Second Derivative Test

This homework set takes you through a proof of the second derivative test for functions of two variables. It will also give you an idea of what a test for functions of three and more variables might look like.

The main tool that we will need is a version of Taylor's theorem for functions of several variables. We start with the theorem in the one variable setting.

1. Taylor's Theorem with the Lagrange form of the remainder. We start with some Calc III material; namely, Taylor's theorem with the integral form of the remainder.

Taylor: If $f^{(n+1)}$ is continuous on an open interval I about the point a, and if x is a point of I, then

$$f(x) \;=\; f(a) + f'(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n$$

where the remainder term is given by the integral

$$R_n = \int_a^x \frac{f^{(n+1)}(t)(x-t)^n}{n!}\,dt.$$

You can find a statement of this in your book. You can also see a well motivated proof of this fact in *Extra Hwk 3* from Math 2423–001, Honors Calculus II in Spring 2008.

In this question you will prove that the following is also an expression for the remainder

$$R_n \;=\; \frac{f^{(n+1)}(\theta)(x-a)^{n+1}}{(n+1)!} \qquad \text{for some } \theta \text{ between } a \text{ and } x.$$

This is called the Lagrange form of the remainder. We prove this in several steps.

(a) First prove the Weighted Mean Value Theorem for integrals. If f and g are continuous on an interval $[a, b]$ and g does not change sign in $[a, b]$, then there exists a number θ in $[a, b]$ so that

$$\int_a^b f(x)g(x)\,dx \;=\; f(\theta)\int_a^b g(x)\,dx.$$

A sketch of the proof is given below. Supply details and write the argument out in full.

(i) Use the Extreme Value Theorem (write down its statement) and continuity of f to argue that there exist numbers m and M so that

$$m \;\le\; f(x) \;\le\; M \qquad \text{for all } x \text{ in } [a, b].$$

(ii) Now multiply this inequality across by $g(x)$ (you will need to use the hypothesis that g doesn't change sign on $[a, b]$ here. Why?), integrate over $[a, b]$, and simplify to get

$$m \;\le\; \frac{\int_a^b f(x)g(x)\,dx}{\int_a^b g(x)\,dx} \;\le\; M$$

There may be two cases to consider. Why? Also, what happens if $\int_a^b g(x)\,dx = 0$?

(iii) Now use continuity of f and the Intermediate Value Theorem (write out its statement), to conclude that there is a number θ in $[a, b]$ so that the ratio of integrals in the previous step is equal to $f(\theta)$. Now we're done. Why?

(b) Now we will use the Weighted Mean Value Theorem to obtain the Lagrange form of the remainder from the integral form of the remainder above. Take $g(t) = \frac{(x-t)^n}{n!}$. Are all the hypotheses of the Weighted Mean Value Theorem satisfied? What is the conclusion? Now finish the proof of the Lagrange form of the remainder.

2. Taylor's Theorem in several variables. If $f(x,y)$ and all its partial derivatives up to order $(n+1)$ are continuous on an open disk around the point (a,b), then for $(a+h,b+k)$ in this disk one has

$$\begin{aligned} f(a+h,b+k) &= f(a,b) + \left(\frac{\partial f}{\partial x}h + \frac{\partial f}{\partial y}k\right) + \frac{1}{2!}\left(\frac{\partial^2 f}{\partial x^2}h^2 + 2\frac{\partial^2 f}{\partial x \partial y}hk + \frac{\partial^2 f}{\partial y^2}k^2\right) + \cdots \\ &\cdots + \frac{1}{n!}\left(\frac{\partial^n f}{\partial x^n}h^n + n\frac{\partial^n f}{\partial x^{n-1}\partial y}h^{n-1}k + \cdots + n\frac{\partial^n f}{\partial x \partial y^{n-1}}hk^{n-1} + \frac{\partial^n f}{\partial y^n}k^n\right) + R_n \end{aligned}$$

where all the partial derivatives are evaluated at (a,b), and the remainder R_n has a Lagrange form

$$R_n = \frac{1}{(n+1)!}\left(\frac{\partial^{n+1} f}{\partial x^{n+1}}h^{n+1} + (n+1)\frac{\partial^{n+1} f}{\partial x^n \partial y}h^n k + \cdots + (n+1)\frac{\partial^{n+1} f}{\partial x \partial y^n}hk^n + \frac{\partial^{n+1} f}{\partial y^{n+1}}k^{n+1}\right)$$

In the remainder term all the partial derivatives are evaluated at some point (c,d) on the line segment from (a,b) to $(a+h,b+k)$.

