MA 200 - Lecture 8

1 Recap

1. We proved Clairaut's theorem (and generalised it to more than two derivativeS).
2. Applications of the chain rule - composition of functions and derivative of the inverse. Started the inverse function theorem for single-variable functions.

The image $f(U)$ is open and hence f^{-1} is continuous near $f(a)$. (why?)
What about f^{-1} being differentiable near $f(a)$? $\frac{f^{-1}(f(b)+h)-b}{h} = \frac{k}{h} = \frac{1}{f'(\theta)} \to \frac{1}{f'(b)}$. Using the formula for the derivative, one can show that f^{-1} is C^1. $\square$
Can we generalise this theorem to multivariable calculus? Other than differentiability of the inverse, is there any other point? The answers are 'yes'. But before that, if f is not C^1, this theorem is false. For instance, $f(x) = x + 2x^2 \sin(1/x)$ when $x \neq 0$ and $f(0) = 0$ is not locally invertible near 0.

Theorem 1. *Inverse function theorem Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be a C^r ($\infty \geq r \geq 1$) function on an open set U. If Df_a is invertible, then there is a neighbourhood V of a such that $f(V)$ is open, $f : V \to f(V)$ is 1-1, onto, and $f^{-1} : f(V) \to V$ is C^r.*

What is the point? Well, even in one-variable, here is an argument to show that for all c sufficiently close to 1, there exists an x such that $x + \frac{1}{100}e^{\sin(x)} - \frac{e^{\sin(1)}}{100} = c$. Not that $f(x) = x + \frac{1}{100}e^{\sin(x)} - \frac{e^{\sin(1)}}{100}$ satisfies $f(1) = 1$ and $f'(1) = 1 + \frac{1}{100}e^{\sin(1)}\cos(1) \neq 0$. Hence by IFT, f is locally invertible and we are done. In higher dimensions, basically, if we want to solve a *nonlinear* system of equations (same number of variables and unknowns) near $f(x_0) = y_0$, then the ability to solve upto first order (that is, *linear* equations) is enough (roughly speaking)!

Proof. Let us first prove the theorem for $r = 1$:

Proof. 1. f is locally $1-1$: Basically, f is locally like multiplication by an invertible matrix and hence $1-1$. We can make this precise by proving a stronger result that there exists an $\alpha > 0$ such that $|f(x_0) - f(x_1)| \geq \alpha |x_0 - x_1|$ for all x_0, x_1 in an open ball centred at a. (Why does this result imply local injectivity?) Zeroethly, why is this true even if $f(x) = Ax$ where A is an invertible matrix? This is because, $f(x_0) - f(x_1) = A(x_0 - x_1)$ and hence $A^{-1}(f(x_0) - f(x_1)) = x_0 - x_1$ and thus $\frac{1}{\|A^{-1}\|_{Frobenius}}|x_0 - x_1| \leq |f(x_0) - f(x_1)|$. Indeed, firstly choose some open ball $B_a(r)$ (and we will shrink this ball if necessary). Now $f(x_1+h) - f(x_1) = Df_{x_1}h + \Delta$ where $x_1, x_1+h \in B_a(r)$. Choose the open ball to be so small that Df_x is invertible throughout the open ball and $\|Df_x^{-1}\| \leq C$ (why is this possible?) throughout the ball (and call the radius of the shrunk ball to be r abusing notation). Thus $Df_{x_1}^{-1}(f(x_1+h)-f(x_1)) = h+Df_{x_1}^{-1}\Delta$. Thus $\|h+Df_{x_1}^{-1}\Delta\| \leq C\|f(x_1+h)-f(x_1)\|$. Now choose r to be so small that $\|h + Df_{x_1}^{-1}\Delta\| \geq \frac{\|h\|}{2}$ for all $\|h\| < 2r$ (HW - why can this be done?). Thus we are done (why?). □(Another way to do this is to take $H(x) = f(x) - Df_a(x)$ and use the mean-value-theorem for each component of H.)

2. The image of a neighbourhood of f is open: We wish to prove that every point near $f(a)$ lies in the image of f, i.e., the image of f contains an open ball $B_{f(a)}(r')$. Then $f^{-1}(B_{f(a)}(r')) \cap B_a(r) = U$ will be the desired neighbourhood whose image is open (and one where f is $1-1$). That is, we want to produce an $r' > 0$ such that for every $b \in B_{f(a)}(r')$, we can solve $f(x) = b$ to get an x. There are two ways of doing this (one of them is the usual way and the other is in the textbook):

   (a) Iteration/contraction mapping principle: We can try Newton's method for finding a "root" of the equation $f(x) = b$, i.e., we choose an initial guess $x_1 = a$, and try the iterative scheme $x_{n+1} = x_n - Df(x_n)^{-1}(f(x_n) - b)$ if the later makes sense. Firstly, recall that on $B_a(r)$, $\|(Df)^{-1}\| \leq C$. If we choose r' (which is $> \|f(a) - b\|$) to be small enough, we claim that all the x_n belong to $B_a(r'')$ where $r'' < r$ and that $f(x_n) \in B_b(r')$. The rough idea is that $f(x_{n+1}) - b \approx (f(x_n) - b) - (f(x_n) - b) = 0$ and hence less than r'. (Moreover, $x_{n+1} - x_n \approx 0$ and hence the geometric series sum will show that x_n is close to a for all n.)
   Firstly, there exists $r'' < r$ and (how? - HW exercise) $y, z \in B_a(r'')$, we have $\|Df(y)(Df)^{-1}(z) - I\|_{Frobenius} < \epsilon = \frac{1}{2}$. (The choice of this ϵ comes from hindsight rather than foresight.)
   Secondly, $x_2 = a - (Df(a))^{-1}(f(a) - b)$ and hence $\|x_2 - a\| \leq C\|f(a) - b\| < Cr' < r''$ if $r' = \frac{r''}{4C}$ (again from hindsight). (To be continued....)

□

□