ANALYTICAL METHODS OF OPTIMIZATION

D. F. LAWDEN
Emeritus Professor
University of Aston in Birmingham, U.K.

DOVER PUBLICATIONS, INC.
Mineola, New York

1 STATIC SYSTEMS

1.1 Statement of the problem

The fundamental problem which we shall study can be expressed in the following terms: Suppose the state of a given physical system can be changed by varying the values of N physical quantities denoted by $u_1, u_2, \ldots, u_N$. This process whereby the system's state is influenced by an external agent or operator will be referred to as *controlling the system* and the quantities u_r ($r = 1,2, \ldots, N$) will be called the *control variables*. The behaviour of the system or the configuration it adopts as a consequence of any particular choice of control variables, we shall designate the system's *response*. It will be desired to control the system in such a manner that its response is as close as possible to a standard the operator has his own reasons for striving to attain. It will always be assumed that the operator's success in achieving this object can be measured by giving the value of a quantity C called the *performance criterion* (or *index*). Thus, a response resulting in a large value of C may be judged superior to any other response which gives a smaller value. In these circumstances, our problem is to choose values for the control variables which maximize C. The system response which results from such a choice will be said to be *optimal* and our problem may accordingly be described as the *optimization of systems*.

In some cases, it is more natural to assess the performance of a system by reference to the cost to the operator of eliciting a particular response. The quantity C will then be termed the *cost* or *penalty index* and the optimization problem will be to minimize its value.

In this chapter, we shall investigate the optimization of systems whose responses to control are of a particularly simple character. Thus, having set the control variables to chosen values, it will be assumed that the system under consideration responds by adopting a corresponding fixed configuration and that the excellence or otherwise of this configuration for the operator's purpose can be assessed by calculation of an index C; i.e. C will be a function of the control variables. For example, if the system is a yacht, the angles made by the sails and boat axis with the wind could be taken as control variables; having set these, the sailing posture becomes determinate and the boat's velocity is calculable; by choosing the upwind component of this velocity as the performance index to be maximized, a simple optimization problem is formulated. Systems of this type will be called *static systems*.

In later chapters, a more complex situation will be analysed. It will be supposed that the control variables are permitted to vary with the time t in any manner and that this causes the system to move continuously through a succession of states. The quality of the overall response of the system during this period of motion will be taken to be measured by an index C, but C will no longer be a simple function of the u_r, since it will depend upon the infinity of values taken by each control variable during its interval of variation. In these circumstances, C is said to be a *functional* of the functions $u_r(t)$. Systems of this type will be termed *dynamic systems*.

1.2 Unconstrained control

We are supposing that C is a known function of the control variables, so that we can write

$$C = C(u_1, u_2, \ldots, u_N). \tag{1.2.1}$$

This function will be assumed to possess continuous first and second order partial derivatives. The variables u_r may be thought of as the components of a vector in an N-dimensional Euclidean space (the *control space*); this vector will be denoted by u and will be termed the *control vector*. We shall employ the usually accepted notation and exhibit the components of u as the elements of a column matrix, thus:

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_N \end{bmatrix}. \tag{1.2.2}$$

If the column is not displayed, it will be convenient to write it in the form $(u_1, u_2, \ldots, u_N)^T$, the superscript T implying that the matrix to which it is attached is to be transposed. Equation (1.2.1) will frequently be abbreviated to read $C = C(u)$.

We shall suppose that the values which can be taken by the u_r are all real, but are not otherwise restricted in any way. The control vector u is then said to be *unconstrained.* In future, unless otherwise stated, $u_1, u_2, \ldots, u_N$ will denote the optimal values of the variables, i.e. the values which maximize or minimize C, depending upon the nature of the problem. To derive necessary conditions to be satisfied by these optimal values, it will be necessary to examine the values taken by C at neighbouring points in the control space; such points will be taken to have coordinates $v_r = u_r + \epsilon\xi_r$. If, now, the ξ_r are given arbitrary fixed values and ϵ is regarded as a variable, $C(u + \epsilon\xi)$ will be a function of ϵ having a minimum (or a maximum) at $\epsilon = 0$. It follows from a well-known theorem from the theory of functions of a single variable that

$$\frac{dC}{d\epsilon} = 0, \qquad \frac{d^2C}{d\epsilon^2} \geqslant 0 \tag{1.2.3}$$

at $\epsilon = 0$. (N.B. For simplicity of statement it will be assumed in future, unless otherwise stated, that C is to be minimized; it will be left for the reader to reverse the appropriate inequalities in case C is to be maximized.) Conditions (1.2.3) are clearly valid for all sets of values of the ξ_r.

Assuming that the function C possesses its differentiability and continuity properties over a domain of the control space including the optimal point, it follows that, for sufficiently small ϵ,

$$\frac{dC}{d\epsilon} = \frac{\partial C}{\partial v_r}\frac{dv_r}{d\epsilon} = \frac{\partial C}{\partial v_r}\xi_r; \tag{1.2.4}$$

the summation with respect to r over its values $1, 2, \ldots, N$ is here indicated by repetition of the index (the repeated index summation convention will be operative throughout the remainder of this text unless otherwise stated). Putting $\epsilon = 0$ in equation (1.2.4), since then $v_r = u_r$ the first of the conditions (1.2.3) takes the form

$$\frac{\partial C}{\partial u_r}\xi_r = 0. \tag{1.2.5}$$

But this condition must be satisfied for all values of the ξ_r and it accordingly follows that

$$\frac{\partial C}{\partial u_r} = 0. \tag{1.2.6}$$

A second differentiation of C with respect to ϵ and the setting of ϵ equal to zero yields the result

$$\frac{d^2C}{d\epsilon^2} = \frac{\partial^2 C}{\partial u_r\, \partial u_s}\, \xi_r \xi_s. \tag{1.2.7}$$

Repetition of both indices r, s implies that a double summation must be carried out on the right-hand member of this equation, which is accordingly a quadratic form in the variables ξ_r. The second of the conditions (1.2.3) requires that this form must be positive semi-definite, i.e. may vanish for a non-null set of values of the ξ_r, but is negative for no set of values of these parameters. A necessary and sufficient condition for this to be true is that the eigenvalues of the symmetric $N \times N$ matrix $A = (a_{rs})$, where $a_{rs} = \partial^2 C/\partial u_r \partial u_s$, should be positive or zero (the eigenvalues of A are defined to be the values of α for which the determinant of $A - \alpha I$ vanishes; I is the unit $N \times N$ matrix).

