, 2x_2]^T, \\ \nabla h(\boldsymbol{x}) &= [2x_1, 4x_2]^T. \end{aligned}$$

Thus,

$$D_{\boldsymbol{x}} l(\boldsymbol{x}, \lambda) = D_{\boldsymbol{x}}[f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})] = [2x_1 + 2\lambda x_1, 2x_2 + 4\lambda x_2],$$

and

$$D_\lambda l(\boldsymbol{x}, \lambda) = h(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1.$$

Setting $D_{\boldsymbol{x}} l(\boldsymbol{x}, \lambda) = \mathbf{0}^T$ and $D_\lambda l(\boldsymbol{x}, \lambda) = 0$ we obtain three equations in three unknowns

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

All feasible points in this problem are regular. From the first of the above equations, we get either $x_1 = 0$ or $\lambda = -1$. For the case where $x_1 = 0$, the second and third equations imply that $\lambda = -1/2$ and $x_2 = \pm 1/\sqrt{2}$. For the case where $\lambda = -1$, the second and third equations imply that $x_1 = \pm 1$ and $x_2 = 0$. Thus, the points that satisfy the Lagrange condition for extrema are

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \boldsymbol{x}^{(2)} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \end{bmatrix}, \quad \boldsymbol{x}^{(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{x}^{(4)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Because

$$f(\boldsymbol{x}^{(1)}) = f(\boldsymbol{x}^{(2)}) = \frac{1}{2}$$

and

$$f(\boldsymbol{x}^{(3)}) = f(\boldsymbol{x}^{(4)}) = 1$$

we conclude that if there are minimizers, then they are located at $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, and if there are maximizers, then they are located at $\boldsymbol{x}^{(3)}$ and $\boldsymbol{x}^{(4)}$. It turns out that, indeed, $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are minimizers and $\boldsymbol{x}^{(3)}$ and $\boldsymbol{x}^{(4)}$ are maximizers. This problem can be solved graphically, as illustrated in Figure 19.14. ∎

In the above example both the objective function f and the constraint function h are quadratic functions. In the next example, we take a closer look at a class

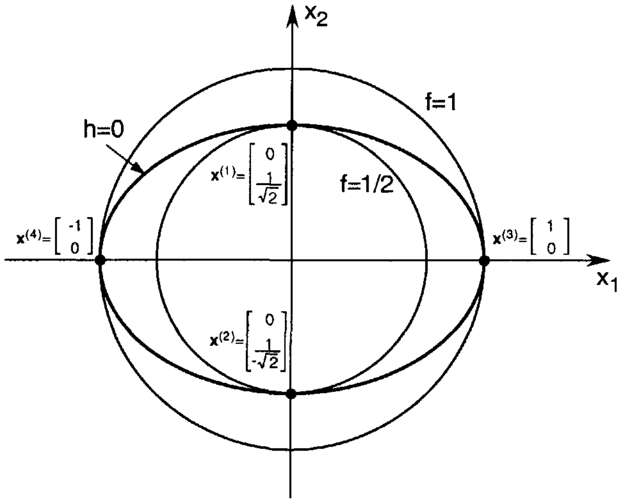


Figure 19.14 Graphical solution of the problem in Example 19.6

of problems where both the objective function f and the constraint h are quadratic functions of n variables.

Example 19.7 Consider the following problem:

$$\text{maximize} \quad \frac{\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}},$$

where $\boldsymbol{Q} = \boldsymbol{Q}^T \geq 0$, and $\boldsymbol{P} = \boldsymbol{P}^T > 0$. Note that if a point $\boldsymbol{x} = [x_1, \ldots, x_n]^T$ is a solution to the problem, then so is any nonzero scalar multiple of it,

$$t\boldsymbol{x} = [tx_1, \ldots, tx_n]^T, \quad t \neq 0.$$

Indeed,

$$\frac{(t\boldsymbol{x})^T\boldsymbol{Q}(t\boldsymbol{x})}{(t\boldsymbol{x})^T\boldsymbol{P}(t\boldsymbol{x})} = \frac{t^2\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}}{t^2\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}} = \frac{\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}}.$$

