MA 105 D1 Lecture 13

Ravi Raghunathan

Department of Mathematics

Autumn 2014, IIT Bombay, Mumbai

Higher Derivatives

Maxima and Minima

The second derivative

Higher partial derivatives

Just as we repeatedly differentiated a function of one variable to get higher derivatives, we can also look at higher partial derivatives.

If we have a function $f(x_1, x_2)$ of two variables, then $\frac{\partial f}{\partial x_1}$ and $\frac{\partial f}{\partial x_2}$ are once again functions of two variables. We may thus take their partial derivatives with respect to either x_1 or x_2 again. There are four such possibilities:

$$\frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_1}, \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1}, \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2}, \quad \text{and} \quad \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_2}.$$

It is customary to write them as

$$\frac{\partial^2 f}{\partial {x_1}^2}, \frac{\partial^2 f}{\partial x_2 \partial x_1}, \frac{\partial^2 f}{\partial x_1 \partial x_2} \quad \text{and} \quad \frac{\partial^2 f}{\partial {x_2}^2}$$

respectively. In each case, as before, we hold one of the variables fixed and treat the two variable function as a function of one variable only.

The order in which differentiation is carried out

However, we now have more freedom. If we have a function $f(x_1, x_2)$ of two variables, we could first take the partial derivative with respect to x_1, then with respect to x_2, then again with respect to x_2, and so on. Does the order in which we differentiate matter?

Theorem 28: Let $f : U \to R$ be a function such that the partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}$ exist and are continuous for every $1 \leq i, j \leq m$. Then,

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Recall that functions $f : U \to \mathbb{R}$ for which the mixed partial derivatives of order 2 (that is, the $\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j}(f)\right)$) are all continuous are called $\mathcal{C}^2$ functions. Theorem 28 says that for $\mathcal{C}^2$ functions, the order in which one takes partial derivatives does not matter.

From now on we will use use the following notation. By

$$\frac{\partial^n f}{\partial x_1^{n_1} \partial x_2^{n_2} \ldots \partial x_k^{n_k}},$$

we mean: first take the partial derivative of f n_k times with respect to x_k, then n_{k-1} times with respect to x_{k-1}, and so on. The number n is nothing but $n_1 + n_2 + \ldots + n_k$. It is called the order of the mixed partial derivative.

Finally, we say that a function is $\mathcal{C}^k$ if all mixed partial derivatives of order k exist and are continuous. A function $f : U \to \mathbb{R}^n$ will be said to be $\mathcal{C}^k$ if each of the functions $f_1, f_2, \ldots, f_n$ are $\mathcal{C}^k$ functions.

From the preceding slide we see the we can talk about $\mathcal{C}^k$ functions for any function from (a subset of) $\mathbb{R}^m$ to $\mathbb{R}^n$. As in the one variable case we can also talk of smooth functions - these are functions for which all partial derivatives of all orders exist. In particular, the notion of a smooth vector field makes perfect sense. We will return to this later.

Recall that we often write f_x instead of $\frac{\partial f}{\partial x}$ for the partial derivative of f with respect to x. Similarly, we will sometimes write f_{xx} for $\frac{\partial^2 f}{\partial x^2}$ or f_{xy} for $\frac{\partial^2 f}{\partial y \partial x}$ etc. when convenient.

Local maxima and minima

As in the one variable case we can define local maxima and minima for a function of two or more variables. These definitions can be made for any function. They do not require us to assume any differentiability properties for the functions. Let $f : U(\subset \mathbb{R}^2) \to \mathbb{R}$ be a function of two variables.

Definition: We will say that the function $f(x, y)$ attains a local minimum at the point (x_0, y_0) (or that (x_0, y_0) is a local minimum point of f) if there is a disc

$$D_r(x_0, y_0) = \{(x, y) \mid \|(x, y) - (x_0, y_0)\| < r\}$$

of radius $r > 0$ around (x_0, y_0) such that $f(x, y) \geq f(x_0, y_0)$ for every point (x, y) in $D_r(x_0, y_0)$.

Similarly, we can define a local maximum point (Do this).

Critical Points

When the function is differentiable we can use the properties of the partial derivatives to find local maxima and minima. As in the one variable situation, we have the first derivative test. This is the analogue of Fermat's theorem. Before formulating the test we need make the following definition.