The point $v = (v_1, v_2, \ldots, v_N)$ will be said to belong to a *neighbourhood* of the optimal point of radius δ if

$$\sqrt{[(v_1 - u_1)^2 + (v_2 - u_2)^2 + \cdots + (v_N - u_N)^2]} < \delta. \tag{1.2.8}$$

Suppose the quadratic form (1.2.7) is positive definite, i.e. positive and non-vanishing for all non-null sets of values of the ξ_r. Then, the eigenvalues of the matrix A will all be positive and non-zero. Since we are assuming the partial derivatives $\partial^2 C/\partial v_r \partial v_s$ to be continuous over a domain containing the optimal point, it will be possible to find a neighbourhood Δ of this point of radius δ within which the eigenvalues associated with the quadratic form $(\partial^2 C/\partial v_r \partial v_s)\xi_r \xi_s$ are also positive and non-vanishing (these eigenvalues being continuous functions of the elements a_{rs} of A). Thus, within Δ, $(\partial^2 C/\partial v_r \partial v_s)\xi_r \xi_s$ is positive definite. Let $u + \eta$ be any point of Δ; then, by Taylor's theorem and using equations (1.2.6),

$$C(u + \eta) = C(u) + \frac{1}{2}\frac{\partial^2 C}{\partial v_r\, \partial v_s}\, \eta_r \eta_s, \tag{1.2.9}$$

where, in the second derivatives, $v = u + \theta\eta$ $(0 < \theta < 1)$. Since the quadratic form in the η_r is positive definite, we can now conclude that

$$C(u + \eta) > C(u), \; u + \eta \in \Delta, \quad \eta \neq 0. \tag{1.2.10}$$

This is to say that C possesses a local minimum at the point u. Thus, we have proved that sufficient conditions for a local minimum are equations (1.2.6) and the positive definiteness of the form (1.2.7).

PROBLEM 1. Calculate the maxima and minima of the function

$$C(x, y) = x^3 - 12xy + y^3 - 63x - 63y.$$

Solution: At a stationary point

$$\frac{\partial C}{\partial x} = 3x^2 - 12y - 63 = 0, \tag{1.2.11}$$

$$\frac{\partial C}{\partial y} = -12x + 3y^2 - 63 = 0. \tag{1.2.12}$$

Hence,

$$3x^2 - 12y = -12x + 3y^2,$$

or

$$(x - y)(x + y) = 4(y - x).$$

Thus, (i) $x = y$ or (ii) $x + y + 4 = 0$. In the first case, substitution for y in equation (1.2.11) leads to $x^2 - 4x - 21 = 0$; i.e. $x = -3, 7$, yielding stationary points $(-3, -3)$, $(7, 7)$. In the second case, substitution for y gives $x^2 + 4x - 5 = 0$; thus, $x = 1, -5$ and the stationary points are $(1, -5)$, $(-5, 1)$.

To determine the nature of the stationary points, we calculate the second derivatives:

$$\frac{\partial^2 C}{\partial x^2} = 6x, \qquad \frac{\partial^2 C}{\partial x\, \partial y} = -12, \qquad \frac{\partial^2 C}{\partial y^2} = 6y.$$

Hence, at the stationary point $(-3, -3)$

$$A = \begin{bmatrix} -18 & -12 \\ -12 & -18 \end{bmatrix}, \qquad A - \alpha I = \begin{bmatrix} -18 - \alpha & -12 \\ -12 & -18 - \alpha \end{bmatrix}.$$

$|A - \alpha I|$ vanishes when $\alpha = -6, -30$. Since both these eigenvalues are negative, $(-3, -3)$ is a local maximum; we find $C_{\max} = 216$.

Similarly, at the point $(7, 7)$, the eigenvalues are found to be $30, 54$ and this point is therefore a local minimum ($C_{\min} = -784$).

At both the remaining stationary points $(1, -5)$, $(-5, 1)$, the eigenvalues are the roots of the equation $\alpha^2 + 24\alpha - 324 = 0$. Since the product of these roots is negative, the eigenvalues have opposite signs and the points are neither maxima nor minima. ●

PROBLEM 2. The force exerted by the wind on the sail of a yacht is proportional to the square of the wind velocity and to the sine of the angle made by the direction of the wind with the sail; its direction is normal to the sail. Assuming that the drag on the boat is proportional to the square of its velocity, calculate the course which must be steered and the setting of the sail if the component of the yacht's velocity in the upwind direction is to be a maximum.

Solution: In Fig. 1.1, θ is the angle made by the axis of the yacht and ϕ is the angle made by the plane of the sail with the upward wind direction. The force exerted by the wind on the sail will be taken to be $F = cW^2 \sin\phi$ (we neglect the effect the yacht's motion has on the velocity with which the wind strikes the sail). If D is the drag force opposing the yacht's motion and V is its velocity, then $D = kV^2$. Assuming uniform motion, the force components along the boat axis must balance and, hence,

$$kV^2 = cW^2 \sin\phi \sin(\theta - \phi).$$

It is required to choose θ and ϕ so that $V \cos\theta$ is maximized. Since

$$V^2 \cos^2\theta = \frac{c}{k} W^2 \sin\phi \sin(\theta - \phi)\cos^2\theta,$$

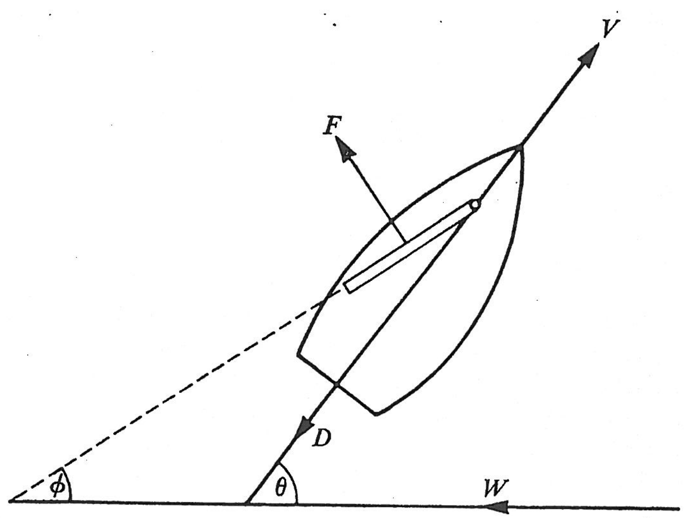

Fig. 1.1

it is more convenient to maximize

$$C = C(\theta, \phi) = \sin\phi \sin(\theta - \phi)\cos^2\theta.$$

Necessary conditions for a maximum are:

$$\frac{\partial C}{\partial \theta} = \sin\phi[\cos(\theta - \phi)\cos^2\theta - 2\sin(\theta - \phi)\cos\theta \sin\theta] = 0,$$

$$\frac{\partial C}{\partial \phi} = [\cos\phi \sin(\theta - \phi) - \sin\phi \cos(\theta - \phi)]\cos^2\theta = 0.$$

