1. Order notation

The following notation is ubiquitous in analyzing many algorithms in numerical computing.

**Definition 1.1. (Order notation.)** Let $\phi$ be a scalar function of a positive variable $h$, let $p$ be fixed, and let $\kappa_u$ and $\kappa_l$ be constants.

- If there exists $\kappa_u > 0$ such that $|\phi(h)| \leq \kappa_u h^p$ for all sufficiently small $h$, we write $\phi = O(h^p)$ and say “$\phi$ is of order $h^p$” (or “$\phi$ is big oh of $h^p$”).

- If $\lim_{h \to 0} \phi(h)/h^p = 0$, then we write $\phi = o(h^p)$ and say “$\phi$ is little oh of $h^p$”.

- If there exists $\kappa_l > 0$ such that $|\phi(h)| \geq \kappa_l h^p$ for all sufficiently small $h$, we write $\phi = \Omega(h^p)$ and say “$\phi$ is omega of $h^p$”.

- If there exist $\kappa_l > 0$ and $\kappa_u > 0$ such that $\kappa_l h^p \leq |\phi(h)| \leq \kappa_u h^p$ for all sufficiently small $h$, we write $\phi = \Theta(h^p)$ and say “$\phi$ is theta of $h^p$”.

In particular, $\phi = O(1)$ means that $\phi$ is bounded as $h \to 0$, and $\phi = o(1)$ means that $\phi \to 0$. Broadly speaking, $\phi = o(h^p)$ means that $\phi$ is going to zero faster than $h^p$, and $\phi = \Theta(h^p)$ means that $\phi$ and $h^p$ are going to zero at the same speed or rate.

2. Scalar sequences

2.1. Convergence of a sequence

Broadly speaking, an infinite sequence of scalars $\{x_k\}, k = 0, \ldots$ (sometimes denoted just by $\{x_k\}$) converges to a limit $x^*$ if $|x_k - x^*|$ is “small” for all sufficiently large values of $k$.

**Definition 2.1. (Convergence of a scalar sequence.)** The sequence $\{x_k\}, k = 0, \ldots$, converges to the value $x^*$ if, for any $\epsilon > 0$, there exists an index $K$ such that for all $k \geq K$,

$$|x_k - x^*| < \epsilon.$$  \hspace{1cm} (2.1)

When a sequence converges, the (necessarily unique) value $x^*$ is called the limit of the sequence, and we write $\lim_{k \to \infty} x_k = x^*$ or $x_k \to x^*$ as $k \to \infty$. If we say simply that a sequence converges, we mean that some limit (whose value is unspecified and may be irrelevant) exists. If the sequence does not have a limit, we say that it diverges, or that it is non-convergent.

**Definition 2.2. (Subsequence.)** A subsequence of $\{x_k\}$ is a new sequence formed by omitting some of the elements of the original sequence, without altering the order of the remaining elements.
Definition 2.3. (Limit point.) A point $x^*$ is a limit point or accumulation point of the sequence $X = \{x_k\}$ if there exists a subsequence of $X$ that converges to $x^*$.

Some examples to illustrate these ideas are:
1. The sequence of integers $x_k = k$ does not converge, has no limit point, and diverges to infinity.
2. The sequence $x_k = 2 + 2^{-k}$ converges to $x^* = 2$.
3. The sequence $X = \{1, -1, 1, -1, \ldots\}$ has two limit points (1 and -1) since the subsequences corresponding to even and odd $k$ converge to these values.

Definition 2.4. (Bounded sequence.) The sequence $\{x_k\}$ is bounded below by the constant $m$ if $x_k \geq m$ for all $k$, and bounded above by $M$ if $x_k \leq M$ for all $k$. If a sequence is bounded both above and below, the sequence is bounded. If there is no constant $M$ such that $x_k \leq M$ for all $k$, $\{x_k\}$ is unbounded above, with an analogous definition for “unbounded below”.

For example, a sequence in which $x_k = k$, $k = 0, 1, \ldots$, is bounded below by zero and unbounded above; the sequence $x_k = 2 + 2^{-k}$ is bounded, with $2 \leq x_k \leq 3$.

