# MA 200 - Lecture 4

## 1 Recap

1. $C^1$ implies differentiability.
2. Proved some properties of derivatives. (A few typos in the product rule but the concepts remain the same.)

## 2 Derivatives

More properties.

1. If $g(a) \neq 0$, then $\frac{f}{g}$ is diff at $a$ with derivative $\frac{g(a)\nabla f(a) - f(a)\nabla g(a)}{g^2(a)}$: WLog $f = 1$ (why?). Now for $\delta$ small enough, we see that for all $\|h\| < \delta$, $g(a+h) \neq 0$.
$$\frac{\left|\frac{1}{g(a+h)} - \frac{1}{g(a)} + \frac{Dg_a(h)}{g^2(a)}\right|}{\|h\|} = \frac{\left|\frac{g(a)-g(a)-Dg_a(h)-\Delta_2}{g(a+h)g(a)} + \frac{Dg_a(h)}{g^2(a)}\right|}{\|h\|}$$
$$\leq \frac{|\Delta_2|}{\|h\|}\frac{1}{|g(a+h)g(a)|} + \|\nabla g(a)\|\frac{1}{|g(a)|}\left|\frac{1}{g(a+h)} - \frac{1}{g(a)}\right| \to 0. \tag{1}$$
$\square$

2. If $f$ is constant, $\nabla f = 0$ (trivial). Conversely, if $U$ is a connected open set and $f$ is differentiable on all of $U$ with $\nabla f = 0$ identically, then $f$ is a constant.
Indeed, fix $a \in U$ and let $S = \{x \in U \mid f(x) = f(a)\}$. By continuity of $f$, $S$ is closed in $U$. It is clearly not empty. If $S$ is proven to be open, then $S = U$ by connectedness (why?).
$S$ is also open: To this end, consider an open ball $B$ around $b \in S$ that is wholly contained in $U$. We shall prove that $B \subset S$. Indeed, let $x \in B$. Then $b+t(x-b) \in B$ for all $t \in [0,1]$. The function $g(t) = f(b+t(x-b))$ is continuous on $[0,1]$ (because it is a composition of continuous functions). It is differentiable on $(0,1)$ and $g' = 0$: $\lim_{h\to 0}\frac{g(t+h)-g(t)}{h} = \lim_{h\to 0}\frac{f(b+t(x-b)+h(x-b)) - f(b+t(x-b))}{h} = \nabla_{x-b}f(b+t(x-b)) = \langle \nabla f(b+t(x-b)), x-b\rangle = 0$. By Lagrange's MVT, $g(1) = g(0)$ and hence $x \in S$. $\square$

3. If $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, then $F$ is differentiable and $DF(h) = F(h)$: $\frac{F(a+h) - F(a) - F(h)}{\|h\|} = 0 \; \forall\, h$. $\square$

Here is a concrete example where using the definition is as efficient as blindly calculating. Prove that $F : Mat_{n\times n} \to Mat_{n\times n}$ given by $F(A) = A^2$ is differentiable everywhere and calculate its derivative: The components of $F$ are polynomials (quadratic ones) in the entries of $A$ and are hence $C^1$ and therefore differentiable. Thus, so is $F$. We can calculate the derivative in two ways:

1. Blindly: $(DF_A(H))_{ij} = \sum_{k,l} \frac{\partial F_{ij}}{\partial A_{kl}} H_{kl} = \sum_{k,l,m} \frac{\partial A_{im}A_{mj}}{\partial A_{kl}} H_{kl} = \sum \delta_{ik}\delta_{ml}A_{mj}H_{kl} + \delta_{mk}\delta_{jl}A_{im}H_{kl} = \sum_{k,l} A_{lj}H_{il} + A_{ik}H_{kj} = \{H, A\}_{ij}$.
2. Using the definition: $F(A+H) - F(A) = (A+H)^2 - A^2 = \{A, H\} + H^2$ and hence $\dfrac{\|F(A+H) - F(A) - \{A,H\}\|_{Frob}}{\|H\|_{Frob}} \leq \|H\|_{Frob} \to 0$.

Here is another interesting example. Prove that $\det : Mat_{n\times n} \to \mathbb{R}$ is differentiable everywhere and that $\nabla_H \det(I) = tr(H)$: The determinant is a polynomial in the entries of the matrix and hence $C^1$ (and differentiable). This can be proven using properties of determinants. (Optional: A fun way to do this is to notice that $\det(I + tH) = (1 + t\lambda_1)(1 + t\lambda_2)\ldots$ where $\lambda_i$ are (possibly complex) eigenvalues of $H$.)