Therefore, to avoid the multiplicity of solutions, we further impose the constraint

$$\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} = 1.$$

The optimization problem becomes

$$\begin{array}{ll} \text{maximize} & \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} = 1. \end{array}$$

Let us write

$$\begin{aligned} f(\boldsymbol{x}) &= \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \\ h(\boldsymbol{x}) &= 1 - \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}. \end{aligned}$$

Any feasible point for this problem is regular (see Exercise 19.9). We now apply Lagrange's method. We first form the Lagrangian function

$$l(\boldsymbol{x}, \lambda) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \lambda(1 - \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x}).$$

Applying the Lagrange condition yields

$$\begin{aligned} D_x l(\boldsymbol{x}, \lambda) &= 2\boldsymbol{x}^T \boldsymbol{Q} - 2\lambda \boldsymbol{x}^T \boldsymbol{P} = \boldsymbol{0}^T, \\ D_\lambda l(\boldsymbol{x}, \lambda) &= 1 - \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} = 0. \end{aligned}$$

The first of the above equations can be represented as

$$\boldsymbol{Q}\boldsymbol{x} - \lambda \boldsymbol{P}\boldsymbol{x} = \boldsymbol{0}$$

or

$$(\lambda \boldsymbol{P} - \boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}.$$

This representation is possible because $\boldsymbol{P} = \boldsymbol{P}^T$ and $\boldsymbol{Q} = \boldsymbol{Q}^T$. By assumption $\boldsymbol{P} > 0$, hence $\boldsymbol{P}^{-1}$ exists. Premultiplying $(\lambda \boldsymbol{P} - \boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}$ by $\boldsymbol{P}^{-1}$, we obtain

$$(\lambda \boldsymbol{I}_n - \boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}$$

or, equivalently,

$$\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{x} = \lambda \boldsymbol{x}.$$

Therefore, the solution, if it exists, is an eigenvector of $\boldsymbol{P}^{-1}\boldsymbol{Q}$, and the Lagrange multiplier is the corresponding eigenvalue. As usual, let $\boldsymbol{x}^*$ and λ^* be the optimal solution. Because $\boldsymbol{x}^{*T}\boldsymbol{P}\boldsymbol{x}^* = 1$, and $\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{x}^* = \lambda^* \boldsymbol{x}^*$, we have

$$\lambda^* = \boldsymbol{x}^{*T}\boldsymbol{Q}\boldsymbol{x}^*.$$

Hence, λ^* is the maximum of the objective function, and therefore is, in fact, the maximal eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$. ∎

In the above problems, we are able to find points that are candidates for extremizers of the given objective function subject to equality constraints. These critical points are the only candidates because they are the only points that satisfy the Lagrange condition. To classify such critical points as minimizers, maximizers, or neither, we need a stronger condition—possibly a necessary and sufficient condition. In the next section, we discuss a second-order necessary condition and a second-order sufficient condition for minimizers.

19.5 SECOND-ORDER CONDITIONS

We assume that $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable, that is, $f, \boldsymbol{h} \in \mathcal{C}^2$. Let

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda_1 h_1(\boldsymbol{x}) + \cdots + \lambda_m h_m(\boldsymbol{x})$$

be the Lagrangian function. Let $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ be the Hessian matrix of $l(\boldsymbol{x}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{x}$, that is,

