MA 200 - Lecture 5

1 Recap

1. Proved the quotient rule and started the proof of the chain rule.

2 Derivatives

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. We stopped at $\|g(f(a+h)) - g(f(a)) - Dg_b Df_a(h)\| \leq \|\Delta_2(Df_a(h) + \Delta_1(h))\| + \|Dg_b\|\|\Delta_1(h)\|$. Therefore, if $\|Df_a(h) + \Delta_1(h)\| < \delta_2 < \delta_1 < 1$, then the first term is less than $\frac{\epsilon}{100+\|Dg_b\|+\|Df_a\|+\|Df_a\|\|Dg_b\|}\|Df_a(h) + \Delta_1(h)\|$. By definition of differentiability of f at a, we see that $\|\Delta_1(h)\| < \frac{\epsilon}{100+\|Dg_b\|+\|Df_a\|+\|Df_a\|\|Dg_b\|}\|h\|$ whenever $\|h\| < \delta_3 < \delta_2 < \delta_1 < 1$. In other words,

$$\begin{aligned} &\|g(f(a+h)) - g(f(a)) - Dg_b Df_a(h)\| \\ &\leq \frac{\epsilon}{100 + \|Dg_b\| + \|Df_a\| + \|Df_a\|\|Dg_b\|}(\|Df_a(h) + \Delta_1(h)\| + \|Dg_b\|\|h\|) \\ &< \epsilon\|h\|. \end{aligned} \tag{1}$$

Why? □

Using the one-variable MVT, we can prove the multivariable one:

Theorem 2. *Let $U \subset \mathbb{R}^n$ be an open set and $a, b \in U$ be points such that the line segment $ta + (1-t)b \ \forall \ t \in [0,1]$ lies in U. Let $f : U \to \mathbb{R}$ be a differentiable function. Then $f(a) - f(b) = \langle \nabla f(\theta a + (1-\theta)b), a - b\rangle$ for some $\theta \in (0,1)$.*

Caution: This theorem is false for vector-valued functions!

Since we have defined differentiability for vector-valued functions, we can ask whether the derivative function Df is differentiable in its own right. We can also separately ask about each partial derivative being further partially differentiable and so on. There are examples (HW) of functions that are C^1, the second partials exist, but the mixed partials, i.e., $D_i D_j f = \frac{\partial^2 f}{\partial x_i \partial x_j}$ are not equal. To this end, we define the notion of being C^2: A function $f : U \to \mathbb{R}^m$ (where U is open) is said to be C^2 on U if

Df is C^1 on U (and hence Df is continuous and hence f is C^1 and continuous). More generally, f is C^r on U if all the $r-1$ partials exist and are C^1 (thus, the $r-1$ partials are differentiable and hence continuous. By induction, all the partials upto order r are continuous and f is r-times differentiable).

Theorem 3. *(Clairaut): If $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is C^2 on U (and U is open), then $D_i D_j f(a) = D_j D_i f(a)$ for all $i \neq j$.*

By the way, it is enough for f to merely be twice-differentiable (without being C^2) but the proof is more complicated.

Proof. Since we are interested in two variables at a time, without loss of generality, we can assume that $f(x, y)$ is a function of merely two variables.

We first prove a second-order mean value theorem. Let $\lambda(h, k) = f(a, b) - f(a+h, b) - f(a, b+k) + f(a+h, b+k)$. We shall prove that $\lambda(h, k) = D_2 D_1 f(p) hk = D_1 D_2 f(q) hk$ where p and q are two points in the rectangle with vertices $(a, b), (a+h, b), (a+h, b+k), (a, b+k)$. By symmetry it is enough to prove one of these equations:
Let $\phi(s) = f(s, b+k) - f(s, b)$. Then $\lambda(h, k) = \phi(a+h) - \phi(a) = \phi'(\theta) h = h(D_1 f(\theta, b+k) - D_1 f(\theta, b)) = hk D_2 D_1 f(\theta, \theta')$.

From these equations, $D_2 D_1 f(p) = D_1 D_2 f(q)$. Taking a sequence of h, k tending to $0, 0$, we see by continuity of the second partials that we are done. □

Proposition 2.1. *As a corollary, if f is C^r, then all of the mixed partials upto order r are equal.*

Proof. We induct on r. $r = 2$ is Clairaut. Assume truth for $2, 3, \ldots, r-1$. Then $D_i D_{j_1 \ldots j_k} f = (D_i D_{j_1})(D_{j_2} \ldots D_{j_k} f)$. Now $(D_{j_2} \ldots D_{j_k} f)$ is C^2 (why?) Thus by the C^2 Clairaut, $D_{j_1}(D_i D_{j_2} \ldots f)$ which means that i can be permuted to whichever position we want. We are done (why?) □

Here is a useful and intuitive result.

Proposition 2.2. *Let $U \subset \mathbb{R}^m, V \subset \mathbb{R}^n$ be open sets. Let $f : U \to V$ and $g : V \to \mathbb{R}^p$ be C^r functions. Then $g \circ f$ is C^r for $0 \leq r \leq \infty$.*

Proof. For $r = 0$ it is a standard property. Let $r = 1$. f, g are C^1 and hence differentiable. By the chain rule, $D(g \circ f)_x = Dg_{f(x)} Df_x$. By the properties of continuity, $D(g \circ f)$ is continuous and hence $g \circ f$ is C^1.

Assume inductively that the theorem has been proven for $1, 2, \ldots, r-1$. Now $D(g \circ f)_x = Dg_{f(x)} Df_x$ is C^{r-1} by the induction hypothesis (why?). Hence, $g \circ f$ is C^r.

If $r = \infty$, then applying the result for each finite r, we see that so $g \circ f$ is C^∞. □

Recall that one of the points of the one-variable chain rule was to calculate the derivatives of inverses if there existed. Indeed, assuming that $\sin^{-1}(x)$ is differentiable, since $\sin(\sin^{-1}(x)) = x$, by the chain rule, $\cos(\sin^{-1}(x))(\sin^{-1}(x))' = 1$. Thus, $(\sin^{-1}(x))' = \frac{1}{\sqrt{1-x^2}}$. Likewise,

Proposition 2.3. *Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be differentiable at an interior point $a \in U$. Suppose $f(a)$ is an interior point of $f(U)$ and $f : V \to f(V)$ (where V is a neighbourhood of a) is $1-1$, onto, $f(V)$ is open, and $f^{-1} : f(V) \to V$ is differentiable at $f(a)$. Then $Df(a)$ is invertible and $Df^{-1}(f(a)) = (Df(a))^{-1}$.*

Proof. Since $f(f^{-1}(x)) = x$, by the chain rule at $f(a)$, $Df|_a Df^{-1}|_{f(a)} = I$. Hence we are done. □