Result 2.1. (Existence of a limit point.) Every bounded sequence has at least one limit point.

2.2. Rate of convergence

An important concept in analyzing iterative algorithms is the rate or order of convergence of the sequence of iterates, which indicates (in some sense) how fast convergence occurs.

Definition 2.5. (Simplified definition of order.) Let $\{x_k\}$ be a scalar sequence that converges to $x^*$, with $x_k \neq x^*$ for any $k$. If there exists a value $q \geq 1$ and a positive finite value $\gamma_q$, such that

$$
\lim_{k \to \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^q} = \gamma_q,
$$

the sequence $\{x_k\}$ is said to converge to $x^*$ with order $q$ and asymptotic error constant $\gamma_q$.

When $q = 2$, the sequence $\{x_k\}$ is said to display quadratic convergence. When $q > 1$, the convergence is said to be superlinear.

When $q = 1$, the sequence is said to converge linearly. There is also an alternative definition of linear convergence.

Definition 2.6. (Linear convergence.) A sequence $\{x_k\}$ converges linearly to $x^*$ if there exists a constant $\beta$ satisfying $0 \leq \beta < 1$ and an integer $K \geq 0$ such that, for all $k \geq K$,

$$
|x_{k+1} - x^*| \leq \beta |x_k - x^*|.
$$

(2.2)

The requirement that $\beta < 1$ ensures that a sequence with linear convergence really does converge, since repeatedly applying (2.2), starting with $k = K$, gives

$$
|x_{k+1} - x^*| \leq \beta^{k+1} |x_K - x^*|,
$$

and $\beta^{k+1} \to 0$ as $k \to \infty$.

Suppose that $X = \{x_k\}$ is a linearly convergent sequence and that $K$ is an integer satisfying the conditions of Definition 2.6; then

$$
\sum_{j=K}^{k} |x_j - x^*| \leq |x_K - x^*|(1 + \beta + \beta^2 + \cdots + \beta^k) = |x_K - x^*| \frac{1 - \beta^{k+1}}{1 - \beta}.
$$

(2.3)

Since $\beta < 1$, this sum is bounded as $k \to \infty$ and $\sum_{j=0}^{\infty} |x_j - x^*| < \infty$.

However, all convergent sequences do not satisfy a condition like (2.3). A standard example from calculus courses is the sequence in which $x_k = 1/k$. Even though $x_k$ converges to zero as $k \to \infty$, so that $x^* = 0$, there is no value of $\beta$ satisfying Definition 2.6, and $\sum_{k=0}^{K} x_k \to \infty$ as $K \to \infty$. 

3. Continuity

A comment on notation: the notation \( f: \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R} \) means that \( f \) is a function that maps a one-dimensional domain \( \mathcal{D} \) (i.e., the set of “input” values of \( f \)) into one real variable.

**Definition 3.1. (Continuity.)** Given a function \( f: \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R} \), \( f \) is continuous at \( x \in \mathcal{D} \) if, given any \( \epsilon > 0 \), there is a \( \delta > 0 \) such that, for any \( y \) in \( \mathcal{D} \) satisfying \( |x - y| \leq \delta \), then \( |f(x) - f(y)| \leq \epsilon \).

A crucial point in this general definition is that, once we are given a point \( x \) and the desired closeness \( \epsilon \) of \( f(y) \) and \( f(x) \), then the value of \( \delta \)—the closeness required of \( y \) and \( x \)—depends on both \( x \) and \( \epsilon \).

The next step up in continuity is called uniform continuity, meaning that, given a desired closeness \( \epsilon \) of \( f \)-values, a single value for the needed closeness of \( x \) and \( y \) works for all \( x \in \mathcal{D} \).