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{F}(\boldsymbol{x}) + \lambda_1 \boldsymbol{H}_1(\boldsymbol{x}) + \cdots + \lambda_m \boldsymbol{H}_m(\boldsymbol{x}),$$

where $\boldsymbol{F}(\boldsymbol{x})$ is the Hessian matrix of f at $\boldsymbol{x}$, and $\boldsymbol{H}_k(\boldsymbol{x})$ is the Hessian matrix of h_k at $\boldsymbol{x}$, $k = 1, \ldots, m$, given by

$$\boldsymbol{H}_k(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 h_k}{\partial x_1^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 h_k}{\partial x_n \partial x_1}(\boldsymbol{x}) \\ \vdots & & \vdots \\ \frac{\partial^2 h_k}{\partial x_1 \partial x_n}(\boldsymbol{x}) & \cdots & \frac{\partial^2 h_k}{\partial^2 x_n}(\boldsymbol{x}) \end{bmatrix}.$$

We introduce the notation $[\boldsymbol{\lambda} \boldsymbol{H}(\boldsymbol{x})]$:

$$[\boldsymbol{\lambda} \boldsymbol{H}(\boldsymbol{x})] = \lambda_1 \boldsymbol{H}_1(\boldsymbol{x}) + \cdots + \lambda_m \boldsymbol{H}_m(\boldsymbol{x}).$$

Using the above notation, we can write

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{F}(\boldsymbol{x}) + [\boldsymbol{\lambda} \boldsymbol{H}(\boldsymbol{x})].$$

Theorem 19.4 ***Second-Order Necessary Conditions.*** *Let $\boldsymbol{x}^*$ be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$, and $f, \boldsymbol{h} \in \mathcal{C}^2$. Suppose $\boldsymbol{x}^*$ is regular. Then, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that*

1. *$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T$; and*

2. *for all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, we have $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{y} \geq 0$.*

□

Proof. The existence of $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T$ follows from Lagrange's theorem. It remains to prove the second part of the result. Suppose $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, that is, $\boldsymbol{y}$ belongs to the tangent space to $S = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ at $\boldsymbol{x}^*$. Because $\boldsymbol{h} \in \mathcal{C}^2$, following the argument of Theorem 19.1, there exists a twice differentiable curve $\{\boldsymbol{x}(t) : t \in (a, b)\}$ on S such that

$$\boldsymbol{x}(t^*) = \boldsymbol{x}^*, \quad \dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$$

for some $t^* \in (a, b)$. Observe that by assumption, t^* is a local minimizer of the function $\phi(t) = f(\boldsymbol{x}(t))$. From the second-order necessary condition for unconstrained minimization (see Theorem 6.2), we obtain

$$\frac{d^2 \phi}{dt^2}(t^*) \geq 0.$$

Using the following formula

$$\frac{d}{dt}(\boldsymbol{y}(t)^T \boldsymbol{z}(t)) = \boldsymbol{z}(t)^T \frac{d\boldsymbol{y}}{dt}(t) + \boldsymbol{y}(t)^T \frac{d\boldsymbol{z}}{dt}(t)$$

and applying the chain rule yields

$$\begin{aligned}\frac{d^2\phi}{dt^2}(t^*) &= \frac{d}{dt}[Df(\boldsymbol{x}(t^*))\dot{\boldsymbol{x}}(t^*)]\\ &= \dot{\boldsymbol{x}}(t^*)^T\boldsymbol{F}(\boldsymbol{x}^*)\dot{\boldsymbol{x}}(t^*) + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*)\\ &= \boldsymbol{y}^T\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{y} + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) \geq 0.\end{aligned}$$

Because $\boldsymbol{h}(\boldsymbol{x}(t)) = \boldsymbol{0}$ for all $t \in (a, b)$, we have

$$\frac{d^2}{dt^2}\boldsymbol{\lambda}^{*T}\boldsymbol{h}(\boldsymbol{x}(t)) = 0.$$

