Optimization of multi-variate functions without constraint

Math 251

September 30, 2015

1 Local extrema

Consider the function $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$. We interpret f as represent the height of the terrain at a point (x,y). Suppose that f is smooth, that is there is no sudden jump or sharp increase or decrease in the terrain height. Then it is of interest for us to determine some criteria for a point (x_0,y_0) such that $f(x,y)$ is highest (or lowest) at (x_0,y_0), compared with the neighboring points that are within a a radius R from (x_0,y_0). Note that we do NOT require (x_0,y_0) to be a *global* maximum or minimum point. We only require that *locally* $f(x_0,y_0)$ is no smaller or no larger than all of its neighbor. The radius R can depend on the point (x_0,y_0) as well.

The terminologies for these points are the so-called *local maximum*:

$$f(x_0,y_0) \geq f(x,y), \text{ for all } x,y \in D((x_0,y_0),R)$$

and *local minimum*

$$f(x_0,y_0) \leq f(x,y), \text{ for all } x,y \in D((x_0,y_0),R).$$

Here $D((x_0,y_0),R)$ refers to the disk of radius R around (x_0,y_0). That is the collection of all points of distance less than or equal to R away from (x_0,y_0).

We also want to mention the *saddle point*. A saddle point is an umbrella term for "all other points" that satisfies the first derivative criterion (see the below section) but fail to be a max or a min.

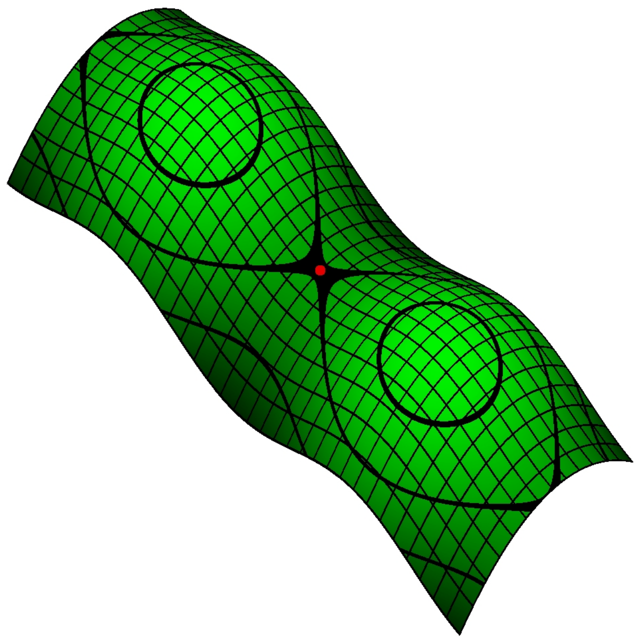

Figure 1.1: A saddle point between two maxima

1.1 The first derivative criterion

Consider a one dimensional function $f(x)$. How do we tell that a point x_0 is a potential extremum? If $f'(x_0)$ exists and is either positive or negative then clearly x_0 can NOT be an extremum. Thus x_0 is a *candidate* for an extremum if we have no information about the *sign* of f' at x_0. That is when either $f'(x_0) = 0$ or $f'(x_0)$ does not exist.

If we fixed a value $y = y_0$ for the function $f(x, y)$ then $f(x, y)$ becomes a one dimensional function $f(x, y_0)$. It is also clear that if f has a local extremum at (x_0, y_0) then $f(x, y_0)$ has a local extremum at $x = x_0$. Applying the same argument, we conclude that either $f_x(x_0, y_0) = 0$ or $f_x(x_0, y_0)$ does not exist. Similar conclusion applies to $f_y(x_0, y_0)$. This is the so-called Fermat's Theorem for local extrema.

