MA1C, WEEK 3: THINGS YOU CAN DO WITH DERIVATIVES.

TA: PADRAIC BARTLETT

These notes, like all future notes, can be found on my website.

1. Last Week's HW

Average: was around 80% for this section. Major issues that cropped up were understanding the last question (Marsden/Tromba, sec. 2.5, question 24 – the question that asked you to spot the flaw in applying the chain rule), basic elements of calculational rigor, understanding the definition of partial derivatives on questions where the function is defined piecewise, and just clarity issues. If you have any questions or desire for clarification, I'm more than happy to meet with people and talk about any questions they have, whether in office hours or at other times.

2. Random Question

Question 2.1. *Consider the sequence of numbers*

- $a_1 = 1$
- $a_2 = 11$
- $a_3 = 21$
- $a_4 = 1211$
- $a_5 = 111221$
- $a_6 = 312211$
- $a_7 = 13112221$
- $\ldots$

Find the pattern in the sequence above; show that it is always increasing; and show that the only digits that ever appear in this sequence are 1,2, and 3.

3. Directional Derivatives

So: for f a function from $\mathbb{R}^n$ to $\mathbb{R}$, recall that we defined

$$\nabla f := Df = \left(\frac{\partial f}{\partial x_1} \ldots \frac{\partial f}{\partial x_n} \right)$$

and called this the "gradient" of f. It turns out that our use of the word "gradient" here is in fact motivated from the colloquial meaning of gradient as a slope or inclination: this is because the gradient of a function is actually a tool we can use to find the directional derivative (i.e. slope in a given direction) of our function at any point!

Explicitly, first recall the following definition:

Definition 3.1. For a function $f : \mathbb{R}^n \to \mathbb{R}$, $x_0 \in \mathbb{R}^n$, and a vector $v \in \mathbb{R}^n$, we defined the **directional derivative** of f at x_0 along v to be

$$\frac{d}{dt} f(x_0 + tv) \bigg|_{t=0}.$$

Then, we have the following theorem:

Theorem 3.2.

$$\frac{d}{dt} f(x_0 + tv)\bigg|_{t=0} = \nabla f(x_0) \cdot v.$$

In other words, the gradient completely determines the directional derivative: to understand a local linear approximation to a function at a given point, it suffices to simply have its gradient. This is kind of neat, and makes calculations rather easy:

Example 3.3. Let $f(x, y, z) = x^2y^2z^2$. Find f's directional derivatives in all directions at all points in $\mathbb{R}^3$.

Proof. Calculating gives us that

- $\frac{\partial f}{\partial x} = 2xy^2z^2$,
- $\frac{\partial f}{\partial y} = 2x^2yz^2$,
- $\frac{\partial f}{\partial z} = 2x^2y^2z$.

Thus, we have that the directional derivative of f at (x_0, y_0, z_0) in the direction (v_1, v_2, v_3) is

$$\nabla f(x, y, z) \cdot (v_1, v_2, v_3) = 2x_0y_0z_0 \cdot (v_1y_0z_0 + v_2x_0z_0 + v_3x_0y_0).$$

□

4. Higher-Order Derivatives

Definition 4.1. So: for a function $f : \mathbb{R}^n \to \mathbb{R}$, we denote the higher-order partial derivatives of f by writing

$$\frac{\partial^m f}{\partial x_{i_1} \dots \partial x_{i_n}},$$

and define them by

$$\frac{\partial^m f}{\partial x_{i_1} \dots \partial x_{i_n}} = \frac{\partial}{\partial x_{i_1}} \left(\frac{\partial}{\partial x_{i_2}} \left(\dots \frac{\partial f}{\partial x_{i_n}} \right) \right),$$

where we calculate each individual partial derivative using our normal definition in terms of limits,

$$\frac{\partial (f)}{\partial x_i}(a_1 \dots a_n) := \lim_{h \to 0} \frac{f(a_1, \dots a_i + h, \dots a_n) - f(a_1 \dots a_n)}{h}.$$

Higher-order partial derivatives have some nice properties. One of the more useful is the following:

Proposition 4.2. *If f is C^2, then*

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

It turns out that this proposition gives as an corollary the following theorem:

Theorem 4.3. *f is C^n, then for any $m \leq n$ and any permutation (i.e. rearrangement) of the variables $x_{i_1} \dots x_{i_m}$, we have that*

$$\frac{\partial^m f}{\partial x_{i_1} \dots \partial x_{i_m}} = \frac{\partial^2 f}{\partial x_{i_{\sigma(1)}} \dots \partial x_{i_{\sigma(m)}}}.$$

In other words, if a function is C^n, for any $m \leq n$ we can take its m-th order derivatives in any order we please.

(You are asked to show a special case of this theorem on your HW, so *don't* use this theorem to prove things on HW #3!)

The way the proof above goes is simply by repeated application of the C^2 case: we prove an illuminating special case of this theorem below.

