1. Partial differentiation, Chain rule, Implicit functions, Jacobian, Differentials: summary

1(A) *Representing and visualising functions of two variables.*

(i) $z = f(x, y)$ is the height above the xy–plane, or the depth below if z is negative.

(ii) Functions can be drawn as perspective plots (almost always done using a package such as MAPLE/Matlab or Python), or as a contour plot. Contour plots are sketches of the level curves $f(x, y) = z = \textit{constant}$ in the xy-plane. Contour maps such as the Ordnance Survey maps are examples of functions of height in terms of x and y coordinates (the grid reference).

(iii) $z = c - x^2 - y^2$ has a local maximum at $x = y = 0$. Its contours are circles centred on the origin. $z = c + x^2 + y^2$ has a local minimum at $x = y = 0$. Note that $z = x^2 - y^2$ also is locally flat at $x = y = 0$, but it has a saddle point there. A saddle is like a pass going from one valley to another, between two mountains.

1(B) *Partial derivatives.*

(i) If $z = z(x, y)$, then $\dfrac{\partial z}{\partial x}$ means differentiate z with respect to x treating y as a constant. Similarly $\dfrac{\partial z}{\partial y}$ means differentiate z with respect to y treating x as a constant.

(ii) Geometrical interpretation: $\dfrac{\partial z}{\partial x}$ is the tangent of the uphill or downhill slope of z as we move in the x–direction holding y constant. Similarly, $\dfrac{\partial z}{\partial y}$ is the slope of z moving in y with x held constant.

(iii) Higher derivatives: $\dfrac{\partial^2 z}{\partial x^2}$ means $\dfrac{\partial}{\partial x}\Big(\dfrac{\partial z}{\partial x}\Big)$. Similarly $\dfrac{\partial^2 z}{\partial y^2}$ means $\dfrac{\partial}{\partial y}\Big(\dfrac{\partial z}{\partial y}\Big)$. The mixed derivative,

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}\Big(\frac{\partial z}{\partial y}\Big) = \frac{\partial}{\partial y}\Big(\frac{\partial z}{\partial x}\Big) = \frac{\partial^2 z}{\partial y \partial x}, \quad \text{cf. Young's theorem.}$$

(iv) Notation: f_x means $\dfrac{\partial f}{\partial x}$, $f_{xx} \equiv \dfrac{\partial^2 f}{\partial x^2}$, $f_{xy} = f_{yx} \equiv \dfrac{\partial^2 f}{\partial x \partial y}$, etc.

1(C) *Total derivatives and the chain rule.*

If $z = z(x, y)$ and $\{x = x(t), y = y(t)\}$ represent a curve in the xy–plane traced out as the parameter t varies, then the chain rule gives

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

Note that if there is a relation between x and y, then $\dfrac{dz}{dx} = \dfrac{\partial z}{\partial x} + \dfrac{\partial z}{\partial y}\dfrac{dy}{dx}$, so the partial derivative is not the same as the total derivative.

If we are given $z = z(x, y)$ and we want to change independent variables from x and y to s and t, and we are given $x = x(s, t)$ and $y = y(s, t)$, then there are formulae for finding $\dfrac{\partial z}{\partial s}$ in terms of $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$. These are the chain rules for several variables,

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$

and

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

These results can be used to generate the second-order derivatives, z_{ss}, z_{st} and z_{tt} in terms of x and y derivatives.

1(D) *Gradient vectors and directional derivatives*

If $f = f(x_1, x_2, \cdots x_n)$, the gradient vector is

$$\nabla f \equiv \Big(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\Big),$$

and the directional derivative formula can then be written as a scalar product $D_{\mathbf{u}} f = \mathbf{u} \cdot \nabla f$. Here $\mathbf{u}$ is a unit vector (i.e. has length 1) and is the direction in which the derivative is taken. In 2D the magnitude of the directional derivative is then greatest when $\mathbf{u}$ is parallel to ∇f, which means that ∇f is the direction of steepest ascent, interpreting $z = f(x, y)$ as height above the xy-plane. Level contours are perpendicular to ∇f.

In three dimensions,

$$\nabla f \equiv \Big(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\Big).$$

$f(x, y, z) = C = \textit{constant}$ is then a level surface and ∇f is a vector in the direction of the normal to this level surface, i.e. perpendicular to the tangent plane at the surface. For all unit vectors with their foot at the point $P = (x_0, y_0, z_0)$ and lying in the tangent plane, $\mathbf{u} \cdot \nabla f = 0$.

1(E) *Implicit differentiation.*

If $f(x, y) = 0$, then $\dfrac{dy}{dx} = -\dfrac{f_x}{f_y}$. This can be generalised to the case where z is given implicitly in terms of x and y through a relation of the form $f(x, y, z) = 0$. Then

$$df = f_x dx + f_y dy + f_z dz = 0$$

can be used to find the partial derivatives, e.g.

$$\left(\frac{\partial z}{\partial x}\right)_y = -\frac{f_x}{f_z},$$

since $dy = 0$ when finding the partial derivative of z with respect to x. Similar formula work for other partial derivatives, and since x, y and z appear symmetrically in $f(x, y, z) = 0$ we can also write

$$\frac{\partial x}{\partial y} = -\frac{f_y}{f_x},$$

where now the partial derivative of x with respect to y means that z is held constant (since $dz = 0$).

1(F) *Differentials.*

Differentials are useful for finding derivatives when variables are linked by more than one relation. For example, if $f(x, y, z) = 0$ and $g(x, y, z) = 0$ you could in principle use $f = 0$ to find how z depends on x and y, and then eliminate z in $g = 0$ to get y as a function of x and then z as a function of x. But doing these eliminations may not be straightforward. A simpler method is to use the chain rule,

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = 0,$$
$$dg = \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy + \frac{\partial g}{\partial z}dz = 0.$$

To get $\dfrac{dy}{dx}$ we could eliminate dz from these two linear equations, and get a relation between dx and dy, which we then arrange to give $\dfrac{dy}{dx}$. Similarly, to find $\dfrac{dz}{dx}$ we would eliminate dy between the two equations to get the relation between dx and dz which gives $\dfrac{dz}{dx}$. Note that $f = 0$ and $g = 0$ define a curve (or possibly a family of curves) in xyz-space. If we want to know the tangent to the curve at a particular point (x_0, y_0, z_0), it will be in the direction $(dx, dy, dz) = dx(1, \dfrac{dy}{dx}, \dfrac{dz}{dx})$ which we can find. To turn this into a unit vector in the direction of the tangent vector, just divide this vector by its magnitude. The dx will then cancel out, giving the required result. The differentials can be interpreted as (infinitesimally) small changes to the values of x, y and z, so if the variables x, y and z are constrained by the relation $f(x, y, z) = \text{constant}$, $g(x, y, z) = \text{constant}$, then by solving the chain rule expressions

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = 0,$$
$$dg = \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy + \frac{\partial g}{\partial z}dz = 0,$$

we can get dx and dy in terms of dz. This then tells us how the variables x and y will change if z is slightly increased or decreased. This method for 3 dimensions can easily and straightforwardly be generalised to work in n dimensions.

MATH2640 Introduction to Optimisation

2. Taylor's theorem, Gradients, Hessians, Extrema, Quadratic Forms: summary

2(A) Taylor's theorem in several variables.

(i) in 1 dimension, Taylor's theorem is $f(x+\delta x) = f(x) + \delta x f_x + \frac{1}{2!}(\delta x)^2 f_{xx} + \cdots$

(ii) In n dimensions, it is $f(\mathbf{X}+\delta\mathbf{x}) = f(\mathbf{X}) + \delta\mathbf{x}\cdot\nabla f + \frac{1}{2!}\delta\mathbf{x}\cdot H\delta\mathbf{x} + \cdots$
where $\delta\mathbf{x} = (\delta x_1, \delta x_2, \cdots \delta x_n)$, $\mathbf{X} = (X_1, X_2, \cdots X_n)$, and H is the Hessian matrix with components $H_{ij} = \dfrac{\partial^2 f}{\partial x_i \partial x_j}$.

(iii) The Taylor series formula is derived by considering $f(\mathbf{X}+\lambda\delta\mathbf{x})$ as a function of λ, and expanding about $\lambda = 0$ using the Taylor expansion formula for a function of one variable, λ, together with the chain rule. Then putting $\lambda = 1$ gives the above formula.

(iv) In the two variable case, the Taylor series formula is

$$f(X+h, Y+k) = f(X,Y) + h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y} + \frac{1}{2}\left(h^2\frac{\partial^2 f}{\partial x^2} + 2hk\frac{\partial^2 f}{\partial x\partial y} + k^2\frac{\partial^2 f}{\partial y^2}\right) + \cdots .$$

The linear term is $hf_x + kf_y$, and the quadratic term is

$$\frac{1}{2}\ (h \quad k)\ \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}\begin{pmatrix} h \\ k \end{pmatrix}, \quad \text{where} \quad H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

is called the Hessian matrix. Because $f_{xy} = f_{yx}$ by Young's theorem, the Hessian matrix is symmetric. In three dimensions, the Hessian matrix is

$$\begin{pmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{pmatrix},$$

which is also symmetric because $f_{xy} = f_{yx}$, $f_{xz} = f_{zx}$, and $f_{yz} = f_{zy}$.

2(B) Gradient vectors, normals and the tangent plane.

(i) $\nabla f(x,y) = (f_x, f_y)$ is in the direction in which f increases most rapidly. The unit vector in this direction is $\dfrac{\nabla f(x,y)}{|\nabla f(x,y)|} = \hat{\mathbf{n}}$. $-\nabla f(x,y)$ is in the direction of the rate of fastest decrease. Since $df = f_x dx + f_y dy$ if (dx, dy) is perpendicular to ∇f, so then $(dx, dy)\cdot\nabla f = 0$, then $df = 0$, i.e. if we move in the direction perpendicular to ∇f, f is constant. The direction perpendicular to ∇f is therefore the direction of the contours of constant f.

(ii) For implicitly defined functions, e.g. $g(x,y,z) = 0$, $\nabla g = (g_x, g_y, g_z)$ is in the direction of the normal to the surface $g(x,y,z) = 0$. The unit normal is $\dfrac{\nabla g(x,y,z)}{|\nabla g(x,y,z)|} = \hat{\mathbf{n}}$. This can be evaluated at any point (x_0, y_0, z_0) lying on the surface $g(x,y,z) = 0$.

(iii) The tangent plane to the surface $g(x,y,z) = 0$ at the point $\mathbf{x_0} = (x_0, y_0, z_0)$ is

$$(\mathbf{x} - \mathbf{x_0})\cdot\nabla g = 0,$$

where ∇g is evaluated at (x_0, y_0, z_0). If a surface is given by $z = f(x,y)$ then the tangent plane at a point where $f_x = 0$ and $f_y = 0$ is of the form $z =$ constant.

2(C) First order conditions (FOC)

(i) Local maxima and minima of $f(x,y)$ occur where the tangent plane is horizontal, so $f_x = 0$ and $f_y = 0$ there. In n dimensions,

$$\frac{\partial f}{\partial x_i} = 0, \quad i = 1, 2, \cdots, n, \quad \text{at} \quad \mathbf{x} = \mathbf{x_0},$$

if $\mathbf{x_0}$ is a local min or max. These n conditions give n equations for the n unknowns $x_1,\ x_2, \cdots x_n$. We solve these equations, which are called the first order conditions (FOC), to find the critical points where maxima and minima occur. The name first order conditions arises from the fact that they come from the linear terms in the Taylor expansion. There may be more than one solution of the FOC, i.e. more than one critical point.

(ii) Not all critical points are maxima or minimima. For example saddle points can occur as well. To classify each critical point we must look at the quadratic terms, i.e. look at the Hessian matrix.

