1 Sets and sequences

For a vector $x \in \mathbb{R}^n$ and a scalar $\epsilon > 0$, we denote the open sphere centered at $x$ with radius $\epsilon$ by $B_\epsilon(x)$, i.e.,

$$B_\epsilon(x) = \{ y : \| y - x \| < \epsilon \}.$$  

Sequences

A sequence $\{x_k | k = 1, 2, \ldots \}$ (or $\{x_k \}$ for short) in $\mathbb{R}^n$ is said to converge to $x^* \in \mathbb{R}^n$ if for every $\epsilon > 0$ there exists a $K$ such that $\|x_k - x^*\| < \epsilon$ for every $k \geq K$. If a sequence $\{x_k \}$ converges to some $x^* \in \mathbb{R}^n$, we denote $x_k \to x^*$ or $\lim_{k \to \infty} x_k = x^*$. A sequence $\{x_k \}$ is called a Cauchy sequence if for every $\epsilon > 0$ there exists a $K$ such that $\|x_l - x_m\| < \epsilon$ for all $l, m \geq K$.

A sequence $\{x_k \}$ is called a bounded if there exists a scalar $M$ such that $\|x_k\| \leq M$ for all $k$. In particular, a sequence $\{s_k \}$ in $\mathbb{R}$ is called a bounded above (below) if there exists a scalar $M$ such that $s_k \leq M$ ($s_k \geq M$) for all $k$. A sequence $\{s_k \}$ in $\mathbb{R}$ is said to be nondecreasing (nonincreasing) if $s_{k+1} \geq s_k$ ($s_{k+1} \leq s_k$) for every $k$.

Exercise 1 Suppose that a nondecreasing sequence $\{s_k \}$ in $\mathbb{R}$ is bounded above. Show that $\{s_k \}$ converges to some $s_* \in \mathbb{R}$.

Exercise 2 Show that every convergent sequence is a Cauchy sequence. Conversely, every Cauchy sequence is a convergent sequence.

The limit superior of $\{s_k \}$, denoted by $\limsup_{k \to \infty} s_k$ or $\lim_{k \to \infty} \sup_{k \geq K} s_k$, equals $\lim_{K \to \infty} \left( \sup_{k \geq K} s_k \right)$. Also, the limit inferior of $\{s_k \}$, denoted by $\liminf_{k \to \infty} s_k$ or $\lim_{k \to \infty} \inf_{k \geq K} s_k$, equals $\lim_{K \to \infty} \left( \inf_{k \geq K} s_k \right)$. When $\{s_k \}$ is unbounded above (below), $\lim_{k \to \infty} s_k = +\infty$ ($\lim_{k \to \infty} s_k = -\infty$). A sequence always has $\lim$ and $\lim$ if $\pm \infty$ is allowed. From the definition, the inequality

$$\lim_{k \to \infty} s_k \leq \limsup_{k \to \infty} s_k$$

holds, and the sequence $\{s_k \}$ has the limit $s_* = \lim_{k \to \infty} s_k$ if and only if $\lim_{k \to \infty} s_k$ is equal to $\limsup_{k \to \infty} s_k$, and these three limits equal each other.
Exercise 3 Show the following.

1. $\lim (s_k + t_k) \leq \lim s_k + \lim t_k$.

2. $\lim (s_k + t_k) \geq \lim s_k + \lim t_k$.

3. When $s_k, t_k \geq 0$ for all $k$, $\lim (s_k t_k) = \lim s_k \lim t_k$ holds. Furthermore, if $\lim t_k$ exists, $\lim (s_k t_k) = \lim s_k \lim t_k$.

Accumulation point

Consider a sequence $\{x^k\}$ in $\mathbb{R}^n$. A vector $\bar{x} \in \mathbb{R}^n$ is said to be an accumulation point of the sequence $\{x^k\}$ if there is a subsequence $\{x^k\}_{k \in K}$ of $\{x^k\}$ such that $\{x^k\}_{k \in K}$ converges to $\bar{x}$. Equivalently $\bar{x}$ is an accumulation point of $\{x^k\}$ if, for any $\epsilon > 0$, $B_\epsilon(\bar{x})$ contains infinitely many point of $\{x^k\}$.

