These notes were prepared by Ariel Goodwin for MATH 248 at McGill University as taught by Pengfei Guan.

1 Higher Order Derivatives

The Mean Value Theorem for functions of one variable is an extremely powerful and fundamental result in analysis, allowing us to understand behaviour of a function on an interval using the behaviour of its derivative. The result generalizes to real-valued functions defined on balls in $\mathbb{R}^n$.

Theorem 1: Mean Value Theorem

Let $B \subseteq \mathbb{R}^n$ be a ball, and suppose $f\colon B \to \mathbb{R}$ is differentiable on B. Then for all $x, y \in B$ there exists $z \in B$ such that

$$f(x) - f(y) = Df(z)(x - y)$$

where $Df(z)$ is the Jacobian of f at z. In particular,

$$|f(x) - f(y)| \leq \|Df(z)\|\|x - y\|$$

This is slightly restrictive since we require the range of f to be $\mathbb{R}$. If we allow the range of f to be $\mathbb{R}^m$ we are no longer guaranteed that the equality in Theorem 1 will hold, but we still have a useful inequality:

Theorem 2: Mean Value Inequality

Let $B \subseteq \mathbb{R}^n$ be a ball, and suppose $f\colon B \to \mathbb{R}^m$ is differentiable on B. Then for all $x, y \in B$ there exists $z \in B$ such that

$$\|f(x) - f(y)\| \leq \|Df(z)\|\|x - y\|$$

Recall that if $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable then it has partial derivatives $\dfrac{\partial f_i}{\partial x_j}$ for all $1 \leq i \leq m$, $1 \leq j \leq n$. We can then discuss the differentiability of these functions, which leads us to the concept of **higher order derivatives**.

Definition 1: C^k

Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with Ω open. We say that $f \in C^k(\Omega)$ for an integer $k \geq 0$ if all of the partial derivatives of f up to order k exist and are continuous.

An important special case: $f\colon \Omega \to \mathbb{R}$, $f \in C^2(\Omega)$, we define the **Hessian** of f

at x to be:

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f_1}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f_1}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f_1}{\partial x_1 \partial x_n}(x) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f_1}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f_1}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f_1}{\partial x_n \partial x_n}(x) \end{pmatrix}$$

Higher order partial derivatives can quickly become unwieldy if we have to consider all the possible orders in which they can be taken. Here we state a simple condition under which we can forget about the order and are guaranteed that the result will be the same no matter what.

Theorem 3: Equality of Mixed Partials

Suppose $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. If $x \in \Omega$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$, $\frac{\partial^2 f}{\partial x_j \partial x_i}$ are continuous at x then

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

In particular, if all of the second order partial derivatives of f are continuous at f then $H_f(x)$ is symmetric.

As a curiousity, the converse is not true. The function $f(x, y) = x^2 y^2 \sin(1/x) \sin(1/y)$ can be continuously extended to the lines $x = 0$ and $y = 0$, and it is an interesting exercise to show that $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ while these functions are not continuous at $(0, 0)$. I found this example here: `https://math.stackexchange.com/questions/2095484/clairauts-theorem-is-reversible`

Taylor's theorem in one variable allows us to understand the local behaviour of a function near a point in terms of a polynomial with a remainder term. This can be generalized to higher dimensions, and the general formula is unwieldy but we will focus mainly on the first-order and second-order approximations (meaning in terms of first and second-order partial derivatives).

Theorem 4: Taylor's Theorem

Suppose $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is in $C^{k+1}(\Omega)$ and $x_0 \in \Omega$. Then

$$\begin{aligned} f(x) &= f(x_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0)(x^i - x_0^i) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)(x^i - x_0^i)(x^j - x_0^j) + \cdots \\ &\quad + \frac{1}{k!} \sum_{\alpha_{i_1} + \cdots + \alpha_{i_k} = k} \frac{\partial^k f}{\partial x_{\alpha_{i_1}} \partial x_{\alpha_{i_k}}}(x_0)(x^{\alpha_{i_1}} - x_0^{\alpha_{i_1}}) \cdots (x^{\alpha_{i_k}} - x_0^{\alpha_{i_k}}) \\ &\quad + R_{k+1}^f(x) \end{aligned}$$

where

$$R_{k+1}^f(x) = \frac{1}{(k+1)!} \sum_{\alpha_{i_1} + \cdots + \alpha_{i_{k+1}} = k+1} \frac{\partial^{k+1} f}{\partial x_{\alpha_{i_1}} \cdots \partial x_{\alpha_{i_{k+1}}}}(x_0)(x^{\alpha_{i_1}} - x_0^{\alpha_{i_1}}) \cdots (x^{\alpha_{i_{k+1}}} - x_0^{\alpha_{i_{k+1}}}) \quad (1)$$

which satisfies

$$\lim_{x \to x_0} \frac{R_{k+1}^f(x)}{\|x - x_0\|^k} = 0$$

The R_k^f term represents the error, and the condition at the bottom essentially tells us that the error goes to zero as we get closer to the point x_0 that we are expanding around.