Example 4.4. If $f : \mathbb{R}^2 \to \mathbb{R}$ is C^4, then

$$\frac{\partial^4 f}{\partial x^2 \partial y^2} = \frac{\partial^4 f}{\partial y^2 \partial x^2}.$$

Proof.

$$\begin{aligned}
\frac{\partial^4 f}{\partial x^2 \partial y^2} &= \frac{\partial}{\partial x}\frac{\partial}{\partial x}\frac{\partial}{\partial y}\frac{\partial f}{\partial y} \\
&= \frac{\partial}{\partial x}\frac{\partial}{\partial y}\frac{\partial}{\partial x}\frac{\partial f}{\partial y}, \text{ because } \frac{\partial f}{\partial y} \text{ is } C^2 \\
&= \frac{\partial}{\partial y}\frac{\partial}{\partial x}\frac{\partial}{\partial x}\frac{\partial f}{\partial y}, \text{ because } \frac{\partial^2 f}{\partial x \partial y} \text{ is } C^2 \\
&= \frac{\partial}{\partial y}\frac{\partial}{\partial x}\frac{\partial}{\partial y}\frac{\partial f}{\partial x}, \text{ because } f \text{ is } C^2 \\
&= \frac{\partial}{\partial y}\frac{\partial}{\partial y}\frac{\partial}{\partial x}\frac{\partial f}{\partial x}, \text{ because } \frac{\partial f}{\partial x} \text{ is } C^2.
\end{aligned}$$

□

The idea above is that we can use the C^2 proof to switch any two pairs of partial derivatives, and then iterate this process to simply jumble things up in any fashion we like.

5. Taylor Series

So, recall the definition of a Taylor series in one dimension from first quarter: for $f : \mathbb{R} \to \mathbb{R}$ an infinitely-differentiable function, and h, x_0 points in $\mathbb{R}$, we have that

$$f(x_0 + h) = f(x_0) + f'(x_0) \cdot h + \frac{f''(x_0)}{2!} \cdot h^2 + \ldots \frac{f^{(n)}(x_0)}{n!} \cdot h^n + R_n(x_0, h),$$

where $R_n(x_0, h)$ is some infintely-differentiable remainder function such that

$$\lim_{n\to\infty} \frac{R_n(x_0, h)}{h^n} = 0.$$

So, we would like to have a similar theorem in higher dimensions: however, as we've discussed before at length, the notion of "derivative" in higher dimensions is a complicated widget, and thus coming up with good analogues to things like "the nth derivative of f" is initially hard. However, we can in fact do this! We describe the case for the second-order Taylor series below:

Definition 5.1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a C^2 function. We define the **Hessian** of f at a point x_0 as the function

$$\begin{aligned}
H(f(x_0))(h) &= \frac{1}{2} \sum_{i,j=1`}^{n} h_i h_j \cdot \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0). \\
&= \frac{1}{2} \cdot (h_1, \ldots h_n) \cdot \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j=1}^{n} \cdot (h_1, \ldots h_n)^T
\end{aligned}$$

We think of this as something which is analogous to the "second derivative" of f; many times, where in the one-dimensional case we would write $f''(x)$, we will use the Hessian of f in the multidimensional case.

Definition 5.2. If $f : \mathbb{R}^n \to \mathbb{R}$ is a C^2 function, the **second-order Taylor series approximation** to f around x_0 is

$$f(x_0) + \nabla(f)(x_0) \cdot h + H(f(x_0))(h) + R_2(x_0, h),$$

where $R_2(x_0, h)$ is some C^2 remainder function that satisfies

$$\lim_{n\to\infty} \frac{R_2(x_0, h)}{||h||^2} = 0.$$

It can be proven that this approximation always exists whenever f is a C^2 function.

So: this allows us to, just as in the one-dimensional case, approximate f by a polynomial, which can allow us to get very good ideas for what f is in certain neighborhoods even if we cannot calculate it directly; this is frequently invaluable in many scientific fields. We calculate a simple example below:

Example 5.3. For $f(x, y) = -\cos(xy)$, calculate the second-order Taylor series approximation for f around $(0, 0)$.

Proof. So: note that

- $\frac{\partial f}{\partial x} = y \sin(xy)$,
- $\frac{\partial f}{\partial y} = x \sin(xy)$,
- $\frac{\partial^2 f}{\partial x^2} = y^2 \cos(xy)$,
- $\frac{\partial^2 f}{\partial y^2} = x^2 \cos(xy)$,
- $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} = \sin(xy) + xy \cos(xy)$,

and thus that the Taylor series for f around the point (x, y) is

$$\begin{aligned}
f(x + h_1, y + h_2) =& f(x, y) + \nabla(f)(x, y) \cdot (h_1, h_2) + H(f(x, y))(h_1, h_2) + R_2((x, y), (h_1, h_2))\\
=& -\cos(xy) + (h_1 \cdot y \sin(xy) + h_2 \cdot x \sin(xy))\\
&+ \frac{1}{2}(h_1^2 \cdot y^2 \cos(xy) + h_2^2 \cdot x^2 \cos(xy) + 2h_1h_2(\sin(xy) + xy \cos(xy))) + R_2((x, y), h);
\end{aligned}$$

plugging in $(x, y) = (0, 0)$ then gives that

$$f(h_1, h_2) = -1 + 0 + 0 + R_2(0, 0, h) = R_2((0, 0), h) - 1.$$

□

6. Extrema

So: it turns out that, just like in the one-dimensional case, we can use our knowledge of derivatives to classify the extremal points (critical points, maxima and minima) of a function. We illustrate the process below:

Proposition 6.1. *For $x_0 \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, x_0 is a* ***critical point*** *of f iff $(Df)(x_0) = 0$, (where by $(Df)(x_0) = 0$ we mean that every entry of the $1 \times n$ matrix $(Df)(x_0)$ is 0.)*

Proposition 6.2. *For $x_0 \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, we have that*

- *f has a local minimum at x_0 iff x_0 is a critical point and $(Hf(x_0))$ is positive-definite, and*
- *f has a local maximum at x_0 iff x_0 is a critical point and $(Hf(x_0))$ is negative-definite.*

*(Recall that a function g is called **positive-definite** iff $g(x) > 0$ for all $x \neq 0$ and $g(0) = 0$, and is called **negative-definite** iff $g(x) < 0$ for all $x \neq 0$ and $g(0) = 0$.)*

The proofs of these propositions are in Marsden/Tromba, or any halfway-decent source on vector calculus: we omit them here, in favor of a pair of examples that illustrate their use.

Example 6.3. Classify all of the extremal points of the "monkey saddle" function $f(x, y) = x^3 - 3xy^2$.

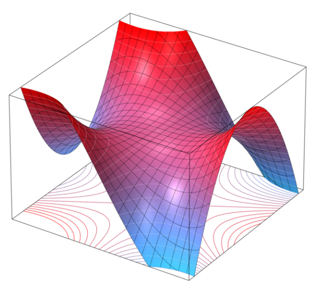
Proof. So: because

- $\frac{\partial f}{\partial x} = 3x^2 - 3y^2$,
- $\frac{\partial f}{\partial y} = -6xy$,

we have that the only critical point of f occurs at $(0, 0)$; then, because

- $\frac{\partial^2 f}{\partial x^2} = 6x$,
- $\frac{\partial^2 f}{\partial y^2} = -6x$,
- $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial y \partial x} = -6y$,

we have that $(Hf(0,0))(x, y) = \frac{1}{2}(x, y) \cdot \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \cdot (x, y)^T = 0$; as a result, $Hf(0, 0)$ is neither positive-definite nor negative-definite, and thus the point at (0,0) is neither a minimum nor a maximum. (It's worth noting that this makes sense, as looking at the graph of the monkey saddle makes it "obvious" that there are no local minima or maxima.)



□

Example 6.4. Classify all of the extremal points of the function $f(x, y) = x^2 + y^2$.

Proof. So: because

- $\frac{\partial f}{\partial x} = 2x$,
- $\frac{\partial f}{\partial y} = 2y$,

we have that the only critical point of f occurs at $(0, 0)$; then, because

- $\frac{\partial^2 f}{\partial x^2} = 2$,
- $\frac{\partial^2 f}{\partial y^2} = 2$,

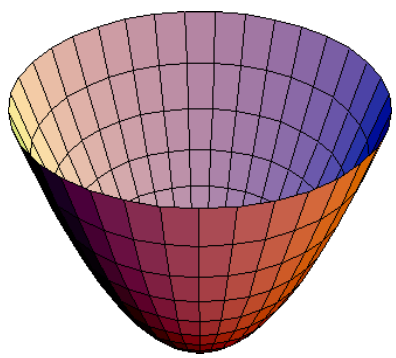
- $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial y \partial x} = 0$,

we have that $(Hf(0,0))(x,y) = \frac{1}{2}(x,y) \cdot \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \cdot (x,y)^T = 2x^2 + 2y^2$; as a result, $Hf(0,0)$ is positive definite, as the quantity $2x^2 + 2y^2$ is larger than 0 whenever both $x, y \neq 0$ and is equal to 0 if $x, y = 0$. This can also be seen by analyzing the eigenvalues of the matrix above – because we know that

- the matrix of second partial derivatives of a C^2 function $f : \mathbb{R}^n \to \mathbb{R}$ is symmetric, as it doesn't matter in which order we take the derivatives, and
- symmetric matrices are positive-definite iff all of their eigenvalues are positive,

we can simply note that the only eigenvalue of the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ is 2, and thus that Hf is positive-definite. So f has a local minimum at $(0,0)$.

Again, this agrees with inspection, as the picture below demonstrates:



$\square$