1

STATIC OPTIMIZATION

In this chapter we discuss optimization when time is not a parameter. The discussion is preparatory to dealing with time-varying systems in subsequent chapters. A reference that provides an excellent treatment of this material is Bryson and Ho (1975), and we shall sometimes follow their point of view.

Appendix A should be reviewed, particularly the section that discusses matrix calculus.

1.1 OPTIMIZATION WITHOUT CONSTRAINTS

A scalar *performance index* $L(u)$ is given that is a function of a *control* or *decision vector* $u \in R^m$. It is desired to determine the value of u that results in a minimum value of $L(u)$.

We proceed to solving this optimization problem by writing the Taylor series expansion for an increment in L as

$$dL = L_u^{\mathrm{T}}\, du + \frac{1}{2}\, du^{\mathrm{T}} L_{uu}\, du + O(3), \tag{1.1-1}$$

where $O(3)$ represents terms of order three. The gradient of L with respect to u is the column vector

$$L_u \triangleq \frac{\partial L}{\partial u}, \tag{1.1-2}$$

and the Hessian matrix is

$$L_{uu} = \frac{\partial^2 L}{\partial u^2}. \tag{1.1-3}$$

L_{uu} is called the *curvature matrix*. For more discussion on these quantities, see Appendix A.

Note. The gradient is defined throughout the book as a *column* vector, which is at variance with some authors, who define it as a row vector.

A *critical* or *stationary point* is characterized by a zero increment dL to first order for all increments du in the control. Hence,

$$L_u = 0 \tag{1.1-4}$$

for a critical point.

Suppose that we are at a critical point, so $L_u = 0$ in (1.1-1). For the critical point to be a local minimum, it is required that

$$dL = \frac{1}{2}\, du^{\mathrm{T}} L_{uu}\, du + O(3) \tag{1.1-5}$$

is positive for all increments du. This is guaranteed if the curvature matrix L_{uu} is positive definite,

$$L_{uu} > 0. \tag{1.1-6}$$

If L_{uu} is negative definite, the critical point is a local maximum; and if L_{uu} is indefinite, the critical point is a saddle point. If L_{uu} is semidefinite, then higher terms of the expansion (1.1-1) must be examined to determine the type of critical point.

The following example provides a tangible meaning to our initial mathematical developments.

Example 1.1-1. Quadratic Surfaces

Let $u \in R^2$ and

$$L(u) = \frac{1}{2}u^{\mathrm{T}} \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} u + [s_1 \quad s_2]\, u \tag{1}$$

$$\triangleq \frac{1}{2}u^{\mathrm{T}} Q u + S^{\mathrm{T}} u. \tag{2}$$

The critical point is given by

$$L_u = Qu + S = 0 \tag{3}$$

and the optimizing control is

$$u^* = -Q^{-1} S. \tag{4}$$

By examining the Hessian

$$L_{uu} = Q \tag{5}$$

one determines the type of the critical point.

The point u^* is a minimum if $L_{uu} > 0$ and it is a maximum if $L_{uu} < 0$. If $|Q| < 0$, then u^* is a saddle point. If $|Q| = 0$, then u^* is a *singular point* and in this case L_{uu} does not provide sufficient information for characterizing the nature of the critical point.

By substituting (4) into (2) we find the extremal value of the performance index to be

$$
\begin{aligned}
L^* \triangleq L(u^*) &= \frac{1}{2} S^{\mathrm{T}} Q^{-1} Q Q^{-1} S - S^{\mathrm{T}} Q^{-1} S \\
&= -\frac{1}{2} S^{\mathrm{T}} Q^{-1} S. \qquad (6)
\end{aligned}
$$

Let

$$
L = \frac{1}{2} u^{\mathrm{T}} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} u + [0 \quad 1]\, u. \qquad (7)
$$

Then

$$
u^* = -\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad (8)
$$

is a minimum, since $L_{uu} > 0$. Using (6), we see that the minimum value of L is $L^* = -\frac{1}{2}$.

