Calculus Review

by Marc S. Paolella

University of Zurich, Swiss Finance Institute

Source: Fundamental Probability: A Computational Approach, by Marc S. Paolella. John Wiley & Sons, 2006.

The Joy of Sets

Sets

A **set** is simply a well-defined collection of objects. For example, if $A = \{n \in \mathbb{N} : n < 7\}$, then A is the set of positive integers less than 7.

Let A and B be two sets.

- the **intersection** of two sets, "A and B", is denoted by $A \cap B$. Each element of $A \cap B$ is contained in A, and contained in B. That is, $A \cap B = \{x : x \in A, x \in B\}$.
- the **union** of two sets, "A or B", is denoted by $A \cup B$ or $A + B$. An element of $A \cup B$ is either in A, or in B, or in both.
- set **subsets**, "A is a subset of B" or "A is contained in B" or "B contains A", denoted $A \subset B$ or $B \supset A$. If every element contained in A is also in B, then $A \subset B$.
- The empty set is denoted by $\emptyset$, and is a subset of every set.

Sets

- set **equality**, "$A = B$", which is true if and only if $A \subset B$ and $B \subset A$.
- the **difference**, or **relative complement**, "B minus A", denoted $B \setminus A$ or $B - A$. It is the set of elements contained in B but not in A.
- If the set B is clear from the context, then it need not be explicitly stated, and the set difference $B \setminus A$ is written as A^c, which is the **complement** of A. Thus, we can write $B \setminus A = B \cap A^c$.
- the **product** of two sets, A and B, consists of all ordered pairs (a, b), such that $a \in A$ and $b \in B$; it is denoted $A \times B$.
- Two sets are **disjoint**, or **mutually exclusive**, if $A \cap B = \emptyset$, i.e., they have no elements in common.

Sets

Some of these set operations are extended to more than two sets in a natural way. For example, with set intersection, if $a \in A_1 \cap A_2 \cap \cdots \cap A_n$, then a is contained in each of the A_i, and is abbreviated by

$$a \in \bigcap_{i=1}^{n} A_i.$$

A similar notation is used for union. To illustrate this for subsets, let

$$A_n = [1/n, 1], n \in \mathbb{N},$$

i.e. $A_1 = \{1\}$, $A_2 = [1/2, 1] = \{x : 1/2 \leq x \leq 1\}$ and so on.
Then $A_1 \subset A_2 \subset \cdots$ and the A_n are said to be **monotone increasing**.
In this case

$$\bigcup_{n=1}^{\infty} A_n = (0, 1] = \{x : 0 < x \leq 1\}.$$

Sets: Basic properties

Sets obey certain rules, such as:

- $A \cup A = A$ (idempotent);
- $(A \cup B) \cup C = A \cup (B \cup C)$ and $(A \cap B) \cap C = A \cap (B \cap C)$ (associative);
- $A \cup B = B \cup A$ and $A \cap B = B \cap A$ (commutative);
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (distributive);
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (distributive);
- $A \cup \emptyset = A$ and $A \cap \emptyset = \emptyset$ (identity);
- and $(A^c)^c = A$ (involution).
- Less obvious are De Morgan's laws, after Augustus De Morgan (1806–1871), which state that $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$. More generally,

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c \quad \text{and} \quad \left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} A_n^c.$$

The indicator function

We will often make use of the so–called **indicator function**. It is defined for every set $M \subseteq \mathbb{R}$ as

$$\mathbb{I}_M(x) = \begin{cases} 1, & \text{if} \quad x \in M, \\ 0, & \text{if} \quad x \notin M. \end{cases}$$

For ease of notation, we will occasionally use the form $\mathbb{I}(x \geq k)$, which is equivalent to $\mathbb{I}_{\{k,k+1,\ldots\}}(x)$ if x takes on only discrete values, and equivalent to $\mathbb{I}_{[k,\infty)}(x)$ if x can assume values from the real line.

Optional Material on Sets

- A set J is an **indexing set** if it contains a set of indices, usually a subset of $\mathbb{N}$, and is used to work with a group of sets A_i, where $i \in J$.
- If A_i, $i \in J$, are such that $\bigcup_{i \in J} A_i \supset \Omega$, then they are said to **exhaust**, or (form a) **cover** (for) the set Ω.
- If sets A_i, $i \in J$, are nonempty, mutually exclusive and exhaust Ω, then they (form a) **partition** (of) Ω.
- For $a, b \in \mathbb{R}$ with $a \leq b$, the interval $(a, b) = \{x \in \mathbb{R} : a < x < b\}$ is said to be an **open interval**, while $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ is a **closed interval**. In both cases, the interval has **length** $b - a$.
- For a set $S \subset \mathbb{R}$, the set of open intervals $\{O_i\}$, for $i \in J$ with J an indexing set, is an **open cover** of S if $\bigcup_{i \in J} O_i$ covers S, i.e., if $S \subset \bigcup_{i \in J} O_i$.

