

Topics in analysis 1 — Real functions

Onno van Gaans

Version May 21, 2008

Contents

| | | |
|----------|--|-----------|
| 1 | Monotone functions | 2 |
| 1.1 | Continuity | 2 |
| 1.2 | Differentiability | 2 |
| 1.3 | Bounded variation | 2 |
| 2 | Fundamental theorem of calculus | 2 |
| 2.1 | Indefinite integrals | 2 |
| 2.2 | The Cantor function | 2 |
| 2.3 | Absolute continuity | 2 |
| 3 | Sobolev spaces | 2 |
| 3.1 | Sobolev spaces on intervals | 3 |
| 3.2 | Application to differential equations | 6 |
| 3.3 | Distributions | 10 |
| 3.4 | Fourier transform and fractional derivatives | 16 |
| 4 | Stieltjes integral | 20 |
| 4.1 | Definition and properties | 20 |
| 4.2 | Langevin's equation | 28 |
| 4.3 | Wiener integral | 32 |
| 5 | Convex functions | 33 |

1 Monotone functions

1.1 Continuity

1.2 Differentiability

1.3 Bounded variation

2 Fundamental theorem of calculus

2.1 Indefinite integrals

2.2 The Cantor function

2.3 Absolute continuity

3 Sobolev spaces

When studying differential equations it is often convenient to assume that all functions in the equation are several times differentiable and that their derivatives are continuous. If the equation is of second order, for instance, one usually assumes that the solution should be two times continuously differentiable. There are several reasons to allow functions with less smoothness. Let us consider two of them.

Suppose that

$$u''(t) + \alpha u'(t) + \beta u(t) = f(t)$$

describes the position $u(t)$ of a particle at time t , where f is some external force acting on the particle. If the force is continuous in t , one expects the solution u to be twice continuously differentiable. If the function f has jumps, the second derivative of u will not be continuous and, more than that, u' will probably not be differentiable at the points where f is discontinuous. Such external forces typically appear in systems such as electrical circuits, where external influences can be switched on or off. There is no harm in such discontinuities. The function u'' still exists almost everywhere and u' will still be the indefinite integral of u'' .

The second reason concerns mathematical rather than modeling issues. Differential equations can often not be solved explicitly. It can then still be useful to know existence of solutions. For instance, if one wants to apply numerical approximation schemes. A functional analytic approach to existence of solutions has become a popular method. Showing existence comes down to finding, among all differentiable functions (or m times differentiable functions) a function which satisfies the equation. The set of all differentiable functions is a *function space*.

The functional analytic approach uses structure of function spaces and maps on them to show existence of solutions. It is almost indispensable that the function space in which the solution is sought is a linear space with a norm and that it is *complete* with respect to the norm. For instance the linear space $C^1[0, 1]$ of all continuously differentiable functions on $[0, 1]$ with the norm

$$\|f\|_{C[0,1]} := |f(0)| + \max_{0 \leq t \leq 1} |f'(t)|$$

is a complete normed space, that is, a *Banach space*. It is often desirable that the space is a *Hilbert space*, that is, a Banach space where the norm is induced by an inner product. It turns out that the space of differentiable functions is not a Hilbert space. However, there are

Hilbert spaces consisting of almost everywhere differentiable functions, which turn out to be very useful for the study of differential equations.

3.1 Sobolev spaces on intervals

The space $C[a, b]$ of all continuous functions from $[a, b]$ to \mathbb{R} is a Banach space with the norm

$$\|f\|_\infty = \max_{a \leq t \leq b} |f(t)|.$$

This space is not a Hilbert space. Let $\mathcal{L}^2[a, b]$ denote the space of all Lebesgue measurable functions $f : [a, b] \rightarrow \mathbb{R}$ such that $\int_{[a, b]} f(t)^2 dt < \infty$. Let $L^2[a, b]$ denote the Banach space of equivalence classes of functions from $\mathcal{L}^2[a, b]$, where we identify two functions that are equal almost everywhere. The space $L^2[a, b]$ is a Hilbert space with the inner product given by

$$\langle f, g \rangle_{L^2} = \int_{[a, b]} f(t)g(t) dt$$

and its induced norm is

$$\|f\|_{L^2} = \left(\int_a^b f(t)^2 dt \right)^{1/2}.$$

A continuously differentiable function can be seen as the indefinite integral of a continuous function. In the hope to obtain a Hilbert space of ‘differentiable’ functions we could consider the functions that are indefinite integrals of functions from $L^2[a, b]$. Recall that every $g \in L^2[a, b]$ is integrable.

Define

$$W^{1,2}[a, b] := \left\{ \alpha + \int_0^t g(s) ds : \alpha \in \mathbb{R}, g \in L^2[a, b] \right\}.$$

First observe that there is no ambiguity in the function $t \mapsto \int_0^t g(s) ds$. If we would take another element \tilde{g} from the same equivalence class as g , then \tilde{g} and g are equal almost everywhere, so that their integrals over any set coincide. Observe further that the set $W^{1,2}[a, b]$ is a vector space with the pointwise addition and scalar multiplication. We have seen that the indefinite integral of an integrable function is absolutely continuous. Therefore the function $f(t) = \alpha + \int_0^t g(s) ds$, where $g \in L^2[a, b]$, is absolutely continuous. Moreover, f is differentiable almost everywhere and its derivative equals g almost everywhere. We will denote the (almost everywhere defined) derivative of an absolutely continuous function f by f' . The set $W^{1,2}$ equals

$$W^{1,2}[a, b] = \left\{ f : [a, b] \rightarrow \mathbb{R} : f \text{ is absolutely continuous and } \int_a^b f'(t)^2 dt < \infty \right\}.$$

Define an inner product on $W^{1,2}[a, b]$ by

$$\langle f, g \rangle_{W^{1,2}} = f(a)g(a) + \int_a^b f'(t)g'(t) dt, \quad f, g \in W^{1,2}.$$

The induced norm is given by

$$\begin{aligned} \|f\|_{1,2} &= \left(f(a)^2 + \int_a^b f'(t)^2 dt \right)^{1/2} \\ &= (f(a)^2 + \|f'\|_{L^2}^2)^{1/2}. \end{aligned}$$

The space $W^{1,2}[a, b]$ endowed with the inner product $\langle \cdot, \cdot \rangle_{W^{1,2}}$ is called a *Sobolev space*. The role of the upper indices 1, 2 becomes clear if we introduce some other Sobolev spaces:

$$W^{1,p}[a, b] = \left\{ \alpha + \int_0^t g(s) ds : \alpha \in \mathbb{R}, g \in L^p[a, b] \right\}$$

and, inductively,

$$W^{m,p}[a, b] = \left\{ \alpha + \int_0^t g(s) ds : \alpha \in \mathbb{R}, g \in W^{m-1,p}[a, b] \right\},$$

where $m \in \mathbb{N}$ and $p \geq 1$. We focus on $W^{1,2}$.

Lemma 3.1. *The Sobolev space $(W^{1,2}[a, b], \langle \cdot, \cdot \rangle_{W^{1,2}})$ is a Hilbert space.*

Proof. It is straightforward to show that $W^{1,2}$ is a vector space, that $\langle \cdot, \cdot \rangle_{W^{1,2}}$ is a symmetric bilinear form on $W^{1,2}$, and that $\langle f, f \rangle_{W^{1,2}} \geq 0$ for every f . If $f \in W^{1,2}$ is such that $\langle f, f \rangle_{W^{1,2}} = 0$, then $\int_a^b f'(t)^2 dt = 0$, so $f'(t) = 0$ for almost every t . Since f is absolutely continuous (as it is an indefinite integral), it follows that f is constant. Since we also have that $f(a) = 0$, the function f must be identically zero.

Next we show that $W^{1,2}[a, b]$ is complete. Let (f_n) be a Cauchy sequence in $W^{1,2}$. As $\|f'_n - f'_m\|_2^2 \leq \|f_n - f_m\|_{W^{1,2}}^2$, the sequence (f'_n) is a Cauchy sequence in $L^2[a, b]$. Because $L^2[a, b]$ is complete, there exists a $g \in L^2[a, b]$ such that $\|f'_n - g\|_{L^2} \rightarrow 0$ as $n \rightarrow \infty$. Also, $|f_n(a) - f_m(a)| \leq \|f_n - f_m\|_{W^{1,2}}$, so $(f_n(a))$ is a Cauchy sequence in \mathbb{R} . Let $\alpha := \lim_{n \rightarrow \infty} f_n(a)$. Define

$$f(t) := \alpha + \int_a^t g(s) ds, \quad t \in [a, b].$$

Then $f \in W^{1,2}$. It remains to show that f is the limit of f_n with convergence in the sense of $W^{1,2}$. Well,

$$\|f - f_n\|_{W^{1,2}}^2 = (\alpha - f_n(a))^2 + \|g - f'_n\|_2^2 \rightarrow 0 \text{ as } n \rightarrow \infty,$$

so f is indeed the limit in $W^{1,2}$ of (f_n) . Hence $W^{1,2}$ is complete. \square

There is a physical interpretation of the space $W^{1,2}$. If f describes the position of a particle, then $f'(t)$ is its velocity at time t . The condition $\int_a^b f'(t)^2 dt < \infty$ then means that the total kinetic energy of the motion from time a to b is finite. So $W^{1,2}$ is the space of ‘curves with finite kinetic energy’.

The essential part of the norm $\|\cdot\|_{W^{1,2}}$ is the term $\int_a^b f'(t)^2 dt$. Only this term would not define a norm as otherwise each constant function would get norm equal to zero. There are different ways to add a term to make it a norm. A common other choice is the following. Define

$$\langle\langle f, g \rangle\rangle = \int_a^b f(t)g(t) dt + \int_a^b f'(t)g'(t) dt, \quad f, g \in W^{1,2}[a, b].$$

It is easy to verify that $\langle\langle \cdot, \cdot \rangle\rangle$ is an inner product on $W^{1,2}$.

