Constrained optimization

The chain rule

The chain rule gives the derivative of a composite function in terms of the derivatives of its component functions.

Start with three functions, x and y, which are functions of one variable, and F, a funcion of two variables.

$$x = x(t) \text{ with derivative } x'(t) = \frac{dx}{dt}(t)$$

$$y = y(t) \text{ with derivative } y'(t) = \frac{dy}{dt}(t)$$

$$z = F(x,y) \text{ with partial derivatives } F'_x(x,y) = \frac{\partial F}{\partial x}(x,y) \text{ and } F'_y(x,y) = \frac{\partial F}{\partial y}(x,y)$$

(With a slight abuse of notation, the letters x and y denote the functions as well as their values.)

From these functions, construct the composite function $f(t)$:

$$f(t) = F(x(t), y(t))$$

The chain rule states that the derivative of the composite function can be expressed in terms of the derivatives of its component functions in the following way.

$$f'(t) = F'_x(x,y)x'(t) + F'_y(x,y)y'(t)$$

Here is a sketch of a proof:

Consider an increment to t, Δt, with corresponding increments to x, y, and f:

$$\Delta x = x(t + \Delta t) - x(t)$$

$$\Delta y = y(t + \Delta t) - y(t)$$

$$\Delta f = f(t + \Delta t) - f(t) = F(x + \Delta x, y + \Delta y) - F(x,y)$$

Adding and subtracting $-F(x, y+\Delta y) + F(x, y+\Delta y) = 0$ to the right hand side of the last expresion gives:

$$\Delta f = (F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y)) + (F(x, y + \Delta y) - F(x,y))$$

Multiplying the first two terms on the right hand side by $\Delta x/\Delta x = 1$ and the second two terms by $\Delta y/\Delta y = 1$ and dividing both sides by Δt yields:

$$\frac{\Delta f}{\Delta t} = \frac{F(x+\Delta x, y+\Delta y) - F(x, y+\Delta y)}{\Delta x}\frac{\Delta x}{\Delta t} + \frac{F(x, y+\Delta y) - F(x,y)}{\Delta y}\frac{\Delta y}{\Delta t}$$

Now take the limit letting $\Delta t \to 0$:

$$\frac{\Delta f}{\Delta t} \to f'(t)$$

$$\frac{\Delta x}{\Delta t} \to x'(t)$$

$$\frac{\Delta y}{\Delta t} \to y'(t)$$

$$\frac{F(x+\Delta x, y+\Delta y) - F(x, y+\Delta y)}{\Delta x} \to F'x(x,y)$$

$$\frac{F(x, y+\Delta y) - F(x,y)}{\Delta y} \to F'y(x,y)$$

as $\Delta t \to 0$

Thus these limits give the desired formula:

$$f'(t) = F'_x(x,y)x'(t) + F'_y(x,y)y'(t)$$

The slope of level curves

A level curve of the function F of two variables x and y, is the set of points (x,y) such that the function values $F(x,y)$ are constant over the set:

$$F(x,y) = k$$

Such a set will usually define a curve in the xy-plane called the level curve of F for the level k. If this curve represents a differentiable function $y(x)$, its derivative y' gives the slope of the level curve, and may be calculated using the chain rule above.

(The conditions under which this will work is given by the Implicit Function Theorem, which is covered by any textbook on the analysis of functions of several variables.)

The point $(x, y(x))$ lies on the level curve, so $F(x, y(x))$ is constant, i.e. $F(x, y(x)) = k$.

The chain rule applied to the composite function

$$f(x) = F(x, y(x)) = k$$

gives:

$$f'(x) = F'_x(x, y)x'(x) + F'_y(x, y)y'(x) = F'_x(x, y) + F'_y(x, y)y'(x) = 0$$

since $x'(x) = 1$ and the derivative of the constant k is 0.

Solving for $y'(x)$ yields the expression for the slope of the level curve:

$$y'(x) = -\frac{F'_x(x, y)}{F'_y(x, y)}$$

Constrained optimization

The problem is to find a point (x, y) maximizing a function

$$F(x, y)$$

subject to the constraint

$$G(x, y) = 0$$

given the two functions F and G of two variables. If these functions satisfy certain regularity and differentiability criteria, the following method works.

The constraint $G(x, y) = 0$ is a level curve of the function G, implicitly defining a function $y = y(x)$ of one variable. Thus, according to what has been said earlier about the slope of level curves,

$$y'(x) = -\frac{G'_x(x, y)}{G'_y(x, y)}$$

Substituting this implicit function $y(x)$ into F, reduces the problem to finding the unconstrained maximum of the following function of one variable:

$$f(x) = F(x, y(x))$$

The first order condition for such a maximum is $f'(x) = 0$ By the chain rule, this amounts to:

$$f'(x) = F'_x(x, y) + F'_y(x, y)y'(x) = 0$$

Substituting the earlier expression for $y'(x)$ into this condition and rearranging yields the following first order condition for the maximization problem:

$$\frac{F'_x(x, y)}{F'_y(x, y)} = \frac{G'_x(x, y)}{G'_y(x, y)}$$

This first order condition together with the constraint is a system of two equations in the two variables x and y, which in well-behaved cases may be solved for the optimum point (x, y).

The first order condition has the following geometric interpretation. The left hand side of the equation expresses $-y(x)$ as the negative of the slope of a level curve of F, while the right hand side similarly expresses $-y(x)$ as the negative of the slope of the level curve $G(x, y) = 0$ to G. Thus the first order condition states that the level curves of F and G are tangent at the optimum point (x, y).

The Lagrange multiplier method

The Lagrange multiplier method is an alternative method to the one shown earlier, to solve constrained maximization problems, i.e. the problem to find a point (x, y) maximizing a function $F(x, y)$, subject to the constraint $G(x, y) = 0$.

To apply the method, form the function

$$L(x, y, \lambda) = F(x, y) + \lambda G(x, y)$$

The function L is called the Lagrangian, and the factor λ the Lagrange multiplier.

Then form the three first order conditions $L'_x = L'_y = L'_\lambda = 0$ as if the problem were to maximize L without constraints. The conditions are:

$$L'_x(x, y, \lambda) = F'x(x, y) + \lambda G'x(x, y) = 0$$

$$L'_y(x, y, \lambda) = F'y(x, y) + \lambda G'y(x, y) = 0$$
$$L'_\lambda(x, y, \lambda) = G(x, y) = 0$$

The last condition, with respect to the Lagrange multiplier, is just the original constraint. Eliminating the multiplier λ from the first two conditions, the ones with respect to x and y, yields:

$$\frac{F'_x(x, y)}{F'_y(x, y)} = \frac{G'_x(x, y)}{G'_y(x, y)}$$

This is just the first order condition for the original constrained maximization problem derived earlier. Together with the constraint it determines the maximum point.

The method with Lagrange multipliers thus give the same solution as the one derived earlier, but may be easier to remember and apply.