Exercise 4 Show the following.

1. A bounded sequence converges if and only if it has a unique accumulation point.

2. Every bounded sequence has at least one accumulation point.

Open, closed and compact sets

A subset $S$ of $\mathbb{R}^n$ is said to be open if for every vector $x \in S$ there is an $\epsilon > 0$ such that $B_\epsilon(x) \subset S$. If $S$ is open and if $x \in S$, then $S$ is sometimes called a neighborhood of $x$. A set $S$ is closed if and only if its complement in $\mathbb{R}^n$ is open. A subset $S$ of $\mathbb{R}^n$ is said to be bounded if there is a number $M > 0$ such that $\|x\| \leq M$ for all $x \in S$. A set $S$ is compact if and only if it is both closed and bounded.

Exercise 5 Are the following sentences true or false? Why?

1. The union of finitely many closed sets is closed.

2. The union of infinitely many closed sets is closed.

3. The intersection of finitely many open sets is open.

4. The intersection of infinitely many open sets is open.

Exercise 6 Show that a subset $S$ is closed if and only if every convergent sequence $\{x^k\}$ in $S$ converges to a point belonging to $S$. Furthermore, if $S$ is compact, then every sequence in $S$ has at least one accumulation point in $S$.

The closure of a set $S$ is the smallest closed set including $S$. This is equivalent to the set containing all accumulation points in $S$. 
2 Functions and Mappings

Let $S$ be a subset of $\mathbb{R}^n$ and let $F$ be a mapping from $S$ into $\mathbb{R}^m$. Then, for given $x \in S$, $F$ is considered an $m$-dimensional column vector whose $i$-th component $F_i(x)$ is a function from $\mathbb{R}^n$ into $\mathbb{R}$, that is

$$F(x) = \begin{pmatrix} F_1(x) \\ \vdots \\ F_m(x) \end{pmatrix}.$$ 

Continuity

Let $S$ be a subset of $\mathbb{R}^n$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be continuous at $x \in S$ if $f(x^k) \to f(x)$ holds for every sequence $\{x^k\}$ of $S$ converging to $x$. Equivalently $f$ is continuous at $x$ if, for any $\epsilon > 0$, there exists a $\delta > 0$ (depending on $x$) such that $|y - x| < \delta$ and $y \in S$ imply $|f(y) - f(x)| < \epsilon$. The function $f$ is said to be continuous on $S$ if it is continuous at every $x \in S$. The function $f$ is said to be uniformly continuous on $S$ if, for any $\epsilon > 0$, there exists $\delta > 0$ (independently of $x$) such that $|y - x| < \delta$ and $y \in S$ imply $|f(y) - f(x)| < \epsilon$.

A mapping $F : S \to \mathbb{R}^m$ is said to be continuous at $x$ if all component functions $F_i$, $i = 1, \ldots, n$, are continuous at $x$. Also $F$ is continuous on $S$ if it is continuous at every $x \in S$.

Exercise 7 Show the following.

1. A norm is a continuous function.
2. A function $f(x) = x^2$ is uniformly continuous on any bounded interval of $\mathbb{R}$, but not uniformly continuous on $\mathbb{R}$.

A function $f : S \to \mathbb{R}$ is said to be lower semicontinuous (upper semicontinuous) at $x \in S$ if $f(x) \leq \liminf_{k \to \infty} f(x^k)$ (respectively, $f(x) \geq \limsup_{k \to \infty} f(x^k)$) for every sequence $\{x^k\}$ converging to $x$. It is also called that $f : S \to \mathbb{R}$ is lower (upper) semicontinuous on $S$ if $f$ is lower (upper) semicontinuous at every $x \in S$.

Exercise 8 Show that a function $f : S \to \mathbb{R}$ is lower semicontinuous if and only if a level set $L(r)$, defined by

$$L(r) = \{x \in S | f(x) \leq r\},$$

is (possibly empty) closed for all $r \in \mathbb{R}$.