We need an analogue of the chain rule to conclude that $\sin(x^2 + y^2)$ is differentiable without using $C^1$ implies differentiability. More importantly, we need a formula for the derivative. This will help us answer two kinds of questions:

1. If a particle is zooming around in a room (whose temperature is $T(x, y, z)$) with trajectory $\vec{r}(t)$, then how fast does the temperature change according to it?
2. To solve partial differential equations that have radial symmetry, it helps to switch to polar coordinates. If we know the derivatives of a function $f(x, y)$ in Cartesian coordinates, how can we calculate those of $\tilde{f}(r, \theta) = f(r\cos(\theta), r\sin(\theta))$ in polar coordinates?

Intuitively, $T(x(t+h), y(t+h), z(t+h)) \approx T(x(t) + hx'(t), y(t) + hy'(t), z(t) + hz'(t)) \approx T(x(t), y(t), z(t)) + hx'(t)\frac{\partial T}{\partial x} + hy'(t)\frac{\partial T}{\partial y} + hz'(t)\frac{\partial T}{\partial z}$. Likewise, $f((r+h)\cos(\theta+k), (r+h)\sin(\theta+k)) - \tilde{f}(r,\theta) \approx f(r\cos(\theta) + h\cos(\theta) - rk\sin(\theta), r\sin(\theta) + h\sin(\theta) + rk\cos(\theta)) - \tilde{f}(r,\theta) \approx f_x(h\cos(\theta) - rk\sin(\theta)) + f_y(r\sin(\theta) + h\sin(\theta) + rk\cos(\theta)) = h(f_x x_r + f_y y_r) + k(f_x x_\theta + f_y y_\theta)$. In other words, we expect that $D\tilde{f} = DfD\vec{r}$ (as matrices). This expectation is very similar to the one-variable chain rule $(f \circ g)'(a) = f'(g(a))g'(a)$.

The rigorous statement of the chain rule is as follows. (One can easily see that the above examples and the one-variable chain rule are special cases of this general formulation.)

**Theorem 1.** *Let $A \subset \mathbb{R}^m, B \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^n$, $g : B \to \mathbb{R}^p$ with $f(A) \subset B$. Suppose $a$ is an interior point of $A$ and $f(a)$ is an interior point of $B$. If $f$ is differentiable at $a$ and $g$ is differentiable at $b$, then $g \circ f$ is differentiable at $a$. Moreover, $D(g\circ f)_a = Dg_{f(a)}Df_a$ (as multiplication of matrices).*

*Proof.* Since $b = f(a)$ is an interior point, $g(y)$ is defined on $|y - b| < \epsilon_1$ for some $\epsilon_1$. Since $f$ is diff, it is continuous. Hence, $|f(x) - b| < \epsilon_1$ for all $x \in A$ such that $|x - a| < \delta_1$

and $\delta_1$ can be chosen to be so small that $|x-a| < \delta_1$ is contained in $A$ (because $a$ is an interior point too).

Let $\Delta_1(h) = f(a+h) - f(a) - Df_a(h)$ and $\Delta_2(\tilde{h}) = g(b+\tilde{h}) - g(b) - Dg_b\tilde{h}$. By definition of the differentiability of $g$ at $b$, we see that for every $1 > \epsilon > 0$ there exists a $1 > \delta_1 > \delta_2 > 0$ such that $\|\Delta_2(\tilde{h})\| < \frac{\epsilon}{100+\|Dg_b\|+\|Df_a\|+\|Df_a\|\|Dg_b\|}\|\tilde{h}\|$ whenever $\|\tilde{h}\| < \delta_2$. Now $g(f(a+h)) - g(f(a)) = g(b + Df_a(h) + \Delta_1(h)) - g(b) = \Delta_2(Df_a(h) + \Delta_1(h)) + Dg_b(Df_a(h) + \Delta_1(h))$. Therefore, $\|g(f(a+h)) - g(f(a)) - Dg_bDf_a(h)\| \leq \|\Delta_2(Df_a(h) + \Delta_1(h))\| + \|Dg_b(\Delta_1(h))\| \leq \|\Delta_2(Df_a(h) + \Delta_1(h))\| + \|Dg_b\|\|\Delta_1(h)\|$.
To be continued.... $\square$