(iii) Sometimes, all the quadratic terms may vanish, i.e. every element in H is zero. If this happens, we have to examine the cubic terms in the Taylor expansion to classify the critical point. In general this is complicated, but it is practical for some simple special cases.

2(D) Quadratic forms

(i) When the FOC are satisfied, the behaviour near the critical point is given by

$$f(\mathbf{X} + \delta \mathbf{x}) = f(\mathbf{X}) + \frac{1}{2!}\delta \mathbf{x} \cdot H \delta \mathbf{x} + \cdots .$$

The expression $\delta \mathbf{x} \cdot H \delta \mathbf{x}$ is called a quadratic form. To study quadratic forms we usually consider just $\mathbf{x} cdot A \mathbf{x} = Q(\mathbf{x})$, where$A$ is a symmetric matrix (for simpicity using $\mathbf{x}$ instead of $\mathbf{x}$).
In the 2-dimensional case,

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad \text{and so} \quad Q = ax^2 + 2bxy + cy^2$$

is the quadratic form. In the 3-dimensional case

$$A = \begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}, \quad \text{and so} \quad Q = ax^2 + by^2 + cz^2 + 2hxy + 2gxz + 2fyz.$$

In any number of dimensions, quadratic forms fall into five classes.
(PD) If $Q > 0$ for all non-zero $\mathbf{x}$, Q is positive definite.
(ND) If $Q < 0$ for all non-zero $\mathbf{x}$, Q is negative definite.
(I) If Q can take either positive or negative values, Q is indefinite.
(PSD) If $Q \geq 0$ for all non-zero $\mathbf{x}$, then Q is positive semidefinite. Note there may be some nontrivial $\mathbf{x}$ which makes $Q = 0$.
(NSD) If $Q \leq 0$ for all non-zero $\mathbf{x}$, then Q is negative semidefinite. Note there may be some nontrivial $\mathbf{x}$ which makes $Q = 0$.
Condition PD implies condition PSD, and ND implies NSD, but not the other way round.
If A is the Hessian matrix, then case (PD) is sufficient to give a local minimum, since the quadratic terms cause f to increase as we move away from $\mathbf{X}$. Similarly, case (ND) is sufficient to give a local maximum. Case (I) is not a maximum or a minimum. In 2-dimensions case (I) corresponds to a saddle point. In case (PSD), if we are at a minimum, PSD must hold, but it may be that PSD holds but it is not a local minimum. Similarly, at a maximum, NSD must hold but NSD is not sufficient to prove a maximum. We say PD and ND are sufficient conditions to establish a min, max respectively, and PSD and NSD are necessary conditions for there to be a min, max respectively.

(ii) To classify the quadratic form we first look at the *leading principal minors*, the LPMs. For a 3×3 symmetric matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

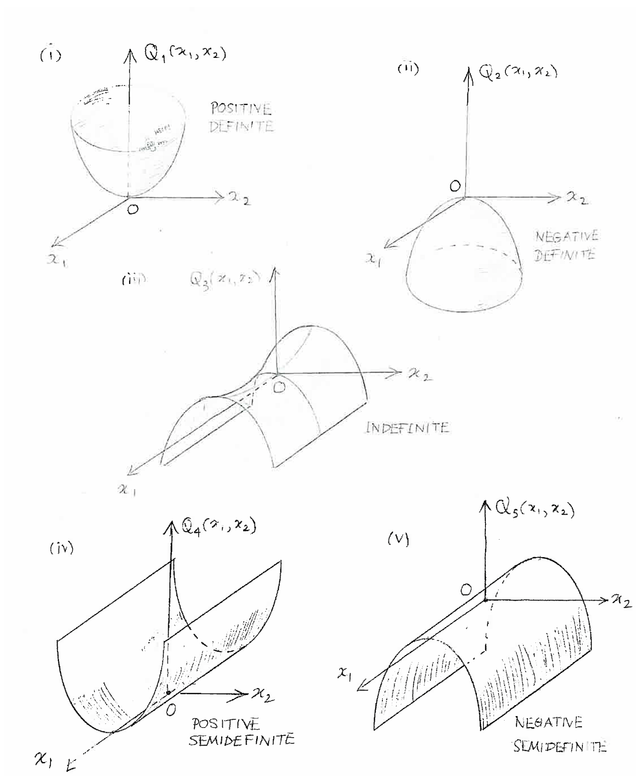


Figure 1: Sketches of quadratic forms. Courtesy Prof. Mark Kelmanson.

There are 3 LPM's, LPM_1, LPM_2 and LPM_3 defined by

$$LPM_1 = a_{11}, \quad LPM_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}^2,$$

$$LPM_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{23}^2) + a_{12}(a_{23}a_{13} - a_{12}a_{33}) + a_{13}(a_{12}a_{23} - a_{22}a_{13}),$$

so the vertical lines mean take the determinant. In the $n \times n$ case, we define n LPM's by extending in the obvious way. Assuming that the determinant $\det(A) \neq 0$, then the LPMs determine the signature of the matrix: *i)* if all $LPM_i > 0$, $i = 1, \ldots, n$, the matrix A is positive definite (PD); *ii)* if $LPM_1 < 0$, and the signs of the subsequent LPM_i alternate, (i.e., all the odd LPM_i are negative and the even LPM_i are positive), then A is (ND); *iii)* if there is any other sign pattern, A is indefinite.

(iii) However, if $LPM_n = \det(A) = 0$ for the $n \times n$ case the situation is more complicated, and A may be semi-definite but not definite. In this case, we have to find all the *principal minors.* A principal minor of order $n-k$ is found by deleting k columns and <u>the same</u> k rows and finding the determinant of what is left. If $k = 0$ there is just principal minor of order n, LPM_n, which is the determinant of A. If $k = 1$, there are n principal minors of order $n - 1$, since we can delete the row and column in n different ways. If **all** the principal minors are ≥ 0 A is positive semi-definite (PSD). If **all** the principal minors of odd order are ≤ 0 and **all** the principal minors of even order are ≥ 0, then A is negative semi-definite (NSD). Otherwise its indefinite.

(iv) Another way to classify quadratic forms is to find the *eigenvalues* of the matrix A. Since A is a symmetric matrix, all the eigenvalues are real. If all the eigenvalues are positive, A is positive definite. If all the eigenvalues are negative, A is ND. If some eigenvalues are zero and the rest positive, A is PSD. If some are zero and the rest negative, A is NSD. If any two eigenvalues have opposite sign, A is indefinite. To find the eigenvalues, we must solve the equation for λ given by

$$A = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{12} & a_{22} - \lambda & a_{23} \\ a_{13} & a_{23} & a_{33} - \lambda \end{vmatrix} = 0$$

in the 3×3 case, with the obvious generalization in the $n \times n$ case. In the 3×3 case this means solving a cubic equation, which will have 3 roots giving 3 eigenvalues. Solving cubic equations (or higher order equations) can be hard, but there are many computer packages which will find the eigenvalues of an $n \times n$ symmetric matrix very quickly.

OB November 26, 2019

MATH2640 Introduction to Optimisation

3. Quadratic forms and Eigenvalues, Economic applications (production functions): summary

3(A) Quadratic forms and eigenvalues

(i) Eigenvalues are the solutions, λ, of the matrix equation

$$A\mathbf{v} = \lambda\mathbf{v} \ , \quad \mathbf{v} \neq \mathbf{0} \ . \tag{3.1}$$

In general, an $n \times n$ matrix has n eigenvalues, but they may be complex valued, and some eigenvalues may coincide. The non-zero vectors $\mathbf{v}$ satisfying this equation are called the eigenvectors. For each eigenvalue λ there is, in fact, an infinity of corresponding eigenvectors $\mathbf{v}$, since any non-zero multiple of an eigenvector is again an eigenvector. Therefore, we sometimes want to restrict ourselves to unit (or normalised) eigenvectors: those with length $\|\mathbf{v}\| = 1$.

(ii) 2×2 case. The quadratic form associated with the symmetric matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad \text{is} \quad Q(x, y) = (x, y) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2bxy + cy^2.$$

To find the eigenvalues, we write equation (3.1) as $(A - \lambda I)\mathbf{x} = 0$, where I is the unit (identity) matrix. Then for this to have nontrivial solutions,

$$\det(A - \lambda I) = \begin{vmatrix} a-\lambda & b \\ b & c-\lambda \end{vmatrix} = 0, \quad \text{so} \quad \lambda^2 - (a + c)\lambda + ac - b^2 = 0$$

is the quadratic equation whose solutions are the two eigenvalues of the matrix A. These eigenvalues are always real for a symmetric matrix. (Prove this!)

(iii) **Normalising eigenvectors**
To find the eigenvector corresponding to a particular eigenvalue, solve $(A - \lambda I)\mathbf{x} = 0$ for that value of λ. You can only solve these equations up to an arbitrary constant k. Example:

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

has eigenvalues $\lambda = 2$ and $\lambda = 4$. To find the eigenvector corresponding to $\lambda = 2$, solve

$$\begin{aligned} x_1 + x_2 = 0 \\ x_1 + x_2 = 0 \end{aligned} \quad \text{giving} \quad (k, -k)$$

as the eigenvector. To normalise this, choose k so that $(k, -k)$ has unit length, $2k^2 = 1$, so the normalised eigenvector is $(1/\sqrt{2}, -1/\sqrt{2})$. The normalised eigenvector corresponding to $\lambda = 4$ is $(1/\sqrt{2}, 1/\sqrt{2})$.

(iv) The same methods work for finding eigenvalues of an $n \times n$ matrix: the characteristic equation

$$\det(A - \lambda I) = (-\lambda)^n + \mathrm{tr}(A)(-\lambda)^{n-1} + \cdots + \det(A) = 0 \ ,$$

(where $\mathrm{tr}(A)$ denotes the trace of the matrix, i.e., the sum of the diagonal entries), is now an n^{th}–order polynomial equation which, in principle, has n (possibly complex valued, possibly coinciding) roots, which are the eigenvalues $\lambda_1, \ldots, \lambda_n$. Having found the eigenvalues, the next step is to find the (unit) eigenvector $\mathbf{v}_i$ corresponding to each eigenvalue λ_i, by solving the homogeneous linear system $(A - \lambda_i I)\mathbf{v}_i = 0$. The latter can be done for instance by elementary row manipulations (e.g, see the first year Linear Algebra module) and will always have a nontrivial solution because of the condition on the λ_i.

Quadratic forms $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$, where $\mathbf{x} = (x_1, \ldots, x_n)$, are always associated with *symmetric* matrices $A = (a_{i,j})_{i,j=1,\ldots,n}$, for which $A = A^T$ (where the symbol T denotes matrix transposition) which means $a_{ij} = a_{ji}$. Such matrices have the following remarkable properties (prove these!):

- all eigenvalues of a (real-valued) symmetric matrix A are real (i.e., are real numbers);
- eigenvectors $\mathbf{v}_i$ of a symmetric matrix for distinct eigenvalues λ_i and λ_j are orthogonal to each other, i.e., $\mathbf{v_i} \cdot \mathbf{v_j} = 0$ for $\lambda_i \neq \lambda_j$;

- a symmetric matrix A can be "diagonalised" by an orthogonal matrix O, i.e., written in the form $A = O\Lambda O^T$ where Λ is the diagonal matrix of eigenvalues.

In the latter case, an *orthogonal* matrix is a matrix possessing the property that $O\,O^T = O^T O = I$, and hence it is invertible with $O^{-1} = O^T$. O can be constructed by creating the matrix with as columns the *normalised* (column) eigenvectors: $O = (\mathbf{v}_1\mathbf{v}_2 \dots \mathbf{v}_n)$.