**Definition 3.2. (Uniform continuity.)** Given a function \( f: \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R} \), \( f \) is said to be uniformly continuous on \( \mathcal{D} \) if, given any \( \epsilon > 0 \), there exists \( \delta > 0 \) such that

\[
|f(x) - f(y)| \leq \epsilon \quad \text{for any } x \text{ and } y \text{ in } \mathcal{D} \text{ satisfying } |x - y| \leq \delta.
\] (3.1)

Uniform continuity is a statement about the behavior of \( f \) everywhere in \( \mathcal{D} \), so specification of the domain is important. To see the added strength of uniform continuity compared to ordinary continuity as well as the role of \( \mathcal{D} \), consider \( f(x) = 1/x \) for \( \mathcal{D} = (0, \infty) \), where (obviously) \( x = 0 \) will cause the difficulty. Although \( f(x) = 1/x \) is continuous on \( \mathcal{D} \), it is not uniformly continuous, since, given \( \epsilon \), no single value of \( \delta \) satisfies (3.1) for all points in \((0, \infty)\). In contrast, when the domain excludes the origin, say \( \mathcal{D} = [1, \infty) \), the same function \( f(x) = 1/x \) is uniformly continuous.

A useful result that guarantees uniform continuity is continuity on a compact (i.e., closed and bounded) set.

**Theorem 3.1.** Given a function \( f: \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R} \), assume that \( f \) is continuous on \( K \), a compact subset of \( \mathcal{D} \). Then \( f \) is uniformly continuous on \( K \).

Beyond uniform continuity, we have the even stronger property of Lipschitz continuity, where the difference between function values at any two points in the domain is bounded by a multiple of the distance between the points.

**Definition 3.3. (Lipschitz continuity.)** Given a function \( f: \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R} \), \( f \) is said to be Lipschitz continuous on \( \mathcal{D} \), or to satisfy a Lipschitz condition on \( \mathcal{D} \), if there exists a (finite) positive constant \( L \) called the Lipschitz constant such that, given any \( x \) and \( y \) in \( \mathcal{D} \),

\[
|f(x) - f(y)| \leq L|x - y|.
\] (3.2)

We sometimes say that \( f \) is Lipschitz at the point \( x \), or locally Lipschitz, if (3.2) applies for all \( y \) in a neighborhood of \( x \).

A few aspects of Lipschitz continuity should be noted. Most obviously, the value of the Lipschitz constant is not unique, since (3.2) holds for any \( L' > L \) if it holds for \( L \). Often just the existence of \( L \) matters, not its value. Further, Lipschitz continuity on a domain implies uniform continuity on that same domain. To see this, suppose that \( \epsilon \) is, as usual, the target value for closeness of \( f \)-values and that \( f \) is Lipschitz continuous. Comparing the definitions (3.2) and (3.1), it is not hard to see that, if we choose \( \delta = \epsilon/L \) (a value independent of \( x \)), then for all \( x, y \in \mathcal{D} \) satisfying \( |x - y| \leq \delta \), it holds that

\[
|f(x) - f(y)| \leq L|x - y| \leq L\delta = \epsilon,
\]

proving uniform continuity.

Going the other direction, a uniformly continuous function is not necessarily Lipschitz continuous, based on the following rough argument. Uniform continuity ensures that, given any \( \epsilon \), \( f(x) \) and \( f(y) \) will be within \( \epsilon \) if \( x \) and \( y \) are within \( \delta \), for any two points \( x, y \in \mathcal{D} \). However, this does not guarantee that the specified \( \epsilon \) is bounded by a constant multiple of the distance between \( x \) and \( y \), as needed for Lipschitz continuity. Consider, for instance, \( f(x) = \sqrt{x} \) with \( \mathcal{D} = [0, 1] \), which is
uniformly continuous on $\mathcal{D}$, but not Lipschitz continuous. In order for (3.2) to hold with $y = 0$, a finite constant $L$ would have to exist satisfying
\[ \sqrt{x} \leq Lx \quad \text{for all } x \in (0, 1], \]
which means that
\[ L \geq \frac{\sqrt{x}}{x} = \frac{1}{\sqrt{x}} \quad \text{for all } x \in (0, 1]. \]
But this inequality means that $L$ is not finite, since $1/\sqrt{x}$ can be made as large as desired by choosing $x$ sufficiently close to 0.