Rejecting the possibility that $\cos\theta$ vanishes, the second of these equations is equivalent to $\tan\phi = \tan(\theta - \phi)$; thus, $\phi = \theta - \phi$, or $\phi = \frac{1}{2}\theta$. Substituting for ϕ in the first equation, it is now easy to show that $\cos\theta = \frac{2}{3}$ (i.e. $\theta = 48°$, $\phi = 24°$ approximately).

For these values of θ and ϕ, it may be verified that the second derivatives of C take the values

$$\frac{\partial^2 C}{\partial \theta^2} = -7, \qquad \frac{\partial^2 C}{\partial \theta\, \partial \phi} = \frac{4}{9}, \qquad \frac{\partial^2 C}{\partial \phi^2} = -\frac{8}{9}.$$

The eigenvalues of the matrix A formed from these elements are both negative. It follows that at least a local maximum of C has been located (it is possible to prove that the maximum is global). ●

1.3 Equality constraints

In this section, we shall suppose that the control vector is subject to one or more equality constraints, i.e. its components u_r are permitted to take only values satisfying equations

$$f_i(u_1, u_2, \ldots, u_N) = c_i, \tag{1.3.1}$$

where $i = 1, 2,, \ldots, M$ and the c_i are constants. If $M = N$, these equations will, normally, possess a finite number of solutions only and if $M > N$, they will be inconsistent; in either case, the optimization problem is empty. We assume, therefore, $M < N$. Any set of values of the u_r which satisfy the constraints (1.3.1) is said to be *admissible* and the corresponding point in control space is termed an *admissible* point.

In this case, in general, it will be possible, in principle, to solve equations (1.3.1) for M of the variables u_r in terms of the remaining $N - M$ (the precise condition for this to be so will be given later).

C can then be expressed as a function of $N - M$ variables whose values are unconstrained. By this means, the problem can be reduced to the one already solved. We shall now study the details of this reduction programme.

Firstly, it will be convenient and in conformity with our later nomenclature to denote the M variables for which the constraints are solved by $x_1, x_2, \ldots, x_M$. The remaining independent variables will continue to be represented by $u_1, u_2, \ldots, u_N$. Thus, the constraints (1.3.1) will now be written

$$f_i(x_1, x_2, \ldots, x_M, u_1, u_2, \ldots, u_N) = c_i, \qquad i = 1, 2, \ldots, M, \tag{1.3.2}$$

and the performance index will be given by an equation of the form

$$C = C(x_1, \ldots, x_M, u_1, \ldots, u_N) = C(x, u). \tag{1.3.3}$$

The functions C, f_i will be assumed to possess continuous partial derivatives of the first and second orders with respect to all their arguments over a domain which includes the optimal point within its interior.

As before, u will be referred to as the control vector and $x = [x_1, x_2, \ldots, x_M]^T$ will be termed the *state vector*. The state vector will be regarded as an element in a Euclidean *state space* having M dimensions, the x_i being the coordinates of a point in this space. Since it is being assumed that the x_i are expressible as functions of the u_r, if the point u is allowed to vary in the control space, a consequent variation of the point x occurs in the state space.

Having substituted for the state variables x_i in terms of the control variables u_r in the function $C(x, u)$, let the resulting function of the u_r be denoted by $D(u)$. Then, the values of the control variables which make C a minimum must satisfy the conditions

$$\frac{\partial D}{\partial u_r} = 0, \qquad \frac{\partial^2 D}{\partial u_r \, \partial u_s} \xi_r \xi_s \geqslant 0, \tag{1.3.4}$$

for arbitrary values of the ξ_r.

Now, the first of these conditions can be expressed in the form

$$\frac{\partial D}{\partial u_r} = \frac{\partial C}{\partial x_i}\frac{\partial x_i}{\partial u_r} + \frac{\partial C}{\partial u_r} = 0, \tag{1.3.5}$$

$r = 1, 2, \ldots, N$. In these equations, the derivatives $\partial x_i/\partial u_r$ are calculable from the constraints; thus, differentiating equations

(1.3.2) partially with respect to u_r (remembering that the x_i can be regarded as functions of the u_r), we obtain

$$\frac{\partial f_i}{\partial x_j}\frac{\partial x_j}{\partial u_r} + \frac{\partial f_i}{\partial u_r} = 0. \tag{1.3.6}$$

This provides us with M equations for the M unknowns $\partial x_1/\partial u_r$, $\partial x_2/\partial u_r, \ldots, \partial x_M/\partial u_r$ and these have a unique solution provided the determinant $|\partial f_i/\partial x_j|$ of the coefficients does not vanish; this is precisely the condition which must be satisfied if the constraints (1.3.2) are to specify the x_i as functions of the u_r and we shall assume it to be satisfied at the optimal point. Equations (1.3.5), (1.3.6) then provide a set of necessary conditions from which possible optimal sets of values of the u_r can be derived.

These equations can be put in a more convenient and symmetric form by introducing the adjoint equations, namely,

$$\frac{\partial f_i}{\partial x_j}\lambda_i + \frac{\partial C}{\partial x_j} = 0. \tag{1.3.7}$$

These M equations determine the M quantities λ_i, called the *Lagrange Multipliers* (the non-vanishing of the determinant $|\partial f_i/\partial x_j|$ is again crucial). Multiplying equation (1.3.6) through by λ_i and summing with respect to i, we find

$$\lambda_i \frac{\partial f_i}{\partial x_j}\frac{\partial x_j}{\partial u_r} + \lambda_i \frac{\partial f_i}{\partial u_r} = 0. \tag{1.3.8}$$

Employing equation (1.3.7), this reduces to the form

$$-\frac{\partial C}{\partial x_j}\frac{\partial x_j}{\partial u_r} + \lambda_i \frac{\partial f_i}{\partial u_r} = 0 \tag{1.3.9}$$

and, by equation (1.3.5), this in turn is seen to be equivalent to

$$\frac{\partial C}{\partial u_r} + \lambda_i \frac{\partial f_i}{\partial u_r} = 0. \tag{1.3.10}$$

We next define the *Hamiltonian* H by the equation

$$H = H(x, \lambda, u) = C + \lambda_i f_i. \tag{1.3.11}$$

Equations (1.3.7), (1.3.10) can then be written in the simple form

$$\frac{\partial H}{\partial x} = 0, \qquad \frac{\partial H}{\partial u_r} = 0. \tag{1.3.12}$$

This set of equations, together with the constraints (1.3.2), provides us with $(2M + N)$ equations for the same number of unknowns x_i, λ_i, u_r, and, hence, serves to locate all possible-optimal points.