Thus, for all $t \in (a, b)$,

$$\begin{aligned}\frac{d^2}{dt^2}\boldsymbol{\lambda}^{*T}\boldsymbol{h}(\boldsymbol{x}(t)) &= \frac{d}{dt}\left[\boldsymbol{\lambda}^{*T}\frac{d}{dt}\boldsymbol{h}(\boldsymbol{x}(t))\right]\\ &= \frac{d}{dt}\left[\sum_{k=1}^{m}\lambda_k^*\frac{d}{dt}h_k(\boldsymbol{x}(t))\right]\\ &= \frac{d}{dt}\left[\sum_{k=1}^{m}\lambda_k^* Dh_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t)\right]\\ &= \sum_{k=1}^{m}\lambda_k^*\frac{d}{dt}(Dh_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t))\\ &= \sum_{k=1}^{m}\lambda_k^*\left[\dot{\boldsymbol{x}}(t)^T\boldsymbol{H}_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t) + Dh_k(\boldsymbol{x}(t))\ddot{\boldsymbol{x}}(t)\right]\\ &= \dot{\boldsymbol{x}}^T(t)[\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}(t))]\dot{\boldsymbol{x}}(t) + \boldsymbol{\lambda}^{*T}D\boldsymbol{h}(\boldsymbol{x}(t))\ddot{\boldsymbol{x}}(t)\\ &= 0.\end{aligned}$$

In particular, the above is true for $t = t^*$, that is,

$$\boldsymbol{y}^T[\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)]\boldsymbol{y} + \boldsymbol{\lambda}^{*T}D\boldsymbol{h}(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) = 0.$$

Adding the above equation to the inequality

$$\boldsymbol{y}^T\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{y} + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) \geq 0$$

yields

$$\boldsymbol{y}^T\left(\boldsymbol{F}(\boldsymbol{x}^*) + [\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)]\right)\boldsymbol{y} + (Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T}D\boldsymbol{h}(\boldsymbol{x}^*))\ddot{\boldsymbol{x}}(t^*) \geq 0.$$

But, by Lagrange's theorem, $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T}D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T$. Therefore,

$$\boldsymbol{y}^T\left(\boldsymbol{F}(\boldsymbol{x}^*) + [\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)]\right)\boldsymbol{y} = \boldsymbol{y}^T\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} \geq 0,$$

which proves the result. ∎

Observe that $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ plays a similar role as the Hessian matrix $\boldsymbol{F}(\boldsymbol{x})$ of the objective function f did in the unconstrained minimization case. However, we now require that $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \geq 0$ only on $T(\boldsymbol{x}^*)$ rather than on $\mathbb{R}^n$.

The above conditions are necessary, but not sufficient, for a point to be a local minimizer. We now present, without a proof, sufficient conditions for a point to be a strict local minimizer.

Theorem 19.5 ***Second-Order Sufficient Conditions.*** *Suppose $f, \boldsymbol{h} \in \mathcal{C}^2$ and there exist a point $\boldsymbol{x}^* \in \mathbb{R}^n$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that*

1. *$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T$; and*

2. *for all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, $\boldsymbol{y} \neq \boldsymbol{0}$, we have $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} > 0$.*

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$. □

Proof. The interested reader can consult [64, p. 307] for a proof of this result. ∎

The above theorem states that if an $\boldsymbol{x}^*$ satisfies the Lagrange condition, and $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is positive definite on $T(\boldsymbol{x}^*)$, then $\boldsymbol{x}^*$ is a strict local minimizer. A similar result to Theorem 19.5 holds for a strict local maximizer, the only difference being that $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ be negative definite on $T(\boldsymbol{x}^*)$. We illustrate this condition in the following example.

Example 19.8 Consider the following problem:

$$\text{maximize} \quad \frac{\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x}},$$

where

$$\boldsymbol{Q} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{P} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

As pointed out earlier, we can represent the above problem in the equivalent form

$$\begin{array}{ll} \text{maximize} & \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} = 1. \end{array}$$

The Lagrangian function for the transformed problem is given by

$$l(\boldsymbol{x}, \lambda) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \lambda(1 - \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x}).$$

The Lagrange condition yields

$$(\lambda \boldsymbol{I} - \boldsymbol{P}^{-1} \boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0},$$

where

$$\boldsymbol{P}^{-1} \boldsymbol{Q} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