1.2 The second derivative criterion

We now have a *necessary condition* for local extrema. Can we say something about a *sufficient condition* for local extrema. From Calculus 1, we know that $f'(a) = 0$ does Not imply a is a local extremum of f (for example $f(x) = x^3$). More importantly, we would like to have a criterion to tell us whether the extrema is a max or a min. The following condition gives the answer: let

$$D = f_{xx}f_{yy} - f_{xy}^2\Big|_{(x,y)=(x_0,y_0)}.$$

(This assumes that all the second order derivatives exist and are continuous near (x_0, y_0), hence we must have $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$.) Then we have

D	f_{xx}	Conclusion
> 0	> 0	local min
> 0	< 0	local max
< 0		saddle point
$= 0$		inconclusive

We remark that the inconclusive case means that we cannot conclude whether (x_0, y_0) is a max, min or saddle point based solely on the first and second derivative information. More specifically, we have $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ and the point (x_0, y_0) can be either a max, min or saddle point.

The second derivative test is similar to the result you have seen in Calculus 1. The reason is because both rely on the Taylor's expansion. First recall the Taylor's expansion in one dimension:

$$f(x) \approx f(x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2,$$

(since $f'(a) = 0$). Thus f is at a local max at a if $f''(x_0) < 0$ and a local min at a if $f''(x_0) > 0$. If $f''(x_0) = 0$ then the test is inconclusive because the Taylor approximation now becomes

$$f(x) \approx f(x_0) + f^{(3)}(x_0)(x - x_0)^3.$$

If $f^{(3)}(x_0) \neq 0$ then we are at a saddle point because $f(x)$ increases in one direction and decreases in another direction away from a. However, it could be the case that $f^{(3)}(x_0) = 0$ and we have to go to the fourth derivative etc. So we can not say much when $f''(x_0) = 0$.

1.3 Two dimensional Taylor expansion

Our table for the second derivative test comes from the two dimensional Taylor expansion for $f(x, y)$:

$$\begin{aligned} f(x,y) &\approx f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) \\ &+ \frac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + \frac{1}{2} f_{yy}(x_0,y_0)(y-b)^2 \\ &+ f_{xy}(x_0,y_0)(x-x_0)(y-y_0). \end{aligned} \tag{1}$$

This formula comes from the one dimensional Taylor epxpansion for $f(x, y)$ in the direction $(a + t\delta, b + t\sigma)$ as a function of t. Here we think of δ, σ as representing the small change of x away from a and y away from b. Writing $\phi(t) = f(a + t\delta, b + t\sigma)$ we have

$$\begin{aligned} \phi'(t) &= \nabla f \cdot \big\langle \delta, \sigma \big\rangle = \delta f_x(x_0 + t\delta, y_0 + t\sigma) + \sigma f_y(a + t\delta, b + t\sigma) \\ \phi''(t) &= \delta \nabla f_x \cdot \big\langle \delta, \sigma \big\rangle + \sigma \nabla f_y \cdot \big\langle \delta, \sigma \big\rangle \\ &= \delta(\delta f_{xx} + \sigma f_{xy}) + \sigma(\delta f_{xy} + \sigma f_{yy}). \end{aligned}$$

Since for t small

$$\phi(t) \approx \phi(0) + \phi'(0)t + \frac{1}{2}\phi''(0)t^2,$$

replacing $\phi(t)$ and its derivatives with the corresponding expressions in f we have

$$\begin{aligned} f(a + t\delta, b + \tau\epsilon) &\approx f(x_0,y_0) + f_x(x_0,y_0)\delta t + f_y(x_0,y_0)\sigma t \\ &+ \frac{1}{2}\big(f_{xx}(x_0,y_0)\delta^2 + f_{yy}(x_0,y_0)\sigma^2 + 2 f_{xx} f_{yy}(x_0,y_0)\delta\sigma\big)t^2. \end{aligned}$$

Replacing δt with $(x - x_0)$ and σt with $(y - y_0)$ gives us the result.