To apply the second necessary condition (1.3.4), we require to calculate the second derivatives $\partial^2 D/\partial u_r \partial u_s$. Differentiating equations (1.3.5), (1.3.6) partially with respect to u_s, the following results are established:

$$\frac{\partial^2 D}{\partial u_r\,\partial u_s} = \frac{\partial C}{\partial x_i}\frac{\partial^2 x_i}{\partial u_r\,\partial u_s} + \frac{\partial^2 C}{\partial x_i\,\partial x_j}\frac{\partial x_i}{\partial u_r}\frac{\partial x_j}{\partial u_s} + \frac{\partial^2 C}{\partial x_i\,\partial u_r}\frac{\partial x_i}{\partial u_s} + \frac{\partial^2 C}{\partial x_i\,\partial u_s}\frac{\partial x_i}{\partial u_r} + \frac{\partial^2 C}{\partial u_r\,\partial u_s}, \tag{1.3.13}$$

$$0 = \frac{\partial f_k}{\partial x_i}\frac{\partial^2 x_i}{\partial u_r\,\partial u_s} + \frac{\partial^2 f_k}{\partial x_i\,\partial x_j}\frac{\partial x_i}{\partial u_r}\frac{\partial x_j}{\partial u_s} + \frac{\partial^2 f_k}{\partial x_i\,\partial u_r}\frac{\partial x_i}{\partial u_s} + \frac{\partial^2 f_k}{\partial x_i\,\partial u_s}\frac{\partial x_i}{\partial u_r} + \frac{\partial^2 f_k}{\partial u_r\,\partial u_s}. \tag{1.3.14}$$

Multiplying equation (1.3.14) through by λ_k, summing over k and adding to equation (1.3.13), it follows that

$$\frac{\partial^2 D}{\partial u_r\,\partial u_s} = \frac{\partial^2 H}{\partial x_i\,\partial x_j}\frac{\partial x_i}{\partial u_r}\frac{\partial x_j}{\partial u_s} + \frac{\partial^2 H}{\partial x_i\,\partial u_r}\frac{\partial x_i}{\partial u_s} + \frac{\partial^2 H}{\partial x_i\,\partial u_s}\frac{\partial x_i}{\partial u_r} + \frac{\partial^2 H}{\partial u_r\,\partial u_s}. \tag{1.3.15}$$

(Note: Use equation (1.3.7).) After calculating the values of the derivatives $\partial x_i/\partial u_r$ from equations (1.3.6), this last equation permits the construction of the quadratic form appearing in the second necessary condition (1.3.4).

PROBLEM 3. A rectangular block is inscribed in the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{1.3.16}$$

Determine its maximum volume.

Solution: The eight corners of the block will lie at points $(\pm x, \pm y, \pm z)$, where x, y, z satisfy the equation of the ellipsoid. The volume of the block is given by $V = 8xyz$, where $x \geqslant 0$, $y \geqslant 0$,

$z \geqslant 0$. Hence, xyz is to be maximized subject to the constraint (1.3.16). Any pair of the variables x, y, z can be regarded as control variables and, then, the third is the state variable determined by the constraint.

The Hamiltonian for the problem is

$$H = xyz + \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\right)$$

and necessary conditions for a maximum are

$$\frac{\partial H}{\partial x} = yz + 2\lambda x/a^2 = 0,$$

$$\frac{\partial H}{\partial y} = zx + 2\lambda y/b^2 = 0,$$

$$\frac{\partial H}{\partial z} = xy + 2\lambda z/c^2 = 0.$$

The first equation yields $x^2/a^2 = -xyz/2\lambda$; y^2/b^2, z^2/c^2 can be found similarly from the other equations. Substitution in the constraint shows that $-3xyz/2\lambda = 1$; i.e. $\lambda = -3xyz/2$. It now follows that $x = a/\sqrt{3}$, $y = b/\sqrt{3}$, $z = c/\sqrt{3}$ and, hence, $\lambda = -abc/2\sqrt{3}$.

Treating z as the state variable and regarding it as a function of x, y, differentiation of the constraints leads to the equations

$$\frac{2x}{a^2} + \frac{2z}{c^2}\frac{\partial z}{\partial x} = \frac{2y}{b^2} + \frac{2z}{c^2}\frac{\partial z}{\partial y} = 0.$$

Thus, at the optimal point, $\partial z/\partial x = -c/a$ and $\partial z/\partial y = -c/b$.

Equation (1.3.15) applied to this problem gives

$$\frac{\partial^2 D}{\partial x\,\partial y} = \frac{\partial^2 H}{\partial z^2}\frac{\partial z}{\partial x}\frac{\partial z}{\partial y} + \frac{\partial^2 H}{\partial x\,\partial z}\frac{\partial z}{\partial y} + \frac{\partial^2 H}{\partial y\,\partial z}\frac{\partial z}{\partial x} + \frac{\partial^2 H}{\partial x\,\partial y} = -\frac{2c}{\sqrt{3}}.$$

Similarly, it is found that $\partial^2 D/\partial x^2 = -4bc/\sqrt{3}a$, $\partial^2 D/\partial y^2 = -4ca/\sqrt{3}b$.

Forming the matrix A, its eigenvalues are found to satisfy the equation

$$\alpha^2 + \frac{4c}{\sqrt{3}}\left(\frac{a}{b} + \frac{b}{a}\right)\alpha + 4c^2 = 0.$$

This, clearly, has negative roots, indicating that V has been maximized. Then, $V_{\max} = 8abc/3\sqrt{3}$. ●

PROBLEM 4. The cost function is given by

$$C = \tfrac{1}{2}p_{ij}x_i x_j + \tfrac{1}{2}q_{rs}u_r u_s.$$

The constraints are linear and given by

$$x_i = b_{ir}u_r + c_i.$$

It is required to minimize C.

