Math 1c | TA: Padraic Bartlett

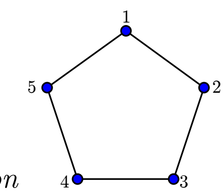
Recitation 3: More Applications of the Derivative

Week 3 | *Caltech 2012*

1 Random Question

Question 1 *A* ***graph*** *consists of the following:*

- *A set V of vertices.*
- *A set E of edges, where each edge consists of a distinct unordered pair of distinct vertices.*

For example, the pentagon  *can be thought of as the graph with*

- $V = \{1, 2, 3, 4, 5\}$,
- $E = \big\{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{5, 1\}\big\}$.

Given a graph G and vertex $v \in V$, we can define the ***degree*** *of v to be the number of edges that have v as one of their endpoints. For example, every vertex in the pentagon above has degree 2, as each vertex is the endpoint for precisely two edges.*

A ***triangle decomposition*** *of a graph G is a way to break its edge set E apart into disjoint triangles. It's not too hard to show that if a triangle decomposition exists, then*

- *every vertex needs to have even degree: this is because each of a triangle's vertices has precisely two edges from the triangle hitting it, so if we've broken all of our edges into triangles, every vertex has degree equal to $2\times$(the number of triangles that use that vertex), which is even.*
- *The number of edges in E has to be a multiple of 3, because each triangle uses three edges and they're all disjoint.*

Suppose that G is a graph on n vertices

1. *(Known, not too hard:) Show that if the degree of every vertex is $n - 1$ (i.e. every vertex is connected to every other vertex,) these conditions are also* ***sufficient****: i.e. if all of the degrees are even and the number of edges is a multiple of 3, a triangle decomposition exists.*
2. *(Known, fairly hard:) Show that if the degree of every vertex is $\geq \left(1 - 10^{-24}\right) n$ – i.e. the vertices are connected to "almost every" other vertex, these conditions are still sufficient.*

3. *(Current research results I'm working on) Show that if the degree of every vertex is* $\geq \left(1 - 10^{-4}\right) n$*, then these conditions are still sufficient.*

4. *(Conjecture, probably wildly difficult) Show that if the degree of every vertex is* $\geq \frac{3}{4}n$*, then these conditions are still sufficient.*

2 Directional Derivatives

In our last class, we spent a lot of time working on the idea of a **derivative** in $\mathbb{R}^n$. Specifically, we introduced the idea of **partial derivatives**, used these to construct the **total derivative** (a "linear approximation" to our function at a point, which was similar to the "linear approximation" idea that we had for the derivative in $\mathbb{R}^1$,) and discussed some of the geometric interpretations and applications of these ideas.

However, these techniques were not the only ways to talk about the derivative! Another, different, way is to generalize the idea of "slope" from $\mathbb{R}^1$ via the concept of the **directional derivative**: we discuss how to do this below.

In $\mathbb{R}^1$, the derivative of a function $f : \mathbb{R} \to \mathbb{R}$ at some point a is the "slope" of the graph $f(x) = y$ at the point $(a, f(a))$: essentially, we are measuring the change in $f(x)$ as we move along the x-axis. However, for functions $\mathbb{R}^n \to \mathbb{R}$ we are no longer restricted to just the x-axis; instead, we can move along **any** vector in $\mathbb{R}^n$! This leads us to define the **directional derivative** of a function $f : \mathbb{R}^n \to \mathbb{R}$ at some point $\mathbf{a}$, *along some direction* $\mathbf{v}$, as the "slope" of f at the point $\mathbf{a}$, as measured in the direction $\mathbf{v}$. More formally:

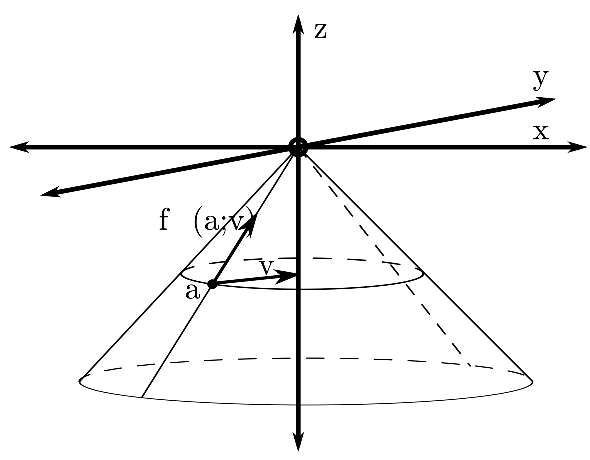
Definition. The **directional derivative** of a function $f : \mathbb{R}^n \to \mathbb{R}$ at some point $\mathbf{a}$ along some direction $\mathbf{v}$ is the derivative

$$f'(\mathbf{a}; \mathbf{v}) := \frac{d}{dt} \left(f(\mathbf{a} + t \cdot \mathbf{v}) \right) \Big|_{t=0}.$$

