Extreme Values for Functions of Two Variables

James K. Peterson

Department of Biological Sciences and Department of Mathematical Sciences
Clemson University

November 18, 2016

Outline

- To understand how to think about finding places where the minimum and maximum of a function to two variables might occur, all you have to do is realize it is a common sense thing.
- We already know that the tangent plane attached to the surface which represents our function of two variables is a way to approximate the function near the point of attachment. We have seen in our pictures what happens when the tangent plane is **flat**. This flatness occurs at the minimum and maximum of the function.
- It also occurs in other situations, but we will leave that more complicated event for other courses. The functions we want to deal with are quite nice and have great minima and maxima. However, we do want you to know there are more things in the world and we will touch on them only briefly.

- To see what to do, just recall the equation of the tangent plane error to our function of two variables $f(x, y)$.

$$f(x,y) = f(x_0,y_0) + \nabla(f)(x_0,y_0)[x-x_0, y-y_0]^T$$
$$+(1/2)[x-x_0, y-y_0]H(x_0+c(x-x_0), y_0+c(y-y_0))[x-x_0, y-y_0]^T$$

where c is some number between 0 and 1 that is different for each x.

- We also know that the equation of the tangent plane to $f(x, y)$ at the point $x_0, y_0)$ is

$$z \quad = \quad f(x_0, y_0) + < \boldsymbol{\nabla} f(x_0, y_0), \boldsymbol{X} - \boldsymbol{X_0} > .$$

where $\boldsymbol{X} - \boldsymbol{X_0} = [x - x_0 . y - y_0]^T$.

- Now let's assume the tangent plane is flat at (x_0, y_0). Then the gradient $\boldsymbol{\nabla} f(x_0, y_0)$ is the zero vector and we have $\frac{\partial f}{\partial x}(x_0, y_0) = 0$ and $\frac{\partial f}{\partial y}(x_0, y_0) = 0$. So the tangent plane error equation simplifies to

$$f(x,y) = f(x_0,y_0)$$
$$+(1/2)[x-x_0, y-y_0]H(x_0+c(x-x_0), y_0+c(y-y_0))[x-x_0, y-y_0]^T$$

- Now let's simplify this. The Hessian is just a 2×2 matrix whose components are the second order partials of f. Let

$$\begin{aligned} A(c) &= \frac{\partial^2 f}{\partial x^2}(x_0 + c(x - x_0), y_0 + c(y - y_0)) \\ B(c) &= \frac{\partial^2 f}{\partial x\, \partial y}(x_0 + c(x - x_0), y_0 + c(y - y_0)) \\ &= \frac{\partial^2 f}{\partial y\, \partial x}(x_0 + c(x - x_0), y_0 + c(y - y_0)) \\ D(c) &= \frac{\partial^2 f}{\partial y^2}(x_0 + c(x - x_0), y_0 + c(y - y_0)) \end{aligned}$$

- Then, we have

$$f(x, y) = f(x_0, y_0) + (1/2) \begin{bmatrix} x - x_0 & y - y_0 \end{bmatrix} \begin{bmatrix} A(c) & B(c) \\ B(c) & D(c) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}$$

- We can multiply this out (a nice simple pencil and paper exercise!) to find

$$\begin{aligned} &f(x, y) = f(x_0, y_0) \\ &+1/2 \left(A(c)(x - x_0)^2 + 2B(c)(x - x_0)(y - y_0) + D(c)(y - y_0)^2 \right) \end{aligned}$$

- Now it is time to remember an old technique from high school – completing the square. Remember if we had a quadratic like $u^2 + 3uv + 6v^2$, to complete the square we take half of the number in front of the mixed term uv and square it and add and subtract it times v^2 as follows.

$$u^2 + 3uv + 6v^2 = u^2 + 3uv + (3/2)^2 v^2 - (3/2)^2 v^2 + 6v^2.$$

- Now group the first three terms together and combine the last two terms into one term.

$$u^2 + 3uv + 6v^2 = \left(u^2 + 3uv + (3/2)^2 v^2 \right) + \left(6 - (3/2)^2 \right) v^2.$$

- The first three terms are a *perfect square*, $(u + (3/2)v)^2$. Simplifying, we find

$$u^2 + 3uv + 6v^2 \quad = \quad \left(u + (3/2)v\right)^2 + (135/4)\, v^2.$$

- This is called *completely the square*! Now let's do this with the Hessian quadratic we have. First, factor our the $A(c)$. We will assume it is not zero so the divisions are fine to do. Also, for convenience, we will replace $x - x_0$ by Δx and $y - y_0$ by Δy This gives

$$\begin{aligned} &f(x, y) = f(x_0, y_0) \\ &+ \frac{A(c)}{2} \left((\Delta x)^2 + 2\frac{B(c)}{A(c)} \Delta x \, \Delta y + \frac{D(c)}{A(c)} (\Delta y)^2 \right). \end{aligned}$$