Defintion: A point (x_0, y_0) is called a critical point of $f(x, y)$ if

$$f_x(x_0, y_0) = f_y(x_0, y_0) = 0.$$

What does this say in geometric terms? Recall that the tangent plane to $z = f(x, y)$ at (x_0, y_0) is given by

$$z = z_0 + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Hence, at a critical point, the tangent plane is horizontal, that is, it is parallel to the xy-plane.

The first derivative test

Theorem 29: If (x_0, y_0) is a local extremum point (that is, a minimum or a maximum point) and if $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ exist, then (x_0, y_0) is a critical point.

The proof is similar in the one variable case. If (x_0, y_0) is not a critical point, then at least one of the two partial derivatives must be non-zero. Without loss of generality we can assume that $f_x(x_0, y_0) \neq 0$.

Suppose $f_x(x_0, y_0) > 0$. This means that

$$\lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} > 0.$$

This means that for $|h|$ small enough,

$$\frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} > 0.$$

If $h > 0$, this shows that the numerator is positive. On the other hand, if $h < 0$, the numerator must be negative.

Thus, in any disc $D_r(x_0, y_0)$ there are points (x, y) for which $f(x_0, y_0) < f(x, y)$ and $f(x_0, y_0) > f(x, y)$. The same argument can be repeated if $f_x(x_0, y_0) < 0$, giving a contradiction to the fact that (x_0, y_0) is an extreme point . $\square$

Towards a second derivative test

As in the one variable case, we would like to decide whether a local extremum is a local maximum or a local minimum. In order to this we will need to look a the partial derivatives of order 2. Let us assume that these exist.

We start by defining the Hessian of f. This is the matrix

$$\begin{pmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{pmatrix}.$$

From now on we will assume that f is a $\mathcal{C}^2$ function. Recall that this means that $f_{xy} = f_{yx}$.

The determinant of the Hessian is sometimes called the discriminant and is sometimes denoted D. Explicitly,

$$D = f_{xx}(x_0, y_0) f_{yy}(x_0, y_0) - [f_{xy}(x_0, y_0)]^2.$$

The second derivative test

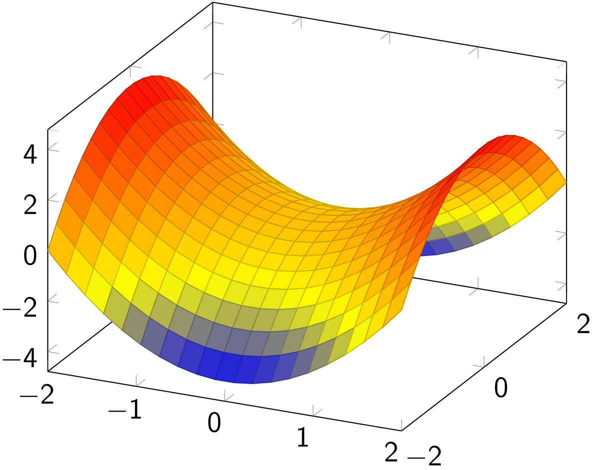
We give a test for finding local maxima and minima below. In the two variable situation, we will also need to understand what a saddle point is. We will explain this after stating the theorem.

Theorem 30: With notation as above:

1. If $D > 0$ and $f_{xx}(x_0, y_0) > 0$, then (x_0, y_0) is a local minimum for f.
2. If $D > 0$ and $f_{xx}(x_0, y_0) < 0$, then (x_0, y_0) is a local maximum for f.
3. If $D < 0$, then (x_0, y_0) is a saddle point for f.
4. If $D = 0$, further examination of the function is necessary.

Saddle points

Since a picture is worth a thousand words, let us start with one.



The point $(0,0)$ is called a saddle point. This is a picture of the graph of $z = x^2 - y^2$.

An example (from Marsden and Tromba)

Example 1: Find the maxima, minima and saddle points of $z = (x^2 - y^2)e^{\frac{(-x^2-y^2)}{2}}$.

