

Chapter 1

Preliminaries From Calculus

Stochastic calculus deals with functions of time t , $0 \leq t \leq T$. In this chapter some concepts of the infinitesimal calculus used in the sequel are given.

1.1 Functions in Calculus

Continuous and Differentiable Functions

A function g is called continuous at the point $t = t_0$ if the increment of g over small intervals is small,

$$\Delta g(t) = g(t) - g(t_0) \rightarrow 0 \text{ as } \Delta t = t - t_0 \rightarrow 0.$$

If g is continuous at every point of its domain of definition, it is simply called continuous.

g is called differentiable at the point $t = t_0$ if at that point

$$\Delta g \sim C \Delta t \text{ or } \lim_{\Delta t \rightarrow 0} \frac{\Delta g(t)}{\Delta t} = C,$$

this constant C is denoted by $g'(t_0)$. If g is differentiable at every point of its domain, it is called differentiable.

An important application of the derivative is a theorem on finite increments.

Theorem 1.1 (Mean Value Theorem) *If f is continuous on $[a, b]$ and has a derivative on (a, b) , then there is c , $a < c < b$, such that*

$$f(b) - f(a) = f'(c)(b - a). \tag{1.1}$$

Clearly, differentiability implies continuity, but not the other way around, as continuity states that the increment Δg converges to zero together with Δt , whereas differentiability states that this convergence is at the same rate or faster.

Example 1.1: The function $g(t) = \sqrt{t}$ is not differentiable at 0, as at this point

$$\frac{\Delta g}{\Delta t} = \frac{\sqrt{\Delta t}}{\Delta t} = \frac{1}{\sqrt{\Delta t}} \rightarrow \infty$$

as $t \rightarrow 0$.

It is surprisingly difficult to construct an example of a continuous function which is not differentiable at *any* point.

Example 1.2: An example of a continuous, nowhere differentiable function was given by the Weierstrass in 1872: for $0 \leq t \leq 2\pi$

$$f(t) = \sum_{n=1}^{\infty} \frac{\cos(3^n t)}{2^n}. \quad (1.2)$$

We don't give a proof of these properties, a justification for continuity is given by the fact that if a sequence of continuous functions converges uniformly, then the limit is continuous; and a justification for non-differentiability can be provided in some sense by differentiating term by term, which results in a divergent series.

To save repetition the following notations are used: *a continuous function f is said to be a C function; a differentiable function f with continuous derivative is said to be a C^1 function; a twice differentiable function f with continuous second derivative is said to be a C^2 function; etc.*

Right and Left-Continuous Functions

We can rephrase the definition of a continuous function: a function g is called continuous at the point $t = t_0$ if

$$\lim_{t \rightarrow t_0} g(t) = g(t_0), \quad (1.3)$$

it is called right-continuous (left-continuous) at t_0 if the values of the function $g(t)$ approach $g(t_0)$ when t approaches t_0 from the right (left)

$$\lim_{t \downarrow t_0} g(t) = g(t_0), \quad \left(\lim_{t \uparrow t_0} g(t) = g(t_0). \right) \quad (1.4)$$

If g is continuous it is, clearly, both right and left-continuous.

The left-continuous version of g , denoted by $g(t-)$, is defined by taking left limit at each point,

$$g(t-) = \lim_{s \uparrow t} g(s). \quad (1.5)$$

From the definitions we have: g is left-continuous if $g(t) = g(t-)$.

The concept of $g(t+)$ is defined similarly,

$$g(t+) = \lim_{s \downarrow t} g(s). \quad (1.6)$$

If g is a right-continuous function then $g(t+) = g(t)$ for any t , so that $g_+ = g$.

Definition 1.2 A point t is called a discontinuity of the first kind or a jump point if both limits $g(t+)$ and $g(t-)$ exist and are not equal. The jump at t is defined as $\Delta g(t) = g(t+) - g(t-)$. Any other discontinuity is said to be of the second kind.

Example 1.3: The function $\sin(1/t)$ for $t \neq 0$ and 0 for $t = 0$ has discontinuity of the second kind at zero, because the limits from the right or the left don't exist.

An important result is that a function can have at most countably many jump discontinuities (see for example Hobson (1921), p.286).

Theorem 1.3 A function defined on an interval $[a, b]$ can have no more than countably many jumps.

A function, of course, can have more than countably many discontinuities, but then they are not all jumps, i.e. would not have limits from right or left.

Another useful result is that a derivative cannot have jump discontinuities at all.

Theorem 1.4 If f is differentiable with a finite derivative $f'(t)$ in an interval, then at all points $f'(t)$ is either continuous or has a discontinuity of the second kind.

PROOF: If t is such that $f'(t+) = \lim_{s \downarrow t} f'(s)$ exists (finite or infinite), then by the Mean Value Theorem the same value is taken by the derivative from the right

$$f'(t) = \lim_{\Delta t \downarrow 0} \frac{f(t + \Delta) - f(t)}{\Delta} = \lim_{\Delta \downarrow 0, 0 < c < \Delta} f'(c) = f'(t+).$$

Similarly for the derivative from the left, $f'(t) = f'(t-)$. Hence $f'(t)$ is continuous at t . The result follows. □

This result explains why functions with continuous derivatives are sought as solutions to ordinary differential equations.