(v) **Normal forms**

When the normalised eigenvectors $\mathbf{v}_i$, $i = 1, \dots, n$, corresponding to the n eigenvalues $\lambda_1, \dots, \lambda_n$, are found, say with components $\mathbf{v_i} = (v_1^{(i)}, v_2^{(i)}, \cdots v_n^{(i)})$, then the quadratic form $Q(\mathbf{x})$ can be re-written, in terms of some new variables $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$, as a sum of squares. In fact, introducing the vector of the new variables as $\tilde{\mathbf{x}} = O^T\mathbf{x}$, we have from the diagonalisation formula (given above):

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{x}^T(O\Lambda O^T)\mathbf{x} = (\mathbf{x}^T O)\Lambda(O^T\mathbf{x}) = \tilde{\mathbf{x}}^T\Lambda\tilde{\mathbf{x}} \ ,$$

using that $\mathbf{x}^T O = (O^T\mathbf{x})^T$. Noting that $\mathbf{x}^T O = \mathbf{x}^T(\mathbf{v}_1, \dots, \mathbf{v}_n) = (\mathbf{x}\cdot\mathbf{v}_1, \dots, \mathbf{x}\cdot\mathbf{v}_n)$ we get the new variables in terms of dot products of the form $\tilde{x}_i = \mathbf{x}\cdot\mathbf{v}_i$, $i = 1, \dots, n$. Thus, the quadratic form can be rewritten as a sum of squares (with coefficients being the eigenvalues) of dot products; in components:

$$Q(x_1, x_2, \dots) = \lambda_1(x_1 v_1^{(1)} + x_2 v_2^{(1)} + \cdots)^2 + \lambda_2(x_1 v_1^{(2)} + x_2 v_2^{(2)} + \cdots)^2 + \cdots \ ,$$

which is said to be the *normal form* of the quadratic form Q. For the above example in part (iii), we get:

$$Q = 3x_1^2 + 2x_1x_2 + 3x_2^2 = 2(\frac{x_1}{\sqrt{2}} - \frac{x_2}{\sqrt{2}})^2 + 4(\frac{x_1}{\sqrt{2}} + \frac{x_2}{\sqrt{2}})^2 = 2\tilde{x}^2 + 4\tilde{x}_2^2.$$

Verify this formula by multiplying it out and comparing it to the original form of Q in this example. It is now obvious that the fact that the matrix A has positive eigenvalues implies that Q is positive definite. In general, writing a quadratic form as a sum of squares in some new variables $\tilde{x}_i$ makes the sign behaviour of these quadratic forms self-evident.

(vi) **Classification rules**

The rules for classifying quadratic forms using the eigenvalues is as follows.

a) If all eigenvalues are > 0, Q is positive definite (PD)
b) If all eigenvalues are < 0, Q is negative definite (ND)
c) If some eigenvalues are ≤ 0 and some are ≥ 0, Q is indefinite (ID).
d) If all eigenvalues are ≥ 0, Q is positive semi-definite (PSD)
e) If all eigenvalues are ≤ 0, Q is negative semi-definite (NSD)

Note that although this classification by means of eigenvalues is more transparent than the principal minor test (see Handout **2**), it requires the finding of the eigenvalues of the matrix A, which in turn requires the solving of a n^{th}–order polynomial equation, whereas the principal minor test only requires the computation of determinants, which in general is much easier to do (and requires less computer time when done numerically).

3(B) Unconstrained Optimisation in Economics.

(i) **Production functions**

A firm produces Q items per year, which sell at a price p per item. The Revenue, $R = pQ$. Q depends on quantities x_1, x_2, etc. which are quantities such as the number of employees, amount spent on equipment, and so on. The vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ is called the input bundle vector. The Cost function measures how much a firm spends on production, and this will also be a function of x_1, x_2 and so on. The profit,

$$\Pi = R - C = pQ(x_1, x_2, \cdots, x_n) - C(x_1, x_2, \cdots, x_n).$$

To maximise the profit, we find the maximum of Π as a function of the inputs $x_1, \cdots, x_n$, leading to

$$p\frac{\partial Q}{\partial x_i} = \frac{\partial C}{\partial x_i}, \quad i = 1, \dots, n.$$

Problems where a firm produces several different products can also lead to optimisation problems.

(ii) **The discriminating monopolist.**

A monopolist produces two products at a rate Q_1 and Q_2. The price obtained is not constant, but reduces if the monopolist floods the market, so in a linear model $p_1 = a_1 - b_1 Q_1$ and $p_2 = a_2 - b_2 Q_2$ are the prices for the two products. The Cost function is then taken as $C = c_1 Q_1 + c_2 Q_2$, being simply proportional to the quantity produced. Here, the a_i, b_i, c_i for $i = 1, 2$ are constants. The profit

$$\Pi = R - C = p_1 Q_1 + p_2 Q_2 - c_1 Q_1 - c_2 Q_2 = (a_1 - c_1) Q_1 + (a_2 - c_2) Q_2 - b_1 Q_1^2 - b_2 Q_2^2,$$

so there is a best choice of Q_1 and Q_2 to maximise profit, when

$$\Pi_{Q_1} = (a_1 - c_1) - 2b_1 Q_1 = 0, \quad \text{and} \quad \Pi_{Q_2} = (a_2 - c_2) - 2b_2 Q_2 = 0.$$

Inspection of the associated Hessian shows that Π has a maximum here provided b_1 and b_2 are positive. (Why?) We then also require $a_1 \geq c_1$ and $a_2 \geq c_2$, such that Q_1, Q_2 are positive at the stationary points. The monopolist then has a *profit maximising strategy* allowing discrimination between the amounts of the two products to be produced. The same approach can be used when the cost function and the pricing are more complicated functions of the Q_i.

(iii) **The Cobb-Douglas production function.**

A commonly used model in microeconomics is the Cobb-Douglas production function

$$Q(x_1, x_2) = x_1^a x_2^b, \quad \text{where} \quad a > 0, \quad b > 0.$$

The input bundle here is then just a two-dimensional vector, and we assume here the price p is independent of Q, unlike in the monopolist problem. If the Cost is linear in the input bundle, $C = w_1 x_1 + w_2 x_2$, then the Profit

$$\Pi = R - C = pQ - C = p x_1^a x_2^b - w_1 x_1 - w_2 x_2$$

and the conditions for a stationary point are

$$\Pi_{x_1} = a p x_1^{a-1} x_2^b = w_1, \quad \Pi_{x_2} = b p x_1^a x_2^{b-1} = w_2$$

giving positive critical values $x_1^* = apQ^*/w_1$ and $x_2^* = bpQ^*/w_2$, where Q^* is the production function at the critical value. (Check these calculations!)

We want to know when this stationary value is a maximum, so we must examine the Hessian

$$H = \begin{pmatrix} a(a-1) p x_1^{a-2} x_2^b & abp x_1^{a-1} x_2^{b-1} \\ abp x_1^{a-1} x_2^{b-1} & b(b-1) p x_1^a x_2^{b-2} \end{pmatrix},$$

which is negative definite provided

$$LPM_1 = a(a-1) p x_1^{a-2} x_2^b < 0 \quad \text{and} \quad LPM_2 = ab(1-a-b) p^2 x_1^{2a-2} x_2^{2b-2} > 0, \tag{1}$$

which requires

$$0 < a < 1 \quad \text{and} \quad 0 < b < 1 \quad \text{and} \quad a + b < 1.$$

If these conditions are satisfied, then the Cobb-Douglas production function gives a *profit maximising strategy*. If they are not satisfied, the profit increases indefinitely as x_1 and x_2 get very large. This is possible economic behaviour: if there is no profit maximising strategy, all it means is that the bigger the firm the more profit it makes, which is of course entirely possible.

Cobb and Douglas (2005) based their analysis and function on data from the USA economy from 1899-1922. Filipe and Adams (2005) revisited these data using modern least-squares techniques, revealing some shortcomings in the data fitting of Cobb and Douglas from 1928, but they also offer some alternatives.

References

`https://en.wikipedia.org/wiki/CobbDouglas_production_function`

Cobb, C. W., Douglas, P. H. 1928: A theory of production. *American Economic Review* **18** (Supplement): 139–165.

Filipe, J., Adams, F.G. 2005: The estimation of the Cobb-Douglas function: a retrospective view. *Eastern Economic Journal* **31**, 427–445.

OB October 30, 2019

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 $LPM_3 = det H_B = -1$, 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.

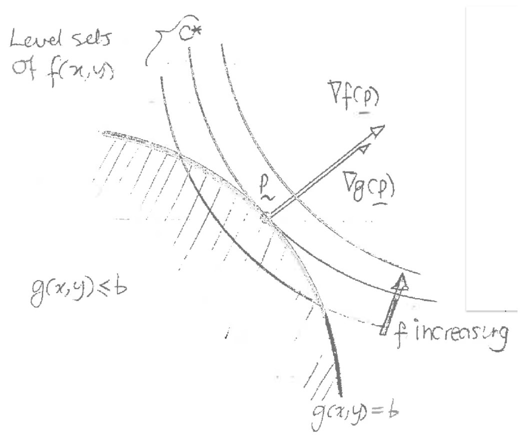
MATH2640 Introduction to Optimisation

5. Inequality Constraints, Complementary slackness condition, Maximisation and Minimisation, Kuhn-Tucker method: summary

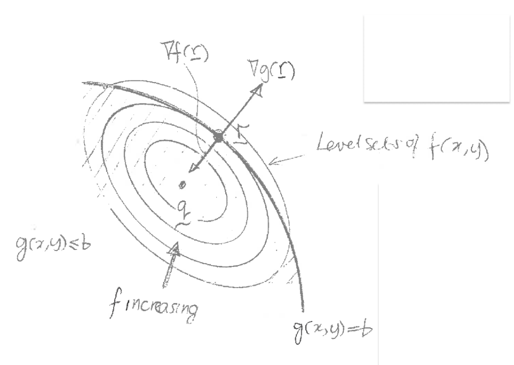
5(A) **Inequality constraints**

(i) Two variables, one inequality. To *maximise* $f(x,y)$ subject to $g(x,y) \leq b$ we look at the boundary of the region allowed by the inequality. By drawing a sketch of lines of constant f and constant g we see that if there is a point where ∇f and ∇g are parallel, then this point will give a maximum of f for the region $g(x,y) \leq b$. It is necessary that ∇f and ∇g are parallel and not anti-parallel, i.e. there must be a positive λ with $\nabla f = \lambda \nabla g$. If this is the case, the inequality constraint is *binding.*

If ∇f and ∇g have opposite signs, then we can increase f by going in the direction of ∇f and we are still in the region $g(x,y) \leq b$. The boundary of the constraint region is then of no particular significance for finding the maximum of f subject to $g(x,y) \leq b$. In this case we say the constraint is not binding, or the constraint is ineffective. The maximum is then found by looking for the *unconstrained* maximum of $f(x,y)$



a)



b)

Figure 1: Sketches for a) binding and b) nonbinding inequality constraints. Courtesy Prof. Mark Kelmanson.

(ii) **Complementary Slackness Condition**

We define a Lagrangian $L(x,y,\lambda) = f(x,y) - \lambda g(x,y)$. If the constraint is binding, then the equations to be solved are

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = 0, \quad g(x,y) = b.$$

If the constraint is not binding, then the equations to be solved are

$$\frac{\partial f}{\partial x} = 0, \quad \frac{\partial f}{\partial y} = 0,$$

and the constraint is ignored. We can neatly capture both of these sets of equations by writing

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = 0, \quad \lambda(g(x,y) - b) = 0.$$

The third equation is called the *complementary slackness condition* (CSC). It can either be solved with $\lambda \neq 0$, in which case we get the binding constraint conditions, or with $\lambda = 0$, in which case the constraint $g(x,y) - b = 0$ does not have to hold, and the Lagrangian $L = f - \lambda g$ reduces to $L = f$. So both cases are taken care of automatically by writing the first order conditions as

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = 0, \quad \lambda(g(x,y) - b) = 0.$$

(iii) **Example:** maximise $f(x,y) = xy$ subject to $x^2+y^2 \le 1$.

$$L = xy - \lambda(x^2+y^2-1).$$

Equalities: $L_x = y - 2\lambda x = 0 \to y = 2\lambda x$, $L_y = x - 2\lambda y = 0 \to x = 2\lambda y$ giving $y(1-4\lambda^2) = 0$ so $y = 0$ or $\lambda = 1/2 > 0$; complementary slackness condition gives $-L_\lambda = \lambda(x^2+y^2-1) = 0$. Inequalities are $\lambda \ge 0$, $x^2+y^2 \le 1$.