Optional Material on Sets

- ▷ Let $S \subset \mathbb{R}$ be such that there exists an open cover $\bigcup_{i \in N} O_i$ of S with a finite or countably infinite number of intervals. Denote the length of each O_i as $\ell(O_i)$. If $\forall \epsilon > 0$, there exists a cover $\bigcup_{i \in N} O_i$ of S such that $\sum_{i=1}^{\infty} \ell(O_i) < \epsilon$, then S is said to have **measure zero**.
- ▷ For our purposes, the most important set with measure zero is any set with a finite or countable number of points. For example, if f and g are functions with domain $I = (a, b) \in \mathbb{R}$, where $a < b$, and such that $f(x) = g(x)$ for all $x \in I$ except for a finite or countably infinite number of points in I, then we say that f **and** g **differ on** I **by a set of measure zero**. As an example from probability, if U is a continuous uniform random variable on $[0, 1]$, then the event that $U = 1/2$ is not impossible, but it has probability zero, because the point $1/2$ has measure zero, as does any finite collection of points, or any countably infinite set of points on $[0, 1]$, e.g., $\{1/n,\ n \in \mathbb{N}\}$.

Limits

- Informally, the limit of a function at a particular point, say x, is the value that $f(x)$ approaches, but need not assume at x.
- For example, $\lim_{x\to 0}(\sin x)/x = 1$, even though the ratio is not defined at $x = 0$.
- More formally, the function $f : A \to \mathbb{R}$ with $A \subset \mathbb{R}$ has the **right-hand limit** L **at** c, if, $\forall\,\epsilon > 0$, $\exists\,\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $x \in (c, c+\delta)$, for which we write $\lim_{x\to c^+} f(x)$. The left-hand limit, $\lim_{x\to c^-} f(x)$, is likewise defined.
- If $\lim_{x\to c^-} f(x)$ and $\lim_{x\to c^+} f(x)$ exist and coincide, then L is the **limit of** f **at** c, and write $\lim_{x\to c} f(x)$.
- We write $\lim_{x\to c^+} f(x) = \infty$ if, $\forall\,M \in \mathbb{R}$, $\exists\,\delta > 0$ such that $f(x) > M$ for every $x \in (c, c+\delta)$; and $\lim_{x\to c^-} f(x) = \infty$ if, $\forall\,M \in \mathbb{R}$, $\exists\,\delta > 0$ such that $f(x) > M$ for every $x \in (c-\delta, c)$. Similar definitions hold for $\lim_{x\to c^+} f(x) = -\infty$ and $\lim_{x\to c^-} f(x) = -\infty$.

Limits

Let f and g be functions whose domain contains the point c and such that $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$. Then, for constant values $k_1, k_2 \in \mathbb{R}$,

- $\lim_{x\to c}[k_1 f(x) + k_2 g(x)] = k_1 L + k_2 M$,
- $\lim_{x\to c} f(x)g(x) = LM$,
- $\lim_{x\to c} f(x)/g(x) = L/M$, if $M \neq 0$,
- if $g(x) \leq f(x)$, then $M \leq L$.
- For the limit of a composition of functions, let $b = \lim_{x\to a} f(x)$ and $L = \lim_{y\to b} g(y)$. Then $\lim_{x\to a} g(f(x)) = L$.

Limits

- Let f be a function with domain A and $a \in A$. If $a \in A$ and $\lim_{x \to a^+} f(x) = f(a)$, then f is said to be **continuous on the right at** a.
- If $a \in A$ and $\lim_{x \to a^-} f(x) = f(a)$, then f is **continuous on the left at** a.
- If both of these conditions hold, then $\lim_{x \to a} f(x) = f(a)$, and f is said to be **continuous at** a.

Limits

- If f is continuous at each point $a \in S \subset A$, then f is **continuous on** S. If f is continuous on (its whole domain) A, then we say f is continuous.
- An important result is the continuity of composite functions: Let $f : A \to B$ and $g : B \to C$ be continuous. Then $g \circ f : A \to C$ is continuous. More precisely, if f is continuous at $a \in A$, and g is continuous at $b = f(a) \in B$, then

$$\lim_{x \to a} g(f(x)) = g\left(\lim_{x \to a} f(x)\right).$$

Limits

- We said that f is continuous at a if $\lim_{x \to a} f(x) = f(a)$.
- An equivalent, but seemingly more complicated, definition of continuity at a is: for a given $\epsilon > 0$, $\exists\ \delta > 0$ such that, if $|x - a| < \delta$ and $x \in A$, then $|f(x) - f(a)| < \epsilon$.
- Its value is seen when contrasting it with the definition of **uniform continuity**:
- Let f be a function with domain A and let $[a, b] \subset A$ be a closed, finite interval. Function f is uniformly continuous on $[a, b]$ if the condition holds: for a given $\epsilon > 0$, $\exists\ \delta > 0$ such that, if $x, y \in [a, b]$, and $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$.
- Note that, with uniform continuity, δ does not depend on the choice of $x \in [a, b]$.