Functions considered in Stochastic Calculus

Functions considered in stochastic calculus are functions without discontinuities of the second kind, that is functions that have both right and left limits at any point of the domain and have one-sided limits at the boundary. These functions are called *regular* functions. It is often agreed to identify functions if they have the same right and left limits at any point.

The class $D = D[0, T]$ of right-continuous functions on $[0, T]$ with left limits has a special name, *càdlàg* functions (which is the abbreviation of “right continuous with left limits” in French). Sometimes these processes are called R.R.C. for regular right continuous. Notice that this class of processes includes C , the class of continuous functions.

Let $g \in D$ be a càdlàg function, then by definition, all the discontinuities of g are jumps. By Theorem 1.3 such functions have no more than countably many discontinuities.

Remark 1.1: In stochastic calculus $\Delta g(t)$ usually stands for the size of the jump at t . In standard calculus $\Delta g(t)$ usually stands for the increment of g over $[t, t + \Delta]$, $\Delta g(t) = g(t + \Delta) - g(t)$. The meaning of $\Delta g(t)$ will be clear from the context.

1.2 Variation of a Function

If g is a function of real variable, its variation over the interval $[a, b]$ is defined as

$$V_g([a, b]) = \sup \sum_{i=1}^n |g(t_i^n) - g(t_{i-1}^n)|, \quad (1.7)$$

where the supremum is taken over partitions:

$$a = t_0^n < t_1^n < \dots < t_n^n = b. \quad (1.8)$$

Clearly, (by the triangle inequality) the sums in (1.7) increase as new points are added to the partitions. Therefore variation of g is

$$V_g([a, b]) = \lim_{\delta_n \rightarrow 0} \sum_{i=1}^n |g(t_i^n) - g(t_{i-1}^n)|, \quad (1.9)$$

where $\delta_n = \max_{1 \leq i \leq n} (t_i - t_{i-1})$. If $V_g([a, b])$ is finite then g is said to be a function of finite variation on $[a, b]$. If g is a function of $t \geq 0$, then the variation function of g as a function of t is defined by

$$V_g(t) = V_g([0, t]).$$

Clearly, $V_g(t)$ is a non-decreasing function of t .

Definition 1.5 g is of finite variation if $V_g(t) < \infty$ for all t . g is of bounded variation if $\sup_t V_g(t) < \infty$, in other words, if for all t , $V_g(t) < C$, a constant independent of t .

Example 1.4:

1. If $g(t)$ is increasing then for any i , $g(t_i) > g(t_{i-1})$ resulting in a telescoping sum, where all the terms excluding the first and the last cancel out, leaving

$$V_g(t) = g(t) - g(0).$$

2. If $g(t)$ is decreasing then, similarly,

$$V_g(t) = g(0) - g(t).$$

Example 1.5: If $g(t)$ is differentiable with continuous derivative $g'(t)$, $g(t) = \int_0^t g'(s)ds$, and $\int_0^t |g'(s)|ds < \infty$, then

$$V_g(t) = \int_0^t |g'(s)|ds.$$

This can be seen by using the definition and the mean value theorem. $\int_{t_{i-1}}^{t_i} g'(s)ds = g'(\xi_i)(t_i - t_{i-1})$, for some $\xi_i \in (t_{i-1}, t_i)$. Thus $|\int_{t_{i-1}}^{t_i} g'(s)ds| = |g'(\xi_i)|(t_i - t_{i-1})$, and

$$\begin{aligned} V_g(t) &= \lim \sum_{i=1}^n |g(t_i) - g(t_{i-1})| = \lim \sum_{i=1}^n \left| \int_{t_{i-1}}^{t_i} g'(s)ds \right| \\ &= \sup \sum_{i=1}^n |g'(\xi_i)|(t_i - t_{i-1}) = \int_0^t |g'(s)|ds. \end{aligned}$$

The last equality is due to the last sum being a Riemann sum for the final integral.

Alternatively, the result can be seen from the decomposition of the derivative into the positive and negative parts,

$$g(t) = \int_0^t g'(s)ds = \int_0^t [g'(s)]^+ ds - \int_0^t [g'(s)]^- ds.$$

Notice that $[g'(s)]^-$ is zero when $[g'(s)]^+$ is positive, and the other way around. Using this one can see that the total variation of g is given by the sum of the variation of the above integrals. But these integrals are monotone functions with the value zero at zero. Hence

$$\begin{aligned} V_g(t) &= \int_0^t [g'(s)]^+ ds + \int_0^t [g'(s)]^- ds \\ &= \int_0^t ([g'(s)]^+ + [g'(s)]^-) ds = \int_0^t |g'(s)| ds. \end{aligned}$$

Example 1.6: (Variation of a pure jump function).

If g is a regular right-continuous (càdlàg) function or regular left-continuous (càglàd), and changes only by jumps,

$$g(t) = \sum_{0 \leq s \leq t} \Delta g(s),$$

then it is easy to see from the definition that

$$V_g(t) = \sum_{0 \leq s \leq t} |\Delta g(s)|.$$

Example 1.7: The function $g(t) = t \sin(1/t)$ for $t > 0$, and $g(0) = 0$ is continuous on $[0, 1]$, differentiable at all points except zero, but has infinite variation on any interval that includes zero. Take the partition $1/(2\pi k + \pi/2), 1/(2\pi k - \pi/2)$, $k = 1, 2, \dots$

The following theorem gives necessary and sufficient conditions for a function to have finite variation.