Claim 3.2. *The norm $\|\cdot\|$ induced by $\langle\langle \cdot, \cdot \rangle\rangle$ is equivalent to $\|\cdot\|_{W^{1,2}}$.*

Proof. For $x \in [a, b]$, by $(\alpha + \beta)^2 \leq 2\alpha^2 + 2\beta^2$ and Cauchy-Schwartz,

$$\begin{aligned}
f(x)^2 &= \left(f(a) + \int_a^x f'(t) \, ds \right)^2 \\
&\leq 2f(a)^2 + 2 \left(\int_a^x f'(t) 1 \, dt \right)^2 \\
&\leq 2f(a)^2 + 2 \left(\left(\int_a^x f'(t)^2 \, dt \right)^{1/2} \left(\int_a^x 1^2 \, dt \right)^{1/2} \right)^2 \\
&\leq 2f(a)^2 + 2(b-a) \int_a^b f'(t)^2 \, dt,
\end{aligned} \tag{1}$$

so

$$\begin{aligned}
\int_a^b f(x)^2 \, dx &\leq 2(b-a)f(a)^2 + 2(b-a)^2 \int_a^b f'(t)^2 \, dt \\
&\leq \max\{2(b-a), 2(b-a)^2\} \|f\|_{W^{1,2}}^2,
\end{aligned}$$

so

$$|||f||| \leq C_1^{1/2} \|f\|_{W^{1,2}},$$

where $C_1 = \max\{2(b-a), 2(b-a)^2\} + 1$.

For the inequality in the converse direction, for $x \in [a, b]$,

$$\begin{aligned}
f(a)^2 &\leq \left(f(x) - \int_a^x f'(t) \, dx \right)^2 \\
&\leq 2f(x)^2 + 2 \left(\int_a^x f'(t) 1 \, dt \right)^2 \\
&\leq 2f(x)^2 + 2 \left(\int_a^x f'(t)^2 \, dt \int_a^x 1^2 \, dt \right) \\
&\leq 2f(x)^2 + 2 \int_a^x f'(t)^2 \, dt (b-a),
\end{aligned}$$

so

$$\begin{aligned}
f(a)^2 &= \frac{1}{b-a} \int_a^b f(a)^2 \, dt \\
&\leq \frac{2}{b-a} \int_a^b f(x)^2 \, dx + 2 \int_a^b f'(t)^2 \, dt.
\end{aligned}$$

Hence

$$\|f\|_{W^{1,2}} \leq C_2^{1/2} |||f|||,$$

where $C_2 = \max\{\frac{2}{b-a}, 2\} + 1$. Thus the norms $||| \cdot |||$ and $\| \cdot \|_{W^{1,2}}$ are equivalent. \square

There are many other equivalent norms, e.g.,

$$\left(f(b)^2 + \int_a^b f'(t) \, dt \right)^{1/2},$$

$$\left(f(c)^2 + \int_a^b f'(t)^2 dt\right)^{1/2},$$

$$\left(5f(a)^2 + 7f(b)^2 + \pi \int_a^b f'(t)^2 dt\right)^{1/2},$$

etc.

We conclude this section by a useful lemma.

Lemma 3.3. *Let $c \in [a, b]$. Let*

$$\varphi_c(f) := f(c), \quad f \in W^{1,2}[a, b].$$

Then φ_c is a bounded linear functional on $W^{1,2}$.

Proof. Clearly, φ_c is linear. From (1) it follows that

$$|\varphi_c(f)|^2 \leq 2f(a)^2 + 2(b-a) \int_a^b f'(t)^2 dt \leq \max\{2, 2(b-a)\} \|f\|_{W^{1,2}}^2$$

for all $f \in W^{1,2}$, so φ_c is a bounded linear functional. □

From (1) we even obtain that

$$\max_{a \leq t \leq b} |f(t)| \leq C \|f\|_{W^{1,2}}$$

for some constant C . That means that the norm $\|\cdot\|_{W^{1,2}}$ is stronger than $\|\cdot\|_\infty$.

3.2 Application to differential equations

In order to explain the functional analytic approach to differential equations and show the use of Sobolev spaces, we consider two examples.

Consider

$$\begin{cases} u'' - u = f & \text{on } (0, 1), \\ u(0) = u(1) = 0, \end{cases} \quad (2)$$

where the function $f : [0, 1] \rightarrow \mathbb{R}$ is given and we want to find a function $u : [0, 1] \rightarrow \mathbb{R}$ satisfying (2). A *classical* solution of (2) is a twice continuously differentiable function u with (2). From the equation we see that for existence of such a solution it is necessary that f is continuous.

We will show existence of a solution (2) by means of the Sobolev space $W^{1,2}[0, 1]$ and the representation theorem of Riesz:

Theorem 3.4. *Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space and let $\varphi : H \rightarrow \mathbb{R}$ be a bounded linear function. Then there exists a unique $y \in H$ such that*

$$\varphi(u) = \langle u, y \rangle \quad \text{for all } u \in H.$$

First we reformulate (2). If u would satisfy (2), we could multiply both sides of the equation by a $v \in W^{1,2}[0, 1]$ and integrate over $[0, 1]$,

$$\int_0^1 u''(t)v(t) dt - \int_0^1 u(t)v(t) dt = \int_0^1 f(t)v(t) dt$$

and use integration by parts to obtain

$$u'(1)v(1) - u'(0)v(0) - \int_0^1 u'(t)v'(t) dt - \int_0^1 u(t)v(t) dt = \int_0^1 f(t)v(t) dt.$$

If $v(0) = v(1) = 0$, we get

$$\int_0^1 u'(t)v'(t) dt + \int_0^1 u(t)v(t) dt = - \int_0^1 f(t)v(t) dt.$$

Hence (2) implies

$$\int_0^1 u(t)v(t) dt + \int_0^1 u'(t)v'(t) dt = - \int_0^1 f(t)v(t) dt \text{ for all } v \in D, \quad (3)$$

where

$$D = \{v \in W^{1,2}[0, 1] : v(0) = v(1) = 0\}.$$

Equation (3) is called the *weak formulation* of (2). The function $v \in D$ is called a *test function*. The integral at the right hand side of (3) makes sense for any $f \in L^2[0, 1]$. The integrals at the left hand side exist for each $u \in W^{1,2}[0, 1]$. A function $u \in W^{1,2}[0, 1]$ with $u(0) = u(1) = 0$ which satisfies (3) is called a *weak solution* of (2).

Assume $f \in L^2[0, 1]$. We claim that

- (a) D is a closed subspace of $W^{1,2}[0, 1]$ hence a Hilbert space,
- (b) there is a unique $u \in D$ such that (3) holds,
- (c) if f is continuous, then u is twice continuously differentiable.

It follows that (2) has a unique solution if $f \in L^2[0, 1]$ and a classical solution if $f \in C[0, 1]$.

Proof of (a). By Lemma 3.3,

$$\varphi_0(v) := v(0) \quad \text{and} \quad \varphi_1(v) := v(1), \quad v \in W^{1,2}[0, 1],$$

define bounded linear functionals $\varphi_0, \varphi_1 : W^{1,2}[0, 1] \rightarrow \mathbb{R}$. Hence

$$\begin{aligned} D &= \{v \in W^{1,2}[0, 1] : \varphi_0(v) = 0, \varphi_1(v) = 0\} \\ &= \varphi_0^{-1}(\{0\}) \cap \varphi_1^{-1}(\{0\}) \end{aligned}$$

is a closed subspace of $W^{1,2}[0, 1]$.

Proof of (b). Consider the inner product

$$\langle\langle u, v \rangle\rangle = \int_0^1 u(t)v(t) dt + \int_0^1 u'(t)v'(t) dt$$

on $W^{1,2}[0,1]$. As shown in the previous section, its induced norm $||| \cdot |||$ is equivalent to $\| \cdot \|_{W^{1,2}}$, so $W^{1,2}[0,1]$ is also a Hilbert space when endowed with the inner product $\langle \langle \cdot, \cdot \rangle \rangle$. As D is a closed subspace, $(D, \langle \langle \cdot, \cdot \rangle \rangle)$ is a Hilbert space. Define

$$\varphi_f(v) := - \int_0^1 f(t)v(t) dt, \quad v \in D.$$

Then φ_f is linear and

$$\begin{aligned} |\varphi_f(v)| &\leq \left(\int_0^1 f(t)^2 dt \right)^{1/2} \left(\int_0^1 v(t)^2 dt \right)^{1/2} \\ &\leq \|f\|_{L^2} |||v||| \text{ for all } v \in D, \end{aligned}$$

so φ_f is a bounded linear functional on D . Equation (3) can be written as

$$\langle \langle u, v \rangle \rangle = \varphi_f(v) \text{ for all } v \in D.$$

Due to Riesz's representation theorem, there exists a unique $y \in D$ such that $\varphi_f(v) = \langle \langle v, y \rangle \rangle = \langle \langle y, v \rangle \rangle$ for all $v \in D$. This y is then the unique solution in D of (3). In other words, y is the unique weak solution of (2).

Proof of (c). Assume that f is continuous. From (3) and integration by parts we have for every $v \in D$ that

$$\begin{aligned} \int_0^1 u'(t)v'(t) dt &= - \int_0^1 (u(t) + f(t))v(t) dt \\ &= \int_0^1 \int_0^t (u(s) + f(s)) ds v'(t) dt, \end{aligned}$$

so

$$\int_0^1 \left(u'(t) - \int_0^t (u(s) + f(s)) ds \right) v'(t) dt = 0. \quad (4)$$

It follows from the next lemma that there exists a constant C such that

$$u'(t) - \int_0^t (u(s) + f(s)) ds = C \text{ for almost every } t. \quad (5)$$

Lemma 3.5. *If $w \in L^2[a, b]$ is such that*

$$\int_a^b w(t)g(t) dt = 0 \quad \text{for all } g \in L^2[a, b] \text{ with } \int_a^b g(t) dt = 0,$$

then there exists a constant C such that

$$w(t) = C \quad \text{for almost every } t \in [a, b].$$

Moreover,

$$C = \int_a^b w(t) dt.$$

Proof. Let $C := \int_a^b w(t) dt$ and $g(t) := w(t) - C$, $t \in [a, b]$. Then $g \in L^2[a, b]$ and $\int_a^b g(t) dt = 0$, so

$$\int_a^b w(t)g(t) dt = 0.$$

Hence

$$\begin{aligned} \int_a^b g(t)^2 dt &= \int_a^b (w(t) - C)g(t) dt \\ &= \int_a^b w(t)g(t) dt - C \int_a^b g(t) dt = 0. \end{aligned}$$

Hence $g = 0$ a.e. □

Let us see that (4) indeed yields (5). We have that

$$\int_0^1 \left(u'(t) - \int_0^t (u(s) + f(s)) ds \right) g(t) dt = 0 \quad (6)$$

holds for any g of the form v' with $v \in D$. If $g \in L^2[0, 1]$ is such that $\int_0^1 g(s) ds = 0$, then $v(t) := \int_0^t g(s) ds$ is such that $v \in D$ and $v' = g$. Hence (6) holds for all $g \in L^2[0, 1]$ with $\int_0^1 g(t) dt = 0$. Then the lemma yields existence of a constant C such that (5) holds. Consequently, since u is absolutely continuous,

$$u(t) = u(0) + \int_0^t u'(s) ds = Ct + \int_0^t \int_0^s (u(s) + f(s)) ds ds.$$

Because $u + f$ is continuous, we obtain that $u \in C^2[0, 1]$, that is, u is twice continuously differentiable. Moreover, differentiating twice yields that

$$u''(t) = u(t) + f(t), \quad t \in (0, 1).$$

Recall that $u \in D$, so that also the boundary conditions $u(0) = u(1) = 0$ are satisfied.

