MA 200 - Lecture 13

1 Recap

1. Intersections of level sets.
2. Surjective derivative theorem.
3. Global extrema. Started the proof of Lagrange's multipliers.

2 Global extrema

Theorem 1. *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 $a_1, \ldots, a_{n-1}$. Moreover, since g is identically zero, $\frac{\partial g}{\partial x_i} + \frac{\partial g}{\partial x_n}\frac{\partial h}{\partial x_i} = 0$ at $a_1, \ldots, a_{n-1}$. Thus, we are done (why?). □

Here is an example: Prove that for non-negative numbers $x_1, \ldots, x_n$, $AM \geq GM$ with equality holding precisely when they are all equal to each other.
Consider $f(x) = x_1 \ldots x_n$ and $g(x) = x_1 + x_2 + \ldots + x_n - c$ (where $c \geq 0$). f and g are C^1 on all of $\mathbb{R}^n$. We want to find the global maximum of f subject to $g = 0$. Firstly, since $g = 0$, we see that $0 \leq x_i \leq c \; \forall \, i$. Hence we consider the compact rectangle $Q = [0, c]^n$ and f restricted to Q (subject to $g = 0$) achieves a global maximum at some point a (why?) Now either that point lies on the boundary of Q: irrelevant (why?) or it lies in the interior. If it lies in the interior, then since $\nabla g \neq 0$, we can use LM. Thus $\nabla f = \lambda \nabla g$ and we are done (why?) □
What we really needed in the above proof of Lagrange's multipliers is simply the reduction of the problem to an unconstrained problem of fewer variables. For instance, the following theorem can be proven too:

Theorem 2. *Let $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$ be C^1 functions on an open set U. Let $\alpha : V \subset \mathbb{R}^{n-1} \to U$ be a $1-1$ C^1 function such that $g(y) = 0$ on U if and only if $y = \alpha(x)$. Suppose $a \in U$ is a point of global max/min of f when restricted to $g = 0$. Assume that $D\alpha_{\alpha^{-1}(a)}$ is an injective linear map. Then $\nabla f(a)$ and $\nabla g(a)$ are parallel.*

Proof. The point $\alpha^{-1}(a)$ is a point of local extremum of $h = f \circ \alpha$ on V. Thus $\nabla h(\alpha^{-1}(a)) = 0 = Df_a D\alpha_{\alpha^{-1}(a)}$. Moreover, since $g \circ \alpha = 0$ identically on V, $Dg_a D\alpha_{\alpha^{-1}}(a) = 0$. By the nullity-rank theorem, the dimension of the row-space is also $n-1$. Thus, the kernel of $v \to vD\alpha_{\alpha^{-1}(a)}$ is one-dimensional. Since $\nabla f(a)$ and $\nabla g(a)$ are in the kernel, they are parallel. □

In the HW, you will see a generalisation of this method to more than one constraint.

It appears that the key to make the method of Lagrange's multipliers work is the ability to parametrise the level set in a nice way *near* the extremum. Can we always parametrise level sets like this? Unfortunately this cannot be done in C^∞ manner.
For instance, consider the level set $x^2 - y^3 = 0$. Suppose we could regularly smoothly parametrise $(x(t), y(t))$ this level set near 0. Then $2xx' = 3y^2y' = 3x^{4/3}y' \Rightarrow 2x' = 3x^{1/3}y'$. Suppose $x' \neq 0$. Then we have a contradiction near the origin. Suppose $y' \neq 0$ near the origin. Then taking two derivatives enables us to come up with a contradiction.
The following result is useful to understand the hypotheses of the theorem better.

Theorem 3 (The injective derivative theorem)**.** *Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}^p$ be a C^r function ($1 \leq r \leq \infty, n \leq p$). Suppose Df_a is an injective linear map. Then there is an open neighbourhood $A \subset \mathbb{R}^n$ of a, $B \subset \mathbb{R}^{n-p}$ of $f(a)$, and C^r-diffeomorphisms $h : A \to h(A)$, $g : B \to g(B)$ such that $g \circ f \circ h(x_1, \ldots, x_n) = (x_1, \ldots, x_n, 0, 0, \ldots)$ when $x \in A$.*

Proof. By means of translations, we can assume that $a = f(a) = 0$ WLOG. By means of permuting the coordinates, since Df_a is injective, WLOG, we can assume that $\frac{\partial(f_1,\ldots,f_n)}{\partial(x_1,\ldots,x_n)}(a)$ is invertible (why?).
Now by IFT (how), the map $G(x, y) = (f_1(x), \ldots, f_n(x), f_{n+1}(x) - y_1, f_{n+2}(x) - y_2, \ldots)$ is a local C^r diffeomorphisms where $y \in \mathbb{R}^{p-n}$. Now $G^{-1} \circ f = (x_1, \ldots, x_n, 0, 0 \ldots)$. □

That is, "upto diffeomorphisms (change of coordinates)", secretly, every map f whose derivative is injective is simply the inclusion of the coordinate axes.

There is one subtle point though: If you consider $g(y) = 0$ as a 'figure-8', and $\alpha : \mathbb{R} \to \mathbb{R}^2$ with image as the figure-8, then α cannot be a homeomorphism to its image (why not?) We want to avoid such weird level sets (simply to be able to say "near every point, the level set looks like an open subset of $\mathbb{R}^n$).