1.4 Sign of quadratic form

At the point (x_0, y_0) where $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ the RHS of (1) reduces to

$$\frac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + \frac{1}{2} f_{yy}(x_0,y_0)(y-y_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0).$$

Thus the change away from (x_0, y_0) of $f(x, y)$ is captured by

$$\frac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + \frac{1}{2} f_{yy}(x_0,y_0)(y-y_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0).$$

To study its sign, we simplify the expression to considering

$$ax^2 + 2cxy + by^2.$$

(The a, b, c here represent f_{xx}, f_{yy}, f_{xy} respectively. You can think of it as analyzing the expression at $(x_0, y_0) = (0, 0)$.) We first suppose $a \neq 0$ then completing the square gives

$$ax^2 + 2cxy + by^2 = a(x + \frac{c}{a}y)^2 + \frac{ba - c^2}{a}y^2.$$

Thus it is easy to see that if $a > 0$ and $ba - c^2 > 0$ then the form is always positive (corresponding to (x_0, y_0) as a min), as well as if $a < 0$ and $ba - c^2 > 0$ then the form is always negative (corresponding to (x_0, y_0) as a max).

If $ba - c^2 < 0$ then the two expressions a and $\frac{ba-c^2}{a}$ always have opposite signs. Thus the quadratic expression can be positive or negative.

When $ba - c^2 = 0$ then the test is inconclusive. The reason is a bit subtle. First if $a = 0$ then it is clear that the whole expression $ax^2 + 2cxy + by^2 = 0$. And it may seem that if $ba - c^2 = 0$ and $a > 0$ then the quadratic expression is positive so

Remarks:

The equation (1) can be expressed compactly in matrix form:

$$[x - a, y - b] \begin{bmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{xy}(x_0, y_0) & f_{yy}(x_0, y_0) \end{bmatrix} \begin{bmatrix} x - a \\ y - b \end{bmatrix}.$$

This is called a quadratic form with the symmetric matrix

$$A = \begin{bmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{xy}(x_0, y_0) & f_{yy}(x_0, y_0) \end{bmatrix}.$$

You can intuitively believe that the quadratic form's sign depends on the determinant of the matrix A and the sign of its $(1, 1)$ position. For example the quadratic forms $x^2 + y^2$ is non-negative and is associated with the matrix

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that the determinant of A is $1 > 0$ and $A_{11} = 1 > 0$. On the other hand the quadratic form $-x^2 - y^2$ is non-positive and is associated with the matrix

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Note that the determinant of A is $1 > 0$ and $A_{11} = -1 > 0$. We mention this representation because quadratic form is more general than the proof we gave above for the sign of quadractic expression. One can generally study the expression $\mathbf{u}^T A\mathbf{u}$ for a vector $\mathbf{u}$ and a symmeric matrix A.

1.5 Examples of inconclusive second derivative test

Consider

$$\begin{aligned} f^1(x,y) &= x^2 + y^4 \\ f^2(x,y) &= -x^2 - y^4 \\ f^3(x,y) &= x^2 - y^4. \end{aligned}$$

You can verify that in all of these cases

$$\begin{aligned} f^i_{xx}(0,0) &= 2 \\ f^i_{xy}(0,0) &= f^i_{yx}(0,0) = f^i_{yy}(0,0) = 0, \quad i = 1,2,3. \end{aligned}$$

Thus the second order derivative test is inconclusive here. Yet you can see that at $(0,0)$, f^1 has a min, f^2 has a max and f^3 has a saddle point.