4. Differentiability

If $\lim_{t \to 0} (f(x+t) - f(x))/t$ exists, $f$ is said to be differentiable at $x$, and the (unique) value satisfying
\[ \lim_{t \to 0} \frac{f(x+t) - f(x)}{t} = f'(x) \] (4.1)
is known as the first derivative of $f$ at $x$.

A central feature of numerical computing is developing simple models of nonlinear functions. In doing so, we will make heavy use of what are called “mean value theorems” (closely related to Taylor series) which depend on differentiability and allow us to make statements about the relationship between function values at points that lie between other points.

**Theorem 4.1. (Mean value theorem (1).)** If $\psi: \mathcal{D} \subseteq \mathbb{R}^1 \mapsto \mathbb{R}^1$ is differentiable on $(a,b)$ and continuous on $[a,b]$, then there exists $\xi \in (a,b)$ such that
\[ \psi(b) - \psi(a) = \psi'(\xi)(b-a). \]

In addition, differentiability implies continuity.

**Result 4.1.** Given $f: \mathcal{D} \subseteq \mathbb{R}^1 \mapsto \mathbb{R}^1$, if $f'(x)$ exists at $x \in \mathcal{D}$, then $f$ is continuous at $x$. Furthermore, if $f'$ exists at an interior point $\bar{x}$ of $\mathcal{D}$, then there are $\delta > 0$ and $M > 0$ such that
\[ |f(x) - f(\bar{x})| \leq M|x - \bar{x}| \quad \text{if } |x - \bar{x}| \leq \delta. \]

The property of continuous differentiability, defined next, will be used constantly.

**Definition 4.1. (Continuous differentiability.)** The function $f: \mathcal{D} \subseteq \mathbb{R}^1 \mapsto \mathbb{R}^1$ is said to be continuously differentiable or once-continuously differentiable, denoted by $f \in C^1$, on the open set $\mathcal{D}_0 \subseteq \mathcal{D}$ if $f$ is differentiable and the derivative $f'$ is continuous at every point of $\mathcal{D}_0$.

Because differentiability of $f$ implies continuity (Result 4.1), any continuously differentiable function is automatically continuous.

The next theorem provides useful bounds on the error in the local linear model when $f$ is continuously differentiable and when $f'$ is Lipschitz continuous (Definition 3.3)—a stronger condition than simple continuity.

**Theorem 4.2. (Error in linear model (1).)** Suppose that $f: \mathcal{D} \subseteq \mathbb{R}^1 \mapsto \mathbb{R}^1$ is continuously differentiable on an open interval $\mathcal{D}_0 \subset \mathcal{D}$. Then
\[
\begin{align*}
\text{(i) for any } x \text{ and } x + p \in \mathcal{D}_0, & \quad f(x + p) - f(x) - f'(x)p = o(p); \\
\text{(ii) if, in addition, } f' \text{ satisfies a Lipschitz condition on } \mathcal{D}_0 \text{ with Lipschitz constant } L, & \quad \text{then, for any } x \text{ and } x + p \in \mathcal{D}_0, \\
& \quad |f(x + p) - f(x) - f'(x)p| \leq \frac{1}{2}L p^2. \quad (4.2)
\end{align*}
\]
5. Second Derivatives

We turn now to second derivatives. If the derivative \( f' \) is itself differentiable, its derivative \( (f')' \) will be called the second derivative of \( f \) and denoted by \( f'' \).

Let \( f : D \subseteq \mathbb{R}^1 \to \mathbb{R}^1 \) be differentiable on \( D \). If \( f''(\bar{x}) \) (the second derivative of \( f \) at \( \bar{x} \in D \)) exists, it must satisfy (4.1) applied to \( f' \):

\[
\lim_{t \to 0} \frac{f'(\bar{x} + t) - f'(\bar{x})}{t} = f''(\bar{x}).
\] (5.1)

The notation \( C^r \) represents the class of functions that are \( r \) times continuously differentiable.

**Definition 5.1. (The class \( C^r \)).** The function \( f : D \subseteq \mathbb{R}^1 \to \mathbb{R}^1 \) is said to be \( r \)-times continuously differentiable, denoted by \( f \in C^r \), on the open set \( D_0 \subseteq D \) if the \( r \)th derivative of \( f \) exists and is continuous at every point of \( D_0 \).