The above procedure is typical for the functional analytic approach to differential equations. First find a weak formulation of the equation by multiplying by a test function, integrating and using integration by parts. Second, apply abstract functional analytic methods (e.g., Riesz representation theorem) to obtain existence of a weak solution. Third, consider regularity of the solution.

More generally, we consider the *Sturm-Liouville problem*

$$\begin{cases} -(pu')' + qu = f & \text{on } (0, 1) \\ u(0) = u(1) = 0, \end{cases} \quad (7)$$

where the coefficients $p \in C^1[0, 1]$ and $q \in C[0, 1]$ are given with $p(x) > 0$ and $q(x) \geq 0$ for all $x \in [0, 1]$. Let $f \in L^2[0, 1]$. The corresponding weak formulation is

$$\int_0^1 (p(x)u'(x)v'(x) + q(x)u(x)v(x)) dx = \int_0^1 f(x)v(x) dx \text{ for all } v \in D. \quad (8)$$

Define

$$\langle u, v \rangle_{SL} := \int_0^1 p(x)u'(x)v'(x) dx + \int_0^1 q(x)u(x)v(x), \quad u, v \in D.$$

We claim

- (a) $\langle \cdot, \cdot \rangle_{SL}$ is an inner product on D and its induced norm $\| \cdot \|_{SL}$ is equivalent to $\| \cdot \|_{W^{1,2}}$ on D ,
- (b) there exists a unique $u \in D$ which satisfies (8),
- (c) this function u is in $C^2[0, 1]$ if $f \in C[0, 1]$.

Proof of (a). It is straightforward that $\langle \cdot, \cdot \rangle_{SL}$ is bilinear and symmetric. For $u \in D$,

$$\|u\|_{SL}^2 \geq \int_0^1 p(x) u'(x)^2 dx \geq \min_{x \in [0,1]} p(x) \int_0^1 u'(x)^2 dx = \min_{x \in [0,1]} p(x) \|u\|_{W^{1,2}}^2$$

and

$$\begin{aligned} \|u\|_{SL}^2 &\leq \max_{x \in [0,1]} p(x) \int_0^1 u'(x)^2 dx + \max_{x \in [0,1]} |q(x)| \int_0^1 u^2(x) dx \\ &\leq \max\{\max_x p(x), \max_x |q(x)|\} \|u\|^2 \leq C \|u\|_{W^{1,2}}^2. \end{aligned}$$

It follows that $\langle \cdot, \cdot \rangle_{SL}$ is an inner product and that its induced norm is equivalent to $\| \cdot \|_{W^{1,2}}$.

Proof of (b). Due to (a), $(D, \langle \cdot, \cdot \rangle_{SL})$ is a Hilbert space. The functional $\varphi_f(v) := \int_0^1 v(x) f(x) dx$, $v \in D$, is bounded linear, so by Riesz's representation theorem, there is a unique $u \in D$ with

$$\varphi_f(v) = \langle u, v \rangle_{SL} \text{ for all } v \in D.$$

That is, u satisfies (8).

Proof of (c). The weak solution u satisfies

$$\int_0^1 p(x) u'(x) v'(x) dx - \int_0^1 \left(\int_0^x (q(s) u(s) - f(s)) ds \right) v'(x) dx = 0 \text{ for all } v \in D,$$

from which we infer by Lemma 3.5 that there exists a constant C such that

$$p(x) u'(x) - \int_0^x (q(s) u(s) - f(s)) ds = C \text{ for a.e. } x \in [0, 1].$$

Hence

$$u'(x) = \frac{1}{p(x)} \left(C + \int_0^x (q(s) u(s) - f(s)) ds \right),$$

so, because u is absolutely continuous,

$$u(x) = \int_0^x \frac{1}{p(r)} \left(C + \int_0^s (q(r) u(r) - f(r)) dr \right) ds.$$

Hence, if f is continuous, u is C^2 , since q is continuous and p is in $C^1[0, 1]$.

3.3 Distributions

In order to be able to differentiate functions that are not everywhere differentiable, we have considered almost everywhere derivatives. There is an even weaker notion of derivative, which makes even more functions 'differentiable'. The price to pay is that derivatives will no longer be functions. The idea is that instead of describing a function f by all its values $f(x)$, it

can also be described by all values $\int f(x)g(x)dx$, where g runs through a suitable set of test functions.

The *support* of a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is the closed set $\text{supp } f := \overline{\{x \in \mathbb{R} : f(x) \neq 0\}}$. Consider the space of *test functions*

$$\mathcal{D}(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} : f \text{ is infinitely many times differentiable and } \text{supp } f \text{ is compact}\}.$$

On $\mathcal{D}(\mathbb{R})$ we consider a notion of convergence given by

$$g_k \rightarrow g \text{ in } \mathcal{D} \iff \begin{array}{l} \exists N \text{ such that } \text{supp } g_k \subseteq [-N, N] \text{ for all } k \text{ and} \\ g_k^{(m)} \rightarrow g^{(m)} \text{ uniformly as } k \rightarrow \infty \text{ for all } m \geq 0. \end{array} \quad (9)$$

Here we use the convention that $g^{(0)} = g$. This notion of convergence is given by a topology on \mathcal{D} , which is more complicated to describe and of little interest. It can be defined as follows. For $N \in \mathbb{N}$, define

$$\begin{aligned} \mathcal{D}^N &= \{g \in \mathcal{D} : \text{supp } g \subseteq [-N, N]\}, \\ \mathcal{B}^N(f, m, r) &= \{g \in \mathcal{D}^N : \sup_{x \in \mathbb{R}} |g^{(m)}(x) - f^{(m)}(x)| < r\}, \quad f \in \mathcal{D}^N, \quad m \geq 0, \quad r > 0 \\ \mathcal{B}^N &= \{\mathcal{B}^N(f, m, r) : f \in \mathcal{D}^N, \quad m \geq 0, \quad r > 0\}, \\ \mathcal{B}_\cap^N &= \{B_1 \cap \dots \cap B_n : B_i \in \mathcal{B}^N, \quad n \in \mathbb{N}\}, \\ \mathcal{U}^N &= \{\bigcup_{i \in I} U_i : U_i \in \mathcal{B}_\cap^N, \quad I \text{ any set}\}, \quad \text{and} \\ \mathcal{U} &= \{U \subseteq \mathcal{D} : U \cap \mathcal{D}^N \in \mathcal{U}^N \text{ for all } N\}. \end{aligned}$$

Proposition 3.6. (i) \mathcal{U} is a topology on \mathcal{D} .

(ii)

$$g_k \rightarrow g \text{ in } \mathcal{D} \text{ w.r.t. } \mathcal{U} \iff \begin{array}{l} \exists N \text{ such that } \text{supp } g_k \subseteq [-N, N] \text{ for all } k \text{ and} \\ g_k^{(m)} \rightarrow g^{(m)} \text{ uniformly as } k \rightarrow \infty \text{ for all } m \geq 0. \end{array}$$

Proof. (i) It is straightforward to check that $\emptyset, \mathcal{D}^N \in \mathcal{U}^N$, that \mathcal{U}^N is closed under finite intersections and closed under arbitrary unions. It then follows that \mathcal{U} is a topology in \mathcal{D} . (ii) \Rightarrow $g_k \rightarrow g$ w.r.t. \mathcal{U} means that for each $U \in \mathcal{U}$ with $g \in U$ there exists a k_0 such that $g_k \in U$ for all $k \geq k_0$. Take N_0 such that $\text{supp } g \subseteq [-N_0, N_0]$. As $\mathcal{D}^{N_0} \in \mathcal{U}$, there exists a k_0 such that $g_k \in \mathcal{D}^{N_0}$ for all $k \geq k_0$. Choose N_1 such that $\text{supp } g_k \subseteq [-N_1, N_1]$ for $k = 1, \dots, N_0$. Let $N := \max\{N_0, N_1\}$. For every $m \geq 0$ and $\varepsilon > 0$, the set $\mathcal{B}^N(g, m, \varepsilon) \in \mathcal{U}$, so there exists a k_1 such that for all $k \geq k_1$ one has $g_k \in \mathcal{B}^N(g, m, \varepsilon)$, which means that $\sup_{x \in \mathbb{R}} |g_k(x)^{(m)} - g(x)^{(m)}| < \varepsilon$. Hence $g_k^{(m)} \rightarrow g^{(m)}$ uniformly.

\Leftarrow Let $U \in \mathcal{U}$ with $g \in U$. Take N such that $\text{supp } g_k \subseteq [-N, N]$ for all k . Then $U \cap \mathcal{D}^N \in \mathcal{U}^N$, so there are $B_1, \dots, B_n \in \mathcal{B}^N$ with $U \cap \mathcal{D}^N \supseteq B_1 \cap \dots \cap B_n$ and $g \in B_1 \cap \dots \cap B_n$. For each j , the uniform convergence and the triangle inequality yield that $g_k \in B_j$ for k sufficiently large. Therefore $g_k \in U$ for k large. Hence $g_k \rightarrow g$ w.r.t. \mathcal{U} . \square

On \mathcal{D} we will consider the convergence given by (9). Denote

$$\mathcal{D}' = \{\varphi : \mathcal{D} \rightarrow \mathbb{R} : \varphi \text{ is linear and continuous}\}.$$

The elements of \mathcal{D}' are called *distributions*.

Example 3.7. (a) A function $f: \mathbb{R} \rightarrow \mathbb{R}$ is *locally integrable* if it is measurable and $\int_a^b |f(x)| dx < \infty$ for every $a < b$. In other words, f is locally integrable if it is integrable on $[-R, R]$ for every $R > 0$. If f is locally integrable, then

$$\varphi_f(g) = \int_{\mathbb{R}} f(x)g(x) dx, \quad g \in \mathcal{D},$$

defines a distribution $\varphi_f \in \mathcal{D}'$. Indeed, the linearity is clear. If $g_k \rightarrow g$ in \mathcal{D} , then there exists $N \in \mathbb{N}$ such that $\text{supp } g_k \subset [-N, N]$ and $\text{supp } g \subseteq [-N, N]$ and $g_k \rightarrow g$ uniformly. Hence

$$|\varphi_f(g_k) - \varphi_f(g)| \leq \int_{\mathbb{R}} |f(x)(g_k(x) - g(x))| dx \leq \int_{-N}^N |f(x)| dx \sup_x |g_k(x) - g(x)| \rightarrow 0,$$

as $k \rightarrow \infty$.

