Notes for Section 2.6

We are already quite familiar with the gradient, so the beginning of this section should be easy reading for you. Let me move on to directional derivatives.

Suppose $T(x,y)$ is the temperature at the point (x,y) in the plane. Suppose T is differentiable at the point (x_0,y_0). We have seen that the partial derivative $\frac{\partial T}{\partial x}(x_0,y_0)$ tells us the rate of change of temperature as we move parallel to the x-axis starting at (x_0,y_0), and similarly, $\frac{\partial T}{\partial y}(x_0,y_0)$ tells us the rate of change of temperature as we move parallel to the y-axis starting at (x_0,y_0). But what happens if we want to move in the direction of $(1,2)$? The temperature will change then too, but at what rate? Let's we take a small step $h(1,2)$ in the direction of $(1,2)$. So h is a small number. Actually, let us suppose $h>0$ for now. We expect that the value of T will change approximately by $\frac{\partial T}{\partial x}(x_0,y_0)\,h$ because we moved h parallel to the x-axis, and it will change approximately by $\frac{\partial T}{\partial y}(x_0,y_0)\,2h$ because we moved $2h$ parallel to the y-axis. We expect that the change in the value of T as we move from (x_0,y_0) to $(x,y)=(x_0+h,y_0+2h)$ is approximately the sum of these two changes. All we are saying here is that since T is a differentiable function, we know that the linear approximation

$$l(x,y)=\left(\frac{\partial T}{\partial x}(x_0,y_0),\frac{\partial T}{\partial y}(x_0,y_0)\right)\cdot\big((x,y)-(x_0,y_0)\big)+T(x_0,y_0)$$

is a good approximation for T near (x_0,y_0), so the change ΔT in the value of T going from (x_0,y_0) to (x_0+h,y_0+2hy) is well estimated by

$$\begin{aligned}
\Delta T &= T(x,y)-T(x_0,y_0)\\
&\approx l(x,y)-T(x_0,y_0)\\
&=\left(\frac{\partial T}{\partial x}(x_0,y_0),\frac{\partial T}{\partial y}(x_0,y_0)\right)\cdot(h,2h)+T(x_0,y_0)-T(x_0,y_0)\\
&=\left(\frac{\partial T}{\partial x}(x_0,y_0),\frac{\partial T}{\partial y}(x_0,y_0)\right)\cdot(h,2h)\\
&=\frac{\partial T}{\partial x}(x_0,y_0)\,h+\frac{\partial T}{\partial y}(x_0,y_0)\,2h
\end{aligned}$$

That is the change ΔT but what is the rate of change? It is

$$\frac{\Delta T}{||(h,2h)||}.$$

Of course, all of this is approximate. But the approximation should get better as $h\to 0$ because l is a good approximation to T, so the error $E(x,y)=T(x,y)-l(x,y)$ should approach 0 quickly as $h\to 0$. One can work this out using limits too, but let me focus on the conceptual picture here instead of the technical details. So let us accept for now that the approximate rate of change of T as we move in the direction of $(1,2)$ is

$$\begin{aligned}
\frac{\Delta T}{||(h,2h||} &\approx \frac{\frac{\partial T}{\partial x}(x_0,y_0)\,h+\frac{\partial T}{\partial y}(x_0,y_0)\,2h}{\sqrt{h^2+(2h)^2}}\\
&=\frac{\frac{\partial T}{\partial x}(x_0,y_0)\,h+\frac{\partial T}{\partial y}(x_0,y_0)\,2h}{\sqrt{5}|h|}
\end{aligned}$$

If $h>0$, as we assumed, this is

$$\frac{\frac{\partial T}{\partial x}(x_0,y_0)\,h+\frac{\partial T}{\partial y}(x_0,y_0)\,2h}{\sqrt{5}h}=\frac{\frac{\partial T}{\partial x}(x_0,y_0)+2\frac{\partial T}{\partial y}(x_0,y_0)}{\sqrt{5}}$$

which does not depend on h at all. Now, as $h \to 0$, the approximation we used $T(x, y) \approx l(x, y)$ at $(x, y) = (x_0 + h, y_0 + 2h)$ should become accurate. So we are justified in saying that the rate of change of T as we move in the direction of $v = (1, 2)$ is

$$\frac{\frac{\partial T}{\partial x}(x_0, y_0) + 2\frac{\partial T}{\partial y}(x_0, y_0)}{\sqrt{5}} = \left(\frac{\partial T}{\partial x}(x_0, y_0), \frac{\partial T}{\partial y}(x_0, y_0)\right) \cdot \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) = \nabla T(x_0, y_0) \cdot u$$

where $u = \frac{v}{||v||}$. That is to find the rate of change of T in the direction v, all we need to do is take the dot product of the gradient of T with the unit vector u that points in the same direction as v. This actually makes good sense. If you take a step of length 1 in the direction of v, then your x-coordinate changes by $1/\sqrt{5}$ and your y-coordinate changes by $2/\sqrt{5}$. So it makes sense that the temperature should change by $\nabla T(x_0, y_0) \cdot u$. Well approximately, but we are talking rates anyway, and because the size of our step was 1, this is not just the change but the rate of change of T in the direction of v as well. There is one little issue. This does not work if $h < 0$. We get the opposite of the rate of change we just found. This makes sense because if $h < 0$ then we are really moving in the opposite direction, in the direction of $-v = (-1, -2)$, so we would expect that the rate of change of T would be the opposite too. Indeed, but we still cannot have two different rates of change at the same point. The way to deal with this ambiguity is to define the rate of change of T in the direction of $u = \frac{v}{||v||}$ as