Solution: Let us first find the critical points:

$$\frac{\partial z}{\partial x} = [2x - x(x^2 - y^2)]e^{\frac{(-x^2-y^2)}{2}} \quad \text{and}$$

$$\frac{\partial z}{\partial y} = [-2y - y(x^2 - y^2)]e^{\frac{(-x^2-y^2)}{2}}.$$

Hence the critical points are the simultaneous solutions of

$$x[2 - (x^2 - y^2)] = 0 \quad \text{and} \quad y[-2 - (x^2 - y^2)] = 0$$

The critical points thus lie at

$$(0,0), \quad (\pm\sqrt{2}, 0), \quad \text{and} \quad (0, \pm\sqrt{2})$$

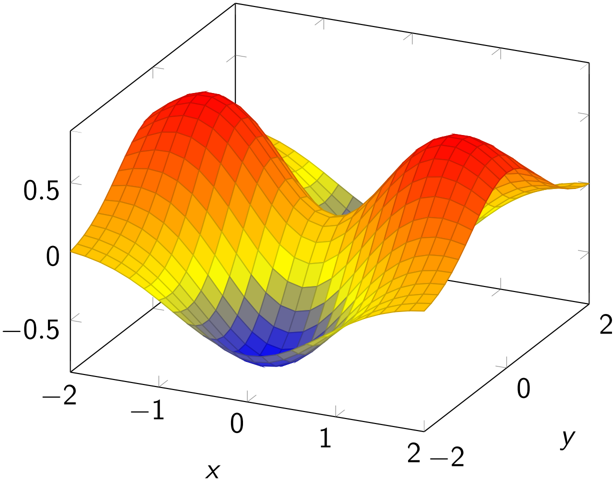
Next we have to find the partial derivatives of order 2. We have

$$\frac{\partial^2 z}{\partial x^2} = [2 - 5x^2 + x^2(x^2 - y^2) + y^2]e^{\frac{(-x^2-y^2)}{2}},$$
$$\frac{\partial^2 z}{\partial x \partial y} = xy(x^2 - y^2)e^{\frac{(-x^2-y^2)}{2}} \quad \text{and}$$
$$\frac{\partial^2 z}{\partial y^2} = [5y^2 - 2 + y^2(x^2 - y^2) - x^2]e^{\frac{(-x^2-y^2)}{2}}.$$

Using the second derivative test we obtain the following table:

Point	f_{xx}	f_{xy}	f_{yy}	D	Type
$(0,0)$	2	0	-2	-4	Saddle
$(\sqrt{2},0)$	$-4/e$	0	$-4/e$	$16/e^2$	Maximum
$(-\sqrt{2},0)$	$-4/e$	0	$-4/e$	$16/e^2$	Maximum
$(0,\sqrt{2})$	$4/e$	0	$4/e$	$16/e^2$	Minimum
$(0,-\sqrt{2})$	$4/e$	0	$4/e$	$16/e^2$	Minimum

The previous example in a picture



This is the graph of $z = (x^2 - y^2)e^{\frac{-x^2-y^2}{2}}$.

Quadratic functions in two variables

Consider functions of the form

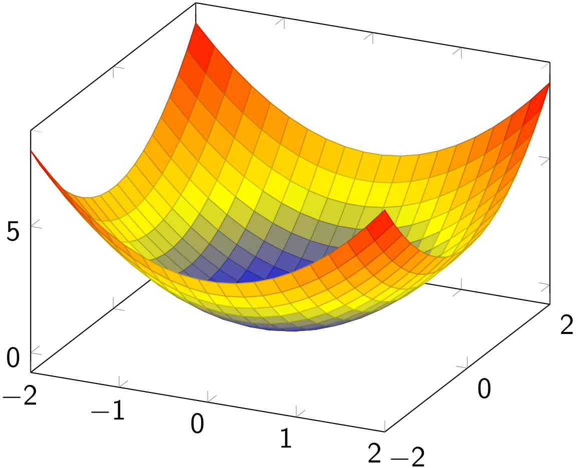
$$z = g(x, y) = Ax^2 + 2Bxy + Cy^2.$$

Notice that $(0, 0)$ is obviously a critical point for the function $g(x, y)$. With a little bit of work we can show that if $AC - B^2 \neq 0$, then $(0, 0)$ is the only critical point of g.