This can be written more succinctly using D notation. Let D be the partial differential operator

$$D = \left(\frac{\partial}{\partial x}h + \frac{\partial}{\partial y}k\right)$$

and let D^m mean the result of applying D m times to a function. With this notation, the expression in Taylor's theorem for functions of two variables becomes

$$f(a+h,b+k) = f(a,b) + Df(a,b) + \frac{1}{2!}D^2 f(a,b) + \cdots + \frac{1}{n!}D^n f(a,b) + R_n$$

where

$$R_n = \frac{1}{(n+1)!}D^{n+1}f(c,d) \qquad \text{for some } (c,d) \text{ on the line segment from } (a,b) \text{ to } (a+h,b+k).$$

In this question you are asked to give a proof of Taylor's theorem. Use the following hint/outline.

(a) Let $(x(t),y(t)) = (a+th, b+tk)$ be a parameterization of the line segment from (a,b) to $(a+h,b+k)$, and define a function of one variable by $F(t) = f(x(t),y(t))$.
(b) Use Taylor's theorem with the Lagrange form of the remainder from Question 1 above for the function $F(t)$ about the point $t=0$.
(c) Now use the chain rule from earlier this semester to write out the successive derivatives $F'(0)$, $F''(0)$ etc in terms of partial derivatives of f. What expression do you get?

3. The second derivative test in the case of a polynomial. Consider the polynomial function

$$P(x,y) = \frac{1}{2}(Ax^2 + 2Bxy + Cy^2) + Dx + Ey + F.$$

Note that the F term just shifts the graph up or down, and so wont affect critical points. We may assume that $F=0$.

(a) Show that $P_x(0,0) = D$ and that $P_y(0,0) = E$. If $D = 0 = E$, then the origin is a critical point.
(b) So now we are investigating the behavior of the function

$$P(x,y) = \frac{1}{2}(Ax^2 + 2Bxy + Cy^2)$$

at the origin. Verify that $P_{xx}(0,0) = A$, $P_{xy}(0,0) = B$, and $P_{yy}(0,0) = C$.
(c) Use the "completing the square" technique to see that $P(x,y)$ can be rewritten as

$$2P(x,y) = A\left[\left(x + \frac{B}{A}y\right)^2 + \frac{AC-B^2}{A^2}y^2\right]$$

(d) If $AC - B^2 < 0$, show that $P(x,y)$ will have opposite signs along two different lines through the origin. (Which two lines did you consider?) This is saddle behavior.
(e) If $AC - B^2 > 0$, show that the sign of $P(x,y)$ is the same as the sign of A (except at $(0,0)$ of course). This tells you that the $(0,0)$ is a local max if $A < 0$, and that it is a local min if $A > 0$.

(f) We are done with this step, but there is one neat inequality that will be important in question 4 below. Show that, if $A > 0$ and $AC - B^2 > 0$, then $A + C > 0$ and

$$2P(x, y) \;\geq\; \frac{AC - B^2}{A + C}(x^2 + y^2).$$

Hint: Write $u = x + By/A$ and $v = \sqrt{AC - B^2}y/A$. Note that the equation in (c) above is now just $2P = A[u^2 + v^2]$. Solve for x and y in terms of u and v. Thinking of these expressions as dot products of certain vectors with the vector $\langle u, v\rangle$, show that

$$x^2 + y^2 \leq \frac{A(A + C)}{AC - B^2}(u^2 + v^2)$$

and hence deduce the inequality asked for above.

4. Taylor reduces the second derivative test to consideration of a polynomial. Here's the idea of this step. From Taylor's theorem we know that

$$f(a + h, b + k) \;=\; f(a, b) + (f_x(a, b)h + f_y(a, b)k) + \frac{1}{2!}\left(f_{xx}(a, b)h^2 + 2f_{xy}(a, b)hk + f_{yy}(a, b)k^2\right) + R_2.$$

We can assume $f(a, b) = 0$ since it only shifts graph up or down, and so wont affect critical point behavior. Also, since (a, b) is a critical point, we have that $f_x(a.b) = 0 = f_y(a, b)$. Thus we have

$$f(a + h, b + k) \;=\; \frac{1}{2}(f_{xx}(a, b)h^2 + 2f_{xy}(a, b)hk + f_{yy}(a, b)k^2) + R_2.$$

So we could say that we are done, by question 3, **provided that we can ignore the effects of R_2.** Indeed, if we can safely ignore R_2, then the $AC - B^2 < 0$ criterion for saddle behavior is exactly $f_{xx}f_{yy} - f_{xy}^2 < 0$. Also the $AC - B^2 > 0$ and $A > 0$ for a local min (respectively $A < 0$ for a local max) is exactly $f_{xx}f_{yy} - f_{xy}^2 > 0$ and $f_{xx} > 0$ (respectively $f_{xx} < 0$).