To illustrate what's going on here, consider the following example:

Question 2 *Consider the function* $f(x, y) = -\sqrt{x^2 + y^2}$*. What is the directional derivative of this function at the point* $(0, -1)$ *in the direction* $(0, 1)$*?*

Solution. First, to get a good idea of what's going on in this problem, we graph our function:



Visually, if we look at the point $(0, -1)$ and its slope in the direction $(0, 1)$, we can see that it **should** be 1, just by examination. So, let's calculate, and see if our visual intuition matches our mathematical definition:

$$
\begin{aligned}
f'((0,-1);(0,1)) &= \frac{d}{dt}\left(f((0,-1)+t\cdot(0,1))\right)\Big|_{t=0}\\
&= \frac{d}{dt}\left(f(0,t-1)\right)\Big|_{t=0}\\
&= \frac{d}{dt}\left(-\sqrt{0^2+(t-1)^2}\right)\Big|_{t=0}\\
&= \frac{d}{dt}\left(-|t-1|\right)\Big|_{t=0}\\
&= \frac{d}{dt}\left(-(-(t-1))\right)\Big|_{t=0}, \quad [\text{ because near } 0, (t-1) \text{ is negative}]\\
&= \frac{d}{dt}\left(t-1\right)\Big|_{t=0}\\
&= 1\Big|_{t=0}\\
&= 1.
\end{aligned}
$$

This matches our visual picture and our intuition.

The following result can make calculating the directional derivative easier, in the case that we already know the gradient of a function:

Theorem 3 *Suppose that f is differentiable at some point $\mathbf{a}$: one notable case where this happens is when all of f's partial derivatives are continuous at $\mathbf{a}$, as we mentioned last class. Then, the **directional derivative** of a function $f : \mathbb{R}^n \to \mathbb{R}$ at some point $\mathbf{a}$ along some direction $\mathbf{v}$ is given by the dot product of the gradient of f at $\mathbf{a}$ with $\mathbf{v}/||\mathbf{v}||$. In other words,*

$$f'(\mathbf{a};\mathbf{v}) := \nabla f\Big|_{\mathbf{a}} \cdot \frac{\mathbf{v}}{||\mathbf{v}||}.$$

To illustrate the use of this theorem, return to our cone problem from earlier. There, we had $f(x, y) = -\sqrt{x^2 + y^2}$; thus, if we hold y constant, we can see that

$$\frac{\partial f}{\partial x} = -\frac{2x}{2\sqrt{x^2+y^2}} = \frac{-x}{\sqrt{x^2+y^2}}.$$

Similarly, by holding y constant, we have

$$\frac{\partial f}{\partial y} = \frac{-y}{\sqrt{x^2+y^2}}.$$

Therefore, we know that the directional derivative of f at $(0,-1)$ in the direction $(0,1)$ is given by

$$\begin{aligned}\left(\frac{\partial f}{\partial x}(0,-1), \frac{\partial f}{\partial y}(0,-1)\right) \cdot (0,1) &= \left(\frac{-(0)}{\sqrt{0^2+(-1)^2}}, \frac{-(-1)}{\sqrt{0^2+(-1)^2}}\right) \cdot (0,1) \\ &= (0,1) \cdot (0,1) \\ &= 1,\end{aligned}$$

which matches our earlier answer.

3 Higher-Order Derivatives and their Applications

Another thing we could want to do with the derivative, motivated by what we were able to do in $\mathbb{R}^1$, is the concept of **higher-order derivatives**. These are relatively easy to define for partial derivatives:

Definition. Given a function $f : \mathbb{R}^n \to \mathbb{R}$, we can define its **second-order partial derivatives** as the following:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right).$$

In other words, the second-order partial derivatives are simply all of the functions you can get by taking two consecutive partial derivatives of your function f.