Theorem 1.6 (Jordan Decomposition) *Any function $g : [0, \infty) \rightarrow \mathbb{R}$ of finite variation can be expressed as the difference of two increasing functions*

$$g(t) = a(t) - b(t).$$

One such decomposition is given by

$$a(t) = V_g(t) \quad b(t) = V_g(t) - g(t). \quad (1.10)$$

It is easy to check that $b(t)$ is increasing, and $a(t)$ is obviously increasing. The representation of a function of finite variation as difference of two increasing functions is not unique. Another decomposition is

$$g(t) = \frac{1}{2}(V_g(t) + g(t)) - \frac{1}{2}(V_g(t) - g(t)).$$

The sum, the difference and the product of functions of finite variation are also functions of finite variation. This is also true for the ratio of two functions of finite variation provided the modulus of the denominator is larger than a positive constant.

The following result follows by Theorem 1.3, and its proof is easy.

Theorem 1.7 *A finite variation function can have no more than countably many discontinuities. Moreover, all discontinuities are jumps.*

PROOF: It is enough to establish the result for monotone functions, since a function of finite variation is a difference of two monotone functions.

A monotone function has left and right limits at any point, therefore any discontinuity is a jump. The number of jumps of size greater or equal to $\frac{1}{n}$ is no more than $(g(b) - g(a))n$. The set of all jump points is a union of the sets of jump points with the size of the jumps greater or equal to $\frac{1}{n}$. Since each such set is finite, the total number of jumps is at most countable. \square

A sufficient condition for a *continuous* function to be of finite variation is given by the following theorem, the proof of which is outlined in Example 1.5.

Theorem 1.8 *If g is continuous, g' exists and $\int |g'(t)|dt < \infty$ then g is of finite variation.*

Theorem 1.9 (Banach) *Let $g(t)$ be a continuous function on $[0, 1]$, and denote by $s(a)$ the number of t 's with $g(t) = a$. Then the variation of g is $\int_{-\infty}^{\infty} s(a)da$.*

Continuous and Discrete Parts of a Function

Let $g(t)$, $t \geq 0$, be a right-continuous increasing function. Then it can have at most countably many jumps, moreover the sum of the jumps is finite over finite time intervals. Define the discontinuous part g^d of g by

$$g^d(t) = \sum_{s \leq t} (g(s) - g(s-)) = \sum_{0 < s \leq t} \Delta g(s), \quad (1.11)$$

and the continuous part g^c of g by

$$g^c(t) = g(t) - g^d(t). \quad (1.12)$$

Clearly, g^d changes only by jumps, g^c is continuous and $g(t) = g^c(t) + g^d(t)$. Since a finite variation function is the difference of two increasing functions, the decomposition (1.12) holds for functions of finite variation. Although representation as the difference of increasing functions is not unique, decomposition (1.12) is essentially unique, in a sense that any two such decomposition differ by a constant. Indeed, if there were another such decomposition $g(t) = h^c(t) + h^d(t)$, then $h^c(t) - g^c(t) = g^d(t) - h^d(t)$, implying that $h^d - g^d$ is continuous. Hence h^d and g^d have the same set of jump points, and it follows that $h^d(t) - g^d(t) = c$ for some constant c .

Quadratic Variation

If g is a function of real variable, define its quadratic variation over the interval $[0, t]$ as the limit (when it exists)

$$[g](t) = \lim_{\delta_n \rightarrow 0} \sum_{i=1}^n (g(t_i^n) - g(t_{i-1}^n))^2, \quad (1.13)$$

where the limit is taken over partitions: $0 = t_0^n < t_1^n < \dots < t_n^n = t$, with $\delta_n = \max_{1 \leq i \leq n} (t_i^n - t_{i-1}^n)$.

Remark 1.2: Similarly to the concept of variation, there is a concept of Φ -variation of a function. If $\Phi(u)$ is a positive function, increasing monotonically with u then the Φ -variation of g on $[0, t]$ is

$$V_\Phi[g] = \sup \sum_{i=1}^n \Phi(|g(t_i^n) - g(t_{i-1}^n)|), \quad (1.14)$$

where supremum is taken over *all* partitions. Functions with finite Φ -variation on $[0, t]$ form a class V_Φ . With $\Phi(u) = u$ one obtains the class VF of functions of finite variation, with $\Phi(u) = u^p$ one obtains the class of functions of p -th finite variation, VF_p . If $1 \leq p < q < \infty$, then finite p -variation implies finite q -variation.

The stochastic calculus definition of quadratic variation is different to the classical one with $p = 2$ (unlike for the first variation $p = 1$, when they are the same). In stochastic calculus the limit in (1.13) is taken over *shrinking* partitions with $\delta_n = \max_{1 \leq i \leq n} (t_i^n - t_{i-1}^n) \rightarrow 0$, and not over all possible partitions. We shall use only the stochastic calculus definition.

Quadratic variation plays a major role in stochastic calculus, but is hardly ever met in standard calculus due to the fact that smooth functions have zero quadratic variation.