$$\lim h \to 0 = \frac{T\big((x_0, y_0) + hu\big) - T(x_0, y_0)}{h}.$$

Note that we have h and not $|h|$ in the denominator. We used u instead of v so that the size of our step hu in the direction of u is really h, and not $h||v||$. What we just defined is called the directional derivative of T at (x_0, y_0) in the direction of u. It is often denoted by $T_u(x_0, y_0)$ or $\mathbf{D}_u\, T(x_0, y_0)$ or $\nabla_u T(x_0, y_0)$. Our textbook uses the notation $\mathbf{D}\, T(x_0, y_0)u$, which basically just writes out what we have just learned about how the directional derivative can be calculated by multiplying the gradient at (x_0, y_0) by the direction vector u. Here is the definition of the directional derivative in general:

Definition 1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function, x_0 a point $\mathbb{R}^n$ and v a vector in $\mathbb{R}^n$. The directional derivative of f at x_0 in the direction of v is

$$\mathbf{D}_v\, f(x_0) = \lim_{h \to 0} \frac{f(x_0 + hv) - f(x_0)}{h}.$$

Note that the directional derivative is only defined for functions whose values are real numbers. The definition does not specify that v must be a unit vector, but it is common to assume that it is, or to say that the direction of v is the unit vector $u = v/||v||$ and use u instead of v. Note that another way we can view this is that if g is the single variable function $g(t) = f(x_0 + tv)$ then the directional derivative is $\mathbf{D}_v\, f(x_0) = g'(0)$. This is how our textbook defines it. Of course, the two definitions are equivalent.

Theorem 12 states (for $n = 3$) what we have already discovered: if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at x_0, then all of its partial derivatives must exist there, and its directional derivative in the direction of v is

$$\mathbf{D}_v\, f(x_0) = \nabla f(x_0) \cdot v.$$

In particular, the directional derivative exists for any vector $v \in \mathbb{R}^n$. Instead of the somewhat informal and somewhat broken argument I gave at the beginning of these notes, the textbook uses the Chain Rule to prove this, which is shorter, slicker, and correct. But I would say it does not explain nearly as well why the directional derivative is the dot product of the gradient and the vector v.

You already know how to calculate the gradient of a function, so calculating directional derivatives should be a breeze. Examples 3 and 4 illustrate this.

The statement in Theorem 13, that value of the directional derivative $\mathbf{D}_v f$ should be largest (i.e. the rate of increase should be the fastest) if v points in the same direction as the gradient of f should be quite obvious and the proof easy to understand.

The connection between the gradient of $f : \mathbb{R}^n \to \mathbb{R}$ and the level sets of f in Theorem 14 may be surprising at first. The proof is short and easy, but perhaps a good way to develop an intuitive feel for why this should be true is to picture yourself using a topographic map. The map shows the contour curves depicting elevation. Suppose you want to climb a hill and you want to waste no time getting to the top. That is you want to find the shortest and consequently steepest path that gets you there. Which direction are you going to head? You will not want to walk along a contour curve. If you do, you are not gaining any elevation. You will want to go in the direction that gets you to next contour curve as quickly as possible. In other words, you want to move away from the current contour curve you are standing on as quickly as possible. Then you will want to head in a direction that is perpendicular to that contour curve.

You can use the fact that the gradient is always perpendicular at the level sets to find the tangent plane to the level surface of a function $f : \mathbb{R}^3 \to \mathbb{R}$ at a point (x_0, y_0, z_0). The gradient $\nabla f(x_0, y_0, z_0)$ is perpendicular to the level surface that passes through (x_0, y_0, z_0). This means exactly that it is perpendicular to any vector that is tangent to the surface. So it is perpendicular to any vector in the tangent plane (x_0, y_0, z_0). Now you have a normal vector to that tangent plane and you know the plane passes through (x_0, y_0, z_0), so you can use what we learned in Section 1.3 to find the equation of that plane. This is illustrated in Example 6. There is of course nothing special about $\mathbb{R}^3$ versus $\mathbb{R}^n$. You can use the same ideas to find the equation of the tangent hyperplane to a level set for any function $f : \mathbb{R}^n \to \mathbb{R}$. Guess what, this even works in 2-dimensions to find the equation of the tangent line to a curve given by an equation with x and y in it. In single variable calculus, you called such curves implicit functions and you learned how to use implicit differentiation to find the slope and the equation of the tangent line at a point. Now you know another way to solve such a problem.

The part about vector fields is a visualization technique to visualize the gradient of a function. The gradient is of course itself a function. But even in the simplest cases, you would have a hard time visualizing its graph because the graph lives in such a high-dimensional space. However, thinking about the gradient as a vector field really makes it possible to visualize it. My topographic map analogy reappears here.

The section ends with three examples. Examples 7 and 9 illustrate applications of the gradient and what we have just learned about the gradient to physics. Example 8 shows you a clever trick to find a normal vector to a surface given by an equation in x, y, and z by treating it as a level surface of a function. This is the 3-dimensional analog of what I mentioned earlier about finding the tangent line to the graph of an implicit function.