A useful theorem for calculating these partial derivatives is the following:

Theorem 4 *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called C^2 at some point if all of its second-order partial derivatives are continuous at that point. If a function is C^2, then the order in which second-order partial derivatives are calculated* ***doesn't matter****: i.e.*

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

for any i, j.

It bears noting that if the conditions of this theorem are **not** met, then the order for computing second-order partial derivatives may actually matter! One such example is the function

$$f(x,y) = \begin{cases} \frac{x^3y - xy^3}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases}$$

At $(0,0)$, you can calculate that $\frac{\partial^2 f}{\partial x \partial y} = 1 \neq -1 = \frac{\partial^2 f}{\partial y \partial x}$: a result that occurs because the second-order partials of this function are not continuous.

However, the interesting aspects of higher-order partial derivatives are not really in their calculation; rather, the **applications** of higher-order partial derivatives are the things worth studying. In $\mathbb{R}$, for example, we could turn the second derivative of a function into a lot of information about that function: in particular, we could use this second derivative to determine

- whether a given critical point was a local minima or maxima,
- whether a function is concave up or down at a given point,
- and what the second-order Taylor approximation to that function was at a point.

Can we do the same for functions from $\mathbb{R}^n$ to $\mathbb{R}$? As it turns out, the answer is yes! The tool with which we do this is called the Hessian, which we define here:

Definition. The Hessian of a function $f : \mathbb{R}^n \to \mathbb{R}$ at some point $\mathbf{a}$, $H(f)\big|_{\mathbf{a}}(\mathbf{h})$, is the following function from $\mathbb{R}^n$ to $\mathbb{R}$:

$$H(f)\big|_{\mathbf{a}}(\mathbf{h}) = \frac{1}{2}(h_1, h_2, \dots h_n) \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{bmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix}$$

The main useful property of the Hessian is the following:

Theorem 5 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function with well-defined second-order partials at some point $\mathbf{a}$, and $H = H(f)\big|_{\mathbf{a}}$ be its Hessian. Pick any two co ordinates x_i, x_j in $\mathbb{R}^n$: then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = \frac{\partial^2 H}{\partial h_i \partial h_j}.$$

In other words, H's second-order partial derivative are precisely the second-order partial derivatives of f at $\mathbf{a}$! So H is basically a function designed to have the same second-order partials as f at $\mathbf{a}$.

One quick thing this theorem suggests is that we could use $H(f)\big|_{\mathbf{a}}$ to create a "second-order" approximation to f at $\mathbf{a}$, in a similar fashion to how we used the derivative to create a linear (i.e. first-order) approximation to f. We define this below:

Theorem 6 *If $f : \mathbb{R}^n \to \mathbb{R}$ is a function with continuous second-order partials, we define the **second-order Taylor approximation** to f at $\mathbf{a}$ as the function*

$$T_2(f)\big|_{\mathbf{a}}(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + (\nabla f)(\mathbf{a}) \cdot \mathbf{h} + H(f)\big|_{\mathbf{a}}(\mathbf{h}).$$

You can think of $f(\mathbf{a})$ as the constant, or zero-th order part, $(\nabla f)(\mathbf{a}) \cdot \mathbf{h}$ as the linear part, and $H(f)\big|_{\mathbf{a}}(\mathbf{h})$ as the second-order part of this approximation.

To illustrate how this process actually creates a pretty decent approximation to f, we calculate an example:

Example. Calculate the second-order Taylor approximation to the function $f(x, y) = e^{xy}$ at the origin.

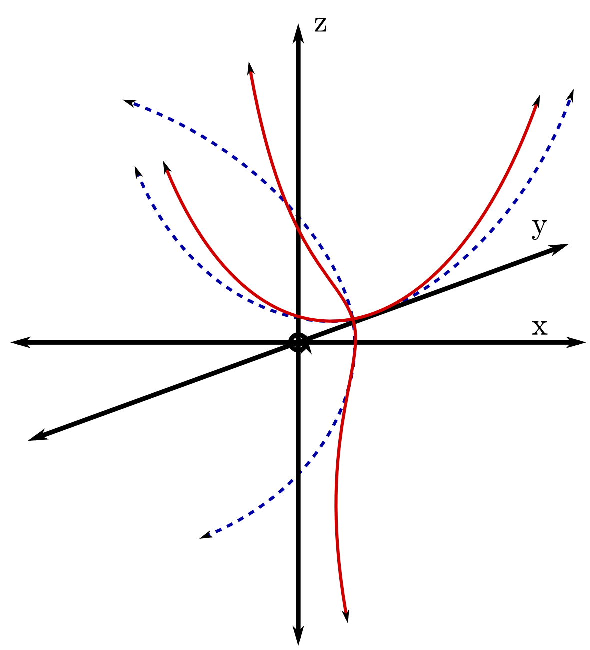
Answer. Calculating the second derivatives of f is pretty straightforward:

