MATH2640 Introduction to Optimisation

Introduction

What is optimisation?

Answer: Optimisation is about "the quest for the best", i.e. designing strategies for obtaining the best result. (Optimum is Latin for best.) In many cases of human endeavour we search for conditions under which we attain the optimum. So often, we are trying to find maxima (or minima).

In economics, this can concern the highest profit, best rewards and maximum growth. In river or coastal flood-mitigation, this concerns mitigation or minimisation of flood damage, expressed in terms of staying below a specified water level at given locations, properties saved or economic damage saved, along a river or a coast. For wave-energy devices, this concerns maximising energy output given incoming waves. In real life, optimisation will include uncertainty, but we will not deal with the effects of uncertainty in this course.

In mathematics, optimisation is the subject of dealing with finding maxima or minima of functions. The definition of a *function* or *functions*: an expression or expressions in terms of variables $x, y, \ldots$ that can be altered, often also under additional conditions. Then we are tasked to find maxima or minima of the expression $f(x, y)$ (or expressions $f(x, y)$ and $g(x, y)$) subject to such additional conditions. Here the functions f, g can describe the quantities we are interested in, such as profit, revenue, tax, energy output, flood damage, etc.

Breakdown of the course

Chapter 1: Several variable calculus ($\sim$ 6 lectures)

- Develop techniques of multivariate functions, partial derivatives, total derivatives, gradients, directional derivatives, implicit differentiation, chain rule (in many variables), Taylor series, Hessian and stationary points.

Chapter 2: Unconstrained optimization ($\sim$ 4 lectures)

- Quadratic forms, eigenvalues, definiteness and principal minor test, stationary points, local extrema, applications to economics (Cobb-Douglas functions).

Chapter 3: Constrained optimization ($\sim$ 10 lectures)

- A) Equality constraints (conditions involving an equality $=$–symbol)
- B) Inequality constraints (conditions involving an inequality $<$–symbol)
- Lagrange multiplier formalism to find:
  - critical points and maximisers;
  - (bordered) Hessians to classify critical points;
  - Kuhn-Tucker theory; and,
  - applications to economics, etc.

Background reading including further exercises

- Enid R. Pinch 2002: *Optimal control and the calculus of variations.* Oxford Science Publications. 234 pp.
- Carl P. Simon and Lawrence Blume 1994: *Mathematics for economists.* Norton. 930 pp.

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,

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

$f(x, y, z) = C = \mathit{constant}$ 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 $f(x, y, z) =$ constant, $g(x, y, z) =$ 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.