There are only two values of λ that satisfy $(\lambda \boldsymbol{I} - \boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}$, namely, the eigenvalues of $\boldsymbol{P}^{-1}\boldsymbol{Q}$: $\lambda_1 = 2$, $\lambda_2 = 1$. We recall from our previous discussion of this problem that the Lagrange multiplier corresponding to the solution is the maximum eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$, namely, $\lambda^* = \lambda_1 = 2$. The corresponding eigenvector is the maximizer, that is, the solution to the problem. The eigenvector corresponding to the eigenvalue $\lambda^* = 2$ satisfying the constraint $\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} = 1$ is $\pm\boldsymbol{x}^*$, where

$$\boldsymbol{x}^* = \left[\frac{1}{\sqrt{2}}, 0\right]^T.$$

At this point, all we have established is that the pairs $(\pm\boldsymbol{x}^*, \lambda^*)$ satisfy the Lagrange condition. We now show that the points $\pm\boldsymbol{x}^*$ are, in fact, strict local maximizers. We do this for the point $\boldsymbol{x}^*$. A similar procedure applies to $-\boldsymbol{x}^*$. We first compute the Hessian matrix of the Lagrangian function. We have

$$\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) = 2\boldsymbol{Q} - 2\lambda\boldsymbol{P} = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}.$$

The tangent space $T(\boldsymbol{x}^*)$ to $\{\boldsymbol{x} : 1 - \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} = 0\}$ is

$$\begin{aligned} T(\boldsymbol{x}^*) &= \{\boldsymbol{y} \in \mathbb{R}^2 : \boldsymbol{x}^{*T}\boldsymbol{P}\boldsymbol{y} = 0\} \\ &= \{\boldsymbol{y} : [\sqrt{2}, 0]\boldsymbol{y} = 0\} \\ &= \{\boldsymbol{y} : \boldsymbol{y} = [0, a]^T,\ a \in \mathbb{R}\}. \end{aligned}$$

Note that for each $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, $\boldsymbol{y} \neq \boldsymbol{0}$,

$$\boldsymbol{y}^T\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*)\boldsymbol{y} = [0, a]\begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} 0 \\ a \end{bmatrix} = -2a^2 < 0.$$

Hence, $\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) < 0$ on $T(\boldsymbol{x}^*)$, and thus $\boldsymbol{x}^* = [1/\sqrt{2}, 0]^T$ is a strict local maximizer. The same is true for the point $-\boldsymbol{x}^*$. Note that

$$\frac{\boldsymbol{x}^{*T}\boldsymbol{Q}\boldsymbol{x}^*}{\boldsymbol{x}^{*T}\boldsymbol{P}\boldsymbol{x}^*} = 2,$$

which, as expected, is the value of the maximal eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$. Finally, we point out that any scalar multiple $t\boldsymbol{x}^*$ of $\boldsymbol{x}^*$, $t \neq 0$, is a solution to the original problem of maximizing $\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}/\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}$. ∎

19.6 MINIMIZING QUADRATICS SUBJECT TO LINEAR CONSTRAINTS

Consider the problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

where $Q > 0$, $A \in \mathbb{R}^{m \times n}$, $m < n$, $\operatorname{rank} A = m$. This problem is a special case of what is called a *quadratic programming* problem (the general form of a quadratic programming problem includes the constraint $x \geq 0$). Note that the constraint set contains an infinite number of points (see Section 2.3). We now show, using Lagrange's theorem, that there is a unique solution to the above optimization problem. Following that, we provide an example illustrating the application of this solution to an optimal control problem.