- One half of the $\Delta x \Delta y$ coefficient is $\frac{B(c)}{A(c)}$ so add and subtract $(B(c)/A(c))^2 (\Delta y)^2$. We find

$$\begin{aligned} &f(x, y) = f(x_0, y_0) \\ &+ \frac{A(c)}{2} \times \left((\Delta x)^2 + 2\frac{B(c)}{A(c)} \Delta x \Delta y + \left(\frac{B(c)}{A(c)}\right)^2 (\Delta y)^2 \right) \\ &+ \frac{A(c)}{2} \times \left(-\left(\frac{B(c)}{A(c)}\right)^2 (\Delta y)^2 + \frac{D(c)}{A(c)} (\Delta y)^2 \right). \end{aligned}$$

- Now group the first three terms together – the perfect square and combine the last two terms into one. We have

$$\begin{aligned} &f(x, y) = f(x_0, y_0) \\ &+ \frac{A(c)}{2} \left(\left(\Delta x + \frac{B(c)}{A(c)} \Delta y \right)^2 + \left(\frac{A(c)\, D(c) - (B(c))^2}{(A(c))^2} \right) (\Delta y)^2 \right). \end{aligned}$$

Now we know if a function g is continuous at a point (x_0, y_0) and positive or negative , then it is positive or negative in a circle of radius r centered at (x_0, y_0). Here is the formal statement.

Theorem

(Nonzero Values and Continuity)
If $f(x_0, y_0)$ is a place where the function is positive or negative in value, then there is a radius r so that $f(x, y)$ is positive or negative in a circle of radius r around the center (x_0, y_0).

Proof

We argue for the case $g(x_0, y_0) > 0$. Let $\epsilon = g(x_0, y_0)/2$. Since g is continuous at (x_0, y_0), there is a $\delta > 0$ so that $\sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$ implies $-g(x_0, y_0)/2 < g(x, y) - g(x_0, y_0) < g(x_0, y_0)/2$. Thus, $g(x, y) > g(x_0, y_0)/2$ if $(x, y) \in B_\delta(x_0, y_0)$. $\square$

- Now getting back to our problem. We don't want to look at functions that are locally constant as they are not very interesting. So we are assuming $f(x, y)$, in addition to being differentiable, is not locally constant. We have at this point where the partials are zero, the following expansion

$$f(x, y) = f(x_0, y_0) + \frac{A(c)}{2}\left(\left(\Delta x + \frac{B(c)}{A(c)}\Delta y\right)^2 + \left(\frac{A(c)\,D(c) - (B(c))^2}{(A(c))^2}\right)(\Delta y)^2\right).$$

- The algebraic sign of the terms after the function value $f(x_0, y_0)$ are completely determined by the terms which are not squared. We have two simple cases:
  - $A(c) > 0$ and $A(c)\,D(c) - (B(c))^2 > 0$ which implies the term after $f(x_0, y_0)$ is positive.
  - $A(c) < 0$ and $A(c)\,D(c) - (B(c))^2 > 0$ which implies the term after $f(x_0, y_0)$ is negative.

- Now let's assume all the second order partials are continuous at (x_0, y_0). We know $A(c) = \frac{\partial^2 f}{\partial x^2}(x_0 + c(x - x_0), y_0 + c(y - y_0))$ and from our theorem, if $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0$, then so is $A(c)$ in a circle around (x_0, y_0).
- The other term $A(c)\, D(c) - (B(c))^2 > 0$ will also be positive is a circle around (x_0, y_0) as long as $\frac{\partial^2 f}{\partial x^2}(x_0, y_0)\, \frac{\partial^2 f}{\partial y^2}(x_0, y_0) - \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) > 0$. We can say similar things about the negative case.
- Now to save typing let $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) = f^0_{xx}$, $\frac{\partial^2 f}{\partial y^2}(x_0, y_0) = f^0_{yy}$ and $\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) = f^0_{xy}$. So we can restate our two cases as
    - $f^0_{xx} > 0$ and $f^0_{xx}\, f^0_{yy} - (f^0_{xy})^2 > 0$ which implies the term after $f(x_0, y_0)$ is positive. This implies that $f(x, y) > f(x_0, y_0)$ in a circle of some radius r which says $f(x_0, y_0)$ is a minimum value of the function locally at that point.
    - $f^0_{xx} < 0$ and $f^0_{xx}\, f^0_{yy} - (f^0_{xy})^2 > 0$ which implies the term after $f(x_0, y_0)$ is negative. This implies that $f(x, y) < f(x_0, y_0)$ in a circle of some radius r which says $f(x_0, y_0)$ is a maximum value of the function locally at that point.