Solutions satisfying these equalities and inequalities, are (a), $\lambda = x = y = 0$, (b) $\lambda = 1/2$, $x = y = 1/\sqrt{2}$, and (c) $\lambda = 1/2$, $x = y = -1/\sqrt{2}$. In solution (a), the constraint is not binding, and the solution is a saddle not a maximum. In both the solutions (b) and (c), the constraint is binding, because $\lambda \neq 0$, and so the constraint g takes its maximum value of 1 there. Both (b) and (c) give local maxima, and both give an equal value of 1/2 for f. Since the constraint is binding, we can use the bordered Hessian method to establish that these solutions really are maxima. Since $\nabla g = (2x, 2y)$, and $L_{xx} = -2\lambda$, $L_{xy} = 1$, $L_{yy} = -2\lambda$, for solution 2, $\lambda = 1/2$, $x = y = 1/\sqrt{2}$, the bordered Hessian is

$$\begin{pmatrix} 0 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & -1 & 1 \\ \sqrt{2} & 1 & -1 \end{pmatrix}$$

which has $LPM_3 = 8$ which has the same sign as $(-1)^n = (-1)^2 = 1$. Since $n = 2$, $m = 1$ we only need to examine LPM_3 and so the bordered Hessian is negative definite, corresponding to a local maximum. Similarly, the third solution, $\lambda = 1/2$, $x = y = -1/\sqrt{2}$ is also a local maximum.

(iv) **Example:** maximise $f(x,y) = 2x^2 + y$ subject to $g(x,y) = x^2 + 9y^2 \le 1 = b_1$.

$$L(\lambda, x, y) = f(x,y) - \lambda\,(g(x,y) - b_1) = 2x^2 + y - \lambda(x^2+9y^2-1).$$

FOCs, CSC and inequalities:

$$L_x = 4x - 2\lambda x = 0 \quad \text{(i)} \tag{1}$$
$$L_y = 1 - 18\lambda y = 0 \quad \text{(ii)} \tag{2}$$
$$\lambda(x^2+9y^2-1) = 0 \quad \text{(iii)} \tag{3}$$
$$\lambda \ge 0, x^2+9y^2 = 1 \le 0. \tag{4}$$

From (i): $x = 0$ or $\lambda = 2$; from (ii) $\lambda \neq 0$, otherwise contradiction as (ii) cannot hold so $x^2+9y^2 = 1$ is binding. Case $x = 0$ gives $y = \pm 1/3, \lambda = \pm 1/6$ so stationary point is $(x^*, y^*, \lambda^*) = (0, 1/3, 1/6)$ since $\lambda > 0$ must hold. Case $\lambda = 2$ gives from (ii) $y = 1/36$ and from (iii) $x = \pm(1/12)\sqrt{143}$ so stationary point is $(x^*, y^*, \lambda^*) = (\pm(1/12)\sqrt{143}, 1/36, 2)$. Bordered Hessian is

$$H_b = \begin{pmatrix} 0 & 2x & 18y \\ 2x & 4-2\lambda & 0 \\ 18y & 0 & -18\lambda \end{pmatrix}. \tag{5}$$

$m = 1, n = 2$ so $2m+1 = 3$ and only $LPM_3 = \det(H_B)$ is needed. $LPM_3 = 72(\lambda - 18y^2)$. For $(0, 1/3, 1/6)$ $sign(LPM_3) = -1 = (-1)^{m=1}$ so PD and a (local) minimum. For $(\pm(1/12)\sqrt{143}, 1/36, 2)$, $LPM_3 = 143 > 0$ so $sign(LPM_3 = (-1)^{n=2}$ so ND and a (local) maximum.

5(B) **Several inequality constraints**

We can generalize in a natural way to the case where there is more than one inequality constraint and n variables $\mathbf{x} = (x_1, \cdots, x_n)$. To maximise $f(\mathbf{x})$ subject to

$$g_1(\mathbf{x}) \le b_1, \quad \cdots, \quad g_k(\mathbf{x}) \le b_k\,,$$

we define the Lagrangian

$$L = f(\mathbf{x}) - \lambda_1 g_1(\mathbf{x}) - \cdots - \lambda_k g_k(\mathbf{x})$$

and solve the n first order equality conditions

$$\frac{\partial L}{\partial x_1} = 0, \cdots, \frac{\partial L}{\partial x_n} = 0$$

together with the k *complementary slackness conditions* (CSCs)

$$\lambda_1(g_1(\mathbf{x}) - b_1) = 0, \cdots, \lambda_k(g_k(\mathbf{x}) - b_k) = 0.$$

These are $n+k$ equations for the $n+k$ unknowns $x_1, \cdots, x_n, \lambda_1, \cdots, \lambda_k$. However, ony solutions which satisfy the $2k$ inequalities

$$\lambda_1 \geq 0, \cdots, \lambda_k \geq 0, \quad g_1(\mathbf{x}) \leq b_1, \quad \cdots, \quad g_k(\mathbf{x}) \leq b_k$$

are admissible. If these do not hold, it means that the stationary point solution of the first order equalities is either not in the allowed region given by the constraints, or has a negative Lagrange multiplier, which is not acceptable for a maximum.

If we have a solution of the first order conditions, we must next check which constraints are binding and which are not. If any $\lambda_i = 0$, in the solution, then that constraint is not binding. Only constraints with $\lambda_i > 0$ correspond to binding constraints. To test whether the stationary point is indeed a maximum, we look at the bordered Hessian constructed from the Lagrangian and the binding constraints. Non-binding constraints are simply ignored when we consider whether the stationary point is a maximum or not. We should also check the nondegenerate constraint conditions for the binding constraints, that is check whether the components of ∇g_i form a set of linearly independent vectors.

Example: Maximise $f(x,y,z) = xyz$ subject to $x+y+z \leq 3$, $-x \leq 0$, $-y \leq 0$ and $-z \leq 0$. Note that we have written the conditions that x, y and z are positive in the standard form for a maximisation problem.

Solution:

$$L = xyz - \lambda_1(x+y+z-3) + \lambda_2 x + \lambda_3 y + \lambda_4 z$$

and the first order equalities give

$$L_x = yz - \lambda_1 + \lambda_2 = 0, \quad L_y = xz - \lambda_1 + \lambda_3 = 0, \qquad L_z = xy - \lambda_1 + \lambda_4 = 0,$$

$$\lambda_1(x+y+z-3) = 0, \quad \lambda_2 x = 0, \quad \lambda_3 y = 0, \quad \lambda_4 z = 0.$$

The eight inequalities are that the four $\lambda_i \geq 0$, and the four constraints already listed. The solutions are analysed by looking first at whether $\lambda_1 = 0$ or $x+y+z-3=0$. Four types of solution are found, (a) $x = y = 0$, $z =$ any positive value less than 3, $\lambda_1 = 0$, (b) $x = z = 0$, $y =$ any positive value less than 3, $\lambda_1 = 0$, (c) $y = z = 0$, $x =$ any positive value less than 3, $\lambda_1 = 0$, (d) $x = y = z = 1, \lambda_1 = 1$. All four solutions have $\lambda_2 = \lambda_3 = \lambda_4 = 0$. The solution (d) is the most interesting, as it provides the local and global maximum; note that the domain of available x, y and z is bounded by the constraints. At solution (d) only the constraint $x+y+z=3$ is binding, all the others being not binding and so having $\lambda_2 = \lambda_3 = \lambda_4 = 0$. To show (d) is a local maximum we can examine the 4×4 bordered Hessian with $m = 1$ constraint, $x+y+z-3 = 0$ (discounting the non-binding constraints) and $n = 3$ variables. We get

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

which has $LPM_4 = -3$ and $LPM_3 = 2$. The signs of the $n - m = 2$ highest $LPM's$ therefore alternate in sign, with the biggest, LPM_4, having the same sign as $(-1)^n$. Therefore H_B is negative definite, and so solution (d) is indeed a local maximum. By inserting solutions (a) to (d) into $f = xyz$ we see that (d) is also the global maximum. Note also that the NDCQ is trivially satisfied at solution (d), because $\nabla g = (1,1,1)$ is non-zero.

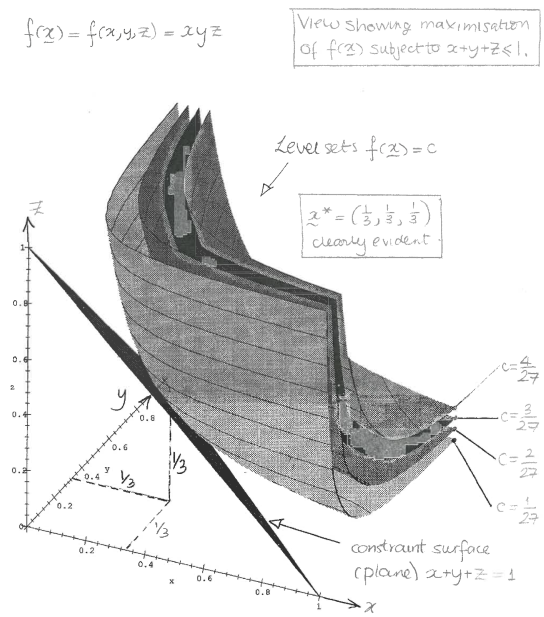


Figure 2: Sketches for nearly equivalent problem with $f(x,y,z) = xyz$ and $x + y + z \le 1$. Adapt the above example accordingly. Courtesy Prof. Mark Kelmanson.

5(C) **The Kuhn-Tucker method**

This is an alternative, and slightly simpler method for dealing with the common case where there are positivity constraints, that is constraints of the form

$$x_i \ge 0 \quad \text{or} \quad -x_i \le 0.$$

This is common in practice, where often variables are only meaningful when positive, e.g. production rates cannot be negative.

The typical problem is maximise $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le b_i$, $i = 1, k$, and $x_i \ge 0$, $i = 1, n$. The standard method for this problem would be to use the $n + k$ Lagrange multipliers corresponding to the $n + k$ constraints, and proceedas before. This works, but is rather cumbersome, and the Kuhn-Tucker (KT) method is neater.

We define the Kuhn-Tucker Lagrangian

$$\bar{L} = f - \sum_{i=1}^{k} \lambda_i (g_i - b_i) \tag{6}$$

just ignoring the positivity constraints. Then the usual Lagrangian

$$L = \bar{L} + \mu_1 x_1 + \cdots + \mu_n x_n$$

with $\mu_i \ge 0$ and $x_i \ge 0$. Now for the classical Lagrangian we know $\dfrac{\partial L}{\partial x_i} = 0$ for stationary points, so $\dfrac{\partial \bar{L}}{\partial x_i} + \mu_i = 0$. The complementary slackness equations for the $-x_i \le 0$ constraints are just $\mu_i x_i = 0$, so we have

$$x_i \frac{\partial \bar{L}}{\partial x_i} + x_u \mu_i = x_i \frac{\partial \bar{L}}{\partial x_i} = 0, \quad i = 1, \cdots, n. \tag{1}$$

The complementary slackness conditions on the remaining λ_i constraints are

$$\lambda_i(g_i - b_i) = 0,$$

which can be written

$$\lambda_i \frac{\partial \bar{L}}{\partial \lambda_i} = 0, \quad i = 1, \cdots, k. \tag{2}$$

Equations (1) and (2) are the $n + k$ first order equalities in the Kuhn-Tucker method. They are nicely symmetric in x_i and λ_i which makes them more memorable. Note also because we do not introduce the μ_i explicitly, we only need $n + k$ equations instead of $2n + k$ equations, a significant saving.