To solve the problem, we first form the Lagrangian function

$$l(x, \lambda) = \frac{1}{2} x^T Q x + \lambda^T (b - Ax).$$

The Lagrange condition yields

$$D_x l(x^*, \lambda^*) = x^{*T} Q - \lambda^{*T} A = 0^T.$$

Rewriting, we get

$$x^* = Q^{-1} A^T \lambda^*.$$

Premultiplying both sides of the above by A gives

$$Ax^* = AQ^{-1} A^T \lambda^*.$$

Using the fact that $Ax^* = b$, and noting that $AQ^{-1}A^T$ is invertible because $Q > 0$ and $\operatorname{rank} A = m$, we can solve for λ^* to obtain

$$\lambda^* = (AQ^{-1}A^T)^{-1} b.$$

Therefore, we obtain

$$x^* = Q^{-1} A^T (AQ^{-1} A^T)^{-1} b.$$

The point x^* is the only candidate for a minimizer. To establish that x^* is indeed a minimizer, we verify that x^* satisfies the second-order sufficient conditions. For this, we first find the Hessian matrix of the Lagrangian function at (x^*, λ^*). We have

$$L(x^*, \lambda^*) = Q,$$

which is positive definite. Thus, the point x^* is a strict local minimizer. We will see in Chapter 21 that x^* is, in fact, a global minimizer.

The special case where $Q = I_n$, the $n \times n$ identity matrix, reduces to the problem considered in Section 12.3. Specifically, the problem in Section 12.3 is to minimize the norm $\|x\|$ subject to $Ax = b$. The objective function here is $f(x) = \|x\|$, which is not differentiable at $x = 0$. This precludes the use of Lagrange's theorem because the theorem requires differentiability of the objective function. We can overcome this difficulty by considering an equivalent optimization problem:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|x\|^2 \\ \text{subject to} & Ax = b. \end{array}$$

The objective function $\|x\|^2/2$ has the same minimizer as the previous objective function $\|x\|$. Indeed, if x^* is such that for all $x \in \mathbb{R}^n$ satisfying $Ax = b$, $\|x^*\| \le \|x\|$, then $\|x^*\|^2/2 \le \|x\|^2/2$. The same is true for the converse. Because the problem of minimizing $\|x\|^2/2$ subject to $Ax = b$ is simply the problem considered above with $Q = I_n$, we easily deduce the solution to be $x^* = A^T(AA^T)^{-1}b$, which agrees with the solution in Section 12.3.

Example 19.9 Consider the discrete-time linear system model

$$x_k = ax_{k-1} + bu_k, \qquad k \ge 1,$$

with initial condition x_0 given. We can think of $\{x_k\}$ as a discrete-time signal that is controlled by an external input signal $\{u_k\}$. In the control literature, x_k is called the *state* at time k. For a given x_0, our goal is to choose the control signal $\{u_k\}$ so that the state remains "small," over a time interval $[1, N]$, but at the same time the control signal is "not too large." To express the desire to keep the state $\{x_k\}$ small, we choose the control sequence to minimize

$$\frac{1}{2}\sum_{i=1}^{N} x_i^2.$$

On the other hand, maintaining a control signal that is not too large, we minimize

$$\frac{1}{2}\sum_{i=1}^{N} u_i^2.$$

The above two objectives are conflicting in the sense that they cannot, in general, be simultaneously achieved—minimizing the first may result in large control effort, while minimizing the second may result in large states. This is clearly a problem that requires compromise. One way to approach the problem is to minimize a weighted-sum of the above two functions. Specifically, we can formulate the problem as:

$$\begin{array}{ll} \text{minimize} & \displaystyle\frac{1}{2}\sum_{i=1}^{N} \left(qx_i^2 + ru_i^2\right) \\ \text{subject to} & x_k = ax_{k-1} + bu_k,\ k = 1, \ldots, N,\ x_0 \text{ given,} \end{array}$$

where the parameters q and r reflect the relative importance of keeping the state small versus keeping the control effort not too large. This problem is an instance of the *linear quadratic regulator* (LQR) problem (see, e.g., [11], [15], [62], [63], or [71]).