Solution: The Hamiltonian is

$$H = \tfrac{1}{2}p_{ij}x_i x_j + \tfrac{1}{2}q_{rs}u_r u_s + \lambda_i(x_i - b_{ir}u_r)$$

and (assuming $p_{ij} = p_{ji}$, $q_{rs} = q_{sr}$) necessary conditions for a minimum are

$$\frac{\partial H}{\partial x_i} = p_{ij}x_j + \lambda_i = 0,$$

$$\frac{\partial H}{\partial u_r} = q_{rs}u_s - \lambda_i b_{ir} = 0.$$

It is convenient to write these as matrix equations, namely,

$$Px + \lambda = 0, \qquad Qu - B^T\lambda = 0,$$

where x, λ, u are columns, P is the matrix whose ijth element is p_{ij}, Q has rsth element q_{rs} and B has irth element b_{ir}. P, Q are symmetric square matrices of orders M, N respectively and B is of type $M \times N$.

Writing the constraints in the matrix form $x = Bu + c$, we can now eliminate x and λ to yield the equation

$$(Q + B^TPB)u + B^TPc = 0.$$

We shall now assume that $Q + B^TPB$ is positive definite and, hence, that the inverse $(Q + B^TPB)^{-1}$ exists. Thus,

$$u = -(Q + B^TPB)^{-1}B^TPc.$$

x and λ can now be found in the forms

$$x = [I - B(Q + B^TPB)^{-1}B^TP]c,$$

$$\lambda = [PB(Q + B^TPB)^{-1}B^TP - Pc].$$

It follows from the constraints that $\partial x_i/\partial u_r = b_{ir}$. Equation (1.3.15) applied to this problem accordingly gives the result

$$\frac{\partial^2 D}{\partial u_r\,\partial u_s} = p_{ij}b_{ir}b_{js} + q_{rs}.$$

The condition for a minimum is that the quadratic form

$$\frac{\partial^2 D}{\partial u_r\,\partial u_s}\xi_r\xi_s = p_{ij}b_{ir}b_{js}\xi_r\xi_s + q_{rs}\xi_r\xi_s = \xi^T(B^TPB + Q)\xi$$

must be positive definite, and this we have already assumed to be the case.

Finally,

$$\begin{aligned} C_{\min} &= \tfrac{1}{2}x^TPx + \tfrac{1}{2}u^TQu \\ &= -\tfrac{1}{2}x^T\lambda + \tfrac{1}{2}u^TB^T\lambda \\ &= -\tfrac{1}{2}(x - Bu)^T\lambda \\ &= -\tfrac{1}{2}c^T\lambda \\ &= \tfrac{1}{2}c^T[P - PB(Q + B^TPB)^{-1}B^TP]c. \end{aligned}$$ ●

1.4 The Lagrange multipliers

Suppose that the constants c_i appearing in the constraints (1.3.2) have their values altered slightly to $c_i + \delta c_i$. A new optimization problem then arises. Let the optimal values of the state and control variables for this new problem be denoted by $x_i + \delta x_i$, $u_r + \delta u_r$. These must satisfy the equations

$$f_i(x + \delta x, u + \delta u) = c_i + \delta c_i \tag{1.4.1}$$

and, working to the first order of small quantities, this equation is equivalent to

$$\frac{\partial f_i}{\partial x_j}\delta x_j + \frac{\partial f_i}{\partial u_r}\delta u_r = \delta c_i; \tag{1.4.2}$$

(remember $f_i(x, u) = c_i$). Multiplying this last equation through by λ_i and summing over i, we get

$$\lambda_i\frac{\partial f_i}{\partial x_j}\delta x_j + \lambda_i\frac{\partial f_i}{\partial u_r}\delta u_r = \lambda_i\delta c_i. \tag{1.4.3}$$

Using equations (1.3.7), (1.3.10), this can be written

$$-\frac{\partial C}{\partial x_j}\,\delta x_j - \frac{\partial C}{\partial u_r}\,\delta u_r = \lambda_i \delta c_i. \tag{1.4.4}$$

To emphasize that optimal values are being taken everywhere, a subscript zero will be attached to C. Then, if C_0 denotes the optimal cost for the original problem and $C_0 + \delta C_0$ the optimal cost for the new problem, equation (1.4.4) indicates that

$$\delta C_0 = -\lambda_i \delta c_i. \tag{1.4.5}$$

Thus, in the limit as $\delta c_i \to 0$, we have

$$\frac{\partial C_0}{\partial c_i} = -\lambda_i. \tag{1.4.6}$$

This equation reveals the mathematical significance of the Lagrange multipliers.

1.5 Inequality constraints

Suppose that $C(u)$ is to be minimized over all vectors u satisfying conditions

$$f_i(u) \geqslant c_i, \qquad i = 1, 2, \ldots, M. \tag{1.5.1}$$

The functions $f_i(u)$ will be assumed to possess continuous second order partial derivatives throughout the control space. Admissible points in this space will determine a closed region R (i.e. including the boundary points) bounded by the hypersurfaces $f_i = c_i$.

If the optimal point is an interior point of R, none of the constraints is operative in a sufficiently small neighbourhood of this point and they can accordingly be ignored. The conditions derived in section 1.2 are then still applicable.

If, however, the optimal point O is a boundary point lying on the first $Q(\leqslant N)$ hypersurfaces $f_m = c_m$ $(m = 1, 2, \ldots, Q)$, but not on the remainder, the constraints $f_i \geqslant c_i$ $(i = 1, 2, \ldots, Q)$ will now be operative but the constraints $f_i \geqslant c_i$ $(i = Q+1, \ldots, M)$ will still be ineffective over a sufficiently small neighbourhood Δ of O. We shall therefore study the *associated problem* of minimizing C subject to the equality constraints $f_m = c_m$ $(m = 1, 2, \ldots, Q)$. We shall refer to the original problem as $\mathscr{P}$ and the associated problem as $\mathscr{P}'$. Then, the optimal point O for $\mathscr{P}$ must be, at least, a local minimum for $\mathscr{P}'$ and it follows that the necessary conditions derived in section 1.3 for the class of problems of the type $\mathscr{P}'$ must be satisfied at O.

