POL502: Multi-variable Calculus

Kosuke Imai
Department of Politics, Princeton University

December 12, 2005

So far, we have been working with a real-valued function with one variable, i.e., $f : X \mapsto \mathbf{R}$ with $X \subset \mathbf{R}$. In this chapter, we study multi-variable calculus to analyze a real-valued function with multiple variables, i.e., $f : X \mapsto \mathbf{R}$ with $X \subset \mathbf{R}^n$. Given our solid understanding of single-variable calculus, we will skip the proofs for the theorems and focus on the computational aspects.

1 Differential Calculus with Multiple Variables

When there are multiple variables, it is natural to consider how the value of a function changes as we change one variable at a time while keeping the other variables constant. In particular, the *partial derivative* is defined as a derivative with respect to one variable while regarding the other variables as constants.

Definition 1 *Let $f : X \mapsto \mathbf{R}$ be a differentiable function with $f(x)$ and $X \subset \mathbf{R}^n$ where $x = (x_1, \dots, x_n)$. Then, the first order partial derivative of f with respect to x_i at the point $x^* = (x_1^*, \dots, x_n^*)$ is defined as*

$$\frac{\partial}{\partial x_i} f(x^*) \; = \; \lim_{h \to 0} \frac{f(x_1^*, \dots, x_{i-1}^*, x_i^* + h, x_{i+1}^*, \dots, x_n^*) - f(x_1^*, \dots, x_n^*)}{h},$$

This notation contrasts with a notation commonly used for derivatives, $\frac{d}{dx} f(x)$. When $n = 2$, the partial derivative of $f(x, y)$ with respect to x at the point (x_0, y_0) is represented by the slope of the curve $z = f(x, y_0)$ at the point $(x_0, y_0, f(x_0, y_0))$. This means that one can also use partial derivatives to come up with *the standard linear approximation* of a function. For example, we can approximate $f(x, y)$ around a point (x_0, y_0) by

$$f(x, y) \; \approx \; f(x_0, y_0) + \frac{\partial}{\partial x} f(x, y_0)(x - x_0) + \frac{\partial}{\partial y} f(x_0, y)(y - y_0).$$

The last two terms representing the change in the linearization of f is called the *total differential* or *total derivative* of f.

To compute partial derivatives, we only need to apply the rules of differentiations we learned earlier in this course to one variable of interest while regarding the other variables as constants.

Example 1 *Find the partial derivatives of the following functions, $f : \mathbf{R}^2 \mapsto \mathbf{R}$*

1. $f(x_1, x_2) = 3x_1^2 + x_1 x_2 + 4x_2^2$.
2. $f(x_1, x_2) = \frac{3x_1 - 2}{x_1^2 + 3x_2}$.

As before, we can differentiate a function many times to obtain higher order partial derivatives. It turns out that the order you take the partial derivative does not matter. That is, calculating the partial derivative first with respect to x_i and then x_j gives you the same answer as calculating the partial derivative first with respect to x_j and then x_i.

Theorem 1 (Euler's Theorem) *Let $f : X \mapsto \mathbf{R}$ be a differentiable function with $f(x)$ and $X \subset \mathbf{R}^n$ where $x = (x_1, \ldots, x_n)$. Then, for any i, j,*

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x^*) \;=\; \frac{\partial^2}{\partial x_j \partial x_i} f(x^*)$$

provided that the relevant partial derivatives exist at a point $x^* = (x_1^*, \ldots, x_n^*)$.

Try Example 1 to see if this is indeed the case. The Chain rule also can be applied to the function with multiple variables.

Theorem 2 (The Chain Rule) *Let $f : X \mapsto \mathbf{R}$ be a differentiable function with $f(x)$ and $X \subset \mathbf{R}^n$ where $x = (x_1, \ldots, x_n)$. If $x_1, \ldots, x_{n-1}$ and x_n are differentiable functions of t, then $f(x)$ is also differentiable function of t and*

$$\frac{d}{dt} f(x) \;=\; \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(x) \frac{dx_i}{dt}, \quad \text{and} \quad \frac{\partial}{\partial t} f(x) \;=\; \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(x) \frac{\partial x_i}{\partial t}.$$

Let's try an example.

Example 2 *Consider a function $f(x, y, z) = x + 2y + z^2$ with $x = r/s$, $y = r^2 + \log s$, and $z = 2r$. Compute $\frac{\partial}{\partial r} f(x, y, z)$ and $\frac{\partial}{\partial s} f(x, y, z)$.*

Now, we can compute the rate of change of a function with multiple variables at a given point x^* in any given direction, which is represented by a unit vector v whose length is 1.

Definition 2 *Let $f : X \mapsto \mathbf{R}$ be a differentiable function with $f(x)$ and $X \subset \mathbf{R}^n$ where $x = (x_1, \ldots, x_n)$. Then, the derivative of f at $x^* = (x_1^*, \ldots, x_n^*)$ in the direction of a unit vector $v = (v_1, \ldots, v_n)$ is called the directional derivative and is defined by*

$$D_v f(x^*) \;=\; \lim_{h \to 0} \frac{f(x_1^* + hv_1, \ldots, x_n^* + hv_n) - f(x_1^*, \ldots, x_n^*)}{h}.$$

It is awkward to calculate the directional derivative using its definition. Fortunately, the next theorem provides an easy way to get around this problem.