What we need to say precisely is that we can ignore R_2. This requires a little thought, since both $\frac{1}{2}(f_{xx}(a, b)h^2 + 2f_{xy}(a, b)hk + f_{yy}(a, b)k^2)$ and R_2 are tending to 0 as $(h, k) \to (0, 0)$. We really want to say that in some small neighborhood of (a, b), the values of $|R_2|$ are less than (say half) of those of the $|\frac{1}{2}(f_{xx}(a, b)h^2 + 2f_{xy}(a, b)hk + f_{yy}(a, b)k^2)|$ term, so that the sign of f is the same as that of the $\frac{1}{2}(f_{xx}(a, b)h^2 + 2f_{xy}(a, b)hk + f_{yy}(a, b)k^2)$ term. Let's write $f_{xx}(a, b) = A$, $f_{xy}(a, b) = B$ and $f_{yy}(a, b) = C$, so that

$$f(a + h, b + k) \;=\; P(h, k) + R_2$$

in the notation of question 3.

Recall that

$$R_2 \;=\; \frac{1}{3!}(f_{xxx}(c, d)h^3 + 3f_{xxy}(c, d)h^2k + 3f_{xyy}(c, d)hk^2 + f_{yyy}(c, d)k^3)$$

for some (c, d) on the segment between (a, b) and $(a + h, b + k)$.

(a) Suppose we know that the third partial derivatives of f are all continuous on a closed ϵ–ball about (a, b) for some $\epsilon > 0$. By the extreme value theorem (for functions of 2 variables) there is a number $M > 0$ so that all 3rd partial derivatives of f are bounded above in absolute value by M in this ϵ–ball about (a, b).

Now argue that $|R_2| \leq \frac{M}{2}(|h| + |k|)(h^2 + k^2)$ on this neighborhood.

(b) In the case that $A > 0$ and $AC - B^2 > 0$, we know from 3(f) that

$$P(h, k) \;\geq\; \frac{AC - B^2}{2(A + C)}(h^2 + k^2)$$

so this term will be much bigger than R_2 provided (h, k) is close enough to $(0, 0)$. Show that picking points within a ball of radius $\min\{\epsilon, \frac{AC-B^2}{10M(A+C)}\}$ about (a, b) for example guarantees this.

The case $AC - B^2 > 0$, $A < 0$ and the case $AC - B^2 < 0$ are handled in a similar fashion, but you've endured enough to believe that the "ignore R_2" strategy can be made very precise.

5. More variables. Finally, one could consider a function of 3–variables $f(x, y, z)$. Again there is a Taylor's theorem which will tell us that, at a critical point where $f(a, b, c) = 0$, we can write

$$f(a+h, b+k, c+l) \;=\; \frac{1}{2!}(f_{xx}h^2 + 2f_{xy}hk + 2f_{xz}hl + 2f_{yz}kl + f_{yy}k^2 + f_{zz}l^2) + R_2$$

where all the 2nd order partial derivatives are evaluated at (a, b, c) and the remainder R_2 is negligible in comparison to the quadratic term.

Thus we are led to consider critical points for the "quadratic form" function

$$Q(x, y, z) \;=\; Ax^2 + By^2 + Cz^2 + 2Dxy + 2Eyz + 2Fxz$$

One can complete the squares to get $Q(x, y, z) = A[(x + Dy/A + Ez/A)^2 + Q_2(y, z)]$ where $Q_2(y, z)$ is some quadratic form function in y and z. Further completing the squares in Q_2 gives an expression for $Q(x, y, z)$ which is easily seen to give positive outputs for all inputs $(x, y, z) \neq (0, 0, 0)$ provided that $A > 0$, $AB - D^2 > 0$ and $ABC - CD^2 - BE^2 - AF^2 + 2DEF > 0$. Do this!

These are seen to be the determinants of the successively larger matrices

$$(A)$$

and

$$\begin{pmatrix} A & D \\ D & B \end{pmatrix}$$

and

$$\begin{pmatrix} A & D & F \\ D & B & E \\ F & E & C \end{pmatrix}.$$

We met determinant language in the computation of cross products of vectors in Calculus III. You may find a discussion of quadratic form functions of n variables in a *linear algebra* book.