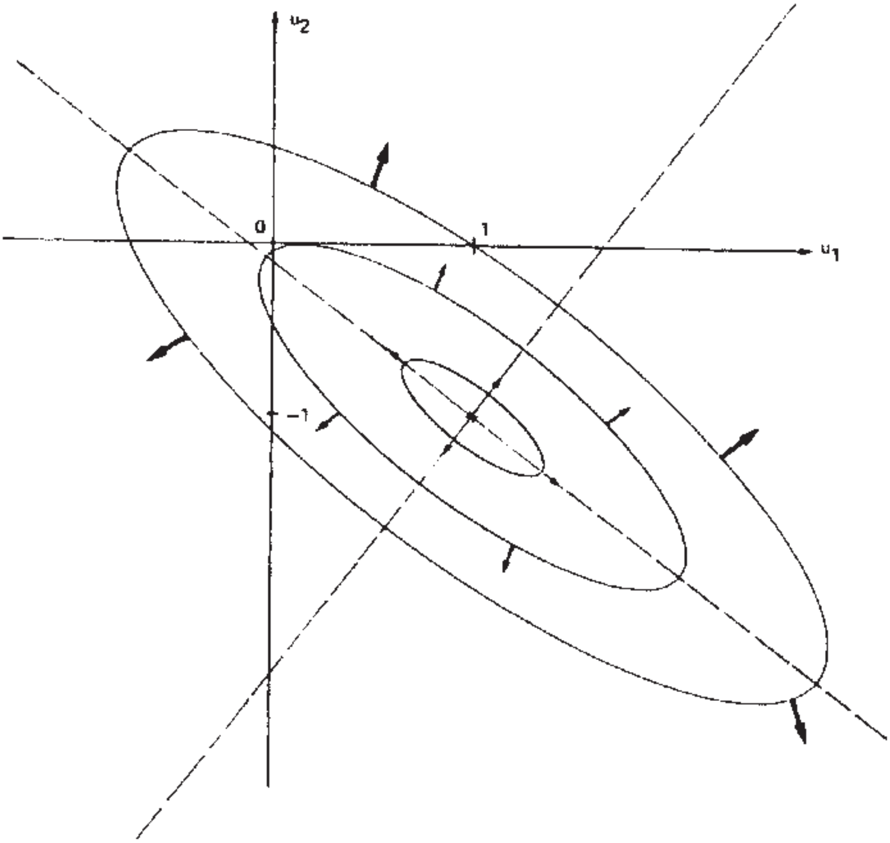

FIGURE 1.1-1 Contours and the gradient vector.

The contours of the $L(u)$ in (7) are drawn in Fig. 1.1-1, where $u = [u_1 \quad u_2]^T$. The arrows represent the gradient

$$L_u = Qu + S = \begin{bmatrix} u_1 + u_2 \\ u_1 + 2u_2 + 1 \end{bmatrix}. \tag{9}$$

Note that the gradient is always perpendicular to the contours and pointing in the direction of increasing $L(u)$.

We shall use an asterisk to denote optimal values of u and L when we want to be explicit. Usually, however, the asterisk will be omitted. ■

Example 1.1-2. Optimization by Scalar Manipulations

We have discussed optimization in terms of vectors and the gradient. As an alternative approach, we could deal entirely in terms of scalar quantities. To demonstrate, let

$$L(u_1, u_2) = \frac{1}{2}u_1^2 + u_1u_2 + u_2^2 + u_2, \tag{1}$$

where u_1 and u_2 are scalars. A critical point is present where the derivatives of L with respect to *all arguments* are equal to zero:

$$\frac{\partial L}{\partial u_1} = u_1 + u_2 = 0,$$

$$\frac{\partial L}{\partial u_2} = u_1 + 2u_2 + 1 = 0. \tag{2}$$

Solving this system of equations yields

$$u_1 = 1, \quad u_2 = -1; \tag{3}$$

thus, the critical point is $(1, -1)$. Note that (1) is an expanded version of (7) in Example 1.1-1, so we have just derived the same answer by another means.

Vector notation is a tool that simplifies the bookkeeping involved in dealing with multidimensional quantities, and for that reason it is very attractive for our purposes. ■

1.2 OPTIMIZATION WITH EQUALITY CONSTRAINTS

Now let the scalar performance index be $L(x, u)$, a function of the control vector $u \in R^m$ and an *auxiliary* (state) *vector* $x \in R^n$. The optimization problem is to determine the control vector u that minimizes $L(x, u)$ and at the same time satisfies the *constraint equation*

$$f(x, u) = 0. \tag{1.2-1}$$

The auxiliary vector x is determined for a given u by the relation (1.2-1). For a given u, (1.2-1) defines a set of n scalar equations.

To find necessary and sufficient conditions for a local minimum that also satisfies $f(x, u) = 0$, we proceed exactly as we did in the previous section, first expanding dL in a Taylor series and then examining the first- and second-order terms. Let us first gain some insight into the problem, however, by considering it from three points of view (Bryson and Ho 1975, Athans and Falb 1966).

Lagrange Multipliers and the Hamiltonian

Necessary Conditions At a stationary point, dL is equal to zero in the first-order approximation with respect to increments du when df is zero. Thus, at a critical point the following equations are satisfied:

$$dL = L_u^{\mathrm{T}}\, du + L_x^{\mathrm{T}}\, dx = 0 \tag{1.2-2}$$

and

$$df = f_u\, du + f_x\, dx = 0. \tag{1.2-3}$$

Since (1.2-1) determines x for a given u, the increment dx is determined by (1.2-3) for a given control increment du. Thus, the Jacobian matrix f_x is nonsingular and one can write

$$dx = -f_x^{-1} f_u\, du. \tag{1.2-4}$$

Substituting this into (1.2-2) yields

$$dL = \left(L_u^{\mathrm{T}} - L_x^{\mathrm{T}} f_x^{-1} f_u\right) du. \tag{1.2-5}$$