Univariate Calculus: Differentiation

- Let $f \in \mathcal{C}^0(I)$, where I is an interval of nonzero length. If the **Newton quotient**

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

exists for $x \in I$, then f is **differentiable** at x, the limit is the **derivative** of f at x, and is denoted $f'(x)$ or df/dx.

- Similar to the notation for continuity, if f is differentiable at each point in I, then f is differentiable on I, and if f is differentiable on its domain, then f is differentiable.

Univariate Calculus: Differentiation

- If f is differentiable and $f'(x)$ is a continuous function of x, then f is **continuously differentiable**.
- Observe that, for h small,

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} \quad \text{or} \quad f(x+h) \approx f(x) + hf'(x),$$

  which, for constant x, is a linear function in h. By letting $h = y - x$, this can be equivalently written as

$$\lim_{y \to x} \frac{f(y) - f(x)}{y - x}$$

  which is sometimes more convenient to work with.
- For $f(x) = x^r$ with $r \in \mathbb{R}$ (and $x \neq 0$ if $r < 0$), $f'(x) = rx^{r-1}$.

Univariate Calculus: Differentiation

▷ Assume for functions f and g defined on I that $f'(x)$ and $g'(x)$ exist on I. Then

$$\begin{aligned}
&\textbf{(sum rule)} && (f+g)'(x) = f'(x) + g'(x)\,,\\
&\textbf{(product rule)} && (fg)'(x) = f(x)\,g'(x) + g(x)\,f'(x)\,,\\
&\textbf{(quotient rule)} && (f/g)'(x) = \frac{g(x)\,f'(x) - f(x)\,g'(x)}{[g(x)]^2}, \quad g(x) \neq 0,\\
&\textbf{(chain rule)} && (g \circ f)'(x) = g'(f(x))\,f'(x)\,.
\end{aligned}$$

Higher Order Derivatives

- The second derivative of f, if it exists, is the derivative of f', and denoted f'' or $f^{(2)}$, and likewise for higher order derivatives.
- If $f^{(r)}$ exists, then f is said to be r^{th} **order differentiable**, and if $f^{(r)}$ exists for all $r \in \mathbb{N}$, then f is **infinitely differentiable**, or **smooth**. Let $f^{(0)} \equiv f$.
- As differentiability implies continuity, it follows that, if f is r^{th} order differentiable, then $f^{(r-1)}$ is continuous, and that smooth functions and all their derivatives are continuous.
- If f is r^{th} order differentiable and $f^{(r)}$ is continuous, then f is **continuously** r^{th} **order differentiable**, and f is of **class** $\mathcal{C}^r$.
- An infinitely differentiable function is of **class** $\mathcal{C}^\infty$.

Higher Order Derivatives

- Let f be a strictly increasing continuous function on a closed interval I. The **inverse function** g is defined as the function such that $g \circ f(x) = x$ and $f \circ g(y) = y$. It is also continuous and strictly increasing.
- If f is also differentiable in the interior of I with $f'(x) > 0$, then a fundamental result is that

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{f'(g(y))}.$$

Exponential and Logarithm

- Consider a function f such that

$$\text{(i) } f'(x) = f(x) \quad \text{and} \quad \text{(ii) } f(0) = 1.$$

- From the product and chain rules, and (i),

$$\begin{aligned}[f(x)f(-x)]' &= -f(x)f'(-x) + f(-x)f'(x) \\ &= -f(x)f(-x) + f(-x)f(x) = 0,\end{aligned}$$

  so that $f(x)f(-x)$ is constant, and from (ii), equals 1.
- Thus, $f(x) \neq 0$ and $f(-x) = 1/f(x)$.
- From (ii), $f'(0) > 0$ and, as $f(x) \neq 0$, it follows that $f(x) > 0$ for all x and is strictly increasing.

Exponential and Logarithm

- It is also straightforward to see that

$$f\,(x+y) = f\,(x)\,f\,(y) \quad \text{and} \quad f\,(nx) = [f\,(x)]^n\,, \quad n \in \mathbb{N}.$$

- The latter also holds for $n \in \mathbb{R}$. Function f is the **exponential**, written $\exp(\cdot)$.
- Also, defining $e = f\,(1)$, we can write $f\,(x) = \exp(x) = e^x$.
- As $f(x)$ is strictly increasing and $f(0) = 1$, $f(1) = e > 1$.