Let C_0 be the value of C at O and let λ_m be the multipliers arising in the solution of $\mathscr{P}'$. Consider a third problem $\mathscr{P}''$ in which C is to be minimized subject to the perturbed constraints $f_m = c_m + \delta c_m$ $(\delta c_m \geqslant 0)$. Assuming the coordinates of O are continuously dependent on the c_m, if the δc_m are sufficiently small, a point O' lying in Δ can be found which is a local minimum for $\mathscr{P}''$. Since $O' \in \Delta$, the constraints $f_i \geqslant c_i$ $(i = Q+1, \ldots, M)$ are all satisfied at O'. Further, since the δc_m are positive, the remaining constraints $f_m \geqslant c_m$ are also satisfied at O'. Hence, $O' \in R$. Let C_0, $C_0 + \delta C_0$ be the values of the cost at O, O' respectively. Then, since O is the optimal point in R, $\delta C_0 \geqslant 0$. But, by equation (1.4.5), $\delta C_0 = -\lambda_m \delta c_m$, to the first order in the δc_m. Since the δc_m are arbitrary positive quantities, we now conclude that

$$\lambda_m \leqslant 0, \qquad m = 1, 2, \ldots, Q. \tag{1.5.2}$$

It has now been proved that, at an optimal point O where the first Q of the constraints (1.5 1) are effective (i.e. O lies on the hypersurfaces $f_m = c_m$, $m = 1, 2, \ldots, Q$), the following conditions are necessarily satisfied.

$$\frac{\partial C}{\partial u_r} + \lambda_m \frac{\partial f_m}{\partial u_r} = 0, \qquad \lambda_m \leqslant 0, \qquad f_m = c_m. \tag{1.5.3}$$

Putting

$$H = C + \lambda_i f_i, \tag{1.5.4}$$

these conditions can be written in the alternative form

$$\left.\begin{aligned} \frac{\partial H}{\partial u_r} &= 0, \\ \lambda_i \leqslant 0 &\text{ if } f_i(u) = 0, \\ \lambda_i = 0 &\text{ if } f_i(u) > 0. \end{aligned}\right\} \tag{1.5.5}$$

We have just established necessary conditions to be satisfied at O. By the usual slight strengthening of these conditions, they become sufficient conditions for O to be a local minimum. Thus, suppose it is known that O is a local minimum for the problem $\mathscr{P}'$ and that $\lambda_m < 0$ $(m = 1, 2, \ldots, Q)$; sufficient conditions for the validity of the first part of this hypothesis have been given in section 1.3; the second part is the strengthened condition. Let P be any point belonging to $\Delta \cap R$ and denote the values of $f_m(u)$ at P by $c_m + \delta c_m$; then $\delta c_m \geqslant 0$. Let $C_0 + \delta C$ be the cost at P. Then P is an admissible point for the problem $\mathscr{P}''$ and, since O' is a local minimum for this problem, it follows that $C_0 + \delta C \geqslant C_0 + \delta C_0$, i.e. $\delta C \geqslant \delta C_0$. But, to the first order, $\delta C_0 = -\lambda_m \delta c_m$ and, therefore, $\delta C \geqslant -\lambda_m \delta c_m$. If all the λ_m are negative and non-zero, we can now conclude that $\delta C > 0$, except possibly when the δc_m all vanish. In this latter event, however, P is an admissible point for $\mathscr{P}'$ and our assumption that O is a local minimum for this problem then gives $\delta C \geqslant 0$. Thus, in all circumstances, $\delta C \geqslant 0$ and this proves that O is a local minimum.

If the original problem is amended by reversal of the inequalities appearing in some or all of the constraints (1.5.1), the corresponding inequalities satisfied by the λ_m must also be reversed. If the problem is amended by the omission of equalities from all of these constraints, all boundary points become *ipso facto* inadmissible. Hence, if an optimal point exists, it must be an interior point at which the conditions derived in section 1.2 apply. However, in such a case, it is possible that the original problem $\mathscr{P}$ has a boundary point solution and, hence, that C can be made to approach this minimum value as close as we please without violating the amended constraints. In these circumstances, the amended problem has no solution.

PROBLEM 5. Maximize xyz subject to the constraint

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leqslant 1.$$

Solution: Clearly, no interior point of the ellipsoidal region R can be a maximum point for xyz. We therefore assume the point required lies on the boundary of R and solve the associated problem. This has already been done in Problem 3, where it was shown that a local maximum occurs on the ellipsoid at the point $(a/\sqrt{3}, b/\sqrt{3}, c/\sqrt{3})$; the associated value of the multiplier λ was calculated to be $-abc/2\sqrt{3}$. This is negative, indicating that this point is also a local maximum for the present problem (Note: C is being maximized and the inequality constraint is the reverse of that assumed in equation (1.5.1).) ●

1.6 Linear programming

A special case of the type of problem studied in the last section is of great practical importance. This is when

$$C = a_r u_r \tag{1.6.1}$$

is to be minimized subject to the constraints

$$b_{ir} u_r \geqslant c_i, \qquad i = 1, 2, \ldots, M. \tag{1.6.2}$$

In this case the boundary hypersurfaces of the region of admissible points in the control space are hyperplanes.

If an interior point of R were minimal, $\partial C/\partial u_r$ $(= a_r)$ would have to vanish for all values of r, i.e. C would vanish identically. Rejecting this case, the minimal point, if it exists, must lie on the boundary. Suppose it lies on all the hyperplanes

$$b_{mr} u_r = c_m, \qquad m = 1, 2, \ldots, Q. \tag{1.6.3}$$

Since, in general, more than N such hyperplanes will have no point in common, we shall assume $Q \leqslant N$. The Hamiltonian for the associated problem is given by

$$H = a_r u_r + \lambda_m b_{mr} u_r \tag{1.6.4}$$

and, at the minimal point, it is necessary that

$$\frac{\partial H}{\partial u_r} = a_r + b_{mr}\lambda_m = 0, \tag{1.6.5}$$

$r = 1, 2, \ldots, N$. If $Q < N$, there are insufficient unknowns λ_m to permit these equations to be satisfied (we consider only the general case). Hence $Q = N$ and the minimal point must lie at the intersection of exactly N of the boundary hyperplanes. Since $Q \leqslant M$, this is only possible if $M \geqslant N$; if $M < N$, the problem has no solution.

Assuming, therefore, $Q = N \leqslant M$, the sets of equations (1.6.3), (1.6.5) can be solved for u, λ respectively. For the point u to be minimal, the λ_r must all be negative. If they are not, another set of N boundary hyperplanes must be selected to replace the set (1.6.3) and u, λ recalculated. If a minimal point exists, a finite number of trials of this type will locate it.