The derivative of L with respect to u holding f constant is therefore given by

$$\left.\frac{\partial L}{\partial u}\right|_{df=0} = \left(L_u^{\mathrm{T}} - L_x^{\mathrm{T}} f_x^{-1} f_u\right)^{\mathrm{T}} = L_u - f_u^{\mathrm{T}} f_x^{-\mathrm{T}} L_x, \tag{1.2-6}$$

where $f_x^{-\mathrm{T}}$ means $(f_x^{-1})^{\mathrm{T}}$. Note that

$$\left.\frac{\partial L}{\partial u}\right|_{dx=0} = L_u. \tag{1.2-7}$$

Thus, for dL to be zero in the first-order approximation with respect to arbitrary increments du when $df = 0$, we must have

$$L_u - f_u^{\mathrm{T}} f_x^{-\mathrm{T}} L_x = 0. \tag{1.2-8}$$

This is a necessary condition for a minimum. Before we derive a sufficient condition, let us develop some more insight by examining two more ways to obtain (1.2-8). Write (1.2-2) and (1.2-3) as

$$\begin{bmatrix} dL \\ df \end{bmatrix} = \begin{bmatrix} L_x^{\mathrm{T}} & L_u^{\mathrm{T}} \\ f_x & f_u \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} = 0. \tag{1.2-9}$$

This set of linear equations defines a stationary point, and it must have a solution $[dx^{\mathrm{T}}\ du^{\mathrm{T}}]^{\mathrm{T}}$. The critical point is obtained only if the $(n+1) \times (n+m)$ coefficient matrix has rank less than $n+1$. That is, its rows must be linearly dependent so there exists an n vector λ such that

$$[1 \quad \lambda^{\mathrm{T}}] \begin{bmatrix} L_x^{\mathrm{T}} & L_u^{\mathrm{T}} \\ f_x & f_u \end{bmatrix} = 0. \tag{1.2-10}$$

Then

$$L_x^{\mathrm{T}} + \lambda^{\mathrm{T}} f_x = 0, \tag{1.2-11}$$

$$L_u^{\mathrm{T}} + \lambda^{\mathrm{T}} f_u = 0. \tag{1.2-12}$$

Solving (1.2-11) for λ gives

$$\lambda^{\mathrm{T}} = -L_x^{\mathrm{T}} f_x^{-1}, \tag{1.2-13}$$

and substituting in (1.2-12) again yields the condition (1.2-8) for a critical point.

Note. The left-hand side of (1.2-8) is the transpose of the Schur complement of L_u^{T} in the coefficient matrix of (1.2-9) (see Appendix A for more details).

The vector $\lambda \in R^n$ is called a *Lagrange multiplier*, and it will turn out to be an extremely useful tool for us. To give it some additional meaning now, let $du = 0$ in (1.2-2), (1.2-3) and eliminate dx to get

$$dL = L_x^{\mathrm{T}} f_x^{-1}\, df. \tag{1.2-14}$$

Therefore,

$$\left.\frac{\partial L}{\partial f}\right|_{du=0} = \left(L_x^{\mathrm{T}} f_x^{-1}\right)^{\mathrm{T}} = -\lambda, \tag{1.2-15}$$

so that $-\lambda$ is the partial of L with respect to the constraint holding the control u constant. It shows the effect on the performance index of holding the control constant when the constraints are changed.

As a third method of obtaining (1.2-8), let us develop the approach we shall use for our analysis in subsequent chapters. Include the constraints in the performance index to define the *Hamiltonian* function

$$H(x, u, \lambda) = L(x, u) + \lambda^{\mathrm{T}} f(x, u), \tag{1.2-16}$$

where $\lambda \in R^n$ is an as yet undetermined Lagrange multiplier. To determine x, u, and λ, which result in a critical point, we proceed as follows.

Increments in H depend on increments in x, u, and λ according to

$$dH = H_x^{\mathrm{T}}\, dx + H_u^{\mathrm{T}}\, du + H_\lambda^{\mathrm{T}} d\lambda. \tag{1.2-17}$$

Note that

$$H_\lambda = \frac{\partial H}{\partial \lambda} = f(x, u), \tag{1.2-18}$$

so suppose we choose some value of u and demand that

$$H_\lambda = 0. \tag{1.2-19}$$

Then x is determined for the given u by $f(x, u) = 0$, which is the constraint relation. In this situation the Hamiltonian equals the performance index:

$$H|_{f=0} = L. \tag{1.2-20}$$

Recall that if $f = 0$, then dx is given in terms of du by (1.2-4). We should rather not take into account this coupling between du and dx, so it is convenient to choose λ so that

$$H_x = 0. \tag{1.2-21}$$

Then, by (1.2-17), increments dx do not contribute to dH. Note that this yields a value for λ given by

$$\frac{\partial H}{\partial x} = L_x + f_x^{\mathrm{T}} \lambda = 0 \tag{1.2-22}$$

or (1.2-13).

If (1.2-19) and (1.2-21) hold, then

$$dL = dH = H_u^{\mathrm{T}}\, du, \tag{1.2-23}$$

since $H = L$ in this situation. To achieve a stationary point, we must therefore finally impose the *stationarity condition*

$$H_u = 0. \tag{1.2-24}$$

In summary, necessary conditions for a minimum point of $L(x, u)$ that also satisfies the constraint $f(x, u) = 0$ are

$$\frac{\partial H}{\partial \lambda} = f = 0, \tag{1.2-25a}$$

$$\frac{\partial H}{\partial x} = L_x + f_x^{\mathrm{T}} \lambda = 0, \tag{1.2-25b}$$

$$\frac{\partial H}{\partial u} = L_u + f_u^{\mathrm{T}} \lambda = 0, \tag{1.2-25c}$$

with $H(x, u, \lambda)$ defined by (1.2-16). The way we shall often use them, these three equations serve to determine x, λ, and u in that respective order. The last two of these equations are (1.2-11) and (1.2-12). In most applications determining the value of λ is not of interest, but this value is required, since it is an intermediate variable that allows us to determine the quantities of interest, u, x, and the minimum value of L.