Theorem 3 (Directional Derivative and Gradient) *Let $f : X \mapsto \mathbf{R}$ be a differentiable function with $f(x)$ and $X \subset \mathbf{R}^n$ where $x = (x_1, \ldots, x_n)$. Assume that the partial derivatives of f are defined at a point $x^* = (x_1^*, \ldots, x_n^*)$. Then*

$$D_v f(x^*) \;=\; \nabla f(x^*) \cdot v$$

where $\cdot$ represents the inner product, and $\nabla f(x^*)$ is called a gradient (or gradient vector) which is an n dimensional column vector defined as,

$$\nabla f(x) \;=\; \begin{pmatrix} \frac{\partial}{\partial x_i} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix}.$$

The theorem implies that the directional derivative can be computed by $D_v f(x^*) = \sum_{i=1}^n \frac{\partial}{\partial x_i} f(x^*)\, v_i$. Let's do some examples.

Example 3 *Compute the directional derivative of* $f(x, y) = y^2 e^{3x} + xy$ *at the point* $(4, -2)$ *in the direction of* $(1, 2)$.

The gradient vector can also be seen as the partial derivative of a function $f(x_1, \ldots, x_n)$ with respect to a vector of variables x where $x = (x_1, \ldots, x_n)$,

$$\frac{\partial}{\partial x} f(x) \;=\; \begin{pmatrix} \frac{\partial}{\partial x_i} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix}.$$

All the algebraic operations we learned for the derivatives apply to the gradients as well.

Theorem 4 (Algebraic Operations of Gradients) *Let* $f : X \mapsto \mathbf{R}$ *and* $g : X \mapsto \mathbf{R}$ *be a differentiable function with* $f(x), g(x)$, *and* $X \subset \mathbf{R}^n$ *where* $x = (x_1, \ldots, x_n)$. *Assume that the gradients of these functions exist at a point* $x^* = (x_1^*, \ldots, x_n^*)$ *and denote them by* $\nabla f(x^*)$ *and* $\nabla g(x^*)$.

1. $\nabla(kf(x^*)) = k\nabla f(x^*)$ *for* $k \in \mathbf{R}$.
2. $\nabla(f(x^*) + g(x^*)) = \nabla f(x^*) + \nabla g(x^*)$.
3. $\nabla(f(x^*)g(x^*)) = g(x^*)\nabla f(x^*) + f(x^*)\nabla g(x^*)$.
4. $\nabla \frac{f}{g}(x^*) = \frac{g(x^*)\nabla f(x^*) - f(x^*)\nabla g(x^*)}{g(x^*)^2}$.

Using partial derivatives, we can find local maximum and minimum of a real-valued function with multiple variables. The first and second order conditions are completely analogous to those for a function with a single variable.

Theorem 5 (First and Second Order Conditions) *Let* $f : X \mapsto \mathbf{R}$ *be a differentiable function with* $f(x)$ *and* $X \subset \mathbf{R}^n$ *where* $x = (x_1, \ldots, x_n)$.

1. *If* x^* *is a local maximum or minimum of* f *and is an interior point (i.e., not boundary points) of* X, *then*
$$\frac{\partial}{\partial x_i} f(x^*) \;=\; 0 \quad \text{for all} \quad i = 1, \ldots, n.$$
If this condition is met, x^* *is said to be a critical (or stationary) point.*

2. *Suppose* x^* *is a critical point. Then,*
   (a) $f(x^*)$ *is a local maximum if* $H(x^*)$ *is a negative definite symmetric matrix.*
   (b) $f(x^*)$ *is a local minimum if* $H(x^*)$ *is a positive definite symmetric matrix.*
   (c) $f(x^*)$ *is a saddle (or inflection) point if* $H(x^*)$ *is indefinite.*

   where $H(x^*)$ *is called the Hessian (or Hessian matrix) of* $f(x)$ *and is defined as,*
$$H(x^*) \;=\; \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f(x) \end{bmatrix}.$$

Let's try an example.

Example 4 *Compute the critical points of $f(x, y) = x^3 - y^3 + 9xy$ and determine whether they are local minimum, local maximum, or a saddle point.*

Finally, we conclude this section by listing some useful differentiation rules for vector and matrices. When taking the derivative with respect to a vector, we compute a partial derivative with respect to each element of the vector.

Theorem 6 (Vector Differentiation) *Let x be an n dimensional vector of variables. Then,*

1. $\frac{\partial a^\top x}{\partial x} = a$ *for any n dimensional column vector of real numbers a.*
2. $\frac{\partial Ax}{\partial x} = A^\top$ *and* $\frac{\partial Ax}{\partial x^\top} = A$ *for any $m \times n$ matrix A.*
3. $\frac{\partial x^\top Ax}{\partial x} = (A + A^\top)x$ *for any $n \times n$ square matrix A.*

Here is an example we encounter in POL 571.