In addition, since the $\mu_i \geq 0$, we must have the inequalities

$$-\mu_i = \frac{\partial \bar{L}}{\partial x_i} \leq 0,$$

so the full set of $2n + 2k$ inequalities that a solution of the first-order conditions must satisfy are

$$\frac{\partial \bar{L}}{\partial x_i} \leq 0, \quad \frac{\partial \bar{L}}{\partial \lambda_i} = -\left(g_i - b_i\right) \geq 0, \quad x_i \geq 0, \quad i = 1, \cdots, n; \quad \lambda_i \geq 0, \quad i = 1, \cdots, k. \tag{3}$$

Eqs. (1),(2),(3) comprise the full set of KT relations needed for the analysis.

In summary, to solve the problem of maximising the function subject to $k = 1, \ldots, k$ inequality constraints $g_i(\mathbf{x}) \leq b_i$ and non-negativity constraints (NNCs) $x_i \geq 0, i = 1, \ldots, n$, it suffices to construct the KT Lagrangian which then leads to $n + k$ equalities and $2(n + k)$ inequalities, as follows

$$\bar{L}(\boldsymbol{\lambda}, \mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^{k} \lambda_i \left(g_i(\mathbf{x}) - b_i\right) \tag{7a}$$

$$x_i \frac{\partial \bar{L}}{\partial x_i} = 0, \quad i = 1, \ldots, n; \quad \lambda_i \frac{\partial \bar{L}}{\partial \lambda_i} = 0, \quad i = 1, \ldots, k \tag{7b}$$

$$\frac{\partial \bar{L}}{\partial x_i} \leq 0, \quad i = 1, \ldots, n; \quad \frac{\partial \bar{L}}{\partial \lambda_i} \geq 0, \quad i = 1, \ldots, k \tag{7c}$$

$$x_i \geq 0, \quad i = 1, \ldots, n \quad \lambda_i \geq 0, \quad i = 1, \ldots, k. \tag{7d}$$

Example: Maximise $f = x - y^2$ subject to $x^2 + y^2 \leq 4$, $x \geq 0$, $y \geq 0$ using the Kuhn-Tucker method.

Solution: $\bar{L} = x - y^2 - \lambda(x^2 + y^2 - 4)$. The Kuhn-Tucker equalities are

$$x \frac{\partial \bar{L}}{\partial x} = x - 2\lambda x^2 = 0, \tag{i}$$

$$y \frac{\partial \bar{L}}{\partial y} = -2y^2 - 2\lambda y^2 = 0, \tag{ii}$$

$$\lambda(x^2 + y^2 - 4) = 0. \tag{iii}$$

The Kuhn-Tucker inequalities are

$$\frac{\partial \bar{L}}{\partial x} = 1 - 2\lambda x \leq 0, \quad \frac{\partial \bar{L}}{\partial y} = -2y - 2\lambda y \leq 0, \quad \frac{\partial \bar{L}}{\partial \lambda_i} = -(x^2 + y^2 - 4) \geq 0, \quad x \geq 0, \quad y \geq 0, \quad \lambda \geq 0.$$

Analysis of these shows that $x = 2$, $y = 0$, $\lambda = 1/4$ is the only admissible solution, and is the local and global maximum.

Example: Maximise $f(x, y, z) = xyz$ subject to $g(x, y, z) = x + y + z - 3 \leq 0$, $x \geq 0$, $y \geq 0$, $z \geq 0$ using the Kuhn-Tucker method. (Done earlier using the standard method.)

The KT Lagrangian is:

$$\bar{L} = xyz - \lambda(x + y + z - 3).$$

KT equalities and inequalities:

$$
\begin{aligned}
x\bar{L}_x =& x(yz-\lambda)=0 \quad \text{(i)} && (8a)\\
y\bar{L}_y =& y(xz-\lambda)=0 \quad \text{(ii)} && (8b)\\
z\bar{L}_z =& z(xy-\lambda)=0 \quad \text{(iii)} && (8c)\\
\lambda\bar{L}_\lambda =& \lambda(x+y+z-3)=0 \quad \text{(iv)} && (8d)\\
\bar{L}_x =& yz-\lambda \le 0, \quad \bar{L}_y = xz-\lambda \le 0, \quad \bar{L}_z = zy-\lambda \le 0 \quad \text{(v,vi,vii)} && (8e)\\
\bar{L}_\lambda =& -(x+y+z-3) \ge 0 \quad \text{(viii)} && (8f)\\
x \ge& 0, y\ge 0, z\ge 0, \lambda \ge 0. && (8g)
\end{aligned}
$$

together with $x \ge 0, y \ge 0, z \ge 0, \lambda \ge 0$.
Solution: from (iv), $\lambda = 0$ or $x+y+z=3$.
Case $\lambda = 0$. From (i)–(iii) we have $xyz = 0$. From (v)–(viii): $yz \le 0, xz \le 0, xy \le 0$ so at least two variables zero; so $(x,0,0),(0,y,0),(0,0,z)$ with $0 \le x \le 3$, $0 \le y \le 3$, $0 \le z \le 3$.
Case $\lambda \ne 0, x+y+z=3$. From (i), $x=0$ or $yz=\lambda$. From (ii) $y=0$ or $xz=\lambda$. From (iii), $z=0$ or $xy=\lambda$. If $x=0$ then (ii) gives $y=0$ (since $\lambda \ne 0$); (iii) gives $z=0$ but $(0,0,0)$ violates constraint. Similarly $x>0, y>0, z>0$. Hence, $xy=xz=yz=\lambda>0$ so $x=y=z=\lambda=1$ so stationary point $(x^*,y^*,z^*,\lambda^*)=)=(1,1,1,1)$ and this is a maximum. So we recover results obtained earlier.

Example from economics (maximising sales with advertising): A firm wants to maximise sales, i.e. the revenue, via advertsing without letting profit fall below some fixed level. Here $\Pi = \Pi(y,a)$ is profit wity a the cost of advertising and y the level of production. Also $R = R(y,a)$ is the revenue obtained and $C(y)$ is the cost of production. So we wish to maximise revenue $R(y,a)$ subject to the constraint that the profit $\pi(y,a) = R(y,a) - C(y) - a \ge M$ is above a fixed level M with $y \ge 0, a \ge 0$.

We solve this type of problem via the KT-method. KT Lagrangian is

$$\bar{L}(y,a) = R(y,a) - (M - R(y,a) + C(y) + a)\,,$$

since $\Pi \ge M$ we have constraint $g(y,a) - b = M - \Pi(y,a) \le 0$. The KT-relations are

$$
\begin{aligned}
y\bar{L}_y =& y(R_y + \lambda R_y - \lambda C_y) = 0 \quad \text{(i)} && (9a)\\
a\bar{L}_a =& a(R_a + \lambda R_a - \lambda) = 0 \quad \text{(ii)} && (9b)\\
\lambda\bar{L}_\lambda =& -\lambda(M - R + C + a) = 0 \quad \text{(iii)} && (9c)
\end{aligned}
$$

with inequalities

$$
\begin{aligned}
\bar{L}_y =& (R_y + \lambda R_y - \lambda C_y) \le 0 \quad \text{(iv)} && (9d)\\
\bar{L}_a =& (R_a + \lambda R_a - \lambda) \le 0 \quad \text{(v)} && (9e)\\
\bar{L}_\lambda =& -(M - R + C + a) \ge 0 \Longrightarrow \Pi \ge M. \quad \text{(vi)} && (9f)
\end{aligned}
$$

while $\lambda \ge 0, a \ge 0, y \ge 0$. We are really interested in models where advertusing increases the revenue, so where $R(y,a) > 0$ (strict inequality). From (v) and this requirement, $0 < R_a \ge \lambda/(1+\lambda) \ge 0$ such that $\lambda > 0$. The with $\lambda > 0$ (ii) yields $M - R - C + a = 0$ and the production $\Pi = M$ is equal to M, the rpoduction remains at a minimum level. How can this be alleviated?

Specific model: Take a Cobb-Douglas model with $R = 3y^{1/4}a^{1/4}, C = y, M = 1$. Hence, production $\Pi(y,a) = 3y^{1/4}a^{1/4} - y - a \ge 1$

$$\bar{L}(y,a) = 3y^{1/4}a^{1/4} - \lambda\left(1 + y + a - 3y^{1/4}a^{1/4}\right).$$

KT relations

$$
\begin{aligned}
y\bar{L}_y =& (1+\lambda)\frac{3}{4}y^{1/4}a^{1/4} - \lambda y = 0 \quad \text{(i)} && (10a)\\
a\bar{L}_a =& (1+\lambda)\frac{3}{4}y^{1/4}a^{1/4} - \lambda a = 0 \quad \text{(ii)} && (10b)\\
\lambda\bar{L}_\lambda =& -\lambda\left(1 + y + a - 3y^{1/4}a^{1/4}\right) = 0. \quad \text{(iii)} && (10c)\\
& && (10d)
\end{aligned}
$$

Note that > 0 as long as $y > 0, a > 0$. Now $\Pi \geq 1$ yields $3y^{1/4}a^{1/4} \geq 1$, $3y^{1/4}a^{1/4} \geq 1 + y + a > 0$ so $R_a >!$ From (i) and (ii) $\lambda y/(1+\lambda) = \lambda a/(1+\lambda)$ so $y = a$ and we know $\lambda > 0$ so constraint is binding. Hence, constraint yields $-1 - 2y + 3y^{1/2} = (2y^{1/2} - 1)(y^{1/2} - 1) = 0$ so $y = a = 1/4$ or $y == a = 1$ and $\lambda = -3$ (discarded) or $\lambda = 3$. So $y = a = 1, \lambda = 3; C = 1, R = 3, \Pi = 1$.

So the *advantage* of the KT-method is that there are fewer Lagrange multipliers and more symmetric relations. In addition, no bordered Hessian analysis is required.

5(D) **Mixed equality and inequality constraints**

In general we can have both inequality constraints and equality constraints. To maximise $f(\mathbf{x})$ subject to

$$g_1(\mathbf{x}) \leq b_1, \quad \cdots, \quad g_k(\mathbf{x}) \leq b_k\,,$$

and

$$h_1(\mathbf{x}) = c_1, \quad \cdots, \quad h_m(\mathbf{x}) = c_m\,,$$

we use the complementary slackness conditions to provide the equations for the Lagrange multipliers corresponding to the inequalities, and the usual constraint equations to give the Lagrange multipliers corresponding to the equality constraints. Thus

$$L = f - \sum_{i=1}^{k} \lambda_i (g_i - b_i) - \sum_{i=1}^{m} \mu_i (h_i - c_i)$$

and the equalities are

$$\frac{\partial L}{\partial x_1} = 0, \cdots, \frac{\partial L}{\partial x_n} = 0$$

together with the k complementary slackness conditions

$$\lambda_1(g_1(\mathbf{x}) - b_1) = 0, \cdots, \lambda_k(g_k(\mathbf{x}) - b_k) = 0,$$

and the m equality constraints

$$\frac{\partial L}{\partial \mu_1} = 0, \cdots, \frac{\partial L}{\partial \mu_m} = 0$$

This gives $n+m+k$ equations for n variables $x_1, \cdots, x_n$, k Lagrange multipliers $\lambda_1, \cdots, \lambda_k$ and m Lagrange multipliers $\mu_1, \cdots, \mu_m$. These equality conditions go along with $2k$ inequalities that any maximum must satisfy, $\lambda_i \geq 0$ and $g_i \leq b_i$. These equations constitute the first order conditions. The second order conditions are found from the bordered Hessian, where any non-binding inequality constraint is ignored, and the binding inequality constraints are treated in the same way as the equality constraints.

