MATH2640 Introduction to Optimisation

4. Constrained Optimisation, Lagrangians & Lagrange multipliers, Bordered Hessians: summary

4(A) Constrained Optimisation: Lagrange multipliers

Normally, economic activity is constrained, e.g. expenditure is constrained by available income. The constraints are applied to the input bundle, and may take many different forms. Often constraints are in the form of inequalities, e.g. we can spend up to so much on equipment but no more. However, it is often simpler to solve problems with equality constraints, e.g. we are allowed to spend 1000 pounds on equipment, and we will spend it all, so the equipment expenditure $x_1 = 1000$ is an equality constraint. In general, we are trying to maximise the *objective function*

$$f(x_1, x_2, \cdots, x_n)$$

subject to the k inequality constraints

$$g_i(x_1, x_2, \cdots, x_n) \le b_i, \quad 1 \le i \le k$$

and subject to the M equality constraints

$$h_i(x_1, x_2, \cdots, x_n) = c_i, \quad 1 \le i \le m.$$

(i) **Single Equality constraint**

Maximise $f(\mathbf{x})$ subject to $h(\mathbf{x}) = c$. At a stationary point of this problem, the constraint surface must touch the level surfaces of f. To see this, note that if $\mathbf{dx}$ is a small displacement from the critical point $\mathbf{x}^*$, then if the small displacement lies in the constraint surface $\mathbf{dx} \cdot \nabla h = 0$. But if this displacement is at a critical point, f must not change as we move from $\mathbf{x}^*$ to $\mathbf{x}^* + \mathbf{dx}$, so we must have $\mathbf{dx} \cdot \nabla f = 0$ also. Since this has to be true for all small displacements on the constraint surface, ∇h is parallel to ∇f at $\mathbf{x}^*$. When two vectors are parallel, there is a scalar factor λ such that

$$\nabla f = \lambda \nabla h, \quad \text{or} \quad \nabla(f - \lambda h) = 0.$$

The constant λ is called the **Lagrange multiplier**. $L = f - \lambda h$ is called the **Lagrangian**.
To find the stationary (critical) points of $f(\mathbf{x})$ subject to $h(\mathbf{x}) - c = 0$, we solve the $n+1$ equations

$$\frac{\partial f}{\partial x_i} - \lambda \frac{\partial h}{\partial x_i} = 0, \quad i = 1, n$$

$$h(\mathbf{x}) = c,$$

for the $n+1$ unknowns x_1, $x_2 \cdots x_n$, λ. As usual, we must be very careful to get all the solutions of these $n+1$ equations, and not to miss some (or include bogus solutions).

Example:
Maximise $f(x, y) = xy$ subject to $h(x, y) = x + 4y - 16 = 0$.
Solution: The Lagrangian $L = f - \lambda h = xy - \lambda(x + 4y - 16)$. The equations we need are

$$L_x = 0, \quad \text{or} \quad y - \lambda = 0 \qquad (A)$$

$$L_y = 0, \quad \text{or} \quad x - 4\lambda = 0 \qquad (B)$$

$$x + 4y = 16. \qquad (C)$$

The only critical point is $x = 8$, $y = 2$ and $\lambda = 2$. At this point, $f = 16$, and this is the maximum value of f subject to this constraint. Strictly speaking, we do not yet know whether it is a maximum, because we have not examined the Hessian (see the section on Bordered Hessians below). However, if we look at other values of x and y that satisfy the constraint, such as $x = 0$, $y = 4$, we see they give lower values of f, which strongly suggests that the critical point is a maximum.

(ii) **The non-degenerate constraint qualification, NDCQ**

The above algorithm for finding stationary points normally works well, but the argument breaks down if $\nabla h = 0$ at the critical point. In this case, the constraint surface does not have a uniquely defined normal, so we cannot say that ∇f must be parallel to ∇h. We should therefore check that at each critical point we find the constraint qualification, usually called the non-degenerate constraint qualification NDCQ holds, i.e. $\nabla h \neq 0$ at each critical point.

What happens if it doesn't hold, i.e. $\nabla h = 0$ at the critical point? Usually the Lagrange equations can't be solved.

Example: Maximise $f = x$ subject to $h = x^3 + y^2 = 0$.

Solution: $L = x - \lambda(x^3 + y^2)$, so $L_x = 0$ and $L_y = 0$ give

$$1 = 3\lambda x^2, \quad (A); \quad \lambda y = 0, \quad (B); \quad x^3 = -y^2 \quad (C).$$

From (A), $\lambda \neq 0$, so from (B), $y = 0$. Then from (C), $x = 0$. $x = y = 0$ is actually the maximising point, as we can see from the constraint (C), as $-y^2$ must be ≤ 0, so $x \leq 0$, so $f = 0$ is the largest value of f. However, $x = y = 0$ is not a well-defined solution of (A), (B) and (C) because if $x = 0$ equation (A) reads $1 = 0$! The problem is that $\nabla h = (3x^2, 2y)$ happens to be zero at the critical point $(0, 0)$ so the NDCQ is **not** satisfied at this critical point.