The usefulness of the Lagrange-multiplier approach can be summarized as follows. In reality dx and du are not independent increments, because of (1.2-4). By introducing an undetermined multiplier λ, however, we obtain an extra degree of freedom, and λ can be selected to make dx and du behave *as if* they were independent increments. Therefore, setting independently to zero the gradients of H with respect to *all arguments* as in (1.2-25) yields a critical point. By introducing Lagrange multipliers, the problem of minimizing $L(x, u)$ subject to the constraint $f(x, u) = 0$ is replaced with the problem of *minimizing the Hamiltonian* $H(x, u, \lambda)$ *without constraints*.

Sufficient Conditions Conditions (1.2-25) determine a stationary (critical) point. We are now ready to derive a test that guarantees that this point is a minimum. We proceed as we did in Section 1.1.

Write Taylor series expansions for increments in L and f as

$$dL = \begin{bmatrix} L_x^{\mathrm{T}} & L_u^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} + \frac{1}{2} \begin{bmatrix} dx^{\mathrm{T}} & du^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} L_{xx} & L_{xu} \\ L_{ux} & L_{uu} \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} + O(3), \tag{1.2-26}$$

$$df = \begin{bmatrix} f_x & f_u \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} + \frac{1}{2} \begin{bmatrix} dx^{\mathrm{T}} & du^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xu} \\ f_{ux} & f_{uu} \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} + O(3), \tag{1.2-27}$$

where

$$f_{xu} \triangleq \frac{\partial^2 f}{\partial u\, dx}$$

and so on. (What are the dimensions of f_{xu}?) To introduce the Hamiltonian, use these equations to see that

$$\begin{bmatrix} 1 & \lambda^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} dL \\ df \end{bmatrix} = \begin{bmatrix} H_x^{\mathrm{T}} & H_u^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} + \frac{1}{2} \begin{bmatrix} dx^{\mathrm{T}} & du^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} + O(3). \tag{1.2-28}$$

A critical point requires that $f = 0$, and also that dL is zero in the first-order approximation for all increments dx, du. Since f is held equal to zero, df is also zero. Thus, these conditions require $H_x = 0$ and $H_u = 0$ exactly as in (1.2-25).

To find sufficient conditions for a minimum, let us examine the second-order term. First, it is necessary to include in (1.2-28) the dependence of dx on du. Hence, let us suppose we are at a critical point so that $H_x = 0$, $H_u = 0$, and $df = 0$. Then by (1.2-27)

$$dx = -f_x^{-1} f_u\, du + O(2). \tag{1.2-29}$$

Substituting this relation into (1.2-28) yields

$$dL = \frac{1}{2} du^{\mathrm{T}} \begin{bmatrix} -f_u^{\mathrm{T}} f_x^{-\mathrm{T}} & I \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} -f_x^{-1} f_u \\ I \end{bmatrix} du + O(3). \tag{1.2-30}$$

To ensure a minimum, dL in (1.2-30) should be positive for all increments du. This is guaranteed if the *curvature matrix with constant f equal to zero*

$$\begin{aligned} L_{uu}^f \triangleq L_{uu}|_f &= \begin{bmatrix} -f_u^{\mathrm{T}} f_x^{-\mathrm{T}} & I \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} -f_x^{-1} f_u \\ I \end{bmatrix} \\ &= H_{uu} - f_u^{\mathrm{T}} f_x^{-\mathrm{T}} H_{xu} - H_{ux} f_x^{-1} f_u + f_u^{\mathrm{T}} f_x^{-\mathrm{T}} H_{xx} f_x^{-1} f_u \end{aligned} \tag{1.2-31}$$

is positive definite. Note that if the constraint $f(x, u)$ is identically zero for all x and u, then (1.2-31) reduces to L_{uu} in (1.1-6). If (1.2-31) is negative definite (indefinite), then the stationary point is a constrained maximum (saddle point).

Examples

To gain a feel for the theory we have just developed, let us consider some examples. The first example is a geometric problem that allows easy visualization, while the second involves a quadratic performance index and linear constraint. The second example is representative of the case that is used extensively in controller design for linear systems.

Example 1.2-1. Quadratic Surface with Linear Constraint

Suppose the performance index is as given in Example 1.1-1:

$$L(x, u) = \frac{1}{2}\begin{bmatrix} x & u \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}, \tag{1}$$

where we have simply renamed the old scalar components u_1, u_2 as x, u, respectively. Let the constraint be

$$f(x, u) = x - 3 = 0. \tag{2}$$

The Hamiltonian is

$$H = L + \lambda^{\mathrm{T}} f = \frac{1}{2}x^2 + xu + u^2 + u + \lambda(x - 3), \tag{3}$$

where λ is a scalar. The conditions for a stationary point are (1.2-25), or

$$H_\lambda = x - 3 = 0, \tag{4}$$

$$H_x = x + u + \lambda = 0, \tag{5}$$

$$H_u = x + 2u + 1 = 0. \tag{6}$$

Solving in the order (4), (6), (5) yields $x = 3$, $u = -2$, and $\lambda = -1$. The stationary point is therefore