(b) The Dirac distribution δ_a at $a \in \mathbb{R}$ is defined by

$$\delta_a(g) = g(a), \quad g \in \mathcal{D}.$$

(c) If μ is any finite measure on the Borel σ -algebra of \mathbb{R} ,

$$\varphi_\mu(g) = \int_{\mathbb{R}} g(x) d\mu(x), \quad g \in \mathcal{D},$$

defines $\varphi_\mu \in \mathcal{D}'$.

Lemma 3.8. If f_1 and f_2 are locally integrable and $\varphi_{f_1}(g) = \varphi_{f_2}(g)$ for all $g \in \mathcal{D}$, then $f_1 = f_2$ a.e.

Proof. (Sketch) Denote $f = f_1 - f_2$. Then $\varphi_f(g) = 0$ for all $g \in \mathcal{D}$. Let

$$h(x) = c \begin{cases} e^{\frac{1}{x^2-1}}, & -1 < x < 1, \\ 0, & |x| \geq 1, \end{cases}$$

where c is chosen such that $\int_{\mathbb{R}} h(x) dx = 1$. Then $h \in \mathcal{D}$. Let

$$h_m(x) = mh(mx), \quad x \in \mathbb{R}.$$

Let $c < d$. Let

$$g_m(x) = \int_{[c-1/m, d+1/m]} h_m(x-y) dy.$$

Then $g_m \in \mathcal{D}$ for all m and $g_m(x) \rightarrow 1_{[c,d]}(x)$ for all $x \in \mathbb{R}$. By Lebesgue's dominated convergence theorem,

$$\int_{[c,d]} f(x) dx = \lim_{m \rightarrow \infty} \int_{\mathbb{R}} f(x)g_m(x) dx = 0.$$

It follows that $f = 0$ a.e. □

It follows from the lemma that there is an injection $f \mapsto \varphi_f$ from the locally integrable functions into \mathcal{D}' , if we identify functions that are equal almost everywhere. In other words, every locally integrable function f corresponds to a distribution φ_f . Instead of considering the function f we can as well consider the distribution φ_f .

Distributions on intervals. More generally we can consider distributions on any open interval $I \subseteq \mathbb{R}$. Let

$$\mathcal{D}(I) = \{g : I \rightarrow \mathbb{R} : g \text{ is infinitely many times differentiable and } \text{supp } g \subset I\}.$$

Observe that $\text{supp } g \subset I$ means that there is a closed interval $[c, d] \subset I$ such that $g = 0$ outside $[c, d]$. In particular g vanishes near the boundary of I . On $\mathcal{D}(I)$ we introduce the following convergence.

$$g_k \rightarrow g \text{ in } \mathcal{D}(I) \iff \begin{array}{l} \exists \text{ closed interval } [c, d] \text{ such that } \text{supp } g_k \subseteq [c, d] \text{ for all } k \text{ and} \\ g_k^{(m)} \rightarrow g^{(m)} \text{ uniformly as } k \rightarrow \infty \text{ for all } m \geq 0. \end{array}$$

Define the distributions on I by

$$\mathcal{D}'(I) = \{\varphi : \mathcal{D} \rightarrow \mathbb{R} : \varphi \text{ is linear and continuous}\}.$$

As before, we can view locally integrable functions on I as distributions, by

$$\varphi_f(g) = \int_I f(x)g(x) \, dx.$$

If φ is a distribution, we will say that $\varphi \in C(I)$ (or $L^p(I)$ or ...) if there exists a function $f \in C(I)$ (or $L^p(I)$ or ...) such that $\varphi = \varphi_f$. **Derivatives of distributions.** If f is C^1 and

$g \in \mathcal{D}$, then by integration by parts,

$$\int_I f'(x)g(x) \, dx = - \int_I f(x)g'(x) \, dx.$$

That is,

$$\varphi_{f'}(g) = -\varphi_f(g') \quad \text{for all } g \in \mathcal{D}.$$

Definition 3.9. The *derivative* of a distribution $\varphi \in \mathcal{D}'$ is the distribution $D\varphi \in \mathcal{D}'$ defined by

$$D\varphi(g) = -\varphi(g'), \quad g \in \mathcal{D}.$$

It is not difficult to prove that $D\varphi$ is indeed an element of \mathcal{D}' . It follows that every locally integrable function and even every distribution has a derivative.

Lemma 3.10. *If $f : I \rightarrow \mathbb{R}$ is locally integrable, then*

$$f \text{ is absolutely continuous} \iff D\varphi_f \text{ is integrable.}$$

Proof. \Rightarrow) Let $g \in \mathcal{D}$. We first show that fg is absolutely continuous. If (a_i, b_i) , $i = 1, \dots, n$, are mutually disjoint subintervals of I , then

$$\begin{aligned} & \sum_{i=1}^n |(fg)(b_i) - (fg)(a_i)| \\ & \leq \sum_{i=1}^n (|f(b_i)g(b_i) - f(b_i)g(a_i)| + |f(b_i)g(a_i) - f(a_i)g(a_i)|) \\ & \leq \max_{x \in [a, b]} |f(x)| \sum_{i=1}^n |g(b_i) - g(a_i)| + \max_{x \in [a, b]} |g(x)| \sum_{i=1}^n |f(b_i) - f(a_i)|. \end{aligned}$$

Moreover, g is C^1 hence absolutely continuous. Therefore, it follows that fg is absolutely continuous. (In fact we have just shown that the product of two absolutely continuous functions is absolutely continuous.) Then, for $a \in I$,

$$\int_a^x (fg)'(t) dt = (fg)(x) - (fg)(a) \text{ for all } x \in I, x \geq a.$$

Since $\text{supp } g \subset I$ and I is open, g vanishes near the boundaries of I . Hence we obtain

$$\int_I (fg)'(t) dt = 0.$$

That is,

$$\int_I f'(t)g(t) dt + \int_I f(t)g'(t) dt = 0.$$

Hence $D\varphi_f = \varphi_{f'}$. As f' is integrable, $D\varphi_f$ is integrable.

\Leftarrow) By definition, there exists an integrable $h : I \rightarrow \mathbb{R}$ such that $D\varphi_f = \varphi_h$. That is,

$$-\int_I f(t)g'(t) dt = \int_I h(t)g(t) dt.$$

Let a be the left end point of I (possibly $-\infty$) and let

$$H(x) = \int_a^x h(t) dt.$$

Then H is absolutely continuous, so Hg is absolutely continuous, so

$$\begin{aligned} H(x)g(x) &= \int_a^x (Hg)'(t) dt \\ &= \int_a^x H'(t)g(t) dt + H(t)g'(t) dt. \end{aligned}$$

Hence (with $x = b$)

$$0 = \int_I h(t)g(t) dt + \int_I H(t)g'(t) dt,$$

so

$$\int_I f(t)g'(t) dt = - \int_I h(t)g(t) dt = \int_I H(t)g'(t) dt,$$

so

$$\int_I (f(t) - H(t))g'(t) dt = 0 \text{ for all } g \in \mathcal{D}.$$

From this we want to conclude that there exists a constant C such that $f(t) - H(t) = C$ for almost every $t \in I$. Since $H + C$ is absolutely continuous, it then follows that f is absolutely continuous. If $w \in \mathcal{D}$ is such that $\int_I w(s) ds = 0$, then $g(x) := \int_a^x w(s) ds$ defines an element g in \mathcal{D} with $g' = w$. So $\int_I (f(t) - H(t))w(t) dt = 0$. By a suitable approximation argument it can then be shown that $\int_I (f(t) - H(t))w(t) dt = 0$ holds for any $w \in L^2(I)$ with $\int_I w(s) ds = 0$, so that there exists a constant C with $f(t) - H(t) = C$ for almost every t . \square

Corollary 3.11.

$$W^{1,2}[a, b] = \{f \in L^2[a, b]: D\varphi_f \in L^2[a, b]\}.$$

Similarly,

$$\begin{aligned} W^{m,p}[a, b] &= \{f \in L^p[a, b]: D\varphi_f \in W^{m-1,p}[a, b]\} \\ &= \{f \in L^p[a, b]: D\varphi_f, \dots, D^m\varphi_f \in L^p[a, b]\}. \end{aligned}$$

Multidimensional distributions. Let $d \in \mathbb{N}$ and let Ω be an open subset of \mathbb{R}^d . Let

$$\begin{aligned} \mathcal{D}(\Omega) &:= \{g: \Omega \rightarrow \mathbb{R}: \text{all partial derivatives of } g \text{ of all orders exist and are continuous and} \\ &\quad \exists \text{ compact } K \subseteq \Omega \text{ such that } \text{supp } g \subseteq K\}. \end{aligned}$$

Introduce a notion of convergence in \mathcal{D} by

$$g_n \rightarrow g \iff \begin{aligned} &\exists \text{ compact } K \subseteq \Omega \text{ such that } \text{supp } g_n \subseteq K \text{ for all } n \text{ and for every} \\ &m_1, \dots, m_d \geq 0, \frac{\partial^{m_1+\dots+m_d} g_n}{\partial^{m_1} x_1 \dots \partial^{m_d} x_d} \rightarrow \frac{\partial^{m_1+\dots+m_d} g}{\partial^{m_1} x_1 \dots \partial^{m_d} x_d} \text{ uniformly} \end{aligned}$$

The set of *distributions* is given by

$$\mathcal{D}'(\Omega) := \{\varphi: \mathcal{D} \rightarrow \mathbb{R}: \varphi \text{ is linear and continuous}\}.$$

A function $f: \Omega \rightarrow \mathbb{R}$ is *locally integrable* if

$$\int_B |f(x_1, \dots, x_d)| dx_1 \dots dx_d < \infty$$

for every bounded $B \subseteq \Omega$. The functional

$$\varphi_f(g) = \int_{\Omega} f(x_1, \dots, x_d)g(x_1, \dots, x_d) dx_1 \dots dx_d, \quad g \in \mathcal{D},$$

is then a distribution.