Theorem 1.10 *If g is continuous and of finite variation then its quadratic variation is zero.*

PROOF:

$$\begin{aligned} [g](t) &= \lim_{\delta_n \rightarrow 0} \sum_{i=0}^{n-1} (g(t_{i+1}^n) - g(t_i^n))^2 \\ &\leq \lim_{\delta_n \rightarrow 0} \max_i |g(t_{i+1}^n) - g(t_i^n)| \sum_{i=0}^{n-1} |g(t_{i+1}^n) - g(t_i^n)| \\ &\leq \lim_{\delta_n \rightarrow 0} \max_i |g(t_{i+1}^n) - g(t_i^n)| V_g(t). \end{aligned}$$

Since g is continuous, it is uniformly continuous on $[0, t]$, hence $\lim_{\delta_n \rightarrow 0} \max_i |g(t_{i+1}^n) - g(t_i^n)| = 0$, and the result follows. \square

Remark that there are functions with zero quadratic variation and infinite variation (called functions of *zero energy*).

Define the quadratic covariation (or simply covariation) of f and g on $[0, t]$ by the following limit (when it exists)

$$[f, g](t) = \lim_{\delta_n \rightarrow 0} \sum_{i=0}^{n-1} (f(t_{i+1}^n) - f(t_i^n)) (g(t_{i+1}^n) - g(t_i^n)), \quad (1.15)$$

when the limit is taken over partitions $\{t_i^n\}$ of $[0, t]$ with $\delta_n = \max_i(t_{i+1}^n - t_i^n)$.

The same proof as for Theorem 1.10 works for the following result

Theorem 1.11 *If f is continuous and g is of finite variation, then their covariation is zero $[f, g](t) = 0$.*

Let f and g be such that their quadratic variation is defined. By using simple algebra, one can see that covariation satisfies

Theorem 1.12 (Polarization identity)

$$[f, g](t) = \frac{1}{2} ([f + g, f + g](t) - [f, f](t) - [g, g](t)), \quad (1.16)$$

It is obvious that covariation is symmetric, $[f, g](t) = [g, f](t)$, it follows from(1.16) that it is linear, that is, for any constants α and β

$$[\alpha f + \beta g, h](t) = \alpha [f, h](t) + \beta [g, h](t). \quad (1.17)$$

Due to symmetry it is bilinear, that is, linear in both arguments. Thus the quadratic variation of the sum can be opened similarly to multiplication of sums $(\alpha_1 f + \beta_1 g)(\alpha_2 h + \beta_2 k)$. It follows from the definition of quadratic variation, that it is a non-decreasing function in t , and consequently it is of finite variation. By the polarization identity, covariation is also of finite variation. More about quadratic variation is given in the Stochastic Calculus chapter.

1.3 Riemann Integral and Stieltjes Integral

Riemann Integral

The Riemann Integral of f over interval $[a, b]$ is defined as the limit of Riemann sums

$$\int_a^b f(t) dt = \lim_{\delta \rightarrow 0} \sum_{i=1}^n f(\xi_i^n) (t_i^n - t_{i-1}^n), \quad (1.18)$$

where t_i^n 's represent partitions of the interval,

$$a = t_0^n < t_1^n < \dots < t_n^n = b, \delta = \max_{1 \leq i \leq n} (t_i^n - t_{i-1}^n), \text{ and } t_{i-1}^n \leq \xi_i^n \leq t_i^n.$$

It is possible to show that Riemann Integral is well defined for continuous functions, and by splitting up the interval, it can be extended to functions which are discontinuous at finitely many points.

Calculation of integrals is often done by using the antiderivative, and is based on the the following result.

Theorem 1.13 (The fundamental theorem of calculus) *If f is differentiable on $[a, b]$ and f' is Riemann integrable on $[a, b]$ then*

$$f(b) - f(a) = \int_a^b f'(s) ds.$$

In general, this result cannot be applied to discontinuous functions, see example below. For such functions a jump term must be added, see (1.20).

Example 1.8: Let $f(t) = 2$ for $1 \leq t \leq 2$, $f(t) = 1$ for $0 \leq t < 1$. Then $f'(t) = 0$ at all $t \neq 1$. $\int_0^t f'(s) ds = 0 \neq f(t)$. f is continuous and is differentiable at all points but one, the derivative is integrable, but the function does not equal the integral of its derivative.

Main tools for calculations of Riemann integrals are change of variables and integration by parts. These are reviewed below in a more general framework of the Stieltjes integral.

Stieltjes Integral

The Stieltjes Integral is an integral of the form $\int_a^b f(t) dg(t)$, where g is a function of finite variation. Since a function of finite variation is a difference of two increasing functions, it is sufficient to define the integral with respect to monotone functions.

Stieltjes Integral with respect to Monotone Functions

The Stieltjes Integral of f with respect to a monotone function g over an interval $(a, b]$ is defined as

$$\int_a^b f dg = \int_a^b f(t) dg(t) = \lim_{\delta \rightarrow 0} \sum_{i=1}^n f(\xi_i^n) (g(t_i^n) - g(t_{i-1}^n)), \quad (1.19)$$

with the quantities appearing in the definition being the same as above for the Riemann Integral. This integral is a generalization of the Riemann Integral, which is recovered when we take $g(t) = t$. This integral is also known as the Riemann-Stieltjes integral.

Particular Cases

If $g'(t)$ exists, and $g(t) = g(0) + \int_0^t g'(s)ds$, then it is possible to show that

$$\int_a^b f(t)dg(t) = \int_a^b f(t)g'(t)dt.$$

If $g(t) = \sum_{k=a}^{[t]} h(k)$ (a integer, and $[t]$ stands for the integer part of t) then

$$\int_a^b f(t)dg(t) = \sum_a^b f(k)h(k).$$

This property allows us to represent sums as integrals.