If the NDCQ constraint is not satisfied at the critical point, we have to use a different method to find the maximum, other than Lagrange's method. In this simple problem, we could just eliminate x using the constraint, so $x = (-y^2)^{1/3}$, and sketch $f = (-y^2)^{1/3}$ as a function of y to find the minimum. In general, however, the problem is much more difficult if the NDCQ is not satisfied.

(iii) **Several equality constraints**

If we have m equality constraints,

$$h_i(\mathbf{x}) - c_i = 0, \quad 1 \leq i \leq m,$$

then we need m Lagrange multipliers, and the Lagrangian L is

$$L = f - \lambda_1 h_1 - \lambda_2 h_2 - \cdots - \lambda_m h_m.$$

Then the $n + m$ equations to be solved are

$$\frac{\partial L}{\partial x_i} = 0, \quad i = 1, n$$

$$h_i(\mathbf{x}) = c_i, \quad i = 1, m,$$

for the $n + m$ unknowns $x_1, \cdots, x_n$ and $\lambda_1, \cdots, \lambda_m$.

If any constraint is simple, it may well be easier to use that constraint to eliminate a variable before finding the stationary points.

Example: Maximise $f = x + y + z$ subject to the constraints $x^2 + y^2 = 1$ and $z = 2$.

Solution: Just using the general Lagrange method mechanically,

$$L = x + y + z - \lambda_1(x^2 + y^2) - \lambda_2(z - 2)$$

and the five equations we need to solve are

$$1 - 2\lambda_1 x = 0, \quad 1 - 2\lambda_1 y = 0, \quad 1 - \lambda_2 = 0, \quad x^2 + y^2 = 1, \quad z = 2.$$

It is easily seen that the two solutions are $x = 1/\sqrt{2}, y = 1/\sqrt{2}, z = 2, \lambda_1 = 1/\sqrt{2}, \lambda_2 = 1$ and $x = -1/\sqrt{2}, y = -1/\sqrt{2}, z = 2, \lambda_1 = -1/\sqrt{2}, \lambda_2 = 1$.

However, it is much simpler to use $z = 2$ to eliminate z and just maximise $f = x+y+2$ subject to $x^2+y^2 = 1$, which only requires three equations for three unknowns, and gives the same answer more quickly.

If there are several equality constraints, the non-degenerate constraint qualification, NDCQ, is that the m rows of the Jacobian derivative matrix,

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \cdots & \frac{\partial h_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix},$$

are linearly independent, that is that none of the rows can be expressed as a linear combination of the other rows. If this is the case, we say the Jacobian derivative matrix is of rank m.

4(B) **Constrained Optimisation: Second order conditions**

The Lagrange multiplier method gives the first-order conditions that determine the stationary points, and we can sometimes tell whether the stationary point is a maximum or a minimum without examining the Hessian Matrix. In more complicated examples, we need a method to determine whether a stationary point is a local maximum or a local minimum, or neither.

Note that having the constraint completely changes the nature of the quadratic expression. For example $Q(x,y) = x^2 - y^2$ is indefinite, but if we apply the constraint $y = 0$ then $Q(x,y) = Q(x,0) = x^2$ which is positive definite and has a local minimum at $x = y = 0$. We must therefore replace our Leading Principal Minor test with something else.

The Lagrangian is $L = f - \lambda_1 h_1 - \lambda_2 h_2, \cdots, -\lambda_m h_m$, with $h_1 = c_1, \cdots, h_m = c_m$ being the constraints. Suppose we have a stationary point at $\mathbf{x}^*$ with Lagrange multipliers λ_i^*, Using the Taylor series,

$$L(\mathbf{x}^* + \mathbf{dx}) = L(\mathbf{x}^*) + \mathbf{dx} \cdot \nabla L + \frac{1}{2}\mathbf{dx}^T H \mathbf{dx} + \quad \text{higher order terms.}$$

The term $\mathbf{dx} \cdot \nabla L = 0$ at a stationary point, because at the stationary points $\nabla L = 0$. So the stationary point is a minimum if $\mathbf{dx}^T H \mathbf{dx} > 0$ and a maximum if $\mathbf{dx}^T H \mathbf{dx} < 0$. Here H is the Hessian based on the second derivatives of the Lagrangian, so the i, j component is

$$L_{x_i x_j} = \frac{\partial^2}{\partial x_i \partial x_j} L = \frac{\partial^2}{\partial x_i \partial x_j}(f - \lambda_1^* h_1 - \lambda_2^* h_2 + \cdots).$$

Now the allowed displacements $\mathbf{dx}$ are restricted by the constraints to satisfy $\mathbf{dx} \cdot \nabla h_i = 0$ for every constraint $i = 1, \ldots, m$.