For $m_1, \dots, m_d \geq 0$, define the *derivative* of $\varphi \in \mathcal{D}'$ of order (m_1, \dots, m_d) by

$$D^{m_1, \dots, m_d} \varphi(g) := (-1)^{m_1+\dots+m_d} \varphi \left(\frac{\partial^{m_1+\dots+m_d}}{\partial^{m_1} x_1 \dots \partial^{m_d} x_d} g \right), \quad g \in \mathcal{D}.$$

Then $D^{m_1, \dots, m_d} \varphi \in \mathcal{D}'$. By means of these multidimensional distributional derivatives we can define Sobolev spaces on Ω as follows:

$$W^{m,p}(\Omega) := \{f \in L^p(\Omega) : D^{m_1, \dots, m_d} \varphi_f \in L^p(\Omega) \text{ for all } m_1, \dots, m_d \geq 0 \text{ with } m_1 + \dots + m_d \leq m\}.$$

If $d = 1$, we have seen that every $f \in W^{1,p}$ is continuous (even absolutely continuous). In higher dimensions elements of $W^{1,p}$ need not be continuous. The following theorem is the famous *Sobolev embedding theorem*. Its proof is beyond the scope of this course.

Theorem 3.12. *If $d \in \mathbb{N}$, $m, k \in \{0, 1, 2, \dots\}$, $\Omega \subseteq \mathbb{R}^d$ is open, and $p > 1$, then*

$$W^{m,p}(\Omega) \subseteq C^k(\Omega) \text{ if } m > k + \frac{d}{p}.$$

If we denote by $H_0^{m,p}(\Omega)$ the closure of $\mathcal{D}(\Omega)$ in $W^{m,p}(\Omega)$, then

- $H_0^{m,p}(\Omega)$ is compactly embedded in $L^q(\Omega)$ if $mp < d$ and $q < \frac{dq}{d-mp}$
- $H_0^{m,p}(\Omega)$ is compactly embedded in $C^k(\overline{\Omega})$ if $mp > d + k$.

Here we say that a normed space X is compactly embedded in a normed space Y if X is a subspace of Y and all elements from the unit ball of X are a relatively compact set in Y .

A standard reference on Sobolev spaces is the book “Sobolev spaces” by R.A. Adams, Academic Press, 1975.

3.4 Fourier transform and fractional derivatives

If one analyses linear operations in \mathbb{R}^d , it is sometimes convenient to choose a new basis. The transformation of coordinates with respect to the old basis to the coordinates for the new basis is a linear bijection. Differentiation is also a linear transformation. It is sometimes convenient to use a transformation of a function space that makes the operation of differentiation simpler. Such a transformation is the Fourier transform. A complete theory of the Fourier transform is quite involved. This section contains a glossary of the main features relevant for Sobolev spaces.

For an integrable function $f: \mathbb{R} \rightarrow \mathbb{R}$ we define the *Fourier transform* of f by

$$(\mathcal{F})(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) e^{-i\omega t} dt, \quad \omega \in \mathbb{R}.$$

Not all mathematicians agree on the choice of the scaling factor $\frac{1}{\sqrt{2\pi}}$. Other common choices are 1 and $\frac{1}{2\pi}$.

The function $\mathcal{F}f$ is often also denoted by \hat{f} . It can be shown that \hat{f} is a continuous function and

$$\hat{f}(\omega) \rightarrow 0 \text{ as } \omega \rightarrow \infty \text{ or } \omega \rightarrow -\infty.$$

If we apply \mathcal{F} to a test function $f \in \mathcal{D}$, the image $\mathcal{F}f$ need not be in \mathcal{D} . If we enlarge the class of test functions somewhat, we get a class that is more compatible with the action of \mathcal{F} . A function $f: \mathbb{R} \rightarrow \mathbb{R}$ is called *rapidly decreasing* if

$$|t|^n f(t) \rightarrow 0 \text{ as } |t| \rightarrow \infty \text{ for every } n \in \mathbb{N},$$

that is, f decays faster than any power of t . For instance, $f(t) = e^{-|t|}$ is rapidly decreasing and every function with a bounded support is rapidly decreasing. The set

$$\mathcal{S}(\mathbb{R}) = \{f \in C^\infty(\mathbb{R}) : f^{(m)} \text{ is rapidly decreasing for all } m \in \mathbb{N}\}$$

is called the *Schwartz space*. Observe that $\mathcal{D}(\mathbb{R}) \subset \mathcal{S}(\mathbb{R})$. The relevance of the Schwartz space for the Fourier transform is shown by the next theorem, which we cite without proof.

Theorem 3.13. *The Fourier transform \mathcal{F} is a bijection from $\mathcal{S}(\mathbb{R})$ onto $\mathcal{S}(\mathbb{R})$.*

Proposition 3.14. *For $f, g \in \mathcal{S}(\mathbb{R})$ we have*

- (a) $\int_{\mathbb{R}} \hat{f}(\omega)g(\omega) d\omega = \int_{\mathbb{R}} f(t)\hat{g}(t) dt$;
- (b) $\mathcal{F}(t \mapsto e^{-\frac{1}{2}t^2}) = \omega \mapsto e^{-\frac{1}{2}\omega^2}$ and, more generally, $\mathcal{F}(t \mapsto e^{-\frac{1}{2}(at)^2}) = \omega \mapsto \frac{1}{a}e^{-\frac{1}{2}(\frac{\omega}{a})^2}$ for every $a \neq 0$;
- (c) $\mathcal{F}(\mathcal{F}f)(\xi) = f(-\xi)$ for all $\xi \in \mathbb{R}$;
- (d) $\int_{\mathbb{R}} \hat{f} \overline{\hat{g}(\omega)} d\omega = \int_{\mathbb{R}} f(t)g(t) dt$.

Proof. (a) By Fubini,

$$\begin{aligned} \int_{\mathbb{R}} \hat{f}(\omega)g(\omega) d\omega &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t)e^{-i\omega t} dt g(\omega) d\omega \\ &= \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(\omega)e^{-i\omega t} d\omega dt \\ &= \int_{\mathbb{R}} f(t)\hat{g}(t) dt. \end{aligned}$$

(b) We have

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(at)^2} e^{-i\omega t} dt &= \frac{1}{\sqrt{2\pi}a} \int_{\mathbb{R}} e^{-\frac{1}{2}s^2} e^{-i\frac{\omega}{a}s} ds \\ &= \frac{1}{\sqrt{2\pi}a} \int_{\mathbb{R}} e^{-\frac{1}{2}(s - \frac{i\omega}{a})^2} ds e^{-\frac{1}{2}(\frac{\omega}{a})^2} \\ &= \frac{1}{a} e^{-\frac{1}{2}(\frac{\omega}{a})^2}. \end{aligned}$$

(c) By (a) and (b),

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\mathcal{F}f)(\omega) e^{-i\omega\xi} e^{-\frac{1}{2}(a\omega)^2} d\omega &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) \mathcal{F}(e^{-i\omega\xi} e^{-\frac{1}{2}(a\omega)^2}) d\omega \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-i\omega\xi} e^{-\frac{1}{2}(a\omega)^2} e^{-i\omega t} d\omega dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) \frac{1}{a} e^{-\frac{1}{2}(\frac{\xi+t}{a})^2} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(as - \xi) e^{-\frac{1}{2}s^2} ds. \end{aligned}$$

If we let $a \rightarrow 0$, Lebesgue's dominated convergence theorem applied to both the left and right hand side yields

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\mathcal{F}f)(\omega) e^{-i\omega\xi} d\omega = f(-\xi),$$

so $\mathcal{F}\mathcal{F}f(\xi) = f(-\xi)$.

(d) Since

$$\begin{aligned} \overline{\hat{g}(\omega)} &= \frac{1}{\sqrt{2\pi}} \overline{\int_{\mathbb{R}} g(t) e^{-i\omega t} dt} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(t) e^{i\omega t} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(-s) e^{-i\omega t} ds = \mathcal{F}(s \mapsto g(-s)), \end{aligned}$$

(a) and (c) yield with $h(t) = g(-t)$ that

$$\begin{aligned} \int_{\mathbb{R}} \hat{f}(\omega) \overline{\hat{g}(\omega)} d\omega &= \int_{\mathbb{R}} \hat{f}(\omega) (\mathcal{F}h)(\omega) d\omega \\ &= \int_{\mathbb{R}} f(t) (\mathcal{F}\mathcal{F}h)(t) dt = \int_{\mathbb{R}} f(t) h(-t) dt \\ &= \int_{\mathbb{R}} f(t) g(t) dt. \end{aligned}$$

□

From part (d) we have

$$\langle \mathcal{F}f, \mathcal{F}g \rangle_{L^2(\mathbb{R})} = \langle f, g \rangle_{L^2(\mathbb{R})}$$

and

$$\|\mathcal{F}f\|_{L^2(\mathbb{R})} = \|f\|_{L^2(\mathbb{R})}.$$

So \mathcal{F} is an isometry with respect to the L^2 -norm on $\mathcal{S}(\mathbb{R})$. It can be shown that $\mathcal{S}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$ and then it follows that \mathcal{F} extends uniquely to an isometry \mathcal{F}_2 from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$. This extension \mathcal{F}_2 is called the *Fourier-Plancherel transformation*. It satisfies

$$\langle \mathcal{F}_2 f, \mathcal{F}_2 g \rangle_{L^2(\mathbb{R})} = \langle f, g \rangle_{L^2(\mathbb{R})} \text{ for all } f, g \in L^2(\mathbb{R}).$$

Moreover, \mathcal{F} is a bijection from $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$.

Fourier transform and derivative. The importance of the Fourier transform in the context of differentiation is that it transforms differentiation to multiplication by $i\omega$, which is an easier operation.

Proposition 3.15. *If $f \in \mathcal{S}(\mathbb{R})$, then*

$$\widehat{f'}(\omega) = i\omega \hat{f}(\omega).$$

Proof.

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(t) e^{-i\omega t} dt &= \lim_{M, N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-M}^N f'(t) e^{-i\omega t} dt \\ &= \lim_{M, N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \left(f(t) e^{-i\omega t} \Big|_{-M}^N - \int_{-M}^N f(t) (-i\omega) e^{-i\omega t} dt \right) \\ &= i\omega \hat{f}(\omega). \end{aligned}$$

□

It can be seen from the Fourier transform of a function whether its distributional derivative is actually a function in $L^2(\mathbb{R})$.