Example 1.9:

$$1. \quad g(t) = 2t^2 \quad \int_a^b f(t)dg(t) = 4 \int_a^b tf(t)dt$$

$$2. \quad g(t) = \begin{cases} 0 & t < 0 \\ 2 & 0 \leq t < 1 \\ 3 & 1 \leq t < 2 \\ 5 & 2 \leq t \end{cases}$$

$$\int_{-\infty}^{\infty} f(t)dg(t) = 2f(0) + f(1) + 2f(2)$$

If, for example, $f(t) = t$ then $\int_{-\infty}^{\infty} tdg(t) = 5$. If $f(t) = (t+1)^2$ then $\int_{-\infty}^{\infty} (t+1)^2 dg(t) = 2 + 4 + 18 = 24$.

Let g be a function of finite variation and

$$g(t) = a(t) - b(t)$$

with $a(t) = V_g(t)$, $b(t) = V_g(t) - g(t)$, which are non-decreasing functions. If

$$\int_0^t |f(s)|da(s) = \int_0^t |f(s)|dV_g(s) := \int_0^t |f(s)||dg(s)| < \infty$$

then f is Stieltjes-integrable with respect to g and its integral is defined by

$$\int_{(0,t]} f(s)dg(s) = \int_{(0,t]} f(s)da(s) - \int_{(0,t]} f(s)db(s).$$

Notation: $\int_a^b f(s)dg(s) = \int_{(a,b]} f(s)dg(s)$.

Note: $\int_{(0,t]} dg(s) = g(t) - g(0)$ and $\int_{(0,t)} dg(s) = g(t-) - g(0)$.

If f is Stieltjes-integrable with respect to a function g of finite variation, then the variation of the integral is

$$V(t) = \int_0^t |f(s)||dg(s)| = \int_0^t |f(s)|dV_g(s).$$

Impossibility of a direct definition of an integral with respect to functions of infinite variation

In stochastic calculus we need to consider integrals with respect to functions of infinite variation. Such functions arise, for example, as models of stock prices. Integrals with respect to a function of infinite variation, cannot be defined as a usual limit of approximating sums. The following result explains, see for example, Protter (1992), p.40.

Theorem 1.14 *Let $\delta_n = \max_i(t_i^n - t_{i-1}^n)$ denote the largest interval in the partition of $[a, b]$. If*

$$\lim_{\delta_n \rightarrow 0} \sum_{i=1}^n f(t_{i-1}^n)[g(t_i^n) - g(t_{i-1}^n)]$$

exists for any continuous function f then g must be of finite variation on $[a, b]$.

This shows that if g has infinite variation then the limit of the approximating sums does not exist for some functions f .

Integration by Parts

Let f and g be functions of finite variation. Denote here $\Delta g(s) = g(s) - g(s-)$, then (with integrals on $(a, b]$)

$$\begin{aligned} f(b)g(b) - f(a)g(a) &= \int_a^b f(s-)dg(s) + \int_a^b g(s-)df(s) + \sum_{a < s \leq b} \Delta f(s)\Delta g(s) \\ &= \int_a^b f(s-)dg(s) + \int_a^b g(s)df(s). \end{aligned} \quad (1.20)$$

The last equation is obtained by putting together the sum of jumps with one of the integrals.

Note that although the sum in (1.20) is written over uncountably many values $a < s \leq b$, it has at most countably many non-zero terms. This is because a finite variation function can have at most a countable number of jumps.

If g is continuous so that $g(s-) = g(s)$ for all s then the formula simplifies and in this case we have the familiar integration by parts formula

$$f(b)g(b) - f(a)g(a) = \int_a^b f(s)dg(s) + \int_a^b g(s)df(s).$$

Example 1.10: Let $g(s)$ be of finite variation, $g(0) = 0$, and consider $g^2(s)$. Using the integration by parts with $f = g$, we have

$$g^2(t) = 2 \int_0^t g(s-) dg(s) + \sum_{s \leq t} (\Delta g(s))^2.$$

In other words,

$$\int_0^t g(s-) dg(s) = \frac{g^2(t)}{2} - \frac{1}{2} \sum_{s \leq t} (\Delta g(s))^2.$$

Now using the formula (1.20) above we also have

$$\int_0^t g(s) dg(s) = g^2(t) - \int_0^t g(s-) dg(s) = \frac{g^2(t)}{2} + \frac{1}{2} \sum_{s \leq t} (\Delta g(s))^2.$$

Thus it follows that

$$\int_0^t g(s-) dg(s) \leq \frac{g^2(t)}{2} \leq \int_0^t g(s) dg(s)$$

Change of Variables

Let f have a continuous derivative ($f \in C^1$) and g be of finite variation and continuous, then

$$f(g(t)) - f(g(0)) = \int_0^t f'(g(s)) dg(s) = \int_{g(0)}^{g(t)} f'(u) du.$$

If g is of finite variation has jumps, and is right-continuous then

$$\begin{aligned} f(g(t)) - f(g(0)) &= \int_0^t f'(g(s-)) dg(s) \\ &+ \sum_{0 < s \leq t} \left(f(g(s)) - f(g(s-)) - f'(g(s-)) \Delta g(s) \right), \end{aligned}$$

where $\Delta g(s) = g(s) - g(s-)$ denotes the jump of g at s . This is known in stochastic calculus as Itô's formula.