So we have come up with a great condition to verify if a place where the partials are zero is a minimum or a maximum.

Theorem

Extrema Test

If the partials of f are zero at the point (x_0, y_0), we can determine if that point is a local minimum or local maximum of f using a second order test. We must assume the second order partials are continuous at the point (x_0, y_0).

- *If $f^0_{xx} > 0$ and $f^0_{xx}\, f^0_{yy} - (f^0_{xy})^2 > 0$ then $f(x_0, y_0)$ is a local minimum.*
- *If $f^0_{xx} < 0$ and $f^0_{xx}\, f^0_{yy} - (f^0_{xy})^2 > 0$ then $f(x_0, y_0)$ is a local maximum.*

We just don't know anything if the test $f^0_{xx}\, f^0_{yy} - (f^0_{xy})^2 = 0$.

Recall the definition of the determinant of a 2×2 matrix:

$$\boldsymbol{A} \;=\; \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Longrightarrow det(\boldsymbol{A}) = ad - bc.$$

So, since we assume the mixed partials match

$$\begin{aligned} \boldsymbol{H}(x,y) &= \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix} \Longrightarrow \\ det(\boldsymbol{H}(x,y)) &= f_{xx}(x,y)\, f_{yy}(x,y) - (f_{xy}(x,y))^2 \end{aligned}$$

and at a critical point (x_0, y_0) where $\nabla f(x_0, y_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ we have

$$det(\boldsymbol{H}(x_0, y_0)) \;=\; f_{xx}(x_0, y_0)\, f_{yy}(x_0, y_0) - (f_{xy}(x_0, y_0))^2$$

Using our usual shorthand notations, we would write this as $det(\boldsymbol{H}^0) = f^0_{xx} f^0_{yy} - (f^0_{xy})^2$.

We can rewrite out second order test for extremal values using this determinant idea:

Theorem

If the partials of f are zero at the point (x_0, y_0), we can determine if that point is a local minimum or local maximum of f using a second order test. We must assume the second order partials are continuous at the point (x_0, y_0).

- *If $f^0_{xx} > 0$ and $det(\boldsymbol{H}^0) > 0$ then $f(x_0, y_0)$ is a local minimum.*
- *If $f^0_{xx} < 0$ and $det(\boldsymbol{H}^0) > 0$ then $f(x_0, y_0)$ is a local maximum.*

We just don't know anything if the test $det(\boldsymbol{H}^0) = 0$.

Proof

We have shown this argument. $\square$

Recall at a critical point (x_0, y_0), we found that

$$\begin{aligned} f(x,y) &= f(x_0,y_0) \\ &+ \frac{A(c)}{2}\left(\left(\Delta x + \frac{B(c)}{A(c)}\Delta y\right)^2 + \left(\frac{A(c)\,D(c) - (B(c))^2}{(A(c))^2}\right)(\Delta y)^2\right). \end{aligned}$$

And we have been assuming $A(c) \neq 0$ here. Now suppose we knew $A(c)\,D(c) - (B(c))^2 < 0$. Then, using the usual continuity argument, we know that there is a circle around the critical point (x_0, y_0) so that $A(c)\,D(c) - (B(c))^2 < 0$ when $c = 0$. This is the same as saying $\det(\boldsymbol{H}(x_0, y_0)) < 0$. But notice that on the line going through the critical point having $\Delta y = 0$, this gives

$$f(x,y) = f(x_0,y_0) + \frac{A(c)}{2}\left(\Delta x\right)^2.$$

and on the line through the critical point with $\Delta x + \frac{B(c)}{A(c)}\Delta y = 0$. we have

$$f(x,y) = f(x_0,y_0) + \frac{A(c)}{2}\left(\frac{A(c)\,D(c) - (B(c))^2}{(A(c))^2}\right)(\Delta y)^2$$

Now, if $A(c) > 0$, the first case gives $f(x,y) = f(x_0,y_0) +$ **a positive number** showing f has a minimum on that trace.