Exponential and Logarithm

- As $f(x)$ is strictly increasing, the inverse function g exists; and as $f(x) > 0$, $g(y)$ is defined for $y > 0$. Function g is the **natural logarithm**, denoted $\log(y)$, $\log y$ or, from the French **logarithm natural**, $\ln y$.
- It follows from (i) and the derivative of the inverse function that

$$g'(y) = \frac{1}{f'(g(y))} = \frac{1}{f(g(y))} = \frac{1}{y}.$$

- It can be shown that

$$\ln(x^p) = p \cdot \ln(x), \quad p \in \mathbb{R},\ x \in \mathbb{R}_{>0}.$$

The Mean Value Theorem

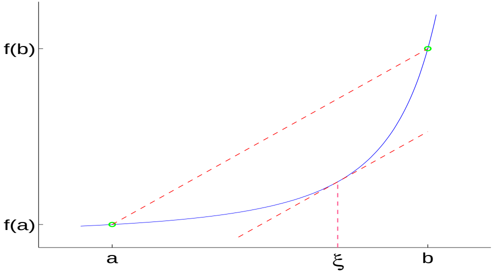
- Let f be a function continuous on its domain $[a, b]$ and differentiable on (a, b). Of great importance is the **Mean Value Theorem (of the Differential Calculus)**, which states that there exists a number $\xi \in (a, b)$ such that $f(b) - f(a) = f'(\xi)(b - a)$.
- This is more easily remembered as

$$\frac{f(b) - f(a)}{b - a} = f'(\xi),$$

for $b - a \neq 0$.

The Mean Value Theorem

The algebraic proof is easy, but informally, this is graphically "obvious" for a differentiable (and, thus, continuous) function, as illustrated in the Figure.



l'Hôpital's Rule

- Of great use is l'Hôpital's rule for evaluating **indeterminate forms or ratios**.
- Let f and g, and their first derivatives, be continuous functions on (a, b).
- If $\lim_{x \to a^+} f(x) = \lim_{x \to a^+} g(x) = 0$ and $\lim_{x \to a^+} f'(x)/g'(x) = L$, then $\lim_{x \to a^+} f(x)/g(x) = L$.

l'Hôpital's Rule

Most students remember this very handy result, but few can intuitively justify it. Here is one: Assume f and g are continuous at a, so that $f(a) = g(a) = 0$. Recall that

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} \quad \text{or} \quad f(x+h) \approx f(x) + hf'(x),$$

Using this gives

$$\begin{aligned}
\lim_{x \to a^+} \frac{f(x)}{g(x)} &= \lim_{h \to 0} \frac{f(a+h)}{g(a+h)} \\
&\approx \lim_{h \to 0} \frac{f(a) + hf'(a)}{g(a) + hg'(a)} \\
&= \frac{f'(a)}{g'(a)} = \lim_{x \to a^+} \frac{f'(x)}{g'(x)}.
\end{aligned}$$

l'Hôpital's Rule

- For example, to evaluate $\lim_{x\to 0^+} x^x$, use l'Hôpital's rule to see that

$$\lim_{x\to 0^+} x\ln x = \lim_{x\to 0^+} \frac{\ln x}{1/x} = \lim_{x\to 0^+} \frac{1/x}{-x^{-2}} = -\lim_{x\to 0^+} x = 0.$$

Then, by the continuity of the exponential function,

$$\begin{aligned}\lim_{x\to 0^+} x^x &= \lim_{x\to 0^+} \exp\left(\ln x^x\right)\\ &= \lim_{x\to 0^+} \exp\left(x\ln x\right)\\ &= \exp\left(\lim_{x\to 0^+} x\ln x\right) = \exp 0 = 1.\end{aligned}$$

Taylor Series

- A **sequence** is a function $f : \mathbb{N} \to \mathbb{R}$, with $f(n)$, $n \in \mathbb{N}$, being the n^{th} **term** of f. We often denote the sequence of f as $\{f_n\}$, where $f_n = f(n)$. Let $\{f_n\}$ be a sequence. If for any given $\epsilon > 0$, $\exists\, a \in \mathbb{R}$ and $\exists\, N \in \mathbb{N}$ such that, $\forall\, n \geq N$, $|a - f_n| < \epsilon$, then the sequence is **convergent**, and **converges to** (the unique value) a. If $\{f_n\}$ converges to a, then we write $\lim_{n\to\infty} f_n = a$. If $\{f_n\}$ does not converge, then it is said to **diverge**.
- Let f_k be a sequence. The sum

$$S = \sum_{k=1}^{\infty} f_k = \lim_{n\to\infty} \sum_{k=1}^{n} f_k = \lim_{n\to\infty} s_n$$

is referred to as a **series** associated with the sequence $\{f_k\}$ and $s_n = \sum_{k=1}^{n} f_k$ is its n^{th} partial sum. Series S **converges** if $\lim_{n\to\infty} s_n$ exists, i.e., if the limit is bounded, and **diverges** if the partial sums are not bounded.