$$\frac{\partial f}{\partial x} = ye^{xy}, \frac{\partial f}{\partial y} = xe^{xy}$$
$$\frac{\partial^2 f}{\partial x^2} = y^2 e^{xy}, \frac{\partial^2 f}{\partial y^2} = x^2 e^{xy}, \frac{\partial^2 f}{\partial x \partial y} = xye^{xy} + e^{xy} = \frac{\partial^2 f}{\partial y \partial x}.$$

If we evaluate these partials at 0 and plug them into the definition above for $T_2(f)\big|_{(0,0)}$, we get

$$\begin{aligned}
T_2(f)\big|_{(0,0)}((0,0) + (h_1, h_2)) &= f(0,0) + (\nabla f)(0,0) \cdot (h_1, h_2) + H(f)\big|_{(0,0)}(h_1, h_2) \\
&= 1 + (0,0) \cdot (h_1, h_2) + \frac{1}{2}(h_1, h_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\
&= 1 + \frac{1}{2}(h_1, h_2) \begin{pmatrix} h_2 \\ h_1 \end{pmatrix} \\
&= 1 + \frac{1}{2}(2h_1 h_2) \\
&= 1 + h_1 h_2.
\end{aligned}$$

So, at the origin, the second-order Taylor approximation for f is just $T_2(f)(x, y) = 1 + xy$. The following graph, with e^{xy} in solid red and T_2 in dashed blue, shows that it's actually a somewhat decent approximation at $(0, 0)$:



As well, we can use the second derivatives to search for and find local minima and maxima! We define these terms here:

Definition. A point $\mathbf{a} \in \mathbb{R}^n$ is called a **local maxima** of a function $f : \mathbb{R}^n \to \mathbb{R}$ iff there is some small value r such that for any point $\mathbf{x}$ in $B_{\mathbf{a}}(r)$ not equal to $\mathbf{a}$, we have $f(\mathbf{x}) \leq f(\mathbf{a})$.

A similar definition holds for local minima.

So: how can we use the derivative to find such local maxima? Well, it's clear that (if our function is differentiable in a neighborhood around this point) that no matter how we move to leave this point, our function must not increase – in other words, for any direction $\mathbf{v} \in \mathbb{R}^n$, the directional derivative $f'(\mathbf{a}, \mathbf{v})$ must be ≤ 0. But this means that in fact all of the directional derivatives must be **equal** to 0!, because if $f'(\mathbf{a}, \mathbf{v})$ was < 0, then $f'(\mathbf{a}, -\mathbf{v})$ would be > 0.

This motivates the following definitions, and basically proves the following theorem:

Definition. A point $\mathbf{a}$ is called a stationary point of some function $f : \mathbb{R}^n \to \mathbb{R}$ iff $\nabla(f)\Big|_{\mathbf{a}} = (0, \ldots, 0)$. A point $\mathbf{a}$ is called a **critical point** if it is a stationary point, or f is not differentiable in any neighborhood of $\mathbf{a}$.

Theorem 7 *A function $f : \mathbb{R}^n \to \mathbb{R}$ attains its local maxima and minima only at critical points.*

However, it bears noting that not every critical or stationary point is a local maxima or minima! A trivial example would be the function $f(x, y) = x^2 - y^2$: the origin is a stationary point, yet neither a local minima or maxima (as $f(0, \epsilon) < 0 < f(\epsilon, 0)$, and thus there are positive and negative values of f attained in any ball around the origin, where it is 0.)

How can we tell which stationary points do what? Well, in one-variable calculus, we used the idea of the "second derivative" to determine what was going on! In specific, we knew that if the second derivative of a function f at some point a was negative, then tiny increases in our variable at that point would cause the first derivative to decrease, and tiny decreases in our variable at that point would cause the negative of the first derivative to increase – i.e. cause the first derivative to decrease, and therefore make the function itself decrease! Therefore, the second derivative being negative at a stationary point implied that that point was a local maxima.