$$(x, u)^* = (3, -2). \tag{7}$$

To verify that (7) is a minimum, find the constrained curvature matrix (1.2-31):

$$L_{uu}^f = 2. \tag{8}$$

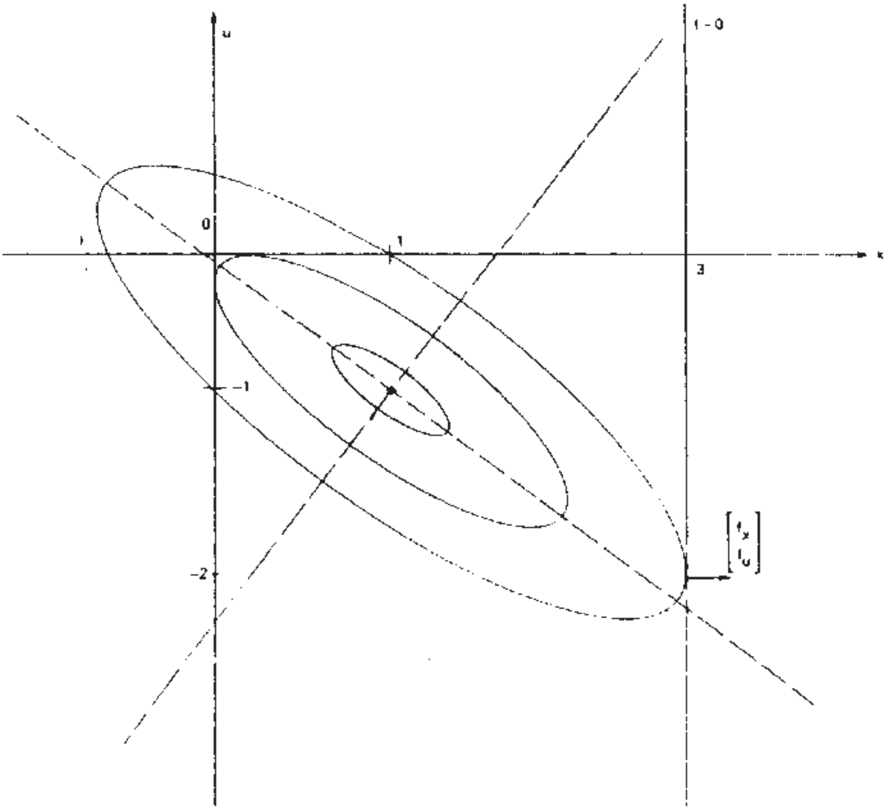

FIGURE 1.2-1 Contours of $L(x, u)$, and the constraint $f(x, u)$.

This is positive, so (7) is a minimum. The contours of $L(x, u)$ and the constraint (2) are shown in Fig. 1.2-1.

It is worthwhile to make an important point. The gradient of $f(x, u)$ in the (x, u) plane is

$$\begin{bmatrix} f_x \\ f_u \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{9}$$

as shown in Fig. 1.2-1. The gradient of $L(x, u)$ in the plane is

$$\begin{bmatrix} L_x \\ L_u \end{bmatrix} = \begin{bmatrix} x + u \\ x + 2u + 1 \end{bmatrix} \tag{10}$$

(cf. (9) in Example 1.1-1). At the constrained minimum $(3, -2)$, this has a value of

$$\begin{bmatrix} L_x \\ L_u \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{11}$$

Note that the gradients of f and L are *parallel* at the stationary point. This means that the constrained minimum occurs where the constraint (2) is tangent to an elliptical contour of L. Moving in either direction along the line $f = 0$ will then increase the value of L. The value of L at the constrained minimum is found by substituting $x = 3$, $u = -2$ into (1) to be $L^* = 0.5$. Since $\lambda = -1$, holding u constant at -2 and changing the constraint by df (i.e., moving the line in Fig. 1.2-1 to the right by df) will result in an *increase* in the value of $L(x, u)$ of $dL = -\lambda\, df = df$ (see (1.2-15)). ■

Example 1.2-2. Quadratic Performance Index with Linear Constraint

Consider the quadratic performance index

$$L(x, u) = \frac{1}{2}x^{\mathrm{T}}Qx + \frac{1}{2}u^{\mathrm{T}}Ru \tag{1}$$

with linear constraint

$$f(x, u) = x + Bu + c = 0, \tag{2}$$

where $x \in R^n$, $u \in R^m$, $f \in R^n$, $\lambda \in R^n$, Q, R, and B are matrices, and c is an n vector. We assume $Q > 0$ and $R > 0$ (with both symmetric). This *static linear quadratic* (LQ) *problem* will be further generalized in Chapters 2 and 3 to apply to time-varying systems.

The contours of $L(x, u)$ are hyperellipsoids, and $f(x, u) = 0$ defines a hyperplane intersecting them. The stationary point occurs where the gradients of f and L are parallel.