Taylor Series

- A series of the form $\sum_{k=0}^{\infty} a_k x^k$ for sequence $\{a_k\}$ is a **power series** in x with coefficients a_k. More generally, $S(x) = \sum_{k=0}^{\infty} a_k (x-c)^k$ is a power series in $(x-c)$, where $c \in \mathbb{R}$.
- Let $f : I \to \mathbb{R}$, where I is an open interval, and let $c \in I$. If $f \in \mathcal{C}^{\infty}(I)$, then the **Taylor series** of f at c is

$$T = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!} (x-c)^k ,$$

and its nth partial T_n sum is called nth order **Taylor polynomial**.
- Observe that $T_n(c) = f(c)$, $T_n'(c) = \left. T_n'(x) \right|_{x=c} = f'(c)$, up to $T_n^{(r)}(c) = f^{(r)}(c)$, so that locally (i.e., for x near c), $T_n(c)$ behaves similarly to $f(x)$ and could be used for effective approximation.

Univariate Integration

- Let $A = [a, b]$ be a bounded interval in $\mathbb{R}$. A **partition** of A is a finite set $\pi = \{x_k\}_{k=0}^{n}$ such that $a = x_0 < x_1 < \cdots < x_n = b$, and its **mesh** (sometimes called the norm, or size), is given by $\mu(\pi) = \max\{x_1 - x_0, x_2 - x_1, \ldots, x_n - x_{n-1}\}$.
- If π_1 and π_2 are partitions of I such that $\pi_1 \subset \pi_2$, then π_2 is a **refinement** of π_1.
- A **selection** associated to a partition $\pi = \{x_k\}_{k=0}^{n}$ is any set $\{\xi_k\}_{k=1}^{n}$ such that $x_{k-1} \leq \xi_k \leq x_k$ for $k = 1, \ldots, n$.

Univariate Integration

- Now let $f : D \to \mathbb{R}$ with $A \subset D \subset \mathbb{R}$, $\pi = \{x_k\}_{k=0}^{n}$ be a partition of A, and $\sigma = \{\xi_k\}_{k=1}^{n}$ a selection associated to π. The **Riemann sum** for function f, with partition π and selection σ, is given by

$$S\left(f, \pi, \sigma\right) = \sum_{k=1}^{n} f\left(\xi_k\right)\left(x_k - x_{k-1}\right).$$

- Observe how S is just a sum of areas of rectangles with heights dictated by f, π and σ. If the Riemann sum converges to a real number as the level of refinement increases, then f is **integrable**.
- Formally, function f is said to be **(Riemann) integrable** over $A = [a, b]$ if there is a number $I \in \mathbb{R}$ such that: $\forall\ \epsilon > 0$, there exists a partition π_0 of A such that, for every refinement π of π_0, and every selection σ associated to π, we have $\left|S\left(f, \pi, \sigma\right) - I\right| < \epsilon$.

Univariate Integration

- The number I is called the **integral** of f over $[a, b]$, and denoted by $\int_a^b f$ or $\int_a^b f(x)\ dx$.
- Integrals can also be taken over infinite intervals; these are referred to as **improper integrals**. If function f is defined on $(-\infty, \infty)$, then

$$\int_{-\infty}^{\infty} f(x)\, dx := \lim_{a \to -\infty} \int_a^t f(x)\, dx + \lim_{b \to \infty} \int_t^b f(x)\, dx,$$

  for any point $t \in \mathbb{R}$, when both limits on the rhs exist.

Fundamental Theorem of Calculus

- Let $f : I \to \mathbb{R}$. The function $F : I \to \mathbb{R}$ is called an **indefinite integral**, or an **antiderivative**, or a **primitive** of f if, $\forall\, x \in I$, $F'(x) = f(x)$.
- The fundamental theorem of calculus, or, in short, FTC, is the link between the differential and integral calculus, of which there are two forms:
  1. Let f be Riemann integrable on $[a, b]$. If F is a primitive of f, then $\int_a^b f = F(b) - F(a)$.
  2. Let f be Riemann integrable on $I = [a, b]$ and define $F(x) = \int_a^x f$, $x \in I$. Then F is continuous on $[a, b]$ and, if f is continuous at $x \in I$, then $F'(x) = f(x)$.

Fundamental Theorem of Calculus

- The proof of the FTC #1 is very simple and worth knowing: Let $\pi = \{x_k\}_{k=0}^{n}$ be a partition of $I = [a, b]$. As $F'(t) = f(t)$ for all $t \in I$, applying the mean value theorem to F implies that

$$F(x_k) - F(x_{k-1}) = F'(\xi_k)(x_k - x_{k-1}) = f(\xi_k)(x_k - x_{k-1})$$

for some $\xi_k \in (x_{k-1},\ x_k)$. The set of ξ_k, $k = 1, \ldots n$ forms a selection, $\sigma = \{\xi_k\}_{k=1}^{n}$, associated to π, so that

$$\begin{aligned} S(f, \pi, \sigma) &= \sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}) \\ &= \sum_{k=1}^{n} (F(x_k) - F(x_{k-1})) \\ &= F(b) - F(a). \end{aligned}$$

This holds for any partition π, so that $\int_a^b f(t)\, dt = F(b) - F(a)$.