If the constraints (1.6.2) are augmented by the requirement that the variables u_r should not be negative, i.e. $u_r \geqslant 0$, $r = 1, 2, \ldots, N$, we have what is termed a *linear programming problem*.

If this problem has a solution, the minimal point will be at the intersection of N hyperplanes, which can now include any of the coordinate hyperplanes $u_r = 0$. Suppose that the nomenclature is chosen such that the minimal point lies on the hyperplanes $b_{ir}u_r = c_i$ $(i = 1, 2, \ldots, Q \leqslant M)$ and the coordinate hyperplanes $u_{Q+1} = u_{Q+2} = \cdots = u_N = 0$. Thus, at the minimal point,

$$b_{mn}u_n = c_m, \quad b_{pn}u_n > c_p, \quad u_l = 0, \quad u_n > 0, \tag{1.6.6}$$

where $m, n = 1, 2, \ldots, Q$, $p = Q + 1, Q + 2, \ldots, M$ and $l = Q + 1, Q + 2, \ldots, N$.

The Hamiltonian for the associated problem is given by

$$H = a_r u_r + \lambda_m b_{mr} u_r + \Lambda_l u_l, \tag{1.6.7}$$

where, for minimality, we must have

$$\lambda_m < 0, \quad \Lambda_l < 0. \tag{1.6.8}$$

The necessary conditions $\partial H/\partial u_r = 0$ separate into two groups, (i) $r = 1, 2, \ldots, Q$ and (ii) $r = Q + 1, Q + 2, \ldots, N$, namely

$$a_n + \lambda_m b_{mn} = 0, \tag{1.6.9}$$

$$a_l + \lambda_m b_{ml} + \Lambda_l = 0. \tag{1.6.10}$$

Solution of the equations (1.6.9) will yield the multipliers λ_m and equations (1.6.10) then give the multipliers Λ_l. These multipliers must all be negative for a minimal point.

The *dual linear programming problem* is that of maximizing

$$C' = c_i u_i' \tag{1.6.11}$$

subject to the constraints

$$b_{ir}u_i' \leqslant a_r, \tag{1.6.12}$$

$$u_i' \geqslant 0. \tag{1.6.13}$$

It will now be proved that, provided equations (1.6.9), (1.6.10) correspond to a minimal point for the original linear programming problem a maximal point P for the dual problem is determined by the equations

$$u_m' = -\lambda_m, \; u'_{Q+1} = u'_{Q+2} = \cdots = u_M' = 0. \tag{1.6.14}$$

Firstly, using equations (1.6.9), (1.6.10) we note that

$$b_{ir}u_i' = -b_{mr}\lambda_m = a_r + \Lambda_q, \tag{1.6.15}$$

where we put $\Lambda_r = 0$ for $r = 1, 2, \ldots, Q$. Then, since Λ_r vanishes for $r = 1, 2, \ldots, Q$ and Λ_r is negative for $r = Q + 1, Q + 2, \ldots, N$, the first set of constraints (1.6.12) is clearly satisfied. Also, since the λ_m are negative, the second set (1.6.13) is also satisfied. Thus, the point P defined by equations (1.6.14) is admissible.

Putting $r = n = 1, 2, \ldots, Q$ in equation (1.6.15), we see that P lies on the hyperplanes $b_{in}u_i' = a_n$; it also lies on the coordinate hyperplanes $u_p' = 0$, $p = Q + 1, Q + 2, \ldots, M$. Thus, the associated problem has Hamiltonian

$$H' = c_i u_i' + \lambda_n' b_{in} u_i' + \Lambda_p' u_p', \tag{1.6.16}$$

with multipliers λ_n', Λ_p'. Necessary conditions to be satisfied at a maximal point are $\partial H'/\partial u_i' = 0$, i.e.

$$c_m + \lambda_n' b_{mn} = 0, \tag{1.6.17}$$

$$c_p + \lambda_n' b_{pn} + \Lambda_p' = 0. \tag{1.6.18}$$

Comparing equations (1.6.17) with the first set of equations (1.6.6), we deduce that $\lambda_n' = -u_n$. It follows that $\lambda_n' < 0$. Also, equations (1.6.18) give

$$\Lambda_p' = -c_p - \lambda_n' b_{pn} = -c_p + b_{pn}u_n > 0, \tag{1.6.19}$$

using the inequalities (1.6.6). Hence the multipliers λ_n', Λ_p' satisfy the sufficiency conditions for a maximal point and the required result is proved.

Further, if we write equations (1.6.6), (1.6.9) in the matrix form

$$Bu = c, \quad B^T\lambda = -a, \tag{1.6.20}$$

where B is the $Q \times Q$ matrix with elements b_{mn} and u, a, c are the columns $[u_1, u_2, \ldots, u_Q]^T$, $[a_1, a_2, \ldots, a_Q]^T$, $[c_1, c_2, \ldots, c_Q]^T$ respectively, then

$$u = B^{-1}c, \quad \lambda = -(B^T)^{-1}a. \tag{1.6.21}$$

Hence

$$u' = -\lambda = (B^T)^{-1}a. \tag{1.6.22}$$

It follows that

$$C_{\min} = a_m u_m = a^T u = a^T B^{-1} c. \tag{1.6.23}$$

$$C'_{\max} = c_m u_m' = u'^T c + a^T B^{-1} c. \tag{1.6.24}$$

Thus $C_{\min} = C'_{\max}$.

It may be proved quite generally that a linear programming problem has a solution if, and only if, the dual problem has a solution and that, if the solutions exist, the corresponding minimum and maximum values are identical.

Problem 6. A manufacturer produces metal boxes and keys. To produce a box, a lathe must operate for 1 minute, a grinder for 3 minutes and a drill for 3 minutes. To produce a key, the respective times are 2 minutes, 1 minute and $1\frac{1}{2}$ minutes. The available weekly machine capacities in minutes are as follows: lathe 40,000, grinder 45,000, drill 48,000. He makes a profit of 10p per box and 15p per key. How many boxes and keys should be produced during a week to maximize his profit.

Solution: Let x, y thousand be the optimal weekly number of boxes and keys respectively. Then the weekly operating time for a lathe is $x + 2y$ (in units of one thousand minutes). Since the total lathe capacity is 40 units this leads to the inequality

$$x + 2y \leqslant 40.$$

Similarly, we derive the inequalities

$$3x + y \leqslant 45, \qquad 3x + \tfrac{3}{2}y \leqslant 48.$$

His net profit is $1000C$, where

$$C = 10x + 15y$$

and this is to be maximized subject to the above inequality constraints and, also $x \geqslant 0, y \geqslant 0$.