5(E) **Constrained Minimisation**

If we want to *minimise* $f(\mathbf{x})$ subject to the inequality constraints, we need to write the inequality constraints in the form

$$g_1(\mathbf{x}) \geq b_1, \quad \cdots, \quad g_k(\mathbf{x}) \geq b_k\,,$$

The key point is the $\geq$, different from the $\leq$ occurring when we want to maximise f. Why do we need to change the sign of the inequality? Because we want to keep the Lagrange multipliers always positive, so we need ∇f and ∇g to point in the same direction. Now at a binding constraint, ∇f points in the direction of increasing f, whereas the minimum lies in the direction of decreasing f. This is why the sign of the inequality must be written the other way round. Draw a few sketches of contours of $f(x, y)$ and $g(x, y)$ to convince yourself of this.

Example: Minimise $f = 2y - x^2$ subject to $x^2 + y^2 \leq 1$, $x \geq 0$, $y \geq 0$.

Solution: we must write the first constraint as $1 - x^2 - y^2 \geq 0$ because we want to minimise f, not maximise it. The other constraints are already in the correct form for minimisation.

$$L = 2y - x^2 - \lambda_1(1 - x^2 - y^2) - \lambda_2 x - \lambda_3 y.$$

The first order equalities are then

$$-2x + 2\lambda_1 x - \lambda_2 = 0, \quad 2 + 2\lambda_1 y - \lambda_3 = 0$$

$$\lambda_1(1 - x^2 - y^2), \quad \lambda_2 x = 0, \quad \lambda_3 y = 0.$$

As usual, all the $\lambda_i \geq 0$. Solutions are $x = y = 0$, $\lambda_1 = \lambda_2 = 0$, $\lambda_3 = 2$ and $x = 1$, $y = 0$, $\lambda_1 = 1$, $\lambda_2 = 0$, $\lambda_3 = 2$. This second solution is a local and global minimum. Note that this second solution has two binding constraints, $x^2 + y^2 \leq 1$ and $y \geq 0$. Since $\nabla(x^2 + y^2) = (2, 0)$ and $\nabla y = (0, 1)$ at the critical point, we note that these are two linearly indepependent vectors (not parallel), so the NDCQ are satisfied here.

OB November 26, 2019

Exam Advice Math2640 **per 13-12-2019**

- The following material is exam material:
  - all five type-set lecture notes posted on Minerva and/or GitHub; the last two sections of set five, not presented in class, are unlikely to lead to exam questions;
  - all five tutorial and assessment questions posted on Minerva and/or GitHub;
  - answers to the tutorial solutions have been posted on Minerva; and,
  - all material provided in the lectures.
- Answers to all five assessment questions will be posed in due course; three solution sets have been posted.
- The exam is a balanced exam and the examples, assessment and tutorial questions presented together form a good reflection of the level of the tentative exam. The exam is similar in nature as previous exams.
- There are (two) past exams with short solutions online; these are provided via the Taught Office; first direct queries about those to the Taught Office.
- One exam and corresponding solutions will be (partially) presented and likely fully posed online.
- Take care that the rules about the character of the Hessians and bordered Hessians are known and that the corresponding conclusions, whether it leads to a minimum, maximum or otherwise, are known.
- The exam may also contain questions about the theory presented.

Finally, good luck with the exam preparations and taking the exam itself.

Onno Bokhove

This question paper consists of
3 printed pages, each of which
is identified by the reference MATH264001.

All calculators must carry
an approval sticker issued
by the School of Mathematics.

School of Mathematics

January 2018

MATH264001

Introduction to Optimisation

Time Allowed: 2 hours

You must attempt to answer 4 questions.
If you answer more than 4 questions, only your best 4 answers will be counted towards your final mark for this exam.
All questions carry equal marks.

1. (a) Consider the function
$$h(x, y, z) = x^2 - 2y^2 + 3z^2 \ ,$$
and its level surface S through the point $(1, 1, 1)$. Determine the shortest distance from S to the origin by considering the function $f(x, y, z) = x^2 + y^2 + z^2$ and writing down the Lagrangian for finding the minimum of f subject to (x, y, z) lying on S. Determine the critical points and by comparison of the values of f at these points determine the minimum distance.

(b) Let $z(x, y)$ be defined implicitly by the relation
$$yz^4 - 2xz^2 + x^2y = 1 \ .$$
Find the partial derivatives z_x and z_y in terms of x, y and z, and evaluate these derivatives at the point $x = 3$, $y = 1$, $z = 2$. If, furthermore, x and y are related by the relation $x^2 - 2y^2 = 7$, find dz/dx at the point $(3, 1, 2)$.

(c) Calculate the gradient of the function h given in part (a), and hence determine the (unit) normal vector to the surface S at $(1, 1, 1)$. Derive the equation of the tangent plane to S at that point, and the rate of increase of h, $D_{\mathbf{u}}h(1, 1, 1)$, in the direction $\mathbf{u}$ parallel to the z-axis.

2. (a) Consider the function
$$f(x, y) = xy - 3x^2y^2$$
with x, y constrained by the relation $h(x, y) = x^2 + y^2 = 2$. Find the partial derivatives f_x, f_y and the total derivatives df/dx and df/dy, and explain the difference between partial and total derivatives geometrically (if possible through a sketch).

(b) Write down the Lagrangian for the problem of finding the critical points of the function f of part (a), subject to the constraint $h = 2$. Write down the first order conditions, and find all 8 stationary points for the function $f(x, y)$ subject to the given constraint.
Hint: It may be useful to consider sums and differences of the first order conditions on the Lagrangian.

(c) Give the general form of the bordered Hessian H_B for the problem of part (b), (i.e., give its entries in terms of x, y and the Lagrange multiplier λ, but without explicitly calculating the Hessian at the stationary points). Assuming that $\det(H_B) \neq 0$, which leading principal minors determine the character of the critical points?

3. (a) Consider the quadratic form
$$Q(x_1, x_2, x_3) = \tfrac{1}{2}(x_1^2 + x_2^2) - 3x_1x_2 + 2(x_1x_3 + x_2x_3) = (x_1, x_2, x_3)\boldsymbol{A}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \ .$$
Find the symmetric matrix $\boldsymbol{A}$ and apply the leading principal minor test to show that Q is indefinite.

(b) Compute the eigenvalues and the corresponding unit eigenvectors, and use them to construct the orthogonal matrix $\boldsymbol{O}$ diagonalising $\boldsymbol{A}$. Give the normal form of

Q and construct the variables in terms of which Q can be written as a sum of squares, i.e., determine variables $\tilde{x}_i$, $i = 1, 2, 3$, such that Q adopts the form

$$Q = \lambda_1 \tilde{x}_1^2 + \lambda_2 \tilde{x}_2^2 + \lambda_3 \tilde{x}_3^2,$$

and confirm from the eigenvalues that Q is indefinite.

(c) Consider next the quadratic form of part (a), but subject to the constraint

$$x_1 + x_2 - x_3 = 0\ .$$

Write down the relevant bordered Hessian, and determine the sign character of the constrained quadratic form using the leading principal minor test for bordered Hessians.

4. (a) Consider the problem of maximising a function $f(x, y, z)$ subject to an inequality constraint $g(x, y, z) \leq b$, with g some function and b a constant value, and subject to the positivity conditions $x \geq 0$, $y \geq 0$, $z \geq 0$. Write down the general form of the Kuhn-Tucker (KT) Lagrangian and all four relevant equality and eight inequality conditions on the variables and Lagrangians.

(b) Consider the function

$$f(x, y, z) = xyz + z$$

subject to the conditions

$$x^2 + y^2 + z^2 \leq 25\ , \quad x \geq 0\ , \quad y \geq 0\ , \quad z \geq 0\ .$$

Use the KT method to find all possible maximiser candidates, and by evaluating the value of f at those points determine the maximum. State which constraints are binding at the maximum.

5. (a) A company produces umbrellas at a rate $Q = x^{1/2}y^{1/2}$ in terms of inputs x and y which are positive quantities. Each umbrella sells at a rate of $£12$, while the production cost is $C(x, y) = x^{3/2} + 12y$ pounds. Write down the revenue and profit functions and compute the critical values of x and y and the stationary values of the profit. By computing the Hessian and using the second order conditions, verify that this stationary value is a maximum.

(b) A company produces a product which sells for $£1$. The production rate Q and cost function are given by

$$Q = 20 - \tfrac{1}{3}x^3 - 2y^2 - z^2\ , \quad C = 6x + 3y + 5z\ ,$$

where x, y and z are input variables which are subject to the constraint $x + y + z = 3$. Write down the profit function and the Lagrangian for the problem of maximising the profit subject to the constraint. Find the values of x, y and z at the stationary point where all input variables are positive.

(c) For the problem in part (b) find the 4×4 bordered Hessian matrix and evaluate the two relevant leading principal minors. Hence, show that the stationary point with x, y and z all positive is a local maximum and find the profit at this point.

End.

Module Title: Introduction to Optimisation **©UNIVERSITY OF LEEDS**

School of Mathematics **Semester One 201819**

Calculator instructions:

- You are allowed to use a calculator which has had an approval sticker issued by the School of Mathematics.

Exam information:

- There are 3 pages to this exam.
- There will be **2 hours** to complete this exam.
- **Answer all questions.**
- All questions are worth equal marks.
- You must show all your calculations

1. (a) Find the unit normal vector to the surface $g(x, y, z) = xy^2 + x^3y - 2z^3 = 0$ at the point $(1, 1, 1)$, and give an equation for the tangent plane to the surface at that point. Find the direction, expressed as a unit vector, along which the function g decreases most rapidly from the point $(1, 1, 1)$. What is the rate of increase of g in the direction $(1, 0, 0)$?

(b) If $f(x, y) = x^2y + xy^3$ and $y^2 + yx^2 = x^3$, find expressions for the partial derivatives $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ and the total derivatives $\dfrac{df}{dx}$ and $\dfrac{df}{dy}$ in terms of x and y. (In the latter you don't need to simplify your answer.)

(c) Let the function $z(x, y)$ be defined implicitly by the relation

$$z^2x - yz + 2xy = 4.$$

Find z_x, z_y and z_{xx} in terms of x, y and z. Show that at $x = y = 1$, $z = 2$ is a value of z. Is this the only possible value of z at $x = y = 1$? Find also the numerical values of z_x, z_y and z_{xx} at $x = y = 1$, $z = 2$.

2. (a) Find the symmetric matrix associated with the quadratic form

$$Q(x_1, x_2, x_3) = 3x_1^2 + \frac{3}{2}x_2^2 + \frac{11}{2}x_3^2 + 4x_1x_2 + 4x_1x_3 + x_2x_3.$$

Find also the eigenvalues and unit eigenvectors of this matrix and hence show Q is positive semi-definite, and confirm this using the Principal Minor Test. From the unit eigenvectors give the normal form of Q.

(b) Now impose on the quadratic form of part (a) the additional constraint $4x_1 - 5x_2 - x_3 = 0$. Use the bordered Hessian to show the sign character of the reduced quadratic form. Explain how the normal form obtained in part (a) reduces under the additional constraint.

3. (a) Write down the Lagrangian, and hence find the two stationary points of the problem

$$f(x, y, z) = \frac{1}{2}x^2 + yz + \frac{1}{3}y^3 - z^2, \quad \text{subject to } h(x, y, z) = x + y + z = 2.$$

(b) Find the bordered Hessian for this problem, and evaluate the required leading principal minors for both solutions. Hence show that one solution is a local maximum and the other is indefinite.