Fundamental Theorem of Calculus

▷ **Example**. Observe that $F(x) = e^{kx}/k + C$ is a primitive of $f(x) = e^{kx}$ for $k \in \mathbb{R} \setminus 0$ and any constant $C \in \mathbb{R}$, because, via the chain rule, $dF(x)/dx = f(x)$. Thus, from the FTC,

$$\int_a^b f = F(b) - F(a) = k^{-1}(e^{kb} - e^{ka}),$$

a result we will make great use of.

▷ **Example**. Let $I(x) = \int_0^{x^2} e^{-t} dt = 1 - e^{-x^2}$, so that $I'(x) = \frac{d}{dx}\left(1 - e^{-x^2}\right) = 2xe^{-x^2}$. Alternatively, let $G(y) = \int_0^y e^{-t} dt$, so that $I(x) = G\left(x^2\right) = G(f(x))$, where $f(x) = x^2$. Then, from the chain rule and the FTC,

$$I'(x) = G'(f(x))\, f'(x) = e^{-x^2} \cdot 2x,$$

as before, but without having to actually evaluate $I(x)$.

Fundamental Theorem of Calculus

- There is also a **Mean Value Theorem for Integrals**. It states: Let f, g be continuous functions on I, for $I = [a, b]$, with g nonnegative. Then $\exists\ c \in I$ such that

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx.$$

- A popular and useful form of the theorem just takes $g(x) \equiv 1$, so that $\int_a^b f = f(c)\,(b - a)$.

Integration Techniques

- The simple technique of **integration by parts** can be invaluable in many situations. It can be written

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du \qquad \text{or} \qquad \int u\,dv = uv - \int v\,du,$$

where $uv|_a^b := u(b)\,v(b) - u(a)\,v(a)$.

- Another vastly useful technique is the **change of variables**, often expressed as

$$\int f[g(x)]g'(x)dx = \int f(u)du, \quad \text{where} \quad u = g(x).$$

Integration Techniques: Example

The natural logarithm is often defined as $\ln x = \int_1^x t^{-1} dt$, from which its most important properties follow, such as $\ln 1 = 0$ and $\ln(xy) = \ln x + \ln y$. For the latter, let $u = t/x$ so that $t = xu$, $dt = xdu$, so that

$$\int_x^{xy} t^{-1} dt = \int_1^y \frac{1}{u} du.$$

Thus,

$$\begin{aligned} \ln(xy) &= \int_1^{xy} t^{-1} dt \\ &= \int_1^x t^{-1} dt + \int_x^{xy} t^{-1} dt \\ &= \int_1^x t^{-1} dt + \int_1^y t^{-1} dt = \ln x + \ln y. \end{aligned}$$

Similarly, $\ln(x/y) = \ln x - \ln y$.

Exercise

Simplify the integral

$$I(z) = \int_0^\infty \exp\left(-(x - z/x)^2\right) dx, \qquad z \geq 0.$$

Hint:

- Start with the substitution $u = z/x$,
- compare the result to $I(z)$,
- let $k = u - z/u$, and
- use the fact that

$$\int_{-\infty}^\infty \exp\left(-k^2\right) dk = \sqrt{\pi}.$$

Solution

Let $u = z/x$ ($x = z/u$ and $dx = -zu^{-2}du$) so that

$$I(z) = -\int_{\infty}^{0} \frac{z}{u^2} \exp\left(-\left(z/u - u\right)^2\right) du$$
$$= \int_{0}^{\infty} \frac{z}{u^2} \exp\left(-\left(u - z/u\right)^2\right) du,$$

which, when compared to $I(z)$, implies that

$$\int_{0}^{\infty} \left(1 + \frac{z}{u^2}\right) \exp\left(-\left(u - z/u\right)^2\right) du = 2I(z).$$

Solution, cont.

Let $k = u - z/u$, so that $dk = \left(1 + zu^{-2}\right) du$. Then, for $u \downarrow 0$, $k \downarrow -\infty$, and the lhs of (40) simplifies to

$$\int_{-\infty}^{\infty} \exp\left(-k^2\right) dk,$$

i.e., from the hint,

$$I(z) = \frac{\sqrt{\pi}}{2},$$

which, interestingly and non-intuitively, does NOT depend on the value of z, only that z it is nonnegative.

Multivariate Calculus: Differentiation

- We use bold face to denote a point in $\mathbb{R}^n$, e.g., $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and also for multivariate functions, e.g., $\mathbf{f} : \mathbb{R} \to \mathbb{R}^m$, $m > 1$.
- Let $\mathbf{f} : A \to \mathbb{R}^m$ with $A \subset \mathbb{R}^n$ $\mathbf{x}_0 \in \overline{A}$ (the closure of A). Paralleling the univariate case, $\lim_{\mathbf{x}\to\mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{b}$ if, $\forall\ \epsilon > 0$, $\exists\ \delta > 0$ such that, when $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ and $\mathbf{x} \in A$, $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\| < \epsilon$.
- Function $\mathbf{f}$ is **continuous** at $\mathbf{a} \in A$ if, for $\mathbf{x} \in A$, $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$.

