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

1 Continuity and Differentiability

Perhaps the most fundamental definitions in this course, the ideas of continuity and differentiability are essential.

Definition 1: Continuity

A function $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **continuous** at $x_0 \in \Omega$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|f(x_0) - f(y)\| < \varepsilon$$

provided that $\|x_0 - y\| < \delta$. We say f is **continuous** on Ω if f is continuous at each $x \in \Omega$.

Definition 2: Differentiability

A function $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** at $x_0 \in \Omega$ if there exists a linear map $L \in \mathbb{R}^{m\times n}$ such that

$$\lim_{x\to x_0} \frac{\|f(x) - f(x_0) - L(x - x_0)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} = 0$$

We say f is **differentiable** on Ω if f is differentiable at each $x \in \Omega$.

Theorem 1: Jacobian Properties

If $f = (f_1, \dots, f_m)\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x_0 \in \Omega$ then f is continuous at x_0, all partial derivatives $\dfrac{\partial f_i}{\partial x_j}(x_0)$ exist for $i = 1, \dots, m$, $j = 1, \dots, n$, and the matrix

$$Df(x_0) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(x_0) & \cdots & \dfrac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(x_0) & \cdots & \dfrac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}$$

of partial derivatives of f at x_0 is the unique linear map L in the definition of the differentiation. This matrix $Df(x_0)$ is called the **Jacobian** or differential or derivative matrix.

Keep in mind that differentiability implies existence of partial derivatives, but existence of partial derivatives does not imply continuity.

Theorem 2: Continuity of Partial Derivatives is Sufficient

If $f = (f_1, \dots, f_m) \colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is such that all partial derivatives $\frac{\partial f_i}{\partial x_j}$ for $i = 1, \dots, m$, $j = 1, \dots, n$ exist and are continuous at x_0, then f is differentiable at x_0.

Do not forget that this condition is not necessary - there are functions that are differentiable at a point and their partial derivatives are not continuous at that point.

Theorem 3: Chain Rule

If $f \colon \Omega_1 \subseteq \mathbb{R}^n \to \Omega_2 \subseteq \mathbb{R}^m$ and $g \colon \Omega_2 \subseteq \mathbb{R}^m \to \mathbb{R}^k$ are such that f is differentiable at $x \in \Omega_1$ and g is differentiable at $f(x) \in \Omega_2$ then $g \circ f$ is differentiable at x and

$$D(g \circ f)(x) = Dg(f(x))Df(x)$$

2 Mean Value Theorem, Taylor's Theorem

The Mean Value Theorem and its corresponding inequality are vital results that allow us to understand function behaviour in terms of derivative behaviour.

Theorem 4: 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,

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

Theorem 5: 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

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

Two important theorems on higher-order derivatives: equality of mixed partials allows us to take partial derivatives in whatever order we please, and Taylor's theorem allows us to approximate functions using polynomials with an explicit form for the error.

Theorem 6: Equality of Mixed Partials

Suppose $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. If $x \in \Omega$ and $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$, $\dfrac{\partial^2 f}{\partial x_j \partial x_i}$ 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 the Hessian $H_f(x)$ is symmetric.

Theorem 7: 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

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

where

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

which satisfies

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

3 Optimization

The theory of critical points and the derivative tests to find and classify critical points is the cornerstone of smooth optimization.

Theorem 8: 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 $\nabla f(x_0) = 0$.

Theorem 9: 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.

Don't forget that saddle points can also occur at critical points.

Definition 3: Saddle Point

Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x_0 \in \Omega$ a critical point of f. If x_0 is neither a maximum nor a minimum we say that x_0 is a **saddle point**. In particular, for all $r > 0$ we can find points x_1 and x_2 in $B_r(x_0)$ with

$$f(x_1) < f(x_0) < f(x_2)$$

Theorem 10: Saddle Point Existence

Let $f\colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be C^3 and x_0 a critical point of f. If $\det(H_f(x_0)) \neq 0$ and $H_f(x_0)$ is neither positive definite nor negative definite then x_0 is a saddle point.

All of the techniques above deal with optimization in open sets. To handle optimization on compact sets that include their boundary, we need some more machinery. There are two main strategies for optimizing the function on the boundary.