(c) A company produces baskets at a rate $Q(x, y) = x^{1/3}y^{1/2}$ which it sells for a price p. The inputs x and y are positive quantities. The cost of production is $C = ax + by$ pounds, where a and b are positive constants. Write down the revenue and profit functions. Write down the Lagrangian for the problem of maximising the profit under the condition that the cost is kept at a constant value C_0. Show that at the critical point (x^*, y^*) of the variables and λ^* of the Lagrange multiplier we have:

$$x^* = \frac{2C_0}{5a}, \quad y^* = \frac{3C_0}{5b}, \quad 1 + \lambda^* = \frac{5p}{6C_0}\left(\frac{2C_0}{5a}\right)^{1/3}\left(\frac{3C_0}{5b}\right)^{1/2}.$$

Give a formula for the profit at the critical point.

Turn the page over

4. Consider maximising the function $f(x, y) = 3x + 2y$ subject to the simultaneous constraints

$$y \leq x + 1 \ , \quad 2y \geq x - 3 \ , \quad \text{and} \quad x \leq 3 \ .$$

(a) Write down the relevant Lagrangian and the corresponding first order equations, including the complementary slackness conditions, and the relevant inequalities. Analyse these to find the unique point that gives the maximiser, and the corresponding values of the Lagrange multipliers (which constraints are binding?). What is the maximum value of f under the constraints?

(b) Sketch a graph indicating the region defined by the inequalities, and explain the solution found in terms of the level curves (i.e., level lines) of f.

(c) Consider the problem of maximising the function $f(x, y, z) = x + y + z$ subject to the constraints

$$x^2 + y^2 + z^2 \leq 9 \ , \quad x \geq 0, \quad y \geq 0, \quad z \geq 0 \ .$$

Write down the Kuhn-Tucker (KT) Lagrangian for this problem, and the corresponding KT equalities and inequalities. Solve this problem and show that there is a unique maximiser. Give the value of the Lagrange multiplier at the maximum and the maximum value of the function.

End.

1. (a) Consider the function

$$h(x, y, z) = x^2 - 2y^2 + 3z^2 \ ,$$

and its level surface S through the point $(1, 1, 1)$. Determine the shortest distance from S to the origin by considering the function $f(x, y, z) = x^2 + y^2 + z^2$ and writing down the Lagrangian for finding the minimum of f subject to (x, y, z) lying on S. Determine the critical points and by comparison of the values of f at these points determine the minimum distance.

Answer: The Lagrangian is

$$L(x, y, z, \lambda) = x^2 + y^2 + z^2 - \lambda(x^2 - 2y^2 + 3z^2 - 2) \ ,$$

with first order conditions:

$$L_x = 2(1 - \lambda)x = 0 \ , \quad L_y = 2(1 + 2\lambda)y = 0 \ , \quad L_z = 2(1 - 3\lambda)z = 0 \ ,$$

$$\text{and} \quad L_\lambda = 2 - x^2 + 2y^2 - 3z^2 = 0 \ .$$

The stationary points are $(\pm\sqrt{2}, 0, 0)$ with $\lambda = 1$ and $(0, 0, \pm\sqrt{\frac{2}{3}})$ with $\lambda = 1/3$. Since $f(\pm\sqrt{2}, 0, 0) = 2$ and $f(0, 0, \pm\sqrt{\frac{2}{3}}) = 2/3$ we have that the latter points are nearest to the surface with distance $\sqrt{2/3}$.

(b) Let $z(x, y)$ be defined implicitly by the relation

$$yz^4 - 2xz^2 + x^2y = 1 \ .$$

Find the partial derivatives z_x and z_y in terms of x, y and z, and evaluate these derivatives at the point $x = 3$, $y = 1$, $z = 2$. If, furthermore, x and y are related by the relation $x^2 - 2y^2 = 7$, find dz/dx at the point $(3, 1, 2)$.

Answer: The z_x and z_y partial derivatives are

$$z_x = \frac{1}{2}\,\frac{xy - z^2}{xz - yz^3} \ , \quad z_x = \frac{1}{4}\,\frac{x^2 + z^4}{xz - yz^3} \ .$$

At $(3, 1, 2)$ they acquire the values: $1/4$ and $-25/8$ respectively. At the curve $x^2 - 2y^2 = 7$ through that point, we have $dy/dx = 3/2$ and $dz/dx = -71/16$.

(c) Calculate the gradient of the function h given in part (a), and hence determine the (unit) normal vector to the surface S at $(1, 1, 1)$. Derive the equation of the tangent plane to S at that point, and the rate of increase of h, $D_{\mathbf{u}}h(1, 1, 1)$, in the direction $\mathbf{u}$ parallel to the z-axis.

Answer: The gradient is $\nabla h = (2x, -4y, 6z)$ and at $(1, 1, 1)$ we have $\nabla h(1, 1, 1) = (2, -4, 6)$. The tangent plane to S at that point has the equation $2x - 4y + 6z = 4$ with unit normal given by $\hat{\mathbf{n}} = (1/\sqrt{14}, -2/\sqrt{14}, 3/\sqrt{14})$. The directional derivative in the direction $\mathbf{u} = (0, 0, 1)$ is $D_{\mathbf{u}}h(1, 1, 1) = 6$.

2. (a) Consider the function

$$f(x,y) = xy - 3x^2y^2$$

with x, y constrained by the relation $h(x,y) = x^2 + y^2 = 2$. Find the partial derivatives f_x, f_y and the total derivatives df/dx and df/dy, and explain the difference between partial and total derivatives geometrically (if possible through a sketch).

Answer: The partial derivatives of f are given by

$$f_x = y - 6xy^2\ , \quad f_y = x - 6x^2y\ ,$$

and from he constraint $h = 2$ by implicit differentiation we get $dy/dx = -x/y$, or equivalently $dx/dy = -y/x$. Thus, the total derivatives are

$$\frac{df}{dx} = y - 6xy^2 - (x - 6x^2y)\frac{x}{y}\ , \quad \frac{df}{dy} = x - 6x^2y - (y - 6xy^2)\frac{y}{x}\ .$$

(b) Write down the Lagrangian for the problem of finding the critical points of the function f of part (a), subject to the constraint $h = 2$. Write down the first order conditions, and find all 8 stationary points for the function $f(x,y)$ subject to the given constraint.
Hint: It may be useful to consider sums and differences of the first order conditions on the Lagrangian.

Answer: The Lagrangian is

$$L(x,y,\lambda) = xy - 3x^2y^2 - \lambda(x^2 + y^2 - 2)\ ,$$

with first order conditions

$$L_x = y - 6xy^2 - 2\lambda x = 0\ , L_y = x - 6x^2y - 2\lambda y = 0\ , \quad L_\lambda = 2 - x^2 - y^2 = 0\ .$$

The stationary points are $(1,1)$ and $(-1,-1)$ both with $\lambda = -5/2$, the points $(1,-1)$ and $(-1,1)$, both with $\lambda = -7/2$, and the additional four points:

$$\left(\sqrt{1+\tfrac{1}{6}\sqrt{35}}, \sqrt{1-\tfrac{1}{6}\sqrt{35}}\right)\ , \quad \left(-\sqrt{1+\tfrac{1}{6}\sqrt{35}}, -\sqrt{1-\tfrac{1}{6}\sqrt{35}}\right)\ ,$$

$$\left(\sqrt{1-\tfrac{1}{6}\sqrt{35}}, \sqrt{1+\tfrac{1}{6}\sqrt{35}}\right)\ , \quad \left(-\sqrt{1-\tfrac{1}{6}\sqrt{35}}, -\sqrt{1+\tfrac{1}{6}\sqrt{35}}\right)\ ,$$

all four with $\lambda = 0$.

(c) Give the general form of the bordered Hessian H_B for the problem of part (b), (i.e., give its entries in terms of x, y and the Lagrange multiplier λ, but without explicitly calculating the Hessian at the stationary points). Assuming that $\det(H_B) \neq 0$, which leading principal minors determine the character of the critical points?

Answer: Since $\nabla h = (2x, 2y)$ and the second order derivatives

$$L_{xx} = -6y^2 - 2\lambda\ , \quad L_{yy} = -6x^2 - 2\lambda\ , m \quad L_{xy} = 1 - 12xy\ ,$$

Turn Over

we have the bordered Hessian

$$H_B = \begin{pmatrix} 0 & 2x & 2y \\ 2x & -6y^2 - 2\lambda & 1 - 12xy \\ 2y & 1 - 12xy & -6x^2 - 2\lambda \end{pmatrix}$$

With $n = 2$, $m = 1$ we only need $n - m = 1$ LPM, namely $LPM_3 = \det(H_B)$.

3. (a) Consider the quadratic form

$$Q(x_1, x_2, x_3) = \tfrac{1}{2}(x_1^2 + x_2^2) - 3x_1x_2 + 2(x_1x_3 + x_2x_3) = (x_1, x_2, x_3)\boldsymbol{A}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} .$$

Find the symmetric matrix $\boldsymbol{A}$ and apply the leading principal minor test to show that Q is indefinite.

Answer: The symmetric matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \begin{pmatrix} 1/2 & -3/2 & 1 \\ -3/2 & 1/2 & 1 \\ 1 & 1 & 0 \end{pmatrix} .$$

With $LPM_3 = \det(\boldsymbol{A}) = -4 < 0$, $LPM_2 = -2 < 0$, $LPM_1 = 1/2 > 0$ the LPM test tells us we have an *indefinite* quadratic form.

(b) Compute the eigenvalues and the corresponding unit eigenvectors, and use them to construct the orthogonal matrix $\boldsymbol{O}$ diagonalising $\boldsymbol{A}$. Give the normal form of Q and construct the variables in terms of which Q can be written as a sum of squares, i.e., determine variables $\bar{x}_i$, $i = 1, 2, 3$, such that Q adopts the form

$$Q = \lambda_1\bar{x}_1^2 + \lambda_2\bar{x}_2^2 + \lambda_3\bar{x}_3^2,$$

and confirm from the eigenvalues that Q is indefinite.

Answer: The eigenvalues are given by $\lambda_1 = 2$, $\lambda_2 = -2$ and $\lambda_3 = 1$. The corresponding unit eigenvectors are

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} , \quad \boldsymbol{v}_2 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} , \quad \boldsymbol{v}_3 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} ,$$

respectively, and hence the diagonalizing orthogonal matrix is given by

$$\boldsymbol{O} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \end{pmatrix} .$$

The normal form for Q is

$$Q(x_1, x_2, x_3) = 2\left(\frac{x_1 - x_2}{\sqrt{2}}\right)^2 - 2\left(\frac{x_1 + x_2 - x_3}{\sqrt{3}}\right)^2 + \left(\frac{x_1 + x_2 + 2x_3}{\sqrt{6}}\right)^2 ,$$

which is manifestly indefinite.

(c) Consider next the quadratic form of part (a), but subject to the constraint

$$x_1 + x_2 - x_3 = 0 .$$

Write down the relevant bordered Hessian, and determine the sign character of the constrained quadratic form using the leading principal minor test for bordered Hessians.

Answer: If we impose the constraint $x_1 + x_2 - x_3$ we observe that the second term in the normal form of part (b) disappears, and the constrained quadratic form reduces to $Q = 2\bar{x}_1^2 + \bar{x}_3^2$ on the constrained variables. Thus, the corresponding reduced quadratic form is positive definite. This is confirmed by considering the bordered Hessian

$$H_B = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 1/2 & -3/2 & 1 \\ 1 & -3/2 & 1/2 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix} ,$$

which has $LPM_4 = \det(H_B) = -6 < 0$, $LPM_3 = -4 < 0$, so that successive LPMs have the same sign and we identify $\mathrm{sgn}(LPM_4) = (-1)^m$ (with $n = 3$, $m = 1$), hence the constrained quadratic form is PD.