From now on we assume that $AC - B^2 \neq 0$. A little more analysis will show the following:

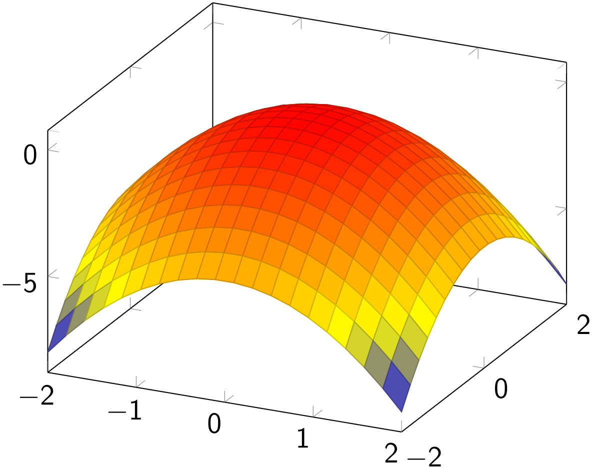
1. If $AC - B^2 > 0$, the function g has a local minimum if $A > 0$ and a local maximum if $A < 0$.
2. If $AC - B^2 < 0$, the function g has a saddle point, that is, in a small disc around the point, the function does not lie on any one side of its tangent plane.

A local minimum



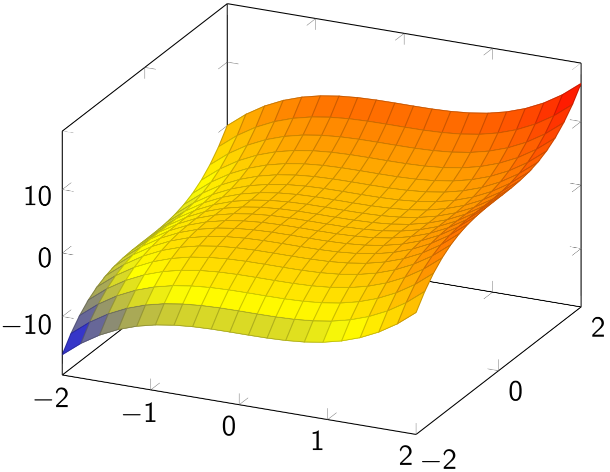
The graph of $x^2 + y^2$ has a local minimum at $(0, 0)$. Clearly $AC - B^2 = 1 > 0$ and $A > 0$.

A local maximum



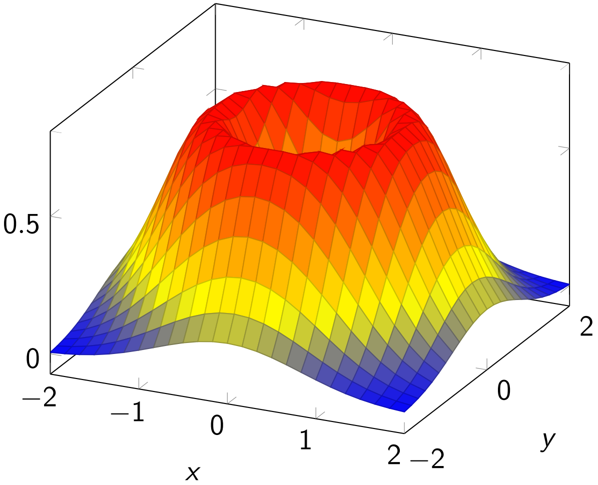
The graph of $-x^2 - y^2$ has a local maximum at $(0, 0)$. Clearly $AC - B^2 = 1 > 0$ and $A < 0$.

Where the test is inconclusive



The graph of $x^3 + y^3$. The test is inconclusive at $(0, 0)$.

The volcano



This is the graph of $z = 2(x^2 + y^2)e^{-x^2-y^2}$. Here the maxima lie on a circle (the rim of the volcano). This sort of behavior cannot arise in a quadratic surface.

Taylor's theorem in two variables

If we look at the quadratic surface $z = Ax^2 + 2Bxy + Cy^2$, we see that $2A = f_{xx}$, $2B = f_{xy}$ and $2C = f_{yy}$. The second derivative test tells us that whatever is true for quadratic surfaces is true in general. Why is this true?
The answer lies in a two variable form of Taylor's Theorem:

Theorem 32: If f is a $\mathcal{C}^2$ function in a disc around (x_0, y_0), then

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + f_x h + f_y k$$
$$+ \frac{1}{2!}[f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2] + \tilde{R}_2(h, k),$$

where $\tilde{R}_2(h, k)/\|(h, k)\|^2 \to 0$ as $\|(h, k)\| \to 0$.