To solve the above problem, we can rewrite it as a quadratic programming problem. Define

$$Q = \begin{bmatrix} qI_N & O \\ O & rI_N \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & & \cdots & 0 & -b & & \cdots & 0 \\ -a & 1 & & \vdots & & -b & & \vdots \\ & \ddots & \ddots & & \vdots & & \ddots & \\ 0 & & -a & 1 & 0 & \cdots & & -b \end{bmatrix}$$

$$b = \begin{bmatrix} ax_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad z = [x_1, \ldots, x_N, u_1, \ldots, u_N]^T.$$

With these definitions, the problem reduces to the previously considered quadratic programming problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} z^T Q z \\ \text{subject to} & Az = b, \end{array}$$

where Q is $2N \times 2N$, A is $N \times 2N$, and $b \in \mathbb{R}^N$. The solution is

$$z^* = Q^{-1} A^T (A Q^{-1} A^T)^{-1} b.$$

The first N components of z^* represent the optimal state signal in the interval $[1, N]$, whereas the second N components represent the optimal control signal.

In practice, the computation of the matrix inverses in the above formula for z^* may be too costly. There are other ways to tackle the problem by exploiting its special structure. This is the study of *optimal control* (see, e.g., [11], [15], [62], [63], or [71]).

∎

The following example illustrates an application of the above discussion.

Example 19.10 *Credit-card holder dilemma.* Suppose we currently have a credit-card debt of \$10,000. Credit-card debts are subject to a monthly interest rate of 2%, and the account balance is increased by the interest amount every month. Each month, we have the option of reducing the account balance by contributing a payment to the account. Over the next 10 months, we plan to contribute a payment every month in such a way as to minimize the overall debt level while at the same time minimize the hardship of making monthly payments.

We solve our problem using the LQR framework as described in Example 19.9. Let the current time be 0, x_k the account balance at the end of month k, and u_k our payment in month k. We have

$$x_k = 1.02 x_{k-1} - u_k, \qquad k = 1, \ldots, 10,$$

that is, the account balance in a given month is equal to the account balance in the previous month plus the monthly interest on that balance minus our payment that month. Our optimization problem is then

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\sum_{i=1}^{10}\left(qx_i^2 + ru_i^2\right) \\ \text{subject to} & x_k = 1.02x_{k-1} - u_k,\ k = 1, \ldots, 10,\ x_0 = 10000, \end{array}$$

which is an instance of the LQR problem. The parameters q and r reflect our priority in trading off between debt reduction and hardship in making payments. The more anxious we are to reduce our debt, the larger the value of q relative to r. On the other hand, the more reluctant we are to make payments, the larger the value of r relative to q.

The solution to the above problem is given by the formula derived in Example 19.9. In Figure 19.15, we plot the monthly account balances and payments over the next 10 months using $q = 1$ and $r = 10$. We can see here that our debt has been reduced to less than \$1,000 after 10 months, but with a first payment close to \$3,000. If we feel that a payment of \$3,000 is too high, then we can try to reduce this amount by increasing the value of r relative to q. However, going too far along these lines can lead to trouble. Indeed, if we use $q = 1$ and $r = 300$ (see Figure 19.16), although the monthly payments do not exceed \$400, the account balance is never reduced by much below \$10,000. In this case, the interest on the account balance eats up a significant portion of our monthly payments. In fact, our debt after 10 months will be higher than \$10,000.

■

EXERCISES

19.1 Find local extremizers for the following optimization problems:

a.

$$\begin{array}{ll} \text{minimize} & x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1 + 5x_2 + 6x_3 \\ \text{subject to} & x_1 + 2x_2 = 3 \\ & 4x_1 + 5x_3 = 6; \end{array}$$

b.

$$\begin{array}{ll} \text{maximize} & 4x_1 + x_2^2 \\ \text{subject to} & x_1^2 + x_2^2 = 9; \end{array}$$

c.

$$\begin{array}{ll} \text{maximize} & x_1x_2 \\ \text{subject to} & x_1^2 + 4x_2^2 = 1. \end{array}$$