Example 5 *Let y be an n dimensional column vector of known constants, X be an $n \times m$ matrix of full column rank, and β be an m dimensional vector of unknown variables. The least squares estimator of β minimizes $f(\beta) = (y - X\beta)^\top(y - X\beta)$. Derive the expression of this estimator.*

2 Integral Calculus with Multiple Variables

Like derivatives, the concepts and techniques we have learned will be easily extended to multiple integrals of the form,

$$\int_{a_n}^{b_n} \cdots \int_{a_2}^{b^2} \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n.$$

All of the basic operations we learned about the integral calculus with a single variable applies to multiple integrals as well. For notational simplicity, we state the next theorem in terms of double indefinite integrals. But, the same results apply to definite integrals and even to triple and other higher order integrals.

Theorem 7 (Algebraic Operations of Multiple Integrals) *Let $f : X \mapsto \mathbf{R}$ be a integrable function with $f(x_1, x_2)$ and $X \subset \mathbf{R}^2$.*

1. $\int\int kf(x_1, x_2)\, dx_1 dx_2 = k \int\int f(x_1, x_2)\, dx_1 dx_2$. *for any $k \in \mathbf{R}$.*
2. $\int\int f(x_1, x_2) + g(x_1, x_2)\, dx_1 dx_2 = \int\int f(x_1, x_2)\, dx_1 dx_2 + \int\int g(x_1, x_2)\, dx_1 dx_2$.
3. $\int\int f(x_1, x_2)\, dx_1 dx_2 \geq \int\int g(x_1, x_2)\, dx_1 dx_2$ *if $f(x_1, x_2) \geq g(x_1, x_2)$.*

The next theorem enables us to interchange the order of integrals.

Theorem 8 (Fubini's Theorem) *Let $f : X \mapsto \mathbf{R}$ with $f(x_1, x_2)$ and $X \subset \mathbf{R}^2$.*

1. *If f is continuous on the rectangular region with $a \leq x_1 \leq b$ and $c \leq x_2 \leq d$, then*

$$\int_c^d \int_a^b f(x_1, x_2)\, dx_1 dx_2 \;=\; \int_a^b \int_c^d f(x_1, x_2)\, dx_2\, dx_1.$$

2. *Suppose f is continuous on the region defined by $a \le x_1 \le b$ and $g_1(x_1) \le x_2 \le g_2(x_2)$. If the same region can also be defined by $c \le x_2 \le d$ and $h_1(x_2) \le x_1 \le h_2(x_2)$, then*

$$\int_c^d \int_{h_1(x_2)}^{h_2(x_2)} f(x_1, x_2)\, dx_1 dx_2 \;=\; \int_a^b \int_{g_1(x_1)}^{g_2(x_1)} f(x_1, x_2)\, dx_2\, dx_1.$$

Here are some examples of multiple integrals. Try computing in two ways and confirm that you get the same answer.

Example 6 *Compute the following integrals.*

1. $\int_{-1}^{1} \int_0^2 (1 - 6x^2 y)\, dx dy$.
2. $\int_0^1 \int_0^x (3 - x - y)\, dy dx$.
3. $\int_0^2 \int_{x^2}^{2x} (4x + 2)\, dy\, dx$.

In the second question of the above example, the key is to find the region defined in two different ways: (1) $0 \le x \le 1$ and $0 \le y \le x$, and (2) $y \le x \le 1$ and $0 \le y \le 1$.

Finally, we give the result concerning the integration by substitution when there are multiple variables.

Theorem 9 (Integration by Substitution) *Let $f : X \mapsto \mathbf{R}$ with $f(x_1, x_2)$ and $X \subset \mathbf{R}^2$. If $x_1 = g(u, v)$ and $x_2 = h(u, v)$, then the change of variable within multiple integrals can be accomplished by the following formula,*

$$\int\int_D f(x_1, x_2)\, dx_1 dx_2 \;=\; \int\int_G f(g(u, v), h(u, v)) |J(u, v)|\, du dv,$$

where $J(u, v)$ is the Jacobian (or Jacobian determinant) defined by

$$J(u, v) \;=\; \begin{vmatrix} \frac{\partial x_1}{\partial u} & \frac{\partial x_1}{\partial v} \\ \frac{\partial x_2}{\partial u} & \frac{\partial x_2}{\partial v} \end{vmatrix} \;=\; \begin{vmatrix} \frac{\partial}{\partial u} g(u, v) & \frac{\partial}{\partial v} g(u, v) \\ \frac{\partial}{\partial u} h(u, v) & \frac{\partial}{\partial v} h(u, v) \end{vmatrix},$$

and D and G correspond to the same region defined by (x_1, x_2) and (u, v), respectively.

The key again is to find the expressions for the corresponding regions D and G. To do this, one needs to derive the boundaries of the integration region in terms of (x_1, x_2) and (u, v).

Example 7 *Use the integration by substitution to compute the following integrals.*

1. $\int_0^1 \int_0^{1-x} \sqrt{x + y} (y - 2x)^2\, dy\, dx$.
2. $\int_0^3 \int_0^4 \int_{y/2}^{y/2+1} \left(\frac{2x - y}{2} + \frac{z}{3} \right)\, dx\, dy\, dz$.