Replacing $\mathbf{dx}$ by $\mathbf{x}$, and ∇h_i by its value $\mathbf{u_i}$ at the stationary point, the problem becomes: classify the quadratic form

$$Q = \mathbf{x}^T H \mathbf{x}, \quad \text{subject to} \quad \mathbf{u_i} \cdot \mathbf{x} = 0 \quad i = 1, \ldots, m.$$

(i) 2×2 **with one linear constraint**

Note that $Q(x,y) = x^2 - y^2$ is indefinite, but if we apply the constraint $y = 0$, i.e. $\mathbf{u} = (0,1)$, then $Q = x^2$ which is positive definite. So a constraint can change the classification of Q.

In general,

$$Q(x,y) = ax^2 + 2bxy + cy^2, \quad \text{with} \quad ux + vy = 0.$$

Eliminate y by setting $y = -ux/v$ to get

$$Q(x) = \frac{x^2}{v^2}(av^2 - 2buv + cu^2)$$

which is positive definite (minimum) if $(av^2 - 2buv + cu^2) > 0$, and negative definite (maximum) if $(av^2 - 2buv + cu^2) < 0$ for nontrivial u and v.

(ii) **Bordered Hessians**

If we have an n variable Lagrangian with m constraints, then the Bordered Hessian is an $(n+m) \times (n+m)$ matrix constructed as follows:

$$H_B = \begin{pmatrix} 0 & B \\ B^T & H \end{pmatrix},$$

where the B part has m rows and n columns, with the rows consisting of the constraint vectors $\mathbf{u}_i,\ i = 1, m$, and the H part is the Hessian based on the Lagrangian. In the case where $m = 2$, $n = 3$ and the two constraints are

$$u_1 x + u_2 y + u_3 z = 0, \quad v_1 x + v_2 y + v_3 z = 0,$$

(u_1, u_2, u_3) being ∇h_1 and (v_1, v_2, v_3) being ∇h_2, both evaluated at stationary point, the 5×5 Bordered Hessian matrix is

$$H_B = \begin{pmatrix} 0 & 0 & u_1 & u_2 & u_3 \\ 0 & 0 & v_1 & v_2 & v_3 \\ u_1 & v_1 & L_{xx} & L_{xy} & L_{xz} \\ u_2 & v_2 & L_{xy} & L_{yy} & L_{yz} \\ u_3 & v_3 & L_{xz} & L_{yz} & L_{zz} \end{pmatrix},$$

where

$$L_{xx} = f_{xx} - \lambda_1^* h_{1_{xx}} - \lambda_2^* h_{2_{xx}},$$

and the other components of the Hessian are defined similarly.

(iii) **Bordered Hessians continued**

Given the Lagrangian

$$L(\boldsymbol{\lambda}, \mathbf{x}) = f(\mathbf{x}) - \boldsymbol{\lambda} \cdot (\mathbf{h}(\mathbf{x}) - \mathbf{c})$$

with $\mathbf{x} = (x_1, \ldots, x_n)$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$, $\mathbf{h} = (h_1, \ldots, h_m))$ and $\mathbf{c} = (c_1, \ldots, c_m)$ we can find the bordered Hessian H_B by taking the second-order derivatives of L with respect to $(\boldsymbol{\lambda}, \mathbf{x})$, to obtain

$$H_B = \begin{pmatrix} 0 & -B \\ -B-^T & H \end{pmatrix},$$

with $H_{ij} = \partial^2 L/\partial x_i \partial x_j = \partial^2 f/\partial x_i \partial x_j - \boldsymbol{\lambda} \cdot \partial^2 \mathbf{h}/\partial x_i \partial x_j$ with $i, j = 1, \ldots, n$ and $B_{ij} = \partial h_i/\partial_j$ with $i = 1, \ldots, m; j = 1, \ldots, n$. Note that for a linear constraint $h_i = \sum_{j=1}^n u_j^{(i)} x_j = \mathbf{u}_i \cdot \mathbf{x}$, we have $B_{ij} = u_j^{(i)}$. Note that in the classification given below there is in the end no contradiction between using either of the two H_B's since multiplication of the first m rows and the first m columns by -1 constitutes an even number of $2m$ multiplications by -1 and, hence, the signs (but not the values) of any LPMs calculated, as required in the classification below, are unaffected.

(iv) **Rules for classification using Bordered Hessians**

We need to classify

$$Q = \mathbf{x}^T H \mathbf{x}, \quad \text{subject to} \quad \mathbf{u_i} \cdot \mathbf{x} = 0, \quad i = 1, \ldots, m.$$

Construct the bordered Hessian as defined above, and evaluate the leading principal minors. We only need the last $(n-m)$ LPM's, starting with LPM_{2m+1} and going up to (and including) LPM_{n+m}. LPM_{n+m} is just the determinant of the whole H_B matrix.