Proposition 3.16. *For $f \in L^2(\mathbb{R})$ one has*

$$f \text{ has distributional derivative in } L^2(\mathbb{R}) \iff \omega \mapsto \omega \hat{f}(\omega) \in L^2(\mathbb{R}).$$

Moreover, in that case the distributional derivative of f equals

$$\mathcal{F}_2^{-1}(\omega \mapsto i\omega \hat{f}(\omega)).$$

Proof. \Leftarrow) As \mathcal{F} is a bijection from $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$, we can define $h = \mathcal{F}_2^{-1}(\omega \mapsto i\omega \hat{f}(\omega)) \in L^2(\mathbb{R})$. For $g \in \mathcal{D}(\mathbb{R})$,

$$\begin{aligned} \int_{\mathbb{R}} h(t)g(t) dt &= \int_{\mathbb{R}} \hat{h}(\omega) \overline{\hat{g}(\omega)} d\omega = \int_{\mathbb{R}} i\omega \hat{f}(\omega) \overline{\hat{g}(\omega)} d\omega \\ &= - \int_{\mathbb{R}} \hat{f}(\omega) \overline{(i\omega) \hat{g}(\omega)} d\omega = - \int_{\mathbb{R}} \hat{f}(\omega) \widehat{g'}(\omega) d\omega = - \int_{\mathbb{R}} f(t)g'(t) dt, \end{aligned}$$

so h is the distributional derivative of f . \Rightarrow) Let $h \in L^2(\mathbb{R})$ be the distributional derivative of f . Then

$$\int_{\mathbb{R}} h(t)g(t) dt = - \int_{\mathbb{R}} f(t)g'(t) dt \text{ for all } g \in \mathcal{D}(\mathbb{R}).$$

Therefore, for every $g \in \mathcal{D}(\mathbb{R})$,

$$\begin{aligned} \int_{\mathbb{R}} \hat{h}(\omega) \overline{\hat{g}(\omega)} d\omega &= \int_{\mathbb{R}} h(t)g(t) dt = - \int_{\mathbb{R}} f(t)g'(t) dt \\ &= - \int_{\mathbb{R}} \hat{f}(\omega) \overline{\widehat{g'}(\omega)} d\omega = - \int_{\mathbb{R}} \hat{f}(\omega) \overline{(i\omega) \hat{g}(\omega)} d\omega \\ &= \int_{\mathbb{R}} i\omega \hat{f}(\omega) \overline{\hat{g}(\omega)} d\omega, \end{aligned}$$

so

$$\int_{\mathbb{R}} (\hat{h}(\omega) - i\omega \hat{f}(\omega)) \overline{\hat{g}(\omega)} d\omega = 0.$$

By a quite involved approximation argument it can be shown that this implies that

$$\hat{h}(\omega) - i\omega \hat{f}(\omega) = 0 \text{ for almost every } \omega.$$

Hence $\omega \mapsto \omega \hat{f}(\omega) = \hat{h}(\omega) \in L^2(\mathbb{R})$. □

We can view a function $f \in L^2[a, b]$ as a function in $L^2(\mathbb{R})$ by extending with zero:

$$\underline{f}(t) = \begin{cases} f(t), & t \in [a, b], \\ 0, & t \in \mathbb{R} \setminus \{0\}. \end{cases}$$

Notice that $\underline{f} \in L^2(\mathbb{R}) \cap L^2(\mathbb{R})$.

The previous proposition yields yet another way to describe the Sobolev space $W^{1,2}[a, b]$.

Corollary 3.17.

$$W^{1,2}[a, b] = \{f \in L^2[a, b] : \omega \mapsto \omega \underline{f}(\omega) \in L^2(\mathbb{R})\}.$$

Fractional derivative. Does there exist an operator $D^{1/2}$ such that applying it twice to a function f yields the derivative f' ? For higher derivatives we see from $\hat{f}'(\omega) = i\omega\hat{f}(\omega)$ that

$$f^{(m)} = \mathcal{F}_2^{-1}(\omega \mapsto (i\omega)^m \hat{f}(\omega)),$$

whenever f is such that $\omega \mapsto (i\omega)^m \hat{f}(\omega)$ is in $L^2(\mathbb{R})$. If $s > 0$ and $f \in L^2(\mathbb{R})$ is such that $\omega \mapsto (i\omega)^s \hat{f}(\omega) \in L^2(\mathbb{R})$, we define

$$D^s f = \mathcal{F}_2^{-1}(\omega \mapsto (i\omega)^s \hat{f}(\omega)).$$

Then the function $D^s f$ is in $L^2(\mathbb{R})$ and it is called the *fractional derivative* of f of order s . Recall that

$$(i\omega)^s = \begin{cases} e^{\frac{1}{2}\pi i s} |\omega|^s, & \omega \geq 0, \\ e^{\frac{3}{2}\pi i s} |\omega|^s, & \omega < 0. \end{cases}$$

By means of the Fourier-Plancherel transform we see that if $f \in L^2(\mathbb{R})$ has distributional derivative in $L^2(\mathbb{R})$, then $D^{1/2} D^{1/2} f = f'$. A short overview with other approaches to fractional derivatives is given on

http://en.wikipedia.org/wiki/Fractional_calculus.

4 Stieltjes integral

4.1 Definition and properties

Riemann's idea to define the integral of a function f on an interval $[a, b]$ was to consider convergence of sums $\sum_i f(s_i)(t_{i+1} - t_i)$, where $a = t_0 \leq t_1 \leq \dots \leq t_n = b$ is a partition and $s_i \in [t_i, t_{i+1}]$. The factor $t_{i+1} - t_i$ is the weight or length of the interval $[t_i, t_{i+1}]$. If μ is an arbitrary measure on $[a, b]$, then the Lebesgue integral of f with respect to μ is constructed by means of sums $\sum_i \alpha_i \mu(A_i)$, where $\sum_i \alpha_i \mathbb{1}_{A_i}$ are step functions approximating f . The factor $\mu(A_i)$ is the weight or measure of the set A_i . The integral of Thomas Stieltjes is both historically and in generality between the Riemann and the Lebesgue integral. It is not as general as Lebesgue's integral but its construction is as explicit as that of Riemann's integral. It is especially convenient in probability theory. Moreover, the stochastic integrals of Wiener and Ito are based on Stieltjes's construction.

A *partition* of an interval $[a, b]$ is a finite set $\underline{t} = \{t_0, t_1, \dots, t_n\}$ of points of $[a, b]$. We will assume that a partition always contains a and b . In proofs and constructions it is often convenient to order the points of a partition. The assumption that the end points are in the partition and the right labeling of the point so that they are ordered can be summarized in the phrase “let $a = t_0 \leq t_1 \leq \dots \leq t_n = b$ be a partition of $[a, b]$ ”. It turns out to be convenient to allow here that some consecutive points are equal. This means in particular that a point may occur several times. The *mesh size* of a partition $a = t_0 \leq t_1 \leq \dots \leq t_n = b$ is defined by

$$\text{mesh}(\underline{t}) = \max_{1 \leq k \leq n} |t_k - t_{k-1}|.$$

If \underline{t} and \underline{s} are two partitions, then \underline{s} is called *finer than* \underline{t} if $\underline{s} \supseteq \underline{t}$. That is, a finer partition is obtained from a coarser one by adding more points. Notice that $\{0, \pi, 4\}$ and $\{1, 2, 4\}$ are both partitions of $[0, 4]$ but none of the two is finer than the other. Given two (or finitely many) partitions, there is always a partition that is finer than each of them, namely the union. Such a finer partition is sometimes called a *common refinement*.

Definition 4.1. Let $f, g: [a, b] \rightarrow \mathbb{R}$. The function f is *Stieltjes integrable with respect to g* if there exists $I \in \mathbb{R}$ such that for every $\varepsilon > 0$ there exists a partition such that for every finer partition $a = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n = b$ and every choice of $s_i \in [t_{i-1}, t_i]$ we have

$$\left| \sum_{k=1}^n f(s_k) (g(t_k) - g(t_{k-1})) - I \right| < \varepsilon.$$

The number I is then called the *Stieltjes integral of f with respect to g* and it is denoted by

$$\int_a^b f(t) dg(t).$$

If $a = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n = b$ is a partition and $s_k \in [t_{k-1}, t_k]$ are intermediate points, we denote the corresponding *Riemann-Stieltjes sum* by

$$S(\underline{s}, \underline{t}) = \sum_{k=1}^n f(s_k) (g(t_k) - g(t_{k-1})).$$

The most famous instance of Stieltjes integrability is the following.

Theorem 4.2. *If $f: [a, b] \rightarrow \mathbb{R}$ is continuous and $g: [a, b] \rightarrow \mathbb{R}$ is of bounded variation, then f is Stieltjes integrable with respect to g .*

Proof. First we compare the Riemann-Stieltjes sums for two partitions, one finer than the other. Let $a = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n = b$ and for each k let $t_{k-1} = \tau_{k,0} \leq \tau_{k,1} \leq \dots \leq \tau_{k,m_k} = t_k$. Let $s_k \in [t_{k-1}, t_k]$ and let $\sigma_{k,i} \in [\tau_{k,i-1}, \tau_{k,i}]$. Then

$$\begin{aligned} & |S(\underline{\sigma}, \underline{\tau}) - S(\underline{s}, \underline{t})| \\ &= \left| \sum_{k=1}^n \sum_{i=1}^{m_k} f(\sigma_{k,i}) (g(\tau_{k,i}) - g(\tau_{k,i-1})) - \sum_{k=1}^n f(s_k) (g(t_k) - g(t_{k-1})) \right| \\ &= \left| \sum_{k=1}^n \sum_{i=1}^{m_k} f(\sigma_{k,i}) (g(\tau_{k,i}) - g(\tau_{k,i-1})) - \sum_{k=1}^n \sum_{i=1}^{m_k} f(s_k) (g(\tau_{k,i}) - g(\tau_{k,i-1})) \right| \\ &= \left| \sum_{k=1}^n \sum_{i=1}^{m_k} (f(\sigma_{k,i}) - f(s_k)) (g(\tau_{k,i}) - g(\tau_{k,i-1})) \right| \\ &\leq \max_k \max_{\sigma, s \in [t_{k-1}, t_k]} |f(\sigma) - f(s)| \sum_{k=1}^n \sum_{i=1}^{m_k} (g(\tau_{k,i}) - g(\tau_{k,i-1})) \\ &\leq \max_k \max_{\sigma, s \in [t_{k-1}, t_k]} |f(\sigma) - f(s)| V_{[a,b]}(g). \end{aligned}$$

Next, choose any sequence of partitions \underline{t}^n with mesh size tending to zero and such that \underline{t}^{n+1} is a refinement of \underline{t}^n for every n . Then for $m \geq n$,

$$|S(\underline{t}^n, \underline{t}^n) - S(\underline{t}^m, \underline{t}^m)| \leq \max_{|s-\sigma| \leq \text{mesh}(\underline{t}^n)} |f(s) - f(\sigma)| V_{[a,b]}(g).$$

As f is continuous on $[a, b]$, it is uniformly continuous, since $[a, b]$ is compact. Hence the previous formula yields

$$|S(\underline{t}^n, \underline{t}^n) - S(\underline{t}^m, \underline{t}^m)| \rightarrow 0 \text{ as } m, n \rightarrow \infty.$$

So $(S(\underline{s}^n, \underline{t}^n))_n$ is a Cauchy sequence and therefore there exists $I \in \mathbb{R}$ such that

$$I = \lim_{n \rightarrow \infty} S(\underline{s}^n, \underline{t}^n).$$

Now we have the candidate integral I .