The Hamiltonian is

$$H = \frac{1}{2}x^{\mathrm{T}}Qx + \frac{1}{2}u^{\mathrm{T}}Ru + \lambda^{\mathrm{T}}(x + Bu + c) \tag{3}$$

and the conditions for a stationary point are

$$H_\lambda = x + Bu + c = 0, \tag{4}$$

$$H_x = Qx + \lambda = 0, \tag{5}$$

$$H_u = Ru + B^{\mathrm{T}}\lambda = 0. \tag{6}$$

To solve these, first use the stationarity condition (6) to find an expression for u in terms of λ,

$$u = -R^{-1}B^{\mathrm{T}}\lambda. \tag{7}$$

According to (5)

$$\lambda = -Qx, \tag{8}$$

and taking into account (4) results in

$$\lambda = QBu + Qc. \tag{9}$$

Using this in (7) yields

$$u = -R^{-1}B^{\mathrm{T}}(QBu + Qc) \tag{10}$$

or

$$\begin{aligned}(I + R^{-1}B^{\mathrm{T}}QB)u &= -R^{-1}B^{\mathrm{T}}Qc,\\ (R + B^{\mathrm{T}}QB)u &= -B^{\mathrm{T}}Qc.\end{aligned} \tag{11}$$

Since $R > 0$ and $B^{\mathrm{T}}QB > 0$, we can invert $R + B^{\mathrm{T}}QB$ and so the optimal control is

$$u = -(R + B^{\mathrm{T}}QB)^{-1}B^{\mathrm{T}}Qc. \tag{12}$$

Using (12) in (4) and (9) gives the optimal-state and multiplier values of

$$x = -(I - B(R + B^{\mathrm{T}}QB)^{-1}B^{\mathrm{T}}Q)c, \tag{13}$$

$$\lambda = (Q - QB(R + B^{\mathrm{T}}QB)^{-1}B^{\mathrm{T}}Q)c. \tag{14}$$

By the matrix inversion lemma (see Appendix A)

$$\lambda = (Q^{-1} + BR^{-1}B^{\mathrm{T}})^{-1}c \tag{15}$$

if $|Q| \neq 0$.

To verify that control (12) results in a minimum, use (1.2-31) to determine that the constrained curvature matrix is

$$L_{uu}^{f} = R + B^{\mathrm{T}}QB, \tag{16}$$

which is positive definite by our restrictions on R and Q. Using (12) and (13) in (1) yields the optimal value

$$L^* = \frac{1}{2}c^{\mathrm{T}}\left[Q - QB(R + B^{\mathrm{T}}QB)^{-1}B^{\mathrm{T}}Q\right]c, \tag{17}$$

$$L^* = \frac{1}{2}c^{\mathrm{T}}\lambda, \tag{18}$$

so that

$$\frac{\partial L^*}{\partial c} = \lambda. \tag{19}$$

■

Effect of Changes in Constraints

Equation (1.2-28) expresses the increment dL in terms of df, dx, and du. In the discussion following that equation we let $df = 0$, found dx in terms of du, and expressed dL in terms of du. That gave us conditions for a stationary point ($H_x = 0$ and $H_u = 0$) and led to the second-order coefficient matrix L_{uu}^f in (1.2-31), which provided a test for the stationary point to be a constrained minimum.

In this subsection we are interested in dL as a function of an increment df in the constraint. We want to see how the performance index L changes in response to changes in the constraint f *if we remain at a stationary point*. We are therefore trying to find stationary points *near* a given stationary point. See Fig. 1.2-2, which shows how the stationary point moves with changes in f.

At the stationary point $(u, x)^*$ defined by $f(x, u) = 0$, the conditions $H_\lambda = 0$, $H_x = 0$, and $H_u = 0$ are satisfied. If the constraint changes by an increment so that $f(x, u) = df$, then the stationary point moves to $(u + du, x + dx)$. The partials in (1.2-25) change by

$$dH_\lambda = df = f_x\,dx + f_u\,du, \tag{1.2-32a}$$

$$dH_x = H_{xx}\,dx + H_{xu}\,du + f_x^{\mathrm{T}} d\lambda, \tag{1.2-32b}$$

$$dH_u = H_{ux}\,dx + H_{uu}\,du + f_u^{\mathrm{T}} d\lambda. \tag{1.2-32c}$$

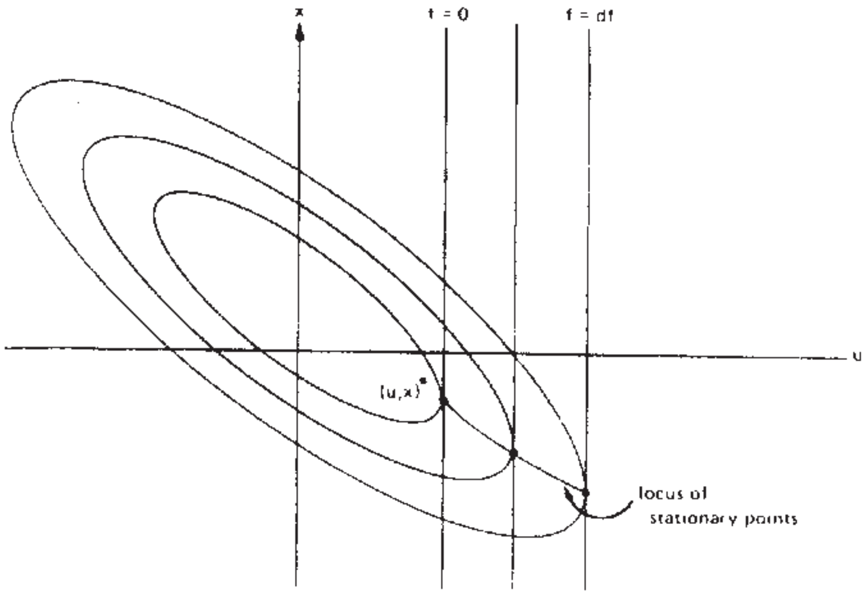

FIGURE 1.2-2 Locus of stationary points as the constraint varies.