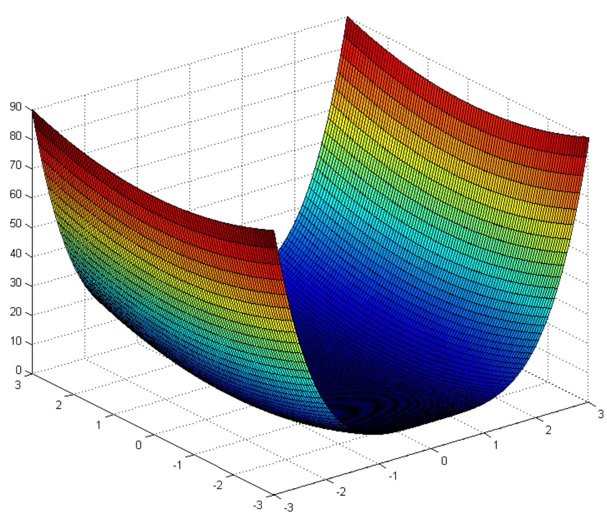

Figure 1.2: The function $x^2 + y^4$

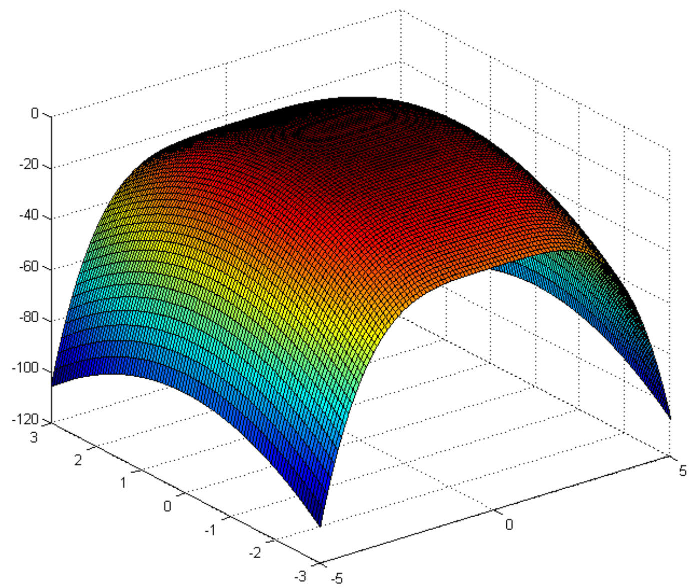

Figure 1.3: The function $-x^2 - y^4$

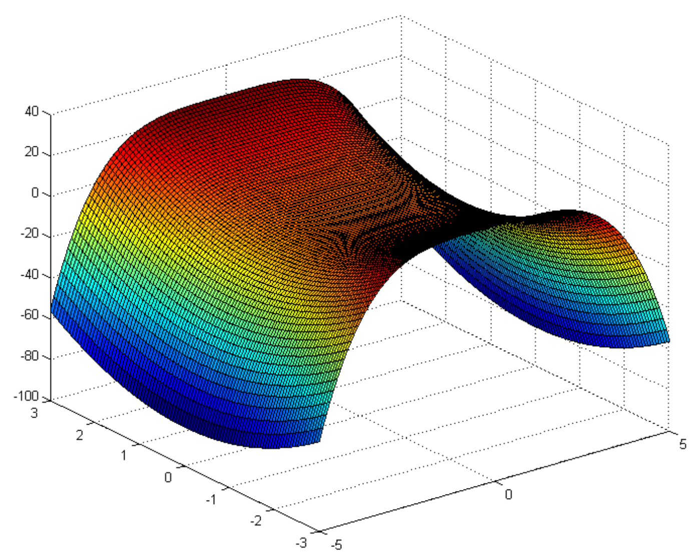

Figure 1.4: The function $x^2 - y^4$

2 Global extrema

If we consider $f(x, y)$ over the whole plane $\mathbb{R}^2$ then f may or may not have a global max or a global min. For example $f(x, y) = x^2 + y^2$ has a global min at $(0, 0)$ but has no global max. On the other hand, if we restrict $f(x, y)$ to a compact region R (say the disk of radius 1 around $(0, 0)$) then f is guaranteed to have a global max and a global min. Global here is with respect to the region R we are discussing.

We outline the steps to find the global max and min over a region R:

a. Find all critical points of f (that is points (x, y) such that $f_x(x, y) = f_y(x, y) = 0$) inside the region R.

b. Consider f on the boundary of the region R (which is a closed curve that contains the region R inside). This reduces f to a one dimensional equation. Find all critical points of f on the boundary of the region R.

c. If the boundary is not smooth (like the boundary of the rectangle $(1, 0), (0, 1), (-1, 0), (0, -1)$) then also consider the corner points of the boundary.

d. Compare the values of f on all of the points found in steps a to d. You will find the global min and the global max of f among them.