Multivariate Calculus: Differentiation

- Let $f : A \to \mathbb{R}$ with $A \subset \mathbb{R}^n$ an open set. For every $\mathrm{x} = (x_1, \ldots, x_n) \in A$ and for each $i = 1, 2, \ldots, n$, the **partial derivative** of f with respect to x_i is defined as

$$\frac{\partial f}{\partial x_i}(\mathrm{x}) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{h},$$

  if the limit exists.
- An alternative notation for the partial derivative is $D_i f(\mathrm{x})$, termed the partial derivative of f with respect to the i^{th} variable.
- The **gradient** of f, denoted $(\operatorname{grad} f)(\mathrm{x})$, or $(\nabla f)(\mathrm{x})$, is the row vector of all partial derivatives:

$$(\operatorname{grad} f)(\mathrm{x}) = (D_1 f(\mathrm{x}), \ldots, D_n f(\mathrm{x})).$$

Multivariate Calculus: Differentiation

- Now consider a multivariate function $\mathrm{f} : A \subset \mathbb{R}^n \to \mathbb{R}^m$ with A an open set, where f is such that $\mathrm{f}(\mathrm{x}) = (f_1(\mathrm{x}), \ldots, f_m(\mathrm{x}))$, $f_i : A \to \mathbb{R}$, $i = 1, \ldots, m$, for all $\mathrm{x} = (x_1, \ldots, x_n)' \in A$. If each partial derivative, $\partial f_i(\mathrm{x}_0)/\partial x_j$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, exists, then the **Jacobian matrix** of f at $\mathrm{x}_0 \in A$ is the $m \times n$ matrix

$$\mathrm{f}'(\mathrm{x}_0) := \mathrm{J}_{\mathrm{f}}(\mathrm{x}_0) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathrm{x}_0) & \ldots & \frac{\partial f_1}{\partial x_n}(\mathrm{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathrm{x}_0) & \ldots & \frac{\partial f_m}{\partial x_n}(\mathrm{x}_0) \end{pmatrix} = \begin{pmatrix} (\mathrm{grad}\, f_1)(\mathrm{x}_0) \\ \vdots \\ (\mathrm{grad}\, f_m)(\mathrm{x}_0) \end{pmatrix}.$$

- When $m = 1$, the total derivative is just the gradient.

Second Order Derivative

- Let $f : A \to \mathbb{R}$ with $A \subset \mathbb{R}^n$ an open set such that the partial derivatives $D_i f(\mathrm{x})$ are continuous at point $\mathrm{x} = (x_1, \ldots, x_n) \in A$, $i = 1, 2, \ldots, n$. As $D_i f$ is a function, its partial derivative may be computed, if it exists, i.e., we can apply the D_j operator to $D_i f$ to get $D_j D_i f$, called the iterated partial derivative of f with respect to i and j.
- The **Hessian matrix**, denoted $\mathrm{hess} f$, collects all second order derivatives into an $n \times n$ matrix:

$$(\mathrm{hess} f)(\mathrm{x}) = \begin{pmatrix} D_1 D_1 f(\mathrm{x}) & \cdots & D_1 D_n f(\mathrm{x}) \\ \vdots & \ddots & \vdots \\ D_n D_1 f(\mathrm{x}) & \cdots & D_n D_n f(\mathrm{x}) \end{pmatrix}.$$

Second Order Derivative

- A fundamental result is that if $D_i f$, $D_j f$, $D_i D_j f$ and $D_j D_i f$ exist and are continuous, then
$$D_i D_j f = D_j D_i f.$$
- Another way of stating this is that if all derivatives exist and are continuous, then the hessian matrix is symmetric.

Integration

- Let $a_1, a_2, b_1, b_2 \in \mathbb{R}$ with $a_1 < b_1$, $a_2 < b_2$, and let the bounded rectangle $D := [a_1, b_1] \times [a_2, b_2]$.
- For a function $f : D \to \mathbb{R}$, we define the Riemann integral of similarly to the one–dimensional case. There, the domain of definition of the function being integrated is split into ever-shorter intervals, but here, one has to partition the *rectangle* into smaller and smaller rectangles.
- If f is Riemann integrable, then the Riemann sums, defined similarly to those in the one–dimensional case, converge to a value called the Riemann integral of f, denoted by $\int_D f$ or $\int_D f(\mathrm{x})\, d\mathrm{x}$.