Theorem 1 (Weierstrass’ Theorem) [Ber95, Proposition A.8] Let $S$ be a nonempty compact subset of $\mathbb{R}^n$ and let $f : S \to \mathbb{R}$ be lower semicontinuous on $S$. Then there exists an $x \in S$ such that $f(x) = \min_{y \in S} f(y)$.

Differentiability

Let $f$ be a function from $\mathbb{R}^n$ into $[-\infty, +\infty]$ and let $x$ be a point where $f$ is finite. We say that $f$ is directionally differentiable at $x$ in the direction $d$ if the limit

$$\lim_{\tau \to 0, \tau > 0} \frac{f(x + \tau d) - f(x)}{\tau}$$

exists.
Exercise 9 Show that a directional derivative is positively homogeneous, i.e., \( f'(x; \alpha d) = \alpha f'(x; d) \) holds for any \( d \in \mathbb{R}^n \) and \( \alpha > 0 \).

When the directional derivative exists in all directions \( d \) at \( x \) and \( f'(x; d) \) is a linear function of \( d \), then \( f \) is said to be Gateaux differentiable (G-differentiable or differentiable for short) at \( x \). \( f \) is G-differentiable at \( x \) if and only if there exists a gradient of \( f \) at \( x \) defined by

\[
\nabla f(x) = \begin{pmatrix} \partial f(x)/\partial x_1 \\ \vdots \\ \partial f(x)/\partial x_n \end{pmatrix}
\]

and the gradient satisfies \( f'(x; d) = \nabla f(x)^T d \) where \( \partial f(x)/\partial x_i \) denotes the \( i \)-th partial derivative of \( f \) at \( x \) defined by

\[
\lim_{\tau \to 0} \frac{f(x + \tau e_i) - f(x)}{\tau}.
\]

We call a function \( f \) differentiable on \( \mathbb{R}^n \) if \( f \) is differentiable at every \( x \in \mathbb{R}^n \).

A differentiable function \( f \) is said to be continuously differentiable at \( x \) if there is a neighbourhood of \( x \) such that the partial derivatives \( \partial f(x)/\partial x_1, \ldots, \partial f(x)/\partial x_n \) are continuous functions of \( x \) over the neighbourhood. A function \( f \) is said to be continuously differentiable on \( \mathbb{R}^n \) if \( f \) is continuously differentiable at every \( x \). A continuously differentiable function \( f \) satisfies

\[
\lim_{d \to 0} \frac{f(x + d) - f(x) - \nabla f(x)^T d}{\|d\|} = 0.
\]

If the second partial derivatives \( \partial^2 f(x)/\partial x_i \partial x_j \) exist for all \( i, j \) and are continuous, then we call \( f \) twice continuously differentiable. The Hessian of \( f \) is defined to be an \( n \times n \) symmetric matrix whose \((i, j)\)-th component is \( \partial^2 f(x)/\partial x_i \partial x_j \).

A mapping \( F : \mathbb{R}^n \to \mathbb{R}^m \) is continuously differentiable if all components \( F_i, i = 1, \ldots, m, \) are continuously differentiable and \( F \) is twice continuously differentiable if all \( F_i \) are twice continuously differentiable. When a mapping \( F \) is continuously differentiable, the derivative is said to be a Jacobian which is an \( n \times m \) matrix defined by

\[
F'(x) = (\nabla F_1(x) \ldots \nabla F_m(x))^T.
\]

Theorem 2 (Chain rule) Let \( f : \mathbb{R}^n \to \mathbb{R}^m \) and \( g : \mathbb{R}^m \to \mathbb{R}^l \) be continuously differentiable functions. Then a composite function \( h(x) = g(f(x)) \) is also continuously differentiable and its Jacobian is

\[
\nabla h(x) = \nabla f(x) \nabla g(f(x)).
\]

For other notions of derivatives and detailed discussions, see [OrR70, Chapter 3].