Let $\varepsilon > 0$. Choose $\delta > 0$ such that

$$|s - \sigma| < \delta \implies |f(s) - f(\sigma)| < \frac{\varepsilon}{2V_{[a,b]}(g) + 1}.$$

(The $+1$ in the denominator is only there to avoid division by 0 in case $V_{[a,b]}(g) = 0$. Alternatively, this trivial case could be treated separately.) Choose N such that

$$\text{mesh}(\underline{t}^N) < \varepsilon/2 \quad \text{and} \quad |S(\underline{s}^N, \underline{t}^N) - I| < \varepsilon/2 \text{ for all } n \geq N.$$

Let $\underline{\tau}$ be any partition finer than \underline{t}^N and $\underline{\sigma}$ any choice of intermediate points for $\underline{\tau}$. Then

$$|S(\underline{\sigma}, \underline{\tau}) - S(\underline{s}^N, \underline{t}^N)| \leq \max_{|s - \sigma| < \text{mesh}(\underline{t}^N)} |f(s) - f(\sigma)| V_{[a,b]}(g) < \varepsilon/2$$

and

$$|S(\underline{s}^N, \underline{t}^N) - I| < \varepsilon/2$$

so

$$|S(\underline{\sigma}, \underline{\tau}) - I| < \varepsilon.$$

Hence f is Stieltjes integrable with respect to g and $\int_a^b f(t) dg(t) = I$. \square

Recall that every increasing or decreasing function g is of bounded variation.

Example 4.3. Let $f: [a, b] \rightarrow \mathbb{R}$ be continuous and $g: [a, b] \rightarrow \mathbb{R}$ be piecewise constant with jumps at u_1, \dots, u_m of size β_1, \dots, β_m , respectively, where $a \leq u_1 < u_2 < \dots < u_m < b$. That is,

$$g = \sum_{i=1}^m \beta_i \mathbb{1}_{[u_i, b]},$$

or, equivalently, $g(t) = \beta_1 + \dots + \beta_\ell$ whenever $t \in [u_\ell, u_{\ell+1})$ (where $u_{m+1} := b$). Then f is Stieltjes integrable with respect to g and

$$\int_a^b f(t) dg(t) = \sum_{\ell=1}^m f(u_\ell) \beta_\ell.$$

Indeed, let $\varepsilon > 0$. Choose $\delta > 0$ such that

$$|s - \sigma| < \delta \implies |f(s) - f(\sigma)| < \frac{\varepsilon}{\sum_{\ell} |\beta_\ell| + 1}$$

and such that $\delta < u_{\ell+1} - u_\ell$ for all ℓ . Let $a = t_0 \leq t_1 \leq \dots \leq t_n = b$ be any partition with $\text{mesh}(\underline{t}) < \delta$ and let \underline{s} be any choice of intermediate points for \underline{t} . Then each interval $[t_{k-1}, t_k]$ contains at most one of the jump points u_ℓ . Further, $\bigcup_{\ell=1}^m [u_\ell, u_{\ell+1}) = [a, b)$, so for each ℓ there is exactly one k_ℓ such that $u_\ell \in [t_{k_\ell+1}, t_{k_\ell})$. Then

$$\sum_{k=1}^n f(s_k) (g(t_k) - g(t_{k-1})) = \sum_{\ell=1}^m f(s_{k_\ell}) (g(t_{k_\ell}) - g(t_{k_\ell-1})) = \sum_{\ell=1}^m f(s_{k_\ell}) \beta_\ell.$$

Since both $s_{k_\ell} \in [t_{k_\ell-1}, t_{k_\ell}]$ and $u_\ell \in [t_{k_\ell} - 1, t_{k_\ell})$, we have $|s_{k_\ell} - u_\ell| < |t_{k_\ell} - t_{k_\ell-1}| < \delta$. Hence

$$|f(s_{k_\ell}) - f(u_\ell)| < \frac{\varepsilon}{\sum_\ell |\beta_\ell| + 1},$$

so

$$\left| \sum_{\ell=1}^m f(s_{k_\ell})\beta_\ell - \sum_{\ell=1}^m f(u_\ell)\beta_\ell \right| \leq \sum_{\ell=1}^m |f(s_{k_\ell}) - f(u_\ell)| |\beta_\ell| < \varepsilon.$$

Therefore f is Stieltjes integrable with respect to g and $\int_a^b f(t) dg(t) = \sum_{\ell=1}^m f(u_\ell)\beta_\ell$.

Example 4.4. The function $f = \mathbb{1}_{[0,1/2)}$ is *not* Stieltjes integrable with respect to $g = \mathbb{1}_{[1/2,1]}$. Indeed, if $0 = t_0 \leq t_1 \leq \dots \leq t_n = 1$ is any partition, then there exists a k^* such that $[t_{k^*-1}, t_{k^*})$ contains $1/2$. Then $\sum_{k=1}^n f(s_k)(g(t_k) - g(t_{k-1})) = f(s_{k^*})$. By choosing $s_{k^*} \in [t_{k^*-1}, t_{k^*}]$ less than $1/2$ or greater or equal $1/2$ we can make the Riemann Stieltjes sum either 0 or 1. Thus f is not Stieltjes integrable with respect to g .

A formula for calculus with Stieltjes integrals. If $g \in C^1[a, b]$ then g is of bounded variation. Indeed, for $a = t_0 \leq t_1 \leq \dots \leq t_n = b$,

$$\sum_{\ell}^n |g(t_k) - g(t_{k-1})| \leq \sum_{k=1}^n \max_{\xi \in [a,b]} |g'(\xi)| |t_k - t_{k-1}| = \max_{\xi \in [a,b]} |g'(\xi)| (b - a),$$

so $V_{[a,b]}(g) \leq \max_{\xi \in [a,b]} |g'(\xi)| (b - a)$.

Theorem 4.5. If $f: [a, b] \rightarrow \mathbb{R}$ is continuous and $g: [a, b] \rightarrow \mathbb{R}$ is continuously differentiable, then

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

Proof. First we show that the continuous differentiability of g implies uniform differentiability in the following sense: for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$0 < |s - t| < \delta \implies \left| \frac{g(s) - g(t)}{s - t} - g'(t) \right| < \varepsilon.$$

Indeed, let

$$h(s, t) := \begin{cases} \frac{g(s) - g(t)}{s - t}, & s \neq t, \\ g'(t), & s = t. \end{cases}$$

Then h is continuous on $[a, b] \times [a, b]$ and therefore uniformly continuous. Hence for $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|(s, t) - (s_0, t_0)\| < \delta \implies |h(s, t) - h(s_0, t_0)| < \varepsilon.$$

Then (with $t = s_0 = t_0$)

$$|s - t| < \delta \implies |h(s, t) - h(t, t)| < \varepsilon,$$

and that is the implication we wanted to show.

Next, let $\varepsilon > 0$. Choose $\delta > 0$ such that

$$0 < |s - t| < \delta \implies \left| \frac{g(s) - g(t)}{s - t} - g'(t) \right| < \frac{\varepsilon}{2(b - a)\|f\|_\infty + 1}.$$

Then

$$0 < |s - t| < \delta \implies |g(s) - g(t) - g'(t)(s - t)| < \frac{\varepsilon}{2(b - a)\|f\|_\infty + 1}|s - t|.$$

Choose a partition \underline{t} such that for every finer partition $a = \tau_0 \leq \tau_1 \leq \dots \leq \tau_n = b$ and for every choice of $\sigma_k \in [\tau_{k-1}, \tau_k]$,

$$\left| \sum_{k=1}^n f(\sigma_k) (g(\tau_k) - g(\tau_{k-1})) - \int_a^b f(t) dg(t) \right| < \varepsilon/2.$$

Without loss of generality we may choose \underline{t} such that $\text{mesh}(\underline{t}) < \delta$. Then also $\text{mesh}(\underline{\tau}) < \delta$, so

$$\begin{aligned} & \left| \sum_{k=1}^n f(\sigma_k) g'(\sigma_k) (\tau_k - \tau_{k-1}) - \int_a^b f(t) dg(t) \right| \\ & < \left| \sum_{k=1}^n f(\sigma_k) g'(\sigma_k) (\tau_k - \tau_{k-1}) - \sum_{k=1}^n f(\sigma_k) (g(\tau_k) - g(\tau_{k-1})) \right| + \varepsilon/2 \\ & \leq \sum_{k=1}^n |f(\sigma_k)| \left| g'(\sigma_k) (\tau_k - \tau_{k-1}) - (g(\tau_k) - g(\tau_{k-1})) \right| + \varepsilon/2 \\ & \leq \sum_{k=1}^n \|f\|_\infty |g'(\sigma_k) (\tau_k - \sigma_k) + g'(\sigma_k) (\sigma_k - \tau_{k-1}) - (g(\tau_k) - g(\sigma_k)) + (g(\tau_{k-1}) - g(\sigma_k))| + \varepsilon/2 \\ & \leq \|f\|_\infty \sum_{k=1}^n \left(\frac{\varepsilon}{2(b - a)\|f\|_\infty + 1} |\tau_k - \sigma_k| + \frac{\varepsilon}{2(b - a)\|f\|_\infty + 1} |\tau_{k-1} - \sigma_k| \right) + \varepsilon/2 \\ & \leq \frac{\varepsilon}{2(b - a)} \sum_{k=1}^n (\tau_k - \sigma_k + \sigma_k - \tau_{k-1}) + \varepsilon/2 \leq \frac{\varepsilon}{2(b - a)} (b - a) + \varepsilon/2 = \varepsilon. \end{aligned}$$

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

Integration by parts. Another important theorem about the Stieltjes integral is the *integration by parts formula*.