4. (a) Consider the problem of maximising a function $f(x, y, z)$ subject to an inequality constraint $g(x, y, z) \leq b$, with g some function and b a constant value, and subject to the positivity conditions $x \geq 0$, $y \geq 0$, $z \geq 0$. Write down the general form of the Kuhn-Tucker (KT) Lagrangian and all four relevant equality and eight inequality conditions on the variables and Lagrangians.

Answer: The general form of the KT Lagrangian in this situation is

$$\bar{L}(x, y, z, \lambda) = f(x, y, z) - \lambda(g(x, y, z) - b) .$$

The KT equalities are:

$$x\frac{\partial \bar{L}}{\partial x} = 0 , \quad y\frac{\partial \bar{L}}{\partial y} = 0 , \quad z\frac{\partial \bar{L}}{\partial z} = 0 , \quad \lambda\frac{\partial \bar{L}}{\partial \lambda} = 0 ,$$

and the inequalities are

$$\frac{\partial \bar{L}}{\partial x} \leq 0 , \quad \frac{\partial \bar{L}}{\partial y} \leq 0 , \quad \frac{\partial \bar{L}}{\partial z} \leq 0 , \quad \frac{\partial \bar{L}}{\partial \lambda} \geq 0 ,$$

together with

$$x \geq 0 , \quad y \geq 0 , \quad z \geq 0 , \quad \lambda \geq 0 .$$

(b) Consider the function

$$f(x, y, z) = xyz + z$$

subject to the conditions

$$x^2 + y^2 + z^2 \leq 25 , \quad x \geq 0 , \quad y \geq 0 , \quad z \geq 0 .$$

Use the KT method to find all possible maximiser candidates, and by evaluating the value of f at those points determine the maximum. State which constraints are binding at the maximum.

Answer: The Lagrangian in this case is:

$$L(x, y, z, \lambda) = xyz + z - \lambda(x^2 + y^2 + z^2 - 25) .$$

Maximisers are $(0, 0, 5)$ with $\lambda = 1/10$ and with $f(0, 0, 5) = 5$, and $(\sqrt{8}, \sqrt{8}, 3)$ with $\lambda = 3/2$ and $f(\sqrt{8}, \sqrt{8}, 3) = 27$. In both cases the constraint $x^2+y^2+z^2 \leq 25$ is binding. The maximum is the latter point.

5. (a) A company produces umbrellas at a rate $Q = x^{1/2}y^{1/2}$ in terms of inputs x and y which are positive quantities. Each umbrella sells at a rate of £12, while the production cost is $C(x, y) = x^{3/2} + 12y$ pounds. Write down the revenue and profit functions and compute the critical values of x and y and the stationary values of the profit. By computing the Hessian and using the second order conditions, verify that this stationary value is a maximum.

Answer: The revenue function is $R(x, y) = 12x^{1/2}y^{1/2}$ and the product function is $\Pi(x, y) = R - C = 12x^{1/2}y^{1/2} - x^{3/2} - 12y$. The critical values of the outputs are $(x^*, y^*) = (4, 1)$ and the critical value of the profit $\Pi^* = 4$. At those values the Hessian is given by

$$H = \begin{pmatrix} -3/4 & 3/2 \\ 3/2 & -6 \end{pmatrix} ,$$

of which the LPMs are: $LPM_2 = 9/4 > 0$, $LPM_1 = -3/4 < 0$, and hence with alternating signs, $\mathrm{sgn}(LMP_2) = (-1)^n$ and hence H is negative definite, imlying the critical point is a maximum.

(b) A company produces a product which sells for £1. The production rate Q and cost function are given by

$$Q = 20 - \tfrac{1}{3}x^3 - 2y^2 - z^2 , \quad C = 6x + 3y + 5z ,$$

where x, y and z are input variables which are subject to the constraint $x+y+z = 3$. Write down the profit function and the Lagrangian for the problem of maximising the profit subject to the constraint. Find the values of x, y and z at the stationary point where all input variables are positive.

Answer: The profit function is $\Pi(x, y, z) = 20 - \frac{1}{3}x^3 - 2y^2 - z^2 - 6x - 3y - 5z$, and the Lagrangian

$$L(x, y, z, \lambda) = 20 - \tfrac{1}{3}x^3 - 2y^2 - z^2 - 6x - 3y - 5z - \lambda(x + y + z - 3) .$$

The stationary point with positive output values is $(1, 1, 1)$ with $\lambda = -7$.

(c) For the problem in part (b) find the 4×4 bordered Hessian matrix and evaluate the two relevant leading principal minors. Hence, show that the stationary point with x, y and z all positive is a local maximum and find the profit at this point.

Answer: The bordered Hessian for problem (b) at the stationary point is given by

$$H_B = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & -2 & 0 & 0 \\ 1 & 0 & -4 & 0 \\ 1 & 0 & 0 & -2 \end{pmatrix},$$

where we need $LPM_4 = \det(H_B) = -20 < 0$ and $LPM_3 = 6 > 0$, which have alternating signs and $\mathrm{sgn}(LMP_4) = (-1)^n$, hence ND. Thus, the critical point is a maximum with $\Pi(1, 1, 1) = 8/3$.

End.

Module Title: Introduction to Optimisation ©**UNIVERSITY OF LEEDS**

School of Mathematics **Semester One 201819**

Exam Checksheet

1. (a) Gradient: $\nabla g = (y^2 + 3x^2 y, 2xy + x^3, -6z^2)$, at $(1,1,1)$, $\nabla g(1,1,1) = (4,3,-6)$,
Unit normal vector: $\hat{\mathbf{n}} = (\frac{4}{\sqrt{61}}, \frac{3}{\sqrt{61}}, -\frac{6}{\sqrt{61}})$,
Equation of the tangent plane: $4x + 3y - 6z = 1$,
Direction of maximum decrease of g: $(-\frac{4}{\sqrt{61}}, -\frac{3}{\sqrt{61}}, \frac{6}{\sqrt{61}})$
Rate of increase of g in direction $(1,0,0)$: 4.

(b) Partial derivatives: $f_x = 2xy + y^3$, $f_y = x^2 + 3xy^2$
Total derivatives:

$$\frac{df}{dx} = 2xy + y^3 + (x^2 + 3xy^2)\,\frac{3x^2 - 2xy}{2y + x^2}\ ,$$

$$\frac{df}{dy} = x^2 + 3xy^2 + (2xy + y^3)\,\frac{2y + x^2}{3x^2 - 2xy}\ .$$

(c) Derivatives:

$$z_x = -\frac{z^2 + 2y}{2xz - y}\ , \quad z_y = -\frac{2x - z}{2xz - y}\ ,$$

$$z_{xx} = -\frac{2x(z^2 + 2y)^2}{(2xz - y)^3} + \frac{4z(z^2 + 2y)}{(2xz - y)^2}\ .$$

At $x = y = 1$ we have $z = 2$ or $z = -1$. At $z = 2$ we have $z_x = -2$, $z_y = 0$ and $z_{xx} = \frac{8}{3}$.

2. (a) Matrix $\mathbf{A} = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3/2 & 1/2 \\ 2 & 1/2 & 11/2 \end{pmatrix}$
Eigenvalues: $\lambda = 3, 0, 7$.
Unit eigenvectors: $\frac{1}{\sqrt{3}}(1,1,-1)$, $\frac{1}{\sqrt{42}}(-4,5,1)$, $\frac{1}{\sqrt{14}}(2,1,3)$ respectively.
Normal form:

$$Q(x_1,x_2,x_3) = 3\left(\frac{x_1 + x_2 - x_3}{\sqrt{3}}\right)^2 + 0\left(\frac{-4x_1 + 5x_2 + x_3}{\sqrt{42}}\right)^2 + 7\left(\frac{2x_1 + x_2 + 3x_3}{\sqrt{14}}\right)^2\ .$$

(b) Bordered Hessian:

$$H_B = \begin{pmatrix} 0 & 4 & -5 & -1 \\ 4 & 3 & 2 & 2 \\ -5 & 2 & 3/2 & 1/2 \\ -1 & 2 & 1/2 & 11/2 \end{pmatrix}$$

Signs with $n = 3, m = 1$: $LPM_4 = -882 < 0, LPM_3 = -179 < 0$, and so same sign and $\operatorname{sgn}(LMP_4) = (-1)^m$, positive definite
When $4x_1 - 5x_2 - x_3 = 0$ the constrained quadratic form $Q = 3\tilde{x}_1^2 + 7\tilde{x}_2^2$ which is positive definite on the constrained set.

3. (a) Lagrangian:

$$L(x, y, z, \lambda) = \frac{1}{2}x^2 + yz + \frac{1}{3}y^3 - z^2 - \lambda(x + y + z - 2)$$

First order conditions

$$L_x = x-\lambda = 0, \quad L_y = z+y^2-\lambda = 0, \quad L_z = y-2z-\lambda = 0, \quad L_\lambda = -(x+y+z)+2 = 0\,.$$

Critical points: $(22, -6, -14)$ with $\lambda = 22$ and $(1, 1, 0)$ with $\lambda = 1$.

(b) Bordered Hessian

$$H_B = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 2y & 1 \\ 1 & 0 & 1 & -2 \end{pmatrix}$$

With $n = 3, m = 1$ we have $LPM_4 = 2y + 5$, $LPM_3 = -(2y + 1)$. Hence, for $y = -6$ H_B negative definite, so $(22, -6, -14)$ local maximum, and for $y = 1$ H_B is indefinite, so $(1, 1, 0)$ an indefinite point.

(c) Revenue $R(x, y) = px^{1/3}y^{1/2}$ and profit $\Pi(x, y) = px^{1/3}y^{1/2} - ax - by$,
Lagrangian: $L(x, y, \lambda) = px^{1/3}y^{1/2} - (1 + \lambda)(ax + by) + \lambda C_0$.
Maximum profit: $\Pi^* = p\left(\frac{2C_0}{5a}\right)^{1/3}\left(\frac{3C_0}{5b}\right)^{1/2} - C_0$.

4. (a) lagrangian:

$$L(x, y, \lambda_1, \lambda_2, \lambda_3) = 3x + 2y - \lambda_1(y - x - 1) - \lambda_2(x - 3 - 2y) - \lambda_3(x - 3).$$

Maximum: $(3, 4)$ with $\lambda_1 = 2, \lambda_2 = 0, \lambda_3 = 5$, so the constraints $y \leq x + 1$ and $x \leq 3$ are binding, maximum value of f: $f(3, 4) = 17$.

(b) Region of interest is a triangle with vertices $(-5, -4)$, $(3, 0)$ and $(3, 4)$, with level contours of f given by lines $3x + 2y = c$, reaching the maximum $c = 17$ at the vertex $(3, 4)$.

(c) KT Lagrangian:

$$\bar{L}(x, y, z, \lambda) = x + y + z - \lambda(x^2 + y^2 + z^2 - 9)\ ,$$

KT equalities:

$$x\bar{L}_x = x(1 - 2\lambda x) = 0\ , \quad y\bar{L}_y = y(1 - 2\lambda y) = 0\ , \quad z\bar{L}_z = z(1 - 2\lambda z) = 0\ ,$$

and $\lambda\bar{L}_\lambda = -\lambda(x^2 + y^2 + z^2 - 9) = 0$
KT inequalities:

$$\bar{L}_x = 1 - 2\lambda x \leq 0\ , \quad \bar{L}_y = 1 - 2\lambda y \leq 0\ , \quad \bar{L}_z = 1 - 2\lambda z \leq 0\ ,$$

and $\bar{L}_\lambda = -(x^2 + y^2 + z^2 - 9) \geq 0$, and $x \geq 0, y \geq 0, z \geq 0, \lambda \geq 0$.
Maximum is the point $(\sqrt{3}, \sqrt{3}, \sqrt{3})$ with $\lambda = 1/\sqrt{12}$ and $f(\sqrt{3}, \sqrt{3}, \sqrt{3}) = 3\sqrt{3}$.