However, the second case gives $f(x,y) = f(x_0,y_0) -$ **a positive number** which shows f has a maximum on that trace.

The fact that f is minimized in one direction and maximized in another direction gives rise to the expression that we consider f to behave like a saddle at this critical point.

The analysis is virtually the same if $A(c) < 0$, except the first trace has the maximum and the second trace has the minimum. Hence, the test for a saddle point is to see if $\det(\boldsymbol{H}(x_0, y_0)) < 0$.

If $A(c) = 0$, we have to argue differently.

We are in the case where $\det(\boldsymbol{H}(x_0, y_0)) < 0$ which we know means we can assume $A(c)D(c) - (B(c))^2 < 0$ also.l If $A(c) = 0$, we must $B(c) \neq 0$. We thus have

$$\begin{aligned} f(x, y) &= f(x_0, y_0) + B(c)\Delta x\, \Delta y + (1/2)D(c)(\Delta y)^2 \\ &= f(x_0, y_0) + (1/2)\Big(2B(c)\Delta x + D(c)\Delta y)\Big)\Delta y \end{aligned}$$

If $D(c) = 0$, $f(x, y) = f(x_0, y_0) + B(c)\Delta x\, \Delta y$ and choosing the paths $\Delta x = \pm\Delta y$, we have $f(x, y) = f(x_0.y_0) \pm B(c)(\Delta y)^2$ which tell us we have a minimum on one path and a maximum on the other path; i.e. this is a saddle.

If $D(c) \neq 0$, since

$$f(x, y) = f(x_0, y_0) + (1/2)\Big(2B(c)\Delta x + D(c)\Delta y)\Big)\Delta y$$

we can choose paths $2B(c)\Delta x = D(c)\Delta y$ and $2B(c)\Delta x = -3D(c)\Delta y$ to give $f(x, y) = f(x_0, y_0) + D(c)(\Delta y)^2$ or
$f(x, y) = f(x_0, y_0) - D(c)(\Delta y)^2$ and again, on one path we have a minimum and on the other a maximum implying a saddle.

Now the second order test fails if $\det(\boldsymbol{H}(x_0, y_0)) = 0$ at the critical point as in that case, the surface can have a minimum, maximum or saddle.

- $f(x, y) = x^4 + y^4$ has a global minimum at $(0, 0)$ but at that point

  $$\boldsymbol{H}(x, y) = \begin{bmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{bmatrix} \Longrightarrow \det(\boldsymbol{H}(x_0, y_0)) = 144x^2y^2.$$

  and hence, $\det(\boldsymbol{H}(x_0, y_0)) = 0$.

- $f(x, y) = -x^4 - y^4$ has a global maximum at $(0, 0)$ but at that point

  $$\boldsymbol{H}(x, y) = \begin{bmatrix} -12x^2 & 0 \\ 0 & -12y^2 \end{bmatrix} \Longrightarrow \det(\boldsymbol{H}(x_0, y_0)) = 144x^2y^2.$$

  and hence, $\det(\boldsymbol{H}(x_0, y_0)) = 0$ as well.

- Finally, $f(x, y) = x^4 - y^4$ has a saddle at $(0, 0)$ but at that point

  $$\boldsymbol{H}(x, y) = \begin{bmatrix} 12x^2 & 0 \\ 0 & -12y^2 \end{bmatrix} \Longrightarrow \det(\boldsymbol{H}(x_0, y_0)) = -144x^2y^2.$$

  and hence, $\det(\boldsymbol{H}(x_0, y_0)) = 0$ again.

Hence, since we have covered all the cases:

Theorem

If the partials of f are zero at the point (x_0, y_0), we can determine if that point is a local minimum or local maximum of f using a second order test. We must assume the second order partials are continuous at the point (x_0, y_0).

- *If $f^0_{xx} > 0$ and $\det(\boldsymbol{H}^0) > 0$ then $f(x_0, y_0)$ is a local minimum.*
- *If $f^0_{xx} < 0$ and $\det(\boldsymbol{H}^0) > 0$ then $f(x_0, y_0)$ is a local maximum.*
- *If $\det(\boldsymbol{H}^0) < 0$, then $f(x_0, y_0)$ is a local saddle.*

We just don't know anything if the test $\det(\boldsymbol{H}^0) = 0$.

Proof

We have shown this argument. □

Example

Use our tests to show $f(x, y) = x^2 + 3y^2$ has a minimum at $(0, 0)$.