The most important special cases are when $k = 0$ and $k = 1$:

$$f(x) = f(x_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0)(x^i - x_0^i) + R_1^f(x)$$

$$f(x) = f(x_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0)(x^i - x_0^i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)(x^i - x_0^i)(x^j - x_0^j) + R_2^f(x)$$

More compactly:

$$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + R_1^f(x)$$

$$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T H_f(x_0)(x - x_0) + R_2^f(x)$$

2 Optimization

An important application of derivatives is for optimizing real-valued functions. Many of the optimization theorems in one dimension have a corresponding result in higher dimensions, and we collect these results here. We will develop more sophisticated methods to solve constrained optimization problems in the coming lectures.

Definition 2: Local Extrema

Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$. We say that $x_0 \in \Omega$ is a **local maximum (minimum) point** if there exists $r > 0$ such that

$$f(x) \leq f(x_0) \quad (f(x) \geq f(x_0) \text{ respectively})$$

for all $x \in B_r(x_0)$.

Definition 3: Critical Points

Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$. We say that $x_0 \in \Omega$ is a **critical point** if f is differentiable at x_0 and $\nabla f(x_0) = 0$.

Theorem 5: First Derivative Test

Suppose $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at x_0. If x_0 is a local maximum or minimum point of f then $\nabla f(x_0) = 0$.

Definition 4: Definite Matrix

A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is called **positive semi-definite**, written $A \succeq 0$, if

$$x^T A x \geq 0$$

for all $x \in \mathbb{R}^n$. We say that A is **positive definite** if the inequality is strict for all non-zero $x \in R^n$. We say that A is **negative (semi)-definite** if $-A \succeq 0$.

Intuitively, positive/negative (semi)-definite matrices are like non-negative/non-positive numbers (strictly non-zero in the definite case). The definiteness of the Hessian plays a role analogous to that of the sign of the second derivative when we are looking at critical points.

A symmetric matrix A is positive/negative (semi)-definite if and only if all the eigenvalues are non-negative/non-positive respectively, and also strictly non-zero in the definite case. This is a nice result but not useful for quick computations, so we point out two useful tricks for checking when small matrices are positive definite.

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \succ 0 \iff \det(A) > 0 \text{ and } a > 0$$

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \succ 0 \iff a > 0 \text{ and } \det \begin{pmatrix} a & b \\ d & e \end{pmatrix} > 0 \text{ and } \det(A) > 0$$

Theorem 6: Second Derivative Test

Suppose $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is C^3 in a neighbourhood of $x_0 \in \Omega$.

a) If x_0 is a local maximum (minimum) point of f then $H_f(x_0)$ is negative (positive) semi-definite.

b) If x_0 is a critical point of f and $H_f(x_0)$ is negative (positive) definite then x_0 is a local maximum (minimum) point of f.

3 Exercises

1. Show that the Mean Value Theorem can fail when the range is no longer one-dimensional. Hint: Consider the function $f(x) = (\cos x, \sin x)$.
2. (Marsden & Tromba 3.1 #32) Let

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

   Compute $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ for $(x,y) \neq (0,0)$, and at the origin. Compute $\dfrac{\partial^2 f}{\partial x \partial y}$ and $\dfrac{\partial^2 f}{\partial y \partial x}$ at $(0,0)$ and compare. Explain the result.
3. (Marsden & Tromba 3.1 #17) Suppose $f(x,y,z)$ is of class C^3. Show that

$$\frac{\partial^3 f}{\partial x \partial y \partial z} = \frac{\partial^3 f}{\partial y \partial z \partial x}$$

4. (Marsden & Tromba 3.2 #11) Let $g(x,y) = \sin(xy) - 3x^2 \log y + 1$. Find the degree 2 polynomial which best approximates g near the point $(\pi/2, 1)$.
5. (Marsden & Tromba 3.3 #5) Find and classify the critical points of $f(x,y) = e^{1+x^2-y^2}$.
6. (Marsden & Tromba 3.3 #17) Find all local extrema for $f(x,y) = 8y^3 + 12x^2 - 24xy$.
7. (Marsden & Tromba 3.3 #18) Let $f(x,y,z) = x^2 + y^2 + z^2 + kyz$. Verify that $(0,0,0)$ is a critical point and determine all k such that $(0,0,0)$ is a local minimum.
8. (Marsden & Tromba 3.3 #46) We say $u \in C^2(\overline{B_1(0)})$ is strictly subharmonic if

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} > 0$$

   Show that u cannot have a maximum point in $B_1(0)$.