Notes for Section 3.3

Much like in single variable calculus, the derivatives of a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ can be used to find the local extrema and sometimes the absolute extrema of such a function. As is often the case, the ideas needed are similar, but somewhat more delicate than in single variable calculus. Start off by understanding the definitions of *local minimum*, *local maximum*, and *critical point*. They are analogous to the definitions in single variable calculus. New in multivariable calculus is the notion of a *saddle point*. The name comes from the shape of the graph of $f(x, y) = x^2 - y^2$ at the critical point $(0, 0)$. Graph this function in one of the multivariable graphing apps and you will understand why it is called saddle point. In fact, the usual definition of a saddle point is some kind of technical description of a point where the shape of the graph is similar to the graph of $f(x, y) = x^2 - y^2$ near $(0, 0)$, that is the section in some plane is a single variable function that has a local minimum at the critical point, while the section in another plane has a local maximum. Our textbook takes a much more general approach and declares any critical point that is neither a local minimum nor a local maximum a saddle point.

In single variable calculus, you learned that any *local extremum* (that is local minimum or maximum) must be a critical point. That is still true (Theorem 4) and the proof of this fact is straightforward enough that I do not think I need to add anything to it. Just like in single variable calculus, this can be used to hunt down the local extrema of a function. We would do this by finding all of the critical points, which can be done by looking for points where f is not differentiable and for points where $\mathbf{D}\, f = \nabla f = 0$. The latter gives a system of equations in n unknowns, which we may be able to solve. Once the critical points have been found, the graph of f can be studied at those points to decide which are local minima, which are local maxima, and which are neither. See Examples 1-4 for how this may work out in practice. Example 4 is quite interesting as it illustrates that a multivariable function can have a lot of local extrema.

The Second Derivative Test for determining which critical points may be local minima, which may be local maxima, and which are neither is the multivariable analog of the Second Derivative Test from single variable calculus: if x is a critical point of f where $f'(x) = 0$ then x is a local minimum if $f''(x) > 0$, x is a local maximum if $f''(x) < 0$, and the test is inconclusive if $f''(x) = 0$. The Second Derivative Test in multivariable calculus is much more complicated. First of all, the role of the second derivative f'' is taken over by the $n \times n$ matrix of all second partial derivatives called the *Hessian* matrix:

$$Hf = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

The idea is that if x_0 is a critical point of f where $\mathbf{D}\, f(x_0) = 0$, and the second partial derivatives exist, we can define a function $g : \mathbb{R}^n \to \mathbb{R}$ by

$$g(h) = \frac{1}{2} h Hf(x_0) h$$

where the first vector h is treated as a row vector and the second vector h is treated as a column vector so that the $hHf(x_0)h$ makes sense as a product of matrices. Then if g has the property that $g(h) \geq 0$ for all $h \in \mathbb{R}^n$, in which case we call g *positive-definite*, then x_0 must be a local minimum of f. On the other hand, if g has the property that $g(h) \leq 0$ for all $h \in \mathbb{R}^n$, in which case we call g *negative-definite*, then x_0 must be a local maximum of f. The number $1/2$ in the definition of g may not seem to make any sense as it really does not affect whether g is positive-definite or negative-definite. It is there for a different reason and has to do with the generalization of Taylor series to multivariable functions. Since we skipped Section 3.2 on Taylor's Theorem, don't spend much time thinking about this detail. In fact, the proof of the Second Derivative Test also refers to Taylor's Theorem. Read it lightly and rather than worrying about what this mysterious Taylor's Theorem might be, view the expression

$$f(x_0 + h) - f(x_0) = g(h) + R_2(x_0, h)$$

with $g(x) = \frac{1}{2}hHf(x_0)h$ as a generalization of the linear approximation

$$l(x) = \mathbf{D}\, f(x_0)(x - x_0) + f(x_0)$$

we talked about in Section 2.3. Here is the connection between these. In Section 2.2, we learned that a function $f : \mathbb{R}^n \to \mathbb{R}$ that is differentiable at x_0 has a good linear approximation

$$l(x) = \mathbf{D}\, f(x_0)(x - x_0) + f(x_0)$$

in the sense that

$$f(x) = l(x) + E(x)$$

where $E(x)$ is an error function which satisfies

$$\lim_{x \to x_0} \frac{E(x)}{||x - x_0||} = 0.$$

So

$$f(x) = \mathbf{D}\, f(x_0)(x - x_0) + f(x_0) + E(x)$$

or

$$f(x) - f(x_0) = \mathbf{D}\, f(x_0)(x - x_0) + E(x).$$

If f is a sufficiently well-behaved function, then it also has a quadratic approximation

$$f(x) = f(x_0) + \mathbf{D}\, f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)Hf(x_0)(x - x_0) + R_2(x)$$

where $(x - x_0)Hf(x_0)(x - x_0)$ is interpreted as a row vector times an $n \times n$ matrix times a column vector and the error function $R_2(x)$ satisfies

$$\lim_{x \to x_0} \frac{R_2(x)}{||x - x_0||^2} = 0.$$

When f is well-behaved enough to have a good quadratic approximation is a tricky issue much like differentiability was. One reasonably general sufficient condition for f to be well-behaved enough is for the third order partial derivatives to be continuous. This is why Theorem 5 requires f to be a C^3 function.