Example 1.11: Take $f(x) = x^2$, then we obtain

$$g^2(t) - g^2(0) = 2 \int_0^t g(s-) dg(s) + \sum_{s \leq t} (\Delta g(s))^2.$$

Remark 1.3: Note that for a continuous f and finite variation g on $[0, t]$ the approximating sums converge as $\delta = \max_i (t_{i+1}^n - t_i^n) \rightarrow 0$,

$$\sum_i f(g(t_i^n))(g(t_{i+1}^n) - g(t_i^n)) \rightarrow \int_0^t f(g(s-)) dg(s).$$

Remark 1.4: One of the shortcomings of Riemann or Stieltjes integrals is that they don't preserve the monotone convergence property, that is, for a sequence of functions $f_n \uparrow f$ does not necessarily follow that their integrals converge. The Lebesgue (or Lebesgue-Stieltjes) Integral preserves this property.

1.4 Lebesgue's Method of Integration

While Riemann sums are constructed by dividing the domain of integration on the x -axis, the interval $[a, b]$, into smaller subintervals, Lebesgue sums are constructed by dividing the range of the function on the y -axis, the interval $[c, d]$, into smaller subintervals $c = y_0 < y_1 < \dots < y_k < \dots < y_n = d$ and forming sums

$$\sum_{k=0}^{n-1} y_k \text{length}(\{t : y_k \leq f(t) < y_{k+1}\}).$$

The Lebesgue Integral is the limit of the above sums as the number of points in the partition increases. It turns out that the Lebesgue Integral is more general than the Riemann Integral, and preserves convergence. This approach also allows integration of functions in abstract probability spaces more general than \mathbb{R} or \mathbb{R}^n ; it requires additional concepts and is made more precise in the next chapter (see Section 2.3).

Remark 1.5: In folklore the following analogy is used. Imagine that money is spread out on a floor. In the Riemann method of integration, you collect the money as you progress in the room. In the Lebesgue method, first you collect \$100 bills everywhere you can find them, then \$50, etc.

1.5 Differentials and Integrals

The differential $df(t)$ of a differentiable function f at t is defined as the *linear in Δt part* of the increment at t , $f(t + \Delta) - f(t)$. If the differential of the independent variable is denoted $dt = \Delta t$, then $f(t + dt) - f(t) = df(t) +$ smaller order terms, and it follows from the existence of the derivative at t , that

$$df(t) = f'(t)dt. \tag{1.21}$$

If g is also a differentiable function of t , then $f(g(t))$ is differentiable, and

$$df(g(t)) = f'(g(t))g'(t)dt = f'(g(t))dg(t), \tag{1.22}$$

which is known as the *chain rule*.

Differential calculus is important in applications because many physical problems can be formulated in terms of differential equations. The main relation between the integral and the differential (or derivative) is given by the fundamental theorem of calculus, Theorem 1.13.

For differentiable functions, differential equations of the form

$$df(t) = \varphi(t)dw(t)$$

can be written in the integral form

$$f(t) = f(0) + \int_0^t \varphi(s)dw(s).$$

In Stochastic Calculus stochastic differentials do not formally exist and the random functions $w(t)$ are not differentiable at any point. By introducing a new (stochastic) integral, stochastic differential equations can be defined, and, by definition, solutions to these equations are given by the solutions to the corresponding stochastic integral equations.

1.6 Taylor's Formula and Other Results

This section contains Taylor's Formula and conditions on functions used in results on differential equations. It may be treated as an appendix, and referred to only when needed.

Taylor's Formula for Functions of One Variable

If we consider the increment of a function $f(x) - f(x_0)$ over the interval $[x_0, x]$, then provided $f'(x_0)$ exists, the differential at x_0 is the linear part in $(x - x_0)$ of this increment and it provides the first approximation to the increment. Taylor's formula gives a better approximation by taking higher order terms of powers of $(x - x_0)$ provided higher derivatives of f at x_0 exist. If f is a function of x with derivatives up to order $n + 1$, then

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \frac{1}{3!}f^{(3)}(x_0)(x - x_0)^3 \\ &+ \dots + \frac{1}{n!}f^{(n)}(x_0)(x - x_0)^n + R_n(x, x_0), \end{aligned}$$

where R_n is the remainder, and $f^{(n)}$ is the derivative of $f^{(n-1)}$. The remainder can be written in the form

$$R_n(x, x_0) = \frac{1}{(n+1)!}f^{(n+1)}(\theta_n)(x - x_0)^{n+1}$$

for some point $\theta_n \in (x_0, x)$.

In our applications we shall use this formula with two terms.

$$f(x) - f(x_0) = f'(x)(x - x_0) + \frac{1}{2}f''(\theta)(x - x_0)^2, \quad (1.23)$$

for some point $\theta \in (x_0, x)$.