In order that we remain at a stationary point, the increments dH_x and dH_u should be zero. This requirement imposes certain relations between the changes dx, du, and df, which we shall use in (1.2-28) to determine dL as a function of df.

To find dx and du as functions of df with the requirement that we remain at an optimal solution, use (1.2-32a) to find

$$dx = f_x^{-1}\, df - f_x^{-1} f_u\, du, \tag{1.2-33}$$

and set (1.2-32b) to zero to find

$$d\lambda = -f_x^{-\mathrm{T}}(H_{xx}\, dx + H_{xu}\, du). \tag{1.2-34}$$

Now use these relations in (1.2-32c) to obtain

$$\begin{aligned} dH_u = {} & \left(H_{uu} - H_{ux} f_x^{-1} f_u - f_u^{\mathrm{T}} f_x^{-\mathrm{T}} H_{xu} + f_u^{\mathrm{T}} f_x^{-\mathrm{T}} H_{xx} f_x^{-1} f_u\right) du \\ & + \left(H_{ux} - f_u^{\mathrm{T}} f_x^{-\mathrm{T}} H_{xx}\right) f_x^{-1}\, df = 0 \end{aligned}$$

so that

$$du = -\left(L_{uu}^f\right)^{-1} \left(H_{ux} - f_u^{\mathrm{T}} f_x^{-\mathrm{T}} H_{xx}\right) f_x^{-1}\, df \overset{\Delta}{=} -C\, df. \tag{1.2-35}$$

Using (1.2-35) in (1.2-33) yields

$$\begin{aligned} dx &= \left[I + f_x^{-1} f_u \left(L_{uu}^f\right)^{-1} \left(H_{ux} - f_u^{\mathrm{T}} f_x^{-1} H_{xx}\right)\right] f_x^{-1}\, df \\ &= f_x^{-1}\left(I + f_u C\right) df. \end{aligned} \tag{1.2-36}$$

Equations (1.2-35) and (1.2-36) are the required expressions for the increments in the stationary values of control and state as functions of df. If $|L_{uu}^f| \neq 0$, then dx and du can be determined in terms of df, and the existence of neighboring optimal solutions as f varies is guaranteed.

To determine the increment dL in the optimal performance index as a function of df, substitute (1.2-35) and (1.2-36) into (1.2-28), using $H_x = 0$, $dH_u = 0$, since we began at a stationary point $(u, x)^*$. The result is found after some work to be

$$dL = -\lambda^{\mathrm{T}}\, df + \frac{1}{2}\, df^{\mathrm{T}} \left(f_x^{-\mathrm{T}} H_{xx} f_x^{-1} - C^{\mathrm{T}} L_{uu}^f C\right) df + O(3). \tag{1.2-37}$$

From this we see that the first and second partial derivatives of $L^*(x, u)$ with respect to $f(x, u)$ under the restrictions $dH_x = 0$, $dH_u = 0$ are

$$\left.\frac{\partial L^*}{\partial f}\right|_{H_x, H_u} = -\lambda, \tag{1.2-38}$$

$$\left.\frac{\partial^2 L^*}{\partial f^2}\right|_{H_x, H_u} = f_x^{-\mathrm{T}} H_{xx} f_x^{-1} - C^{\mathrm{T}} L_{uu}^f C. \tag{1.2-39}$$

Equation (1.2-38) allows a further interpretation of the Lagrange multiplier; it indicates the rate of change of the *optimal* value of the performance index with respect to the constraint.

1.3 NUMERICAL SOLUTION METHODS

Analytic solutions for the stationary point $(u, x)^*$ and minimal value L^* of the performance index cannot be found except for simple functions $L(x, u)$ and $f(x, u)$. In most practical cases, numerical optimization methods must be used. Many methods exist, but *steepest descent* or *gradient* (Luenberger 1969, Bryson and Ho 1975) methods are probably the simplest.

The steps in constrained minimization by the method of steepest descent are (Bryson and Ho 1975)

1. Select an initial value for u.
2. Determine x from $f(x, u) = 0$.
3. Determine λ from $\lambda = -f_x^{-T} L_x$.
4. Determine the gradient vector $H_u = L_u + f_u^T \lambda$.
5. Update the control vector by $\Delta u = -\alpha H_u$, where K is a positive scalar constant (to find a maximum use $\Delta u = \alpha H_u$).
6. Determine the predicted change in the value of L, $\Delta L = H_u^T \Delta u = -\alpha H_u^T H_u$. If ΔL is sufficiently small, stop. Otherwise, go to step 2.

There are many variations to this procedure. If the step-size constant K is too large, then the algorithm may overshoot the stationary point $(u, x)^*$ and convergence may not occur. The step size should usually be reduced as $(u, x)^*$ is approached, and several of the existing variations differ in the approach to adapting K.