Solution

The partials here are $f_x = 2x$ and $f_y = 6y$. These are zero at $x = 0$ and $y = 0$. The Hessian at this critical point is

$$H(x, y) \;=\; \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} = H(0, 0).$$

as H is constant here. Our second order test says the point $(0, 0)$ corresponds to a minimum because $f_{xx}(0, 0) = 2 > 0$ and $f_{xx}(0, 0)\, f_{yy}(0, 0) - (f_{xy}(0, 0))^2 = 12 > 0$.

Example

Use our tests to show $f(x, y) = x^2 + 6xy + 3y^2$ has a saddle at $(0, 0)$.

Solution

The partials here are $f_x = 2x + 6y$ and $f_y = 6x + 6y$. These are zero at when $2x + 6y = 0$ and $6x + 6y = 0$ which has solution $x = 0$ and $y = 0$. The Hessian at this critical point is

$$H(x, y) = \begin{bmatrix} 2 & 6 \\ 6 & 6 \end{bmatrix} = H(0, 0).$$

as H is again constant here. Our second order test says the point $(0, 0)$ corresponds to a saddle because $f_{xx}(0, 0) = 2 > 0$ and $f_{xx}(0, 0)\, f_{yy}(0, 0) - (f_{xy}(0, 0))^2 = 12 - 36 < 0$.

Example

Show our tests fail on $f(x, y) = 2x^4 + 4y^6$ even though we know there is a minimum value at $(0, 0)$.

Solution

For $f(x, y) = 2x^4 + 4y^6$, you find that the critical point is $(0, 0)$ and all the second order partials are 0 there. So all the tests fail. Of course, a little common sense tells you $(0, 0)$ is indeed the place where this function has a minimum value. Just think about how it's surface looks. But the tests just fail. This is much like the curve $f(x) = x^4$ which has a minimum at $x = 0$ but all the tests fail on it also.

Example

Show our tests fail on $f(x,y) = 2x^2 + 4y^3$ and the surface does not have a minimum or maximum at the critical point $(0,0)$.

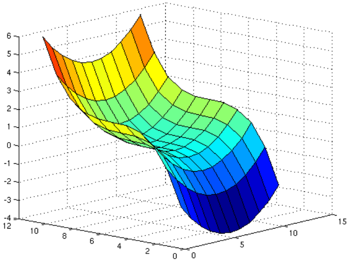
Solution

For $f(x,y) = 2x^2 + 4y^3$, the critical point is again $(0,0)$ and $f_{xx}(0,0) = 4$, $f_{yy}(0,0) = 0$ and $f_{xy}(0,0) = f_{yx}(0,0) = 0$. So $f_{xx}(0,0)\, f_{yy}(0,0) - (f_{xy}(0,0))^2 = 0$ so the test fails. Note the $x = 0$ trace is $4y^3$ which is a cubic and so is negative below $y = 0$ and positive above $y = 0$. Not much like a minimum or maximum behavior on this trace! But the trace for $y = 0$ is $2x^2$ which is a nice parabola which does reach its minimum at $x = 0$. So the behavior of the surface around $(0,0)$ is not a maximum or a minimum. The surface acts a lot like a cubic. Do this in MatLab

Solution

```
>> [X,Y] = meshgrid(-1:.2:1);
>> Z = 2*X.^2 + 4*Y.^3;
>> surf(Z);
```

This will give you the surface. In the plot that is shown go to the tool menu and click of the rotate 3D option and you can spin it around. Clearly like a cubic! You can see the plot in the next slide.



Homework 38

38.1 This is a review of some ideas from statistics. Let $\{X_1, \ldots, X_n\}$ be some data for $N >> 1$. The average of this set of data is $\overline{X} = (1/N) \sum_{i=1}^{N} x_i$. The average of the squares of the data is $\overline{X^2} = (1/N) \sum_{i=1}^{N} x_i^2$. Prove $0 \leq \sum_{i=1}^{N} (x_i - \overline{X})^2 = N\overline{X^2} - N(\overline{X})^2$

38.2 Given data pairs $\{(X_1, Y_1), \ldots, (X_N, Y_N)\}$, the line of regression through this data is the line $y = mx + b$ which minimizes the error function $E(m, b) = \sum_{i=1}^{N} (Y_I - mX_i - b)^2$. Find the slope and intercept (m^*, b^*) which is a critical point for this minimization. The formulae you derive here for m^* and b^* give the optimal slope and intercept for the line of regression that best fits this data. However, the proof of this requires the next problem.

38.3 Use the second order theory for the minimization of a function of two variables to show that the error is a global minimum at the critical point (m^*, b^*).