(A) Q is positive definite (PD) if $\text{sgn}(LPM_{n+m}) = \text{sgn}(\det H_B) = (-1)^m$ and the successive LPM's have the same sign. Here sgn means the sign of, so e.g. $\text{sgn}(-2) = -1$. This case corresponds to a local minimum.

(B) Q is negative definite (ND) if $\text{sgn}(LPM_{n+m}) = \text{sgn}(\det H_B) = (-1)^n$ and the successive LPM's from LPM_{2m+1} to LPM_{n+m} alternate in sign. This case corresponds to a local maximum.

(C) If both (A) and (B) are violated by non-zero LPM's, then Q is indefinite.

The semi-definite conditions are very involved, and we do not go into them here.

Example: If $n = 2$ and $m = 1$, then $n-m = 1$ and we only need examine LPM_3. Suppose this is $LPM_3 = -1$. Then this has the same sign as $(-1)^m$, and so the constrained Q is positive definite. Check: if $Q = x^2 - y^2$ and the constraint is $y = 0$, so $\mathbf{u} = (0, 1)$, the bordered Hessian is

$$H_B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix},$$

which has $LPM_3 = detH_B = -1$, and $Q(x, 0) = x^2$ which is indeed positive definite.

Example: $Q = x^2 - y^2 - z^2$ with constraint $x = 0$, so now $\mathbf{u} = (1, 0, 0)$. Now $n = 3$ and $m = 1$ we only need examine LPM_4 and LPM_3.

$$H_B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

$LPM_4 = -1$, and $LPM_3 = 1$. These alternate in sign so there is a possibility the constrained quadratic is negative definite. To test this against criterion (B) note $LPM_4 = -1$ which has the same sign as $(-1)^n$, and so the constrained Q is negative definite. Check: if $Q = x^2 - y^2 - z^2$ and the constraint is $x = 0$, then $Q = -y^2 - z^2$ which is clearly negative definite.

Example: Use bordered Hessians to classify the following constrained quadratic form $Q = -6y^2 + 3z^2 + 8xy + 2yz - 2xz$ subject to constraint $x + y - z$.

For the classification, one may be tempted to consider the bordered Hessian arising from the Lagrangian

$$L(\lambda, x, y, z) = -6y^2 + 3z^2 + 8xy + 2yz - 2xz - \lambda(x + y - z)$$

giving

$$H_B = \begin{pmatrix} 0 & -1 & -1 & 1 \\ -1 & 0 & 8 & -2 \\ -1 & 8 & -12 & 2 \\ 1 & -2 & 2 & 6 \end{pmatrix},$$

or (by directly using Q and the constraint) the Lagrangian

$$\tilde{L}(\lambda, x, y, z) = -3y^2 + \frac{3}{2}z^2 + 3xy + yz - xz + \tilde{\lambda}(x + y - z)$$

giving

$$H_B = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 4 & -1 \\ 1 & 4 & -6 & 1 \\ -1 & -1 & 1 & 3 \end{pmatrix}.$$

Note, however, that $L(\lambda, x, y, z) = 2\tilde{L}(\tilde{\lambda} = -\lambda/2, x, y, z)$ such that the two representations are leading to the same classications.

(v) **Evaluating determinants**

Evaluating 2×2 and 3×3 determinants is straightforward, but evaluating larger determinants can be difficult. The definition is

$$\det A = \sum (-1)^h A_{1r_1} A_{2r_2} \cdots A_{nr_n},$$

where the summation is made over all permutations $r_1\, r_2\, r_3, \cdots, r_n$. A permutation is just the numbers 1 to n written in a different order, e.g . [1 3 2] is a permutation on the first 3 integers and corresponds to $r_1 = 1$, $r_2 = 3$ and $r_3 = 2$. Permutations are either odd or even. If it takes an odd number of interchanges to restore the permutation to its natural order, the permutation is odd. If it takes an even number, the permutation is even. In the formula for the determinant, $h = 1$ if the permutation is odd, $h = 2$ if it is even. So [2 3 4 1] is an odd permutation, as we must first swap 2 and 1, then 2 and 3, and then 3 and 4, three swaps altogether. Since three is an odd number of swaps, the permutation is odd.

Other useful facts for evaluating determinants are

(i) Swapping any two rows or any two columns reverses the sign of the determinant.

(ii) Adding any multiple of another row to a row leaves the determinant unchanged. Similarly adding any multiple of a column to a different column leaves it unchanged. Using this trick, it is usually simple to generate zero elements which makes evaluating the determinant easier.