MA 200 - Lecture 12

1 Recap

1. Stated and proved the implicit function theorem.
2. Provided examples and non-examples (square roots of matrices, level sets and normals).

2 Implicit function theorem

Examples/Non-examples:

1. More generally, consider $k \leq n$ C^1 functions, $f_i : U \to \mathbb{R}$. Suppose $\nabla f_i(a)$ are all linearly independent. Then we can locally (near a) solve for k variables in terms of the others. Moreover, the vectors $\nabla f_i(a)$ are all normals at a to the resulting set (in the same sense as before). The tangent space is the intersection of the planes $\langle \nabla f_i(a), \vec{r} - \vec{a} \rangle = 0$.

2. Going by the philosophy that diffeomorphisms represent change of coordinates/frames of reference, one may ask what f would "look like" in the "new coordinates" (y, b) (obtained by solving $f(y, x) = b$ for x in terms of y, b), that is,

   Theorem 1 (The surjective derivative theorem)**.** *Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}^p$ be a C^r function $(1 \leq r \leq \infty, 1 \leq p \leq n)$. Suppose $f(a) = 0$ and Df_a has rank p, i.e., it is a surjective linear map. Then there is an open neighbourhood $A \subset U$ of a and a C^r-diffeomorphism $h : A \to h(A)$ such that $f \circ h(x_1, \ldots, x_n) = (x_{n-p+1}, \ldots, x_n)$.*

   Proof. If Df_a is a surjective linear map, then by permuting the coordinates, we can assume WLOG that the rank of the last $p \times p$ minor to be full. Then the implicit function theorem kicks in to show that we can solve $f(y) = b$ for the last p coordinates $y_{n-p+1}, \ldots$ in terms of the first $n - p$ ones $y_1, \ldots, y_{n-p}$ and b in a C^r manner. Now consider $h(y_1, \ldots, y_{n-p}, b) = (y_1, \ldots, y_{n-p}, y_{n-p+1}(y, b), \ldots)$. Now $f \circ h((y, b)) = b$ □

   That is, "upto diffeomorphisms (change of coordinates)", secretly every map whose derivative is surjective is simply a projection.

3 Global extrema

Let $U \subset \mathbb{R}^n$ be an open set such that $\bar{U}$ is compact. Suppose $f : \bar{U} \to \mathbb{R}$ is a continuous function. Then it assumes a global maximum and a global minimum. Our task is to find them (this kind of a question arises in optimisation, in proving inequalities, etc). We have already seen one lemma that helps us: If an extremum occurs at an interior point p, and f is differentiable at p, then $\nabla f(p) = 0$. (This motivates a definition: An interior point p is called a local minimum/maximum if there exists a neighbourhood of p such that f restricted to that neighbourhood assumes a global min/max at p.) Thus, "all" we have to do is to find the 'critical points' (interior points where either f fails to be differentiable or has zero gradient) and look at the extrema of f on the boundary to deduce the global extrema. Here is an example:

Let $f(x, y) = xy$ on $x^2 + y^2 \leq 1$. Firstly, the domain is compact (why?) and the function is continuous (in fact, it is smooth on all of $\mathbb{R}^2$). The derivative is $\nabla f = (y, x) = (0, 0)$ precisely at the origin (where $f(0, 0) = 0$). On the boundary, i.e., $x^2 + y^2 = 1$, $f(x, y) = g(\theta) = \cos(\theta)\sin(\theta)$ over $[0, 2\pi]$. We can find the global extrema of this function (either by using calculus systematically or by cleverness): $-\frac{1}{2} = g(3\pi/4) \leq g(\theta) \leq \frac{1}{2} = g(\pi/4)$.

In other words, finding the global extrema involves *constrained* optimisation, i.e., optimising over level sets $g = 0$. Of course, one can attempt to solve the constraints and therefore reduce the number of variables and inductively attempt to reduce the problem to one dimension (as we did in the example above). However, this sort of a strategy will not always work simply because we cannot always solve the constraints explicitly. But what if we know that there is an *implicit* solution?

Theorem 2. *Let $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$ be C^1 functions (on an open set U). Assume that $a \in U$ is a point of global max/min of f subject to the constraint $g = 0$. Suppose $\frac{\partial g}{\partial x_n}(a) \neq 0$. Then $\nabla f(a) = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$ (called a Lagrange multiplier).*

Note that there is nothing special about x_n. The same theorem would have worked if we could have solved for any of the other variables locally.

Proof. By the implicit function theorem, , there exists a neighbourhood W of $(a_1, \ldots, a_{n-1})$ in $\mathbb{R}^{n-1}$, a neighbourhood $V = W \times (a_n - \epsilon, a_n + \epsilon) \subset U$ of a, and a C^1 function $h(x_1, \ldots, x_{n-1}) : W \to \mathbb{R}$ such that $g(x) = 0$ iff $x_n = h(x_1, \ldots, x_{n-1})$ on V.

The function $s(x) = f(x_1, \ldots, x_{n-1}, h)$ attains a local extremum at $a_1, \ldots, a_{n-1}$. Thus, $\frac{\partial s}{\partial x_i} = \frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial x_n}\frac{\partial h}{\partial x_i} = 0$ at $a_1, \ldots, a_{n-1}$. To be continued...... $\square$