Exercise 10 Let \( f : \mathbb{R}^2 \to \mathbb{R} \) be a function defined by

\[
f(x) = \begin{cases} 
x_1 & \text{if } x_2 = 0 \\
x_2 & \text{if } x_1 = 0 \\
0 & \text{otherwise.}
\end{cases}
\]

Show that \( f \) has partial derivatives \( \partial f(0)/\partial x_1 \) and \( \partial f(0)/\partial x_2 \), but \( f \) is not differentiable at 0.
Exercise 11 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function defined by

$$f(x) = \begin{cases} 
0 & \text{if } x = 0 \\
\frac{x_1 x_2^2}{x_1^2 + x_2^2} & \text{otherwise.}
\end{cases}$$

Show that $f$ is directionally differentiable for every direction $d$ at 0, but $f$ is not differentiable at 0. Show, in addition, that $f$ is not continuous at 0.

Exercise 12 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function defined by

$$f(x) = \begin{cases} 
0 & \text{if } x = 0 \\
\frac{x_2 (x_1^2 + x_2^2)^{3/2}}{(x_1^2 + x_2^2)^2 + x_2^2} & \text{otherwise.}
\end{cases}$$

Show that $f$ is differentiable at 0, but not continuously differentiable at 0.

Lipschitz continuous

Let $S$ be a subset of $\mathbb{R}^n$. A mapping $F : S \to \mathbb{R}^m$ is Lipschitz continuous on $S$ if there is a constant $L > 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\| \text{ for all } x, y \in S.$$ 

The Lipschitz continuity of a Jacobian $\nabla F$ is also defined as

$$\|\nabla F(x) - \nabla F(y)\| \leq L\|x - y\| \text{ for all } x, y \in S,$$

where the norm of the left hand side represents a matrix norm.

Mean value theorems

Let a function $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Then the mean value theorem states that, for any $x, y \in \mathbb{R}^n$, there exists an $\alpha$ with $0 < \alpha < 1$ such that

$$f(y) = f(x) + \nabla f(x + \alpha(y - x))^T(y - x). \quad (1)$$

If, in addition, $f$ is twice continuously differentiable, there exists an $\alpha$ with $0 < \alpha < 1$ such that

$$f(y) = f(x) + \nabla f(x)^T(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x + \alpha(y - x))(y - x).$$

We note that the equation (1) does not necessarily hold for a mapping $F : \mathbb{R}^n \to \mathbb{R}^m$. Instead, we have the mean value theorem of integral form:

$$F(y) = F(x) + \int_0^1 \nabla F(x + \tau(y - x))^T(y - x)d\tau.$$ 

We also note that the above equation holds for a function, that is, $m = 1$.

Exercise 13 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F(x) = \begin{pmatrix} 
x_1^2 \\
x_2^3
\end{pmatrix}.$$ 

Suppose that $x = 0$ and $y = (1, 1)^T$. Then show that there is no $0 < \alpha < 1$ such that

$$F(y) = F(x) + \nabla F(x + \alpha(y - x))^T(y - x).$$
Implicit function theorem

Suppose that $F : \mathbb{R}^{n+m} \to \mathbb{R}^n$ is a continuous mapping and that $F(\bar{x}, \bar{y}) = 0$. If $F$ is continuously differentiable on some neighbourhood of $(\bar{x}^T, \bar{y}^T)^T$ and if $\nabla_x F(\bar{x}, \bar{y})$ is nonsingular, then there exists neighborhoods $S_1 \subset \mathbb{R}^n$ and $S_2 \subset \mathbb{R}^m$ of $\bar{x}$ and $\bar{y}$, respectively, and a continuous function $h : S_2 \to S_1$ such that $F(h(y), y) = 0$ for all $y \in S_2$. The function $h$ is unique in the sense that if $x \in S_1, y \in S_2$ and $F(x, y) = 0$ then $x = h(y)$. Furthermore, $h$ is continuously differentiable on $S_2$ and we have

$$\nabla h(y) = -\nabla_y F(h(y), y)\nabla_x F(h(y), y)^{-1}.$$ 

References