From quadratics surfaces to general surfaces

If (x_0, y_0) is a critical point, Taylor's theorem in a disc around the critical point becomes

$$f(x_0+h, y_0+k) = f(x_0, y_0) + \frac{1}{2!}[f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2] + \tilde{R}_2(h, k),$$

where $\tilde{R}_2(h, k)/\|(h, k)\|^2 \to 0$ as $\|(h, k)\| \to 0$.

Thus, in a small disc around (x_0, y_0) the function $f(x, y)$ looks very much like a quadratic surface, and from the point of view of the critical points there is, in fact, no difference, because the error term can be made as small as we please even after dividing by $\|(h, k)\|^2$. This is why the second derivative test works.

Back to Taylor's Theorem

Suppose $g : [u, v] \to \mathbb{R}$ is a function of one variable. Let us assume that g is twice continuously differentiable on $[u, v]$. For points $a, b \in (u, v)$ we can rewrite Taylor's Theorem as

$$g(b) = g(a) + g'(a)h + \frac{g''(a)}{2!}h^2 + \frac{g''(c) - g''(a)}{2!}h^2,$$

for some c between a and b. Since we have assumed that g'' is continuous we see that $(g''(c) - g''(a)) \to 0$ as $h \to 0$. Thus we can write

$$g(b) = g(a) + g'(a)h + \frac{g''(a)}{2!}h^2 + \tilde{R}_2(h),$$

where $\tilde{R}_2(h)/h^2 \to 0$.

Exercise 1: Let $f(x, y)$ be a $\mathcal{C}^2$ function of two variables. Apply the preceding version of Taylor's Theorem to the function

$$g(t) = f(tx + (1 - t)x_0, ty + (1 - t)y_0),$$

for $0 \leq t \leq 1$. This will give the two variable version of Taylor's Theorem stated above. You can easily generalize this to degree n.

The four squares theorem

We will be studying the method of Lagrange multipliers next in the context of finding constrained optima. But for the moment, I would like to tell (remind?) you of another theorem of Lagrange.

Theorem: Every positive integer can be written as a sum of four squares.

Let us restate the above theorem. Given a natural number n, if there exist integers a_1, a_2, a_3 and a_4 (note that some of these may take the value 0!) such that

$$a_1^2 + a_2^2 + a_3^2 + a_4^2 = n,$$

we say that the function $x_1^2 + x_2^2 + x_3^2 + x_4^2$ in four variables represents the number n. So the theorem above says:

Theorem: The function $x_1^2 + x_2^2 + x_3^2 + x_4^2$ represents every natural number.

Quadratic forms

An *n-ary quadratic form* over the real numbers is a function from $\mathbb{R}^n$ or $\mathbb{Z}^n$ to $\mathbb{R}$ of the form

$$q(x_1, x_2, \ldots, x_n) = \sum_{1 \le i,j \le n} q_{ij} x_i x_j, \quad a_{ij} \in \mathbb{R}.$$

The example $x_1^2 + x_2^2 + x_3^2 + x_4^2$ is an example of a quartenary quadratic form. It is a diagonal form, that is, only square terms appear. Forms need not be diagonal - for instance $q_1(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2$ is a non-diagonal form in two variables. Another example (in three variables) is $q_2(x_1, x_2, x_3) = \sqrt{2}x_1^2 + 5x_2^2 - 7x_3^2 + 2x_1x_3$.

We will be interested only in integral quadratic forms, that is quadratic forms which take $\mathbb{Z}^n$ to $\mathbb{Z}$. Lagrange's quadratic form and q_1 above are integral, but q_2 above is not.

The Bhargava-Hanke Theorem

A quadratic form is called positive definite if $q(x_1, \ldots, x_n) > 0$ for all $(x_1, \ldots, x_n) \neq (0, 0, \ldots, 0)$ in $\mathbb{R}$.

In the examples above, the first two forms are positive definite, while the third is not (easy to check).

Theorem (M. Bhargava, J. Hanke): If a positive definite integral quadratic form represents every number $n \leq 290$, it represents all natural numbers.

This is an absolutely remarkable theorem (guessed at by Conway, a very well-known mathematician at Princeton). It says that you only need to check finitely manly values in order to prove something for infinitely many values!

Manjul Bhargava