Taylor's Formula for Functions of Several Variables

Similarly to the one-dimensional case, Taylor's formula gives successive approximations to the increment of a function. A function of n real variables $f(x_1, x_2, \dots, x_n)$ is differentiable at point $\mathbf{x} = (x_1, x_2, \dots, x_n)$ if the increment at this point can be approximated by a linear part, which is the differential of f at \mathbf{x} .

$$\Delta f(\mathbf{x}) = \sum_{i=1}^n C_i \Delta x_i + o(\rho), \quad \text{when } \rho = \sqrt{\sum_{i=1}^n (\Delta x_i)^2} \text{ and } \lim_{\rho \rightarrow 0} \frac{o(\rho)}{\rho} = 0. \quad (1.24)$$

If f is differentiable at $\mathbf{x} = (x_1, x_2, \dots, x_n)$, then in particular it is differentiable as a function of any one variable x_i at that point, when all the other coordinates are kept fixed. The derivative with respect to x_i is called the partial derivative $\partial f / \partial x_i$. Unlike in the one-dimensional case, the existence of all partial derivatives $\partial f / \partial x_i$ at \mathbf{x} , is necessary but not sufficient for differentiability of f at \mathbf{x} . But if all partial derivatives exist and are continuous at that point, then f is differentiable at that point, moreover, C_i in (1.24) is given by the value of $\partial f / \partial x_i$ at \mathbf{x} . If we define the differential of the independent variable as its increment $dx_i = \Delta x_i$, then we have

Theorem 1.15 *For f to be differentiable at a point, it is necessary that f has partial derivatives at that point, and it is sufficient that it has continuous partial derivatives at that point. If f is differentiable at \mathbf{x} , then its differential at \mathbf{x} is given by*

$$df(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_1, x_2, \dots, x_n) dx_i. \quad (1.25)$$

The first approximation of the increment of a differentiable function is the differential,

$$\Delta f(\mathbf{x}) \approx df(\mathbf{x}).$$

If f possesses higher order partial derivatives, then further approximation is possible and it is given by Taylor's formula. In Stochastic Calculus the second order approximation plays an important role.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be C^2 , ($f(x_1, x_2, \dots, x_n)$ has continuous partial derivatives up to order two), $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $\mathbf{x} + \Delta\mathbf{x} = (x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n)$ then by considering the function of one variable $g(t) = f(\mathbf{x} + t\Delta\mathbf{x})$ for $0 \leq t \leq 1$, the following result is obtained.

$$\begin{aligned} \Delta f(x_1, x_2, \dots, x_n) &= f(\mathbf{x} + \Delta\mathbf{x}) - f(\mathbf{x}) \approx \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_1, x_2, \dots, x_n) dx_i \\ &+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1 + \theta \Delta x_1, \dots, x_n + \theta \Delta x_n) dx_i dx_j, \end{aligned} \quad (1.26)$$

where just like in the case of one variable the second derivatives are evaluated at some ‘‘middle’’ point, $(x_1 + \theta \Delta x_1, \dots, x_n + \theta \Delta x_n)$ for some $\theta \in (0, 1)$, and $dx_i = \Delta x_i$.

Lipschitz and Hölder Conditions

Lipschitz and Hölder conditions describe subclasses of continuous functions. They appear as conditions on the coefficients in the results on the existence and uniqueness of solutions of ordinary and stochastic differential equations.

Definition 1.16 *f satisfies a Hölder condition (Hölder continuous) of order α , $0 < \alpha \leq 1$, on $[a, b]$ (\mathbb{R}) if there is a constant $K > 0$, so that for all $x, y \in [a, b]$ (\mathbb{R})*

$$|f(x) - f(y)| \leq K|x - y|^\alpha. \quad (1.27)$$

A Lipschitz condition is a Hölder condition with $\alpha = 1$,

$$|f(x) - f(y)| \leq K|x - y|. \quad (1.28)$$

It is easy to see that a Hölder continuous of order α function on $[a, b]$ is also Hölder continuous of any lesser order.

Example 1.12: The function $f(x) = \sqrt{x}$ on $[0, \infty)$ is Hölder continuous with $\alpha = 1/2$ but is not Lipschitz, since its derivative is unbounded near zero. To see that it is Hölder, it is enough to show that for all $x, y \geq 0$ the following ratio is bounded,

$$\frac{|\sqrt{x} - \sqrt{y}|}{\sqrt{|x - y|}} \leq K. \quad (1.29)$$

It is an elementary exercise to establish that the left hand side is bounded by dividing through by \sqrt{y} (if $y = 0$, then the bound is obviously one), and applying l'Hôpital's rule. Similarly $|x|^r$, $0 < r < 1$ is Hölder of order r .

A simple sufficient condition for a function to be Lipschitz is to be continuous and piecewise smooth, precise definitions follow.

Definition 1.17 f is smooth on $[a, b]$ if it possesses a continuous derivative f' on (a, b) such that the limits $f'(a+)$ and $f'(b-)$ exist.

Definition 1.18 f is piecewise continuous on $[a, b]$ if it is continuous on $[a, b]$ except possibly a finite number of points at which right-hand and left-hand limits exist.

Definition 1.19 f is piecewise smooth on $[a, b]$ if it is piecewise continuous on $[a, b]$ and f' exists and is also piecewise continuous on $[a, b]$.

Growth Conditions

Linear growth condition also appears in the results on existence and uniqueness of solutions of differential equations. $f(x)$ satisfies the linear growth condition if

$$|f(x)| \leq K(1 + |x|). \quad (1.30)$$

This condition describes the growth of a function for large values of x , and states that f is bounded for small values of x .

Example 1.13: It can be shown that if $f(0, t)$ is a bounded function of t , $|f(0, t)| \leq C$ for all t , and $f(x, t)$ satisfies the Lipschitz condition in x uniformly in t , $|f(x, t) - f(y, t)| \leq K|x - y|$, then $f(x, t)$ satisfies the linear growth condition in x , $|f(x, t)| \leq K_1(1 + |x|)$.