Assuming that \( f \) has a second derivative, there is a further bound on the error in the local linear model; compare with Theorem 4.2.

**Theorem 5.1. (Error in linear model (2).)** Suppose that \( f : D \subseteq \mathbb{R}^1 \to \mathbb{R}^1 \) has a second derivative on the interval \( D_0 \subseteq D \). Then, for any \( x, x + p \in D_0 \),

\[
|f(x + p) - f(x) - f'(x)p| \leq \sup_{0 \leq t \leq 1} |f''(x + tp)| \cdot p^2.
\]

The next mean value theorem is closely related to Taylor series, but involves the second derivative; compare with Theorem 4.1.

**Theorem 5.2. (Mean value theorem (2).)** Let \( f : D \subseteq \mathbb{R}^1 \to \mathbb{R}^1 \) have a second derivative on the open interval \( D_0 \subseteq D \). Then for any \( x, x + p \in D_0 \), there is a point strictly between \( x \) and \( x + p \), i.e., a scalar \( t \) satisfying \( 0 < t < 1 \), such that

\[
f(x + p) - f(x) - f'(x)p = \frac{1}{2}p^2 f''(x + tp),
\]

where \( f' \) and \( f'' \) are the first and second derivatives of \( f \).

6. Miscellaneous Results from Analysis

This is a collection of useful results from analysis.

**Theorem 6.1. (Intermediate Value Theorem.)** Let \( f : \mathbb{R}^1 \to \mathbb{R}^1 \) be a continuous function in \( [a, b] \). If \( f(a) \neq f(b) \), then \( f \) assumes every value between \( f(a) \) and \( f(b) \) in \( [a, b] \).

The Intermediate Value Theorem does not state that \( f \) assumes only values between \( f(a) \) and \( f(b) \), but rather that \( f \) assumes at least this range of values.

**Theorem 6.2. (Rolle’s Theorem.)** Let \( f : \mathbb{R}^1 \to \mathbb{R}^1 \) be differentiable in \( (a, b) \) and assume that \( f \) is continuous at both endpoints \( a \) and \( b \). If \( f(a) = f(b) \), then there is at least one interior point \( \xi \) such that \( f'(\xi) = 0 \).

**Theorem 6.3. (Nonzero Derivative Theorem.)** Let \( f : \mathbb{R}^1 \to \mathbb{R}^1 \) be continuously differentiable on \( (a, b) \), and assume that \( f'(\xi) \neq 0 \) for some \( \xi \in (a, b) \). Then

(i) if \( f'(\xi) > 0 \), there exists \( \delta > 0 \) satisfying \( 0 < \delta < \min(\xi - a, b - \xi) \) such that \( f(x) < f(\xi) < f(z) \) for all \( \xi - \delta < x < \xi \) and \( \xi < z < \xi + \delta \);

(ii) if \( f'(\xi) < 0 \), there exists \( \delta > 0 \) satisfying \( 0 < \delta < \min(\xi - a, b - \xi) \) such that \( f(x) > f(\xi) > f(z) \) for all \( \xi - \delta < x < \xi \) and \( \xi < z < \xi + \delta \).

**Definition 6.1. (Multiplicity of a zero.)** Let \( f \in C^k \) and \( f(x^*) = 0 \). Then \( x^* \) is a zero (root) of \( f \) of multiplicity \( k \) if \( f^{(i)}(x^*) = 0, i = 1, \ldots, k - 1, \) and \( f^{(k)}(x^*) \neq 0 \), where \( f^{(i)} \) denotes the \( i \)th derivative of \( f \).

For example, if \( f(x) = x^2 - 1, x^* = 1 \) is a zero of multiplicity one, since \( f'(x^*) = 2x^* = 2 \neq 0 \). In contrast, if \( f(x) = (x - 1)^2 \), then \( x^* = 1 \) is a zero of multiplicity two, since \( f'(1) = 0 \) but \( f''(1) = 2 \neq 0 \). A zero of multiplicity one is called a simple zero.