Many software routines are available for unconstrained optimization. The numerical solution of the constrained optimization problem of minimizing $L(x, u)$ subject to $f(x, u) = 0$ can be obtained using the MATLAB function *constr.m* available under the Optimization Toolbox. This function takes in the user-defined subroutine *funct.m*, which computes the value of the function, the constraints, and the initial conditions.

PROBLEMS

Section 1.1

1.1-1. Find the critical points u^* (classify them) and the value of $L(u^*)$ in Example 1.1-1 if

a. $Q = \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix}, \quad S^T = [0 \quad 1].$

b. $Q = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}, \quad S^T = [0 \quad 1].$

Sketch the contours of L and find the gradient L_u.

1.1-2. Find the minimum value of

$$L(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 + 3x_1. \tag{1}$$

Find the curvature matrix at the minimum. Sketch the contours, showing the gradient at several points.

1.1-3. Failure of test for minimality. The function $f(x, y) = x^2 + y^4$ has a minimum at the origin.

a. Verify that the origin is a critical point.

b. Show that the curvature matrix is singular at the origin.

c. Prove that the critical point is indeed a minimum.

Section 1.2

1.2-1. Ship closest point of approach. A ship is moving at 10 miles per hour on a course of 30° (measured clockwise from north, which is 0°). Find its closest point of approach to an island that at time $t = 0$ is 20 miles east and 30 miles north of it. Find the distance to the island at this point. Find the time of closest approach.

1.2-2. Shortest distance between two points. Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be two given points. Find the third point $P_3 = (x_3, y_3)$ such that $d_1 = d_2$ is minimized, where d_1 is the distance from P_3 to P_1 and d_2 is the distance from P_3 to P_2.

1.2-3. Meteor closest point of approach. A meteor is in a hyperbolic orbit described with respect to the earth at the origin by

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{1}$$

Find its closest point of approach to a satellite that is in such an orbit that it has a constant position of (x_1, y_1). Verify that the solution indeed yields a minimum.

1.2-4. Shortest distance between a parabola and a point. A meteor is moving along the path

$$y = x^2 + 3x - 6. \tag{1}$$

A space station is at the point $(x, y) = (2, 2)$.

a. Use Lagrange multipliers to find a cubic equation for x at the closest point of approach.

b. Find the closest point of approach (x, y), and the distance from this point to (2, 2).

1.2-5. Rectangles with maximum area, minimum perimeter

a. Find the rectangle of maximum area with perimeter p. That is, maximize

$$L(x, y) = xy \tag{1}$$

subject to

$$f(x, y) = 2x + 2y - p = 0. \tag{2}$$

b. Find the rectangle of minimum perimeter with area a^2. That is, minimize

$$L(x, y) = 2x + 2y \tag{3}$$

subject to

$$f(x, y) = xy - a^2 = 0. \tag{4}$$

c. In each case, sketch the contours of $L(x, y)$ and the constraint. Optimization problems related like these two are said to be *dual*.

1.2-6. Linear quadratic case. Minimize

$$L = \frac{1}{2}x^{\mathrm{T}} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} x + \frac{1}{2}u^{\mathrm{T}} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} u$$

if

$$x = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 0 \end{bmatrix} u.$$

Find x^*, u^*, λ^*, L^*.

1.2-7. Linear quadratic case. In the LQ problem define the *Kalman gain*

$$K \triangleq (B^{\mathrm{T}}QB + R)^{-1}B^{\mathrm{T}}Q \tag{1}$$

a. Express u^*, λ^*, x^*, and L^* in terms of K.

b. Let

$$S_0 \triangleq Q - QB(B^{\mathrm{T}}QB + R)^{-1}B^{\mathrm{T}}Q \tag{2}$$

so that $L^* = c^{\mathrm{T}}S_0c/2$. Show that

$$S_0 = Q(I - BK) = (I - BK)^{\mathrm{T}}Q(I - BK) + K^{\mathrm{T}}RK. \tag{3}$$

Hence, factor L^* as a perfect square. (Let $\sqrt{Q}$ and $\sqrt{R}$ be the square roots of Q and R.)

c. Show that

$$S_0 = (Q^{-1} + BR^{-1}B^{\mathrm{T}})^{-1}. \tag{4}$$

1.2-8. Geometric mean less than or equal to arithmetic mean

a. Show that the minimum value of $x^2y^2z^2$ on the sphere $x^2 + y^2 + z^2 = r^2$ is $(r^2/3)^3$.
b. Show that the maximum value of $x^2 + y^2 + z^2$ on the sphere $x^2y^2z^2 = (r^2/3)^3$ is r^2.
c. Generalize part a or b and so deduce that, for $a_i > 0$,

$$(a_1a_2\cdots a_n)^{1/n} \le (a_1 + a_2 + \cdots + a_n)/n.$$

Note: The problems in parts a and b are dual (Fulks 1967).

1.2-9. Find the point nearest the origin on the line $3x + 2y + z = 1$, $x + 2y - 3z = 4$.

1.2-10. Rectangle inside Ellipse

a. Find the rectangle of maximum perimeter that can be inscribed inside an ellipse. That is, maximize $4(x + y)$ subject to constraint $x^2/a^2 + y^2/b^2 = 1$.
b. Find the rectangle of maximum area $4xy$ that can be inscribed inside an ellipse.