Integration

- If f is continuous on the D, we can use **Fubini's theorem** to calculate its integral:

$$\int_D f(\mathrm{x})\,d\mathrm{x} = \int_{a_2}^{b_2} \left[\int_{a_1}^{b_1} f(x_1, x_2)\,dx_1 \right] dx_2$$
$$= \int_{a_1}^{b_1} \left[\int_{a_2}^{b_2} f(x_1, x_2)\,dx_2 \right] dx_1.$$

- This is just a set of nested *univariate* integrals.

Integration

- The result can be extended in an obvious way to the n–dimensional case.
- Under certain conditions of f, Fubini's theorem still applies for unbounded regions of the form $D_\infty := (-\infty, b_1] \times (-\infty, b_2]$:

$$\int_{D_\infty} f = \int_{-\infty}^{b_2} \left[\int_{-\infty}^{b_1} f(x_1, x_2)\, dx_1 \right] dx_2 = \int_{-\infty}^{b_1} \left[\int_{-\infty}^{b_2} f(x_1, x_2)\, dx_2 \right] dx_1.$$

Exercise

Let $f : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto y^2 e^{2x}$ and $D = [0, 1] \times [0, 1]$. Compute $\int_D f$ by integrating x first, then y. Exchange the order of integration and compare the results.

Solution

$$\int_D f = \int_0^1 \int_0^1 y^2 e^{2x}\, dx\, dy = \int_0^1 \left[y^2 \frac{1}{2} e^{2x} \right]_{x=0}^{x=1} dy$$

$$= \int_0^1 y^2 \frac{1}{2} e^2 - y^2 \frac{1}{2} 1\, dy = \left[\frac{1}{3} y^3 \left(\frac{1}{2} e^2 - \frac{1}{2} \right) \right]_{y=0}^{y=1} = \frac{1}{6}(e^2 - 1).$$

The same result can be easily derived when interchanging the order of integration. However, in this example, the calculations can be simplified by **factorizing** the integrated function:

$$\int_D f = \int_0^1 \int_0^1 y^2 e^{2x}\, dx\, dy = \int_0^1 y^2 \int_0^1 e^{2x}\, dx\, dy$$

$$= \int_0^1 e^{2x}\, dx \int_0^1 y^2\, dy = \left[\frac{1}{2} e^{2x} \right]_0^1 \left[\frac{1}{3} y^3 \right]_0^1 = \frac{1}{2}(e^2 - 1)\frac{1}{3}.$$

Usually, a factorization will not be possible.

Leibniz' Rule

Let $a_1, a_2, b_1, b_2 \in \mathbb{R}$ with $a_1 < b_1$, $a_2 < b_2$, let $D := [a_1, b_1] \times [a_2, b_2]$, and let $f : D \to \mathbb{R}$ be continuous with continuous derivative $\frac{\partial}{\partial x} f(x, y)$. Then the function

$$A(x) := \int_{a_1}^{b_1} f(x, y)\, dy$$

is differentiable for $x \in [a_2, b_2]$, and its derivative is

$$A'(x) = \int_{a_1}^{b_1} \frac{\partial}{\partial x} f(x, y)\, dy,$$

i.e., we can interchange the order of differentiation and integration.

Leibniz' Rule

Now let us consider what happens if the limits of integration themselves depend on x. In particular, let λ and θ be differentiable functions defined on $[a_2, b_2]$ such that $\lambda(x), \theta(x) \in [a_1, b_1]$ for all $x \in [a_2, b_2]$, so that

$$A(x) := \int_{\lambda(x)}^{\theta(x)} f(x, y)\, dy.$$

Leibniz' Rule

- **Leibniz' rule** says that

$$A'(x) = -f\left(x, \lambda(x)\right)\lambda'(x) + f\left(x, \theta(x)\right)\theta'(x) + \int_{\lambda(x)}^{\theta(x)} \frac{\partial f(x,y)}{\partial x}\, dy.$$

- The intuition behind this is as follows: the first term is the contribution from "moving" the lower endpoint, and the second term from "moving" the upper endpoint of integration. The third term is just as before.

Exercise

Consider the function $f(t) := \int_0^t e^{st}\, ds$. Calculate its derivative at $t = 1$ in two ways, first by doing the integral and differentiating afterwards, then by using Leibniz' rule.

Solution

- First way: for $t > 0$,

$$f(t) = \left[\frac{1}{t}e^{st}\right]_0^t = \frac{1}{t}\left(e^{t^2} - 1\right),$$

and

$$f'(t) = \frac{-1}{t^2}\left(e^{t^2} - 1\right) + \frac{1}{t}2te^{t^2} = \frac{-1}{t^2}\left(e^{t^2} - 1\right) + 2e^{t^2},$$

so that $f'(1) = -(e-1) + 2e = 1 + e$.

- Second way: for $t > 0$,

$$f'(t) = \int_0^t se^{st}ds + 1 \cdot e^{t^2},$$

hence

$$f'(1) = \int_0^1 se^{s}ds + e^1 = [(s-1)e^s]_0^1 + e = 0 - (-1) + e = 1 + e.$$