Section 6: Optimization in Higher Dimensions

In the past two weeks, you've been learning how to do optimization for certain classes of functions in higher dimensions. Today, we'll discover some quick proofs about the theorems you've learnt, and develop a method to use Lagrange multipliers.

1. First, we recall **Taylor's theorem** for functions $f : \mathbb{R} \to \mathbb{R}$ that are infinitely differentiable...

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + ... + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_k(x, x_0)$$

where the remainder $R_k(x, x_0) = \frac{f^{k+1}(z)}{(k+1)!}(x - x_0)^{k+1}$ for some z between x_0 and x. Special cases include the first and second-order Taylor polynomial:

$$f(x) = f(x_0) + f'(z)(x - x_0) \tag{1}$$

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(z)}{2}(x - x_0)^2 \tag{2}$$

You already know this.

2. We now generalize Taylor's theorem for twice differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$.

$$f(x_0 + h) = f(x_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0) \cdot h_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \frac{\partial f^2}{\partial x_i \partial x_j}(x_0) + R_2(x_0, h)$$

where the remainder $R_2(x_0, h)$ tends to 0 as $||h||$ approaches 0. As it turns out, the remainder takes on a similar form as in the 1-dimensional case, i.e. in terms of a higher derivative. However, we won't mention its explicit formulation because (a) it's annoying to represent, and (b) we don't require it.

Firstly, note we can re-express this as,

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2} H_{x_0}(h) + R_2(x_0, h)$$

And thereafter, derive the first and second-order Taylor polynomials...

$$f(x_0 + h) = f(x_0) + \nabla f(z) \cdot h \tag{T_1}$$

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2} H_z(h) \tag{T_2}$$

where z is between x_0 and $x_0 + h$. Recall, we define H_x as the Hessian function, which is a quadratic function: it is expressible as $H_x(h) = \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \frac{\partial f^2}{\partial x_i \partial x_j}(x)$. Think of it as the $n-$dimensional analogue of the second derivative.

This is a very useful formula, because **it allows us to approximate a function with its first and second derivatives**. As it turns out, this suffices to prove the theorems we care about.

1. **First derivative test for local extremum:** Show that, if $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and has a local minima at x_0, then $\nabla f(x_0) = 0$.

(A1) For sake of contradiction, assume that $\nabla f(x_0) \neq 0$. The idea is to use the first-order Taylor approximation and gradient descent to show there x_0 is not a local minima.

(A2) Consider the first-order Taylor approximation about x_0...

$$f(x_0 + h) = f(x_0) + \nabla f(z) \cdot h$$

(A3) Since the first-order partial derivatives are assumed continuous, we may find a neighbourhood around x_0 such that entries of $\nabla f(z)$ are close to $\nabla f(x_0)$. This would imply that $\nabla f(z) \cdot \nabla f(x_0) \approx \nabla f(x_0) \cdot \nabla f(x_0) \approx ||f(x_0)||^2 > 0$.

(A4) Now, we do gradient descent. Specifically, we shift to a point $x_1 := x_0 - k\nabla f(x_0)$ in this neighbourhood. What is k? It is any $k > 0$ small enough such that we stay in the neighbourhood from (A3).

(A5) Then, we have by $(A2)$ that,

$$f(x_0 + k\nabla f(x_0)) = f(x_0) - k\nabla f(z) \cdot \nabla f(x_0)$$

But note, $\nabla f(z) \cdot \nabla f(x_0) > 0$ by (A3), for all z between x_0 and x_1. But this means that $f(x_1) < f(x_0)$, which is contradictory.

The same works to show that local maxima implies that $\nabla f(x_0) = 0$.

2. **Second derivative test for local minima:** Show that, if $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$, $\nabla f(x_0) = 0$ for some x_0 and the Hessian H_{x_0} is positive definite, then x_0 is a local minimum.

(A1) This time, we directly show the proof by using the second-order Taylor polynomial w/ remainder.

(A2) Consider

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2} H_z(h)$$

where $H_z(h) = \sum_{i,j=1}^{n} h_i h_j \frac{\partial f^2}{\partial x_i \partial x_j}(z)$ for z between x_0 and $x_0 + h$.

(A3) Since $\nabla f(x_0) = 0$, we rewrite above as...

$$f(x_0 + h) = f(x_0) + \frac{1}{2} H_z(h)$$

(A4) But now, the argument is simple. If the Hessian is positive definite at x_0, then it is positive definite at z sufficiently close to x_0 (by continuity of the second-order partial derivatives). But this means that $H_z(h) > 0$ for z close enough to x_0.

(A5) But by (A4), $f(x_0 + h) > f(x_0)$ for all h such that $||h||$ is sufficiently small, i.e. the point in consideration is close enough to x_0. This proves that x_0 is a local minimum of f.

The same works to show that negative definite Hessian implies that x_0 is a local maximum.

1. **Single constraint** : Find global minima and maxima of $x^4 + y^4$ in the region $x^2 + y^2 \leq 1$.

2. **Multiple Constraints** : Determine the points that are on the cylinder with equation $x^2 + y^2 = 1$and the plane $x + y + z = 1$ whose distance from the origin is maximum or minimum.