Now, suppose x_0 is a critical point of f where $\mathbf{D}\, f(x_0) = 0$. Then the quadratic approximation of f near x_0 becomes

$$f(x) = f(x_0) + \underbrace{\mathbf{D}\, f(x_0)(x - x_0)}_{0} + \frac{1}{2}(x - x_0)Hf(x_0)(x - x_0) + R_2(x),$$

and hence

$$f(x) - f(x_0) = \frac{1}{2}(x - x_0)Hf(x_0)(x - x_0) + R_2(x).$$

Substituting $x = x_0 + h$ in gives

$$f(x_0 + h) - f(x_0) = \underbrace{\frac{1}{2}hHf(x_0)h}_{g(h)} + R_2(x).$$

Now, if g is positive-definite, then $g(h) \geq 0$ for every $h \in \mathbb{R}^n$ and we can require that x be sufficiently close to x_0, i.e. that h be close enough to 0, that the error $R_2(x)$ is guaranteed to be small enough that $g(h) + R_2(x) \geq 0$ for all h as well. Then

$$f(x_0 + h) - f(x_0) \geq 0 \implies f(x_0 + h) \geq f(x_0)$$

for all h close to 0. So $f(x_0)$ is the smallest value of f in a small enough neighborhood of x_0, which means x_0 is a local minimum.

An analogous argument shows that x_0 is a local maximum if $g(h)$ is a negative-definite.

Now that we have fought our way through the general abstraction for functions from $\mathbb{R}^n$ to $\mathbb{R}$, let us see what this means in the more concrete case when $n = 2$ in Example 5. The Hessian matrix is only 2×2 in this case and positive-definiteness and negative-definiteness of g can be phrased in terms of the signs of the upper left entry $\frac{\partial^2 f}{\partial x^2}(x_0, y_0)$ of the Hessian and its determinant. Some relatively straightforward algebra in Lemma 2 shows that g is positive-definite if $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0$ and the determinant of the Hessian at x_0 is positive, g is negative-definite if $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) < 0$ and the determinant of the Hessian at x_0 is positive. I suggest you skip the discussion about the general case of testing for positive-definitess of an $n \times n$ matrix using the determinants of the various square matrices in the upper left corner, and the following discussion about the general Second Derivative Test for functions of n variables, at least for now. Move on to Theorem 6 instead, which recaps what we have learned about finding and classifying critical points of a function $f : \mathbb{R}^2 \to \mathbb{R}$ using the sign of $\frac{\partial^2 f}{\partial x^2}(x_0, y_0)$ and the sign of the determinant of the Hessian at (x_0, y_0). Notice that the determinant of the Hessian is also called the *discriminant* of the Hessian.

You might wonder about the asymmetry of the Second Derivative Test. Why does the sign of $\frac{\partial^2 f}{\partial x^2}$ matter while the sign of $\frac{\partial^2 f}{\partial y^2}$ does not? Actually, you could use either. It is easy to see that if the determinant of the Hessian is positive, then $\frac{\partial^2 f}{\partial x^2}$ and $\frac{\partial^2 f}{\partial y^2}$ are either both positive or both negative.

Enjoy the relative simplicity of using the Second Derivative Test on specific functions of two variables in Examples 6-9. Example 8 shows you how you can use the test to find the absolute minimum of a function, much the same way you learned to find absolute extrema of single variable functions in single variable calculus. Example 10 illustrates how things work with more than two input variables.

Now, you can go back and read about how one can test for the positive-definitness of an $n \times n$ matrix and how the Second Derivative Test for functions of n variables.

The end of the section discusses how the absolute extrema of multivariables functions can be found, at least in certain cases. The results parallel what you learned in single variable calculus about the absolute extrema of single variable functions: a continuous function on a closed interval is always guaranteed to have an absolute minimum and an absolute maximum, and to find these, you need to check at the critical points inside the interval and at the endpoints. Things are again a little more complex for multivariable functions. Example 11 illustrates how these ideas work on a specific multivariable function.