MA 200 - Lecture 4

1 Recap

1. Definition of directional derivative, examples.
2. Differentiability, gradient.

2 Derivatives

Proof of diff implies directional derivatives exist.

Proof. If $v = 0$, then $\nabla_v f = 0 = L(0)$ trivially.

$$\lim_{t\to 0}\left|\frac{f(a+tv)-f(a)-L(tv)}{t}\right| = 0, \tag{1}$$

because of the hypothesis of differentiability (and the fact that $t \to tv$ is a continuous function). Thus, $\nabla_v f(a)$ exists and equals $L(v)$. By linearity of L, $L(v) = \sum_i L(e_i)v_i = \sum_i \nabla_{e_i} f v_i = \sum_i \frac{\partial f}{\partial x_i}(a)v_i$. ☐

We also have a vector-valued version of this proposition (whose proof is almost the same).

Proposition 2.1. *If $f : U \to \mathbb{R}^m$ is differentiable at a, all of its directional derivatives exist at a and $L(h) = \displaystyle\sum_{i,j} \frac{\partial f_i}{\partial x_j}|_{x=a} h_j e_i$.*

The derivative is now a matrix Df whose rows are gradients of the components. *Remark*: If $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at an interior point $a \in U$, then $g : V \to \mathbb{R}^m$ defined by $g(x) = f(x)$ and $a \in V$ (where V is open) is also differentiable with the same derivative (and vice-versa). We can also easily prove that a vector-valued function is differentiable iff each of its components is so (how?).

This notion of differentiability *does* imply continuity. Indeed, $\|f(a+h) - f(a)\| \leq \|f(a+h) - f(a) - L(h)\| + \|L(h)\| < \frac{1}{2}\|h\| + \|L\|\|h\|$ for $\|h\| < \delta_1$. If $\|h\| < \min\left(\delta_1, \frac{\epsilon}{1+\|L\|}\right)$, then we are done. ☐

We define the tangent plane to the graph of $f : \mathbb{R}^n \to \mathbb{R}$ at a differentiable point as follows: $x_{n+1} = f(\vec{a}) + \langle \nabla f(a), \vec{h}\rangle$. The point is that $\frac{|f(\vec{a}+\vec{h}) - \vec{r}(h)|}{\|h\|}$ goes to 0 as $\vec{h} \to \vec{0}$ and hence it is a good approximation of the graph. We shall come back to the tangent plane later (after we discuss the inverse function theorem).

Here is a *practical* criterion to conclude differentiability.

Theorem 1. *If $f : U \to \mathbb{R}^n \to \mathbb{R}^m$ is C^1 at an interior point $a \in U$, i.e., the matrix Df exists in a neighbourhood V of a and is a continuous function in its in own right, then f is differentiable at a (and is hence continuous at a as well).*

Why is this a practical criterion? Here are examples.

1. $f(x,y) = \sin(x^3 e^y) + xe^{y^2}$ on $\mathbb{R}^2$. Note that by one-variable calculus, f_x, f_y exist on all of $\mathbb{R}^2$ and $f_x = \cos(x^3 e^y)3x^2 + e^{y^2}$, $f_y = \cos(x^3 e^y)e^y + 2yxe^{y^2}$. These functions are continuous everywhere because they are sums of products of compositions of continuous functions.

2. $f(x,y,z) = (x^3 z^2, \sin(x^4 yz))$ is C^1 everywhere just as above.

3. Polynomials in any number of variables are C^1 everywhere (and hence differentiable everywhere).

4. Rational functions are differentiable wherever the denominator is not zero (and such points are all interior points. Why?)

Bu this criterion does not work the other way! That is, here is an example of a function that is NOT C^1 and is yet differentiable: $f(x) = x^2 \sin\left(\frac{1}{x}\right)$ when $x \neq 0$ and $f(0) = 0$.
Proof of the criterion:

Proof. Firstly, we can assume without loss of generality that $m = 1$ (because a function is differentiable iff each of its components is so). We prove by induction on the number of variables. For one variable, it is trivial. Suppose this result is true for $1, 2, \ldots, n-1$ variables. Write the point $\vec{a} \in U$ as $(a, \vec{b})$ where $a \in \mathbb{R}$ and $\vec{b} \in \mathbb{R}^{n-1}$. The function $g(\vec{y}) = f(a, \vec{y})$ is C^1 at $\vec{b}$ and is hence differentiable with derivative $L(\vec{k}) = Df_{(a,\vec{b})}(0, \vec{k})$.

$$
\begin{aligned}
\frac{\|f(a+h,\vec{b}+\vec{k}) - f(a,\vec{b}) - Df_{(a,\vec{b})}(h,\vec{k})\|}{\|(h,\vec{k})\|} &\leq \frac{\|f(a+h,\vec{b}+\vec{k}) - f(a,\vec{b}+\vec{k}) - Df_{(a,\vec{b})}(h,0)\|}{\|(h,\vec{k})\|} \\
&+ \frac{\|f(a,\vec{b}+\vec{k}) - f(a,\vec{b}) - Df_{(a,\vec{b})}(0,\vec{k})\|}{\|(h,\vec{k})\|} \\
&\leq \frac{\|f(a+h,\vec{b}+\vec{k}) - f(a,\vec{b}+\vec{k}) - Df_{(a,\vec{b}+\vec{k})}(h,0)\|}{|h|} \\
&+ \frac{\|f(a,\vec{b}+\vec{k}) - f(a,\vec{b}) - Df_{(a,\vec{b})}(0,\vec{k})\|}{\|\vec{k}\|} + \frac{\|Df_{(a,\vec{b}+\vec{k})}(h,0) - Df_{(a,\vec{b})}(h,0)\|}{|h|} \\
&\leq \frac{\|f(a+h,\vec{b}+\vec{k}) - f(a,\vec{b}+\vec{k}) - Df_{(a,\vec{b}+\vec{k})}(h,0)\|}{|h|} \\
&+ \frac{\|f(a,\vec{b}+\vec{k}) - f(a,\vec{b}) - Df_{(a,\vec{b})}(0,\vec{k})\|}{\|\vec{k}\|} + \|Df_{(a,\vec{b}+\vec{k})} - Df_{(a,\vec{b})}\|.
\end{aligned}
$$

By continuity of the derivative at $(a, \vec{b})$ and the induction hypothesis, the last two terms go to zero as $(h, \vec{k}) \to (0, \vec{0})$. For the first term, we use Lagrange's mean value theorem to conclude that it is equal to

$$\frac{\|Df_{(\theta,\vec{b}+\vec{k})}(h,0) - Df_{(a,\vec{b}+\vec{k})}(h,0)\|}{|h|} \leq \|Df_{(\theta,\vec{b}+\vec{k})} - Df_{(a,\vec{b}+\vec{k})}\| \tag{2}$$

where $\theta \in (a-h, a+h)$. By the continuity assumption on Df at $(a, \vec{b})$, we see that this term goes to 0 and hence by the squeeze rule we are done. $\square$