The polynomial growth condition on f is the condition of the form

$$|f(x)| \leq K(1 + |x|^m), \quad \text{for some } K, m > 0. \quad (1.31)$$

Theorem 1.20 (Gronwall's inequality) Let $f(t)$, $g(t)$ and $h(t)$ be non-negative on $[0, T]$, and for all $0 \leq t \leq T$

$$f(t) \leq g(t) + \int_0^t h(s)f(s)ds. \quad (1.32)$$

Then for $0 \leq t \leq T$

$$f(t) \leq g(t) + \int_0^t h(s)g(s) \exp\left(\int_s^t h(u)du\right)ds. \quad (1.33)$$

This form can be found for example, in Dieudonné (1960).

Solution of First Order Linear Differential Equations

Linear differential equations, by definition, are linear in the unknown function and its derivatives. A first order linear equation, in which the coefficient of $\frac{dx(t)}{dt}$ does not vanish, can be written in the form

$$\frac{dx(t)}{dt} + g(t)x(t) = k(t). \quad (1.34)$$

These equations are solved by using the Integrating Factor Method. The integrating factor is the function $e^{G(t)}$, where $G(t)$ is chosen by $G'(t) = g(t)$. After multiplying both sides of the equation by $e^{G(t)}$, integrating, and solving for $x(t)$, we have

$$x(t) = e^{-G(t)} \int_0^t \left(e^{G(s)} k(s) \right) ds + x(0)e^{G(0)-G(t)}. \quad (1.35)$$

The integrating factor $G(t)$ is determined up to a constant, but it is clear from (1.35), that the solution $x(t)$ remains the same.

Further Results on Functions and Integration

Results given here are not required to understand subsequent material. Some of these involve the concepts of a set of zero Lebesgue measure. This is given in the next chapter (see Section 2.2); any countable set has Lebesgue measure zero, but there are also uncountable sets of zero Lebesgue measure. A partial converse to Theorem 1.8 also holds (see, for example, Saks (1964), Freedman (1983) p.209, for the following results).

Theorem 1.21 (Lebesgue) *A finite variation function g on $[a, b]$ is differentiable almost everywhere on $[a, b]$.*

In what follows sufficient conditions for a function to be Lipschitz and not to be Lipschitz are given.

1. If f is continuously differentiable on a finite interval $[a, b]$, then it is Lipschitz. Indeed, since f' is continuous on $[a, b]$, it is bounded, $|f'| \leq K$. Therefore

$$|f(x) - f(y)| = \left| \int_x^y f'(t) dt \right| \leq \int_x^y |f'(t)| dt \leq K|x - y|. \quad (1.36)$$

2. If f is continuous and piecewise smooth then it is Lipschitz, the proof is similar to the above.
3. A Lipschitz function does not have to be differentiable, for example $f(x) = |x|$ is Lipschitz but it is not differentiable at zero.

4. It follows from the definition of a Lipschitz function (1.28), that if it is differentiable, then its derivative is bounded by K .
5. A Lipschitz function has finite variation on finite intervals, since for any partition $\{x_i\}$ of a finite interval $[a, b]$,

$$\sum |f(x_{i+1}) - f(x_i)| \leq K \sum (x_{i+1} - x_i) = K(b - a). \quad (1.37)$$

6. As functions of finite variation have derivatives almost everywhere (with respect to Lebesgue measure), a Lipschitz function is differentiable almost everywhere.

(Note that functions of finite variation have derivatives which are integrable with respect to Lebesgue measure, but the function does not have to be equal to the integral of the derivative.)

7. A Lipschitz function multiplied by a constant, and a sum of two Lipschitz functions are Lipschitz functions. The product of two bounded Lipschitz functions is again a Lipschitz function.
8. If f is Lipschitz on $[0, N]$ for any $N > 0$ but with the constant K depending on N , then it is called locally Lipschitz. For example, x^2 is Lipschitz on $[0, N]$ for any finite N , but it is not Lipschitz on $[0, +\infty)$, since its derivative is unbounded.
9. If f is a function of two variables $f(x, t)$ and it satisfies Lipschitz condition in x for all t , $0 \leq t \leq T$, with same constant K independent of t , it is said that f satisfies Lipschitz condition in x uniformly in t , $0 \leq t \leq T$.

A necessary and sufficient condition for a function f to be Riemann integrable was given by Lebesgue (see, for example, Saks (1964), Freedman (1983) p.208).

Theorem 1.22 (Lebesgue) *A necessary and sufficient condition for a function f to be Riemann integrable on a finite closed interval $[a, b]$ is that f is bounded on $[a, b]$ and almost everywhere continuous on $[a, b]$, that is, continuous at all points except possibly on a set of Lebesgue measure zero.*

Remark 1.6: (This is not used anywhere in the book, and directed only to readers with knowledge of Functional Analysis)

Continuous functions on $[a, b]$ with the supremum norm $\|h\| = \sup_{x \in [a, b]} |h(x)|$ is a Banach space, denoted $C([a, b])$. By a result in Functional Analysis, any linear functional on this space can be represented as $\int_{[a, b]} h(x) dg(x)$ for some function g of finite variation. In this way, the Banach space of functions of finite variation on $[a, b]$ with the norm $\|g\| = V_g([a, b])$ can be identified with the space of linear functionals on the space of continuous functions, in other words, the dual space of $C([a, b])$.