1. If the boundary is sufficiently simple, for example a rectangle or triangle in $\mathbb{R}^2$ that is piecewise linear, then we can restrict our function to the lines defining the boundary and reduce it to a problem of 1-dimensional optimization.

2. If the boundary is given by the form $\{x \in \mathbb{R}^n \mid g(x) = 0\}$ for sufficiently nice g (i.e. $g \in C^1$) then we use the method of **Lagrange multipliers**.

Theorem 11: Lagrange Multiplier Theorem

Let f and g be C^1 functions from $\mathbb{R}^n$ to $\mathbb{R}$. Suppose x_0 belongs to the level set

$$\Gamma = \{x \in \mathbb{R}^n \mid g(x) = 0\}$$

In other words, $g(x_0) = 0$. Suppose also that $Dg(x_0) \neq 0$. If the restricted function $\tilde{f} := f|_\Gamma$ has a maximum or minimum value at x_0 then there exists a constant $\lambda \in \mathbb{R}$ such that

$$\nabla f(x_0) = \lambda \nabla g(x_0)$$

While the statement of this theorem is sophisticated, the key idea is that if we are solving an optimization problem of the form

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0 \end{aligned} \tag{2}$$

then we are searching for optimizers of the function $\tilde{f}$ and the theorem tells us that we can try and study the system of equations

$$\nabla f(x) = \lambda \nabla g(x)$$

$$g(x) = 0$$

to solve our problem. The more general version of this theorem for multiple constraints is given below.

Theorem 12: General Lagrange Multiplier Theorem

Let f and g_i, $i = 1, \ldots, k$ be C^1 functions from $\Omega \subseteq \mathbb{R}^n$ to $\mathbb{R}$. Let $x_0 \in \Omega$ and define

$$\Gamma = \{x \in \Omega \mid g_i(x) = g_i(x_0) \ \forall i = 1, \ldots, k\}$$

Suppose also that $\nabla g_1(x_0), \ldots, \nabla g_k(x_0)$ are linearly independent. If the restricted function $\tilde{f} := f|_\Gamma$ has a maximum or minimum value at x_0 then there exist constants $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, k$ such that

$$\nabla f(x_0) = \sum_{i=1}^{k} \lambda_i \nabla g_i(x_0)$$

We now have a general strategy for solving optimization problems.

1. Locate all critical points of f in Ω.
2. Determine if the boundary $\partial\Omega$ should be dealt with using basic geometry or Lagrange multipliers.
3. Use the method chosen in the previous step to identify all the optimal points on the boundary.
4. Evaluate f at every point found in steps 1 and 3, and the greatest value is the maximum of f while the smallest value is the minimum of f.

4 Exercises

1. Define $f(x,y) = \frac{x^6y^2}{x^8+y^4}$ for $(x,y) \neq (0,0)$ and $f(0,0) = 0$. Determine the differentiability of f at the origin and the continuity of $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}$ at the origin.
2. Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}$ are differentiable. Show that the product function $h(x) = f(x)g(x)$ is differentiable and that
$$Dh(x_0) = f(x_0)Dg(x_0) + g(x_0)Df(x_0)$$
3. Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable, and satisfies
$$f(x) = 0 \ \forall x : \|x\| = 1$$
$$\|\nabla f(x)\| \leq 1 \ \forall x \in B_1(0)$$
Show that $|f(x)| \leq 1$ on $\overline{B_1(0)}$.
4. Find the shortest distance from the point $(3, 2, 1, \dots, 1) \in \mathbb{R}^n$ $(n \geq 3)$ to the hyperplane with equation $x_1 + x_2 + \cdots + x_n = 1$.
5. Determine the nature of the critical points of $f(x,y) = x^3 + y^2 - 6xy + 6x + 3y$.
6. Calculate the second-order Taylor approximation to $f(x,y) = \cos x \sin y$ at $(\pi, \pi/2)$.
7. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a C^2 function with
$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = 0$$
for all $i, j = 1, \dots, n$ and $x \in \mathbb{R}^n$. Show that there exist constants $a_0, a_1, \dots, a_n \in \mathbb{R}$ such that
$$f(x) = a_0 + \sum_{i=1}^{n} a_i x_i$$
for all $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. What are the values of these constants?
8. Design a cylindrical can (with lid) to contain 1L of water using the minimum amount of metal.