In the xy-plane, the boundary lines for the inequality constraints are the straight lines AB, CD, EF (Fig. 1.2). Thus, the point (x, y) is constrained to lie inside or on the polygon $OCQPB$. We know from theory that the maximal point must be one of the vertices of this figure. By calculation, we find that the coordinates of these vertices are as follows: $C(15, 0)$, $Q(13, 6)$, $P(8, 16)$, $B(0, 20)$. At these points, the values of C are 150, 220, 320, 300 respectively.

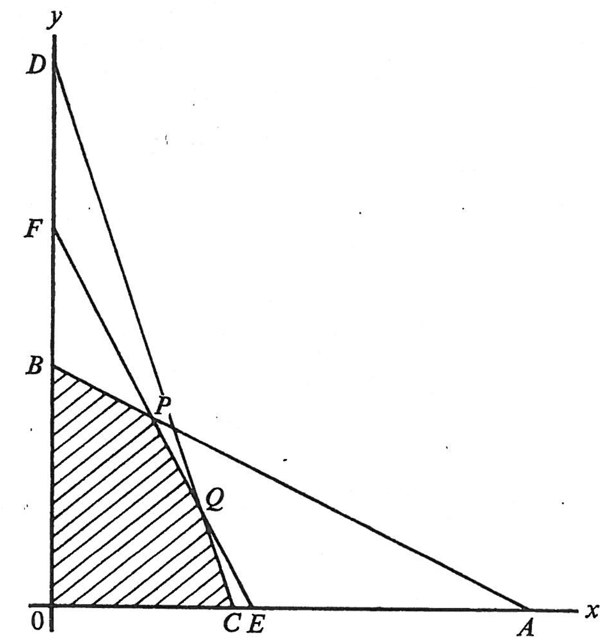

Fig. 1.2

Thus, the maximum profit is 320,000p and occurs when 8,000 boxes and 16,000 keys are made. ●

Problem 7. Formulate the dual of the previous problem and solve it.

Solution: The dual problem is to minimize

$$C' = 40u + 45v + 48w$$

subject to the constraints

$$u + 3v + 3w \geqslant 10,$$

$$2u + v + \tfrac{3}{2}w \geqslant 15,$$

$$u \geqslant 0, \qquad v \geqslant 0, \qquad w \geqslant 0.$$

From the 5 boundary planes in 3-dimensional uvw-space, 10 triads can be selected; each triad of planes has a point in common, the 10 such points having coordinates (20/3, 0, 10/9), (7, 1, 0), (10, 0, 0), (0, 0, 10), (0, 15, 0), (0, −20, 70/3), (0, 0, 10/3), (0, 10/3, 0), (15/2, 0, 0), (0, 0, 0). Only the first five of these points are admissible; the remainder fail to satisfy the constraints. Calculating C' at the admissible points, we obtain the values 320, 325, 400, 480, 675. Thus $C'_{\min} = 320$, the optimal variable values being $u = 20/3$, $v = 0$, $w = 10/9$.

As expected $C'_{\min} = C_{\max}$. ●

Exercises 1

1. Show that the function

$$6x^2 + 8y^3 + 3(2x + y - 1)^2$$

possesses a local minimum value of 11/16.

2. Identical particles are placed at the n points (x_i, y_i) $(i = 1, 2, \ldots, n)$ in the plane of rectangular axes Oxy. The straight line

$$x \cos\theta + y \sin\theta = p$$

is to be constructed so that the moment of inertia of the system of particles about it is a minimum. Show that p and θ must be chosen to satisfy the equations

$$p = \bar{x}\cos\theta + \bar{y}\sin\theta, \qquad \tan 2\theta = \frac{2H}{A - B},$$

where

$$\bar{x} = \frac{1}{n}\sum x_i, \qquad \bar{y} = \frac{1}{n}\sum y_i,$$

$$A = \frac{1}{n}\sum x_i^2 - \bar{x}^2, \qquad B = \frac{1}{n}\sum y_i^2 - \bar{y}^2,$$

$$H = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}.$$

3. The geometric mean of n positive quantities is fixed. Show that their arithmetic mean is a minimum when they are equal.

4. If x, y, z are subject to the constraints

$$x + y + z = 1, \qquad x^2 + y^2 + z^2 = 1,$$

show that $x^3 + y^3 + z^3$ has a minimum value 8/27 and a maximum value 1.

5. Relative to rectangular axes $Oxyz$, A is the point (a, b, c). Write down equations from which can be calculated the coordinates of a point P lying on the surface $f(x, y, z) = 0$, if it is given that the distance AP is stationary with respect to variation of P. Deduce that the line AP has direction ratios $\partial f/\partial x$, $\partial f/\partial y$, $\partial f/\partial z$. Show that, if the surface is the sphere $x^2 + y^2 + z^2 = r^2$, P has two possible positions with coordinates $x = \pm ar/(a^2 + b^2 + c^2)^{\frac{1}{2}}$, etc.

6. Referring to problem 4, suppose the cost function is taken to be

$$C = \tfrac{1}{2}p_{ij}x_i x_j + \tfrac{1}{2}q_{rs}u_r u_s + r_{ir}x_i u_r$$

and it is assumed that the matrix $B^T PB + Q + B^T R + R^T B$ is positive definite. Show that C can be minimized by taking

$$u = (B^T PB + Q + B^T R + R^T B)^{-1}(B^T P + R^T)c$$

and that

$$C_{\min} = \tfrac{1}{2}c^T[(PB + R)(B^T PB + Q + B^T R + R^T B)^{-1}(B^T P + R^T) + P]c.$$

7. If the equality constraints in Ex. 4 above are replaced by the constraints

$$x + y + z \geqslant 1, \qquad x^2 + y^2 + z^2 \geqslant 1,$$

show that the points at which $x^3 + y^3 + z^3$ takes the value 8/27 remain minima. Show also that if the equality constraints are replaced by the constraint $x^2 + y^2 + z^2 \leqslant 1$, the points at which $x^3 + y^3 + z^3$ takes the value +1 remain maxima.

8. x, y, z are positive or zero quantities. Maximize $4x - 2y - z$ subject to the constraints $x + y + z \leqslant 3$, $2x + 2y + z \leqslant 4$, $x - y \leqslant 0$. (Ans. Maximum value 2 occurs at $x = y = 1$, $z = 0$.)