In higher dimensions, things are tricker – at a given point $\mathbf{a}$, we no longer have this idea of a "single" second derivative, but instead have many different second derivatives, like $\frac{\partial^2 f}{\partial x \partial y}(\mathbf{a})$ and $\frac{\partial^2 f}{\partial z^2}(\mathbf{a})$. Yet, we can still use the same ideas as before to figure out what's going on!

In particular, in one dimension, we said that we wanted tiny positive changes of our variables to make the first functions decrease. In other words, given any of the partials $\frac{\partial f}{\partial x_i}$, we want any positive changes in the direction of this partial to make our function decrease – i.e. we want the directional derivative of $\frac{\partial f}{\partial x_i}$ to be negative in any direction $\mathbf{v}$, where all of the coördinates of $\mathbf{v}$ are positive. (Positivity here stems from the same reason that in one dimension, we have that the first derivative is increasing for all of the points to the left of a maxima and decreasing for all of the points to the right of a maxima.)

So: this condition, if we write it out, is just asking that for every i and nonzero $\mathbf{v}$, that

$$\left(\frac{\partial^2 f}{\partial x_1 \partial x_i}(\mathbf{a}), \frac{\partial^2 f}{\partial x_2 \partial x_i}(\mathbf{a}), \ldots \frac{\partial^2 f}{\partial x_n \partial x_i}(\mathbf{a})\right) \cdot (v_1^2, v_2^2, \ldots v_n^2)$$

is negative. If you choose to write this out as a matrix, this actually becomes the claim that for any $\mathbf{v} \neq 0$, we have

$$\mathbf{v}^T \cdot \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{bmatrix} \cdot \mathbf{v} < 0.$$

linear algebra, you may hopefully remember that any matrix satisfying this condition is called being **negative-definite**, and is equivalent to having all n of your eigenvalues existing and being negative. Similarly, if we were looking for a local minima, we would be asking that the above matrix product is always positive: i.e. that the matrix is **positive-definite**, which is equivalent to all of its eigenvalues being positive.

But we've seen this construction before! In particular, this matrix-product thing above is just the Hessian! Based on these observations, we make the following definitions and observations:

Definition. The Hessian $H(f)\big|_{\mathbf{a}}$ of a function f at some point $\mathbf{a}$ is called **positive-definite** if for all $\mathbf{h} \neq 0$,

$$H(f)\big|_{\mathbf{a}}(\mathbf{h}) > 0,$$

and similarly that the Hessian is negative-definite if for all $\mathbf{h} \neq 0$,

$$H(f)\big|_{\mathbf{a}}(\mathbf{h}) < 0.$$

Theorem 8 *The Hessian $H(f)\big|_{\mathbf{a}}$ is positive-definite if and only if the matrix*

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{bmatrix}$$

is positive-definite. (The same relation holds for being negative-definite.)

Theorem 9 *A function $f : \mathbb{R}^n \to \mathbb{R}$ has a local maxima at a stationary point $\mathbf{a}$ if all of its second-order partials exist and are continuous in a neighborhood of $\mathbf{a}$, and the Hesssian of f is negative-definite at $\mathbf{a}$. Similarly, it has a local minima if the Hessian is positive-definite at $\mathbf{a}$. If the Hessian takes on both positive and negative values there, it's a **saddle point**: there are directions you can travel where your function increase, and others where it will decrease. Finally, if the Hessian is identically 0, you have no information as to what your function may be up to: you could be in any of the three above cases.*

A quick example, to illustrate how this gets used:

Example. For $f(x,y) = x^2 + y^2, g(x,y) = -x^2 - y^2$, and $h(x,y) = x^2 - y^2$, find local minima and maxima.

Solution. First, by taking partials, it is clear that the only point at which the gradient of these functions is $(0,0)$ is the origin. There, we have that

$$H(f)\Big|_{(0,0)} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, H(g)\Big|_{(0,0)} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}, H(h)\Big|_{(0,0)} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix},$$

and thus that f is positive-definite at (0,0), g is negative-definite at (0,0), and h is neither at (0,0) by examining the eigenvalues. Thus f has a local minima at $(0,0)$, g has a local maxima at (0,0), and h has a saddle point at (0,0).