Theorem 4.6. *Let $f, g: [a, b] \rightarrow \mathbb{R}$. If f is Stieltjes integrable with respect to g , then g is Stieltjes integrable with respect to f and*

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

Proof. Let $\varepsilon > 0$. Since we know that $\int_a^b f(t) dg(t)$ exists, we can choose a partition $a = t_0 \leq t_1 \leq \dots \leq t_n = b$ such that for every finer partition $a = \tau_0 \leq \tau_1 \leq \dots \leq \tau_m = b$ and every $\sigma_i \in [\tau_{i-1}, \tau_i]$ we have

$$\left| \sum_{k=1}^m f(\sigma_k) (g(\tau_k) - g(\tau_{k-1})) - \int_a^b f(t) dg(t) \right| < \varepsilon.$$

In order to show that $\int_a^b g(t) \, df(t)$ exists and equals $-\int_a^b g(t) \, df(t) + f(b)g(b) - f(a)g(a)$, we consider the partition $a = t_0 \leq t_0 \leq t_1 \leq t_1 \leq t_2 \leq t_2 \leq \dots \leq t_N \leq t_N = b$ (on purpose!). Let $a = \theta_0 \leq \theta_1 \leq \dots \leq \theta_M = b$ be a finer partition and let $\zeta_k \in [\theta_{k-1}, \theta_k]$. Due to the way we have chosen the partition with double t_i s, each t_i will occur as one of the ζ_k . That is, $a = \zeta_1 \leq \zeta_1 \leq \dots \leq \zeta_M = b$ is a partition of $[a, b]$ which is finer than $a = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n = b$. Moreover, $\theta_{k-1} \leq \zeta_k \leq \theta_k \leq \zeta_{k+1}$, so $\theta_k \in [\zeta_k, \zeta_{k+1}]$ for $k = 0, \dots, M-1$, where we set $\zeta_0 := a$. Hence

$$\left| \sum_{k=0}^{M-1} f(\theta_k) (g(\zeta_{k+1}) - g(\zeta_k)) - \int_a^b f(t) \, dg(t) \right| < \varepsilon.$$

Further,

$$\begin{aligned} \sum_{k=0}^{M-1} g(\theta_k) (f(\zeta_{k+1}) - f(\zeta_k)) &= \sum_{k=0}^{M-1} g(\theta_k) f(\zeta_{k+1}) - \sum_{k=0}^{M-1} g(\theta_k) f(\zeta_k) \\ &= \sum_{k=1}^M g(\theta_{k-1}) f(\zeta_k) - \sum_{k=0}^{M-1} g(\theta_k) f(\zeta_k) \\ &= \sum_{k=1}^M g(\theta_{k-1}) f(\zeta_k) - \sum_{k=1}^M g(\theta_k) f(\zeta_k) + g(\theta_0) f(\zeta_0) \\ &= - \sum_{k=1}^M f(\zeta_k) (g(\theta_k) - g(\theta_{k-1})) + f(b)g(b) - f(a)g(a). \end{aligned}$$

So

$$\begin{aligned} &\left| \sum_{k=0}^{M-1} g(\theta_k) (f(\zeta_{k+1}) - f(\zeta_k)) - \left(- \int_a^b f(t) \, dg(t) f(b)g(b) - f(a)g(a) \right) \right| \\ &\leq \left| - \sum_{k=1}^M f(\zeta_k) (g(\theta_k) - g(\theta_{k-1})) + \int_a^b f(t) \, dg(t) \right| < \varepsilon. \end{aligned}$$

Hence g is Stieltjes integrable with respect to f and

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

□

Remark 4.7. The definition given here of Stieltjes integrability is not the original one, which was somewhat more restrictive. Let $f, g: [a, b] \rightarrow \mathbb{R}$. The original definition of “ f is Stieljes integrable with respect to g ” would read

There exists an $I \in \mathbb{R}$ such that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every partition $a = t_0 \leq t_1 \leq \dots \leq t_n = b$ with $\text{mesh}(\underline{t}) < \delta$ and every choice of $s_k \in [t_{k-1}, t_k]$ one has

$$\left| \sum_{k=1}^n f(s_k) (g(t_k) - g(t_{k-1})) - I \right| < \varepsilon.$$

The difference is that in this definition *every* partition of small enough mesh has to be considered, whereas in our definition we only need to consider partitions that are refinement of a given partition. The latter enables us to include some critical points in each of the partitions.

Let us consider an example. Let $f = \mathbb{1}_{[0,1/2]}$ and $g = \mathbb{1}_{[1/2,1]}$. As in Example 4.4 we can show for any partition of which $1/2$ is not one of the points that the Riemann Stieltjes sum can be 1 or 0, depending on the choice of intermediate points. Therefore, f and g do not satisfy the original definition stated above. However, for every partition $0 = \tau_0 \leq \tau_1 \leq \dots \leq \tau_n = 1$ finer than $0 = 0 \leq 1/2 \leq 1 = 1$ and any choice of intermediate points $\sigma_k \in [\tau_{k-1}, \tau_k]$, we have

$$\sum_{k=1}^n f(\sigma_k) \left(g(\tau_k) - g(\tau_{k-1}) \right) = 1.$$

So f is Stieltjes integrable with respect to g according to our definition.

More generally, it can be shown that every càglàd function f is Stieltjes integrable with respect to any càdlàg function g . Here càglàd stands for the french phrase “continu à gauche, limite à droite”, which means left continuous with right limits. It means that f is such that at every point t , both the left and right limits exist and that $f(t)$ equals the left limit. Similarly, càdlàg means right continuous with left limits. This situation, although it seems somewhat technical, is of great importance in the theory of stochastic processes, in particular, integration with respect to semimartingales.

Connection with Lebesgue integral. Let $g: [a, b] \rightarrow \mathbb{R}$ be increasing. Then $\mu((s, t]) := g(t) - g(s)$ extends to a measure μ on the Borel sets of $[a, b]$. If f is continuous, then f is Lebesgue integrable with respect to the bounded measure μ . We will show that

$$\int_a^b f(t) dg(t) = \int_{[a,b]} f(t) d\mu(t).$$

Let $N \in \mathbb{N}$. Choose a partition $a = t_0^N \leq t_1^N \leq \dots \leq t_{n_N}^N = b$ such that $\text{mesh}(t^N) < 1/N$ and

$$\left| \sum_{k=1}^{n_N} f(s_k) \left(g(t_k^N) - g(t_{k-1}^N) \right) - \int_a^b f(t) dg(t) \right| < \frac{1}{N},$$

for any choice of $s_k \in [t_{k-1}^N, t_k^N]$. Choose any $s_k^N \in [t_{k-1}^N, t_k^N]$ and let

$$h_N := \sum_{k=1}^{n_N} f(s_k^N) \mathbb{1}_{[t_{k-1}^N, t_k^N]}.$$

Then

$$\int h_N d\mu(t) = \sum_{k=1}^{n_N} f(s_k^N) \mu((t_{k-1}^N, t_k^N]) = \sum_{k=1}^{n_N} f(s_k^N) \left(g(t_k^N) - g(t_{k-1}^N) \right).$$

Since f is continuous, $h_N(t) \rightarrow f(t)$ as $N \rightarrow \infty$ for every $t \in [a, b]$. Further, $|h_N| \leq \|f\|_\infty$ on $[a, b]$. Due to Lebesgue's dominated convergence theorem we find as $N \rightarrow \infty$

$$\int h_N(t) d\mu(t) \rightarrow \int f d\mu(t).$$

Also,

$$\int h_N(t) d\mu(t) = \sum_{k=1}^{n_N} f(s_k^N) (g(t_k^N) - g(t_{k-1}^N)) \rightarrow \int_a^b f(t) dg(t).$$

Recall that a measure μ on $[a, b]$ is called *absolutely continuous* with respect to the Lebesgue measure λ if $\lambda(A) = 0$ implies $\mu(A) = 0$ for every Borel set $A \subseteq [a, b]$.

Theorem 4.8. *Let $g: [a, b] \rightarrow \mathbb{R}$ be increasing. Let μ be a measure on the Borel sets of $[a, b]$ such that $\mu((s, t]) = g(t) - g(s)$ for all $a \leq s < t \leq b$. Then*

$$\mu \text{ is absolutely continuous with respect to } \lambda \iff g \text{ is absolutely continuous.}$$

Proof. \Leftarrow) Let A be a Borel set with $\lambda(A) = 0$. Let $\varepsilon > 0$. We want to show that $\mu(A) < \varepsilon$. Since g is absolutely continuous, we can choose a $\delta > 0$ such that

$$\sum_{k=1}^n |g(b_k) - g(a_k)| < \varepsilon$$

for every mutually disjoint intervals (a_k, b_k) , $k = 1, \dots, n$, with $\sum_{k=1}^n (b_k - a_k) < \delta$. Since $\lambda(A) = 0$, there are mutually disjoint open intervals (a_i, b_i) , $i \in \mathbb{N}$, such that $A \subseteq \bigcup_{i=1}^{\infty} (a_i, b_i)$ and $\sum_{i=1}^{\infty} (b_i - a_i) < \delta$. Then $\sum_{i=1}^n (b_i - a_i) < \delta$ for each n so $\sum_{i=1}^n |g(b_i) - g(a_i)| < \varepsilon$ for every n . Hence $\sum_{i=1}^{\infty} |g(b_i) - g(a_i)| \leq \varepsilon$. We infer

$$\begin{aligned} \mu(A) &\leq \mu\left(\bigcup_{i=1}^{\infty} (a_i, b_i)\right) = \sum_{i=1}^{\infty} \mu((a_i, b_i)) \leq \sum_{i=1}^{\infty} \mu((a_i, b_i]) \\ &= \sum_{i=1}^{\infty} (g(b_i) - g(a_i)) \leq \varepsilon. \end{aligned}$$

Thus $\mu(A) = 0$.

\Rightarrow) Assume that μ is absolutely continuous with respect to λ . Due to the Radon-Nikodym theorem there exists an integrable function $h: [a, b] \rightarrow \mathbb{R}$ with $h \geq 0$ such that

$$\mu(A) = \int \mathbb{1}_A(s) h(s) d\lambda(s).$$

for all Borel sets A . If we use this for $A = (a, t]$, we find

$$g(t) - g(a) = \mu((a, t]) = \int \mathbb{1}_{(a, t]}(s) h(s) d\lambda(s).$$

Hence

$$g(t) = g(a) + \int_a^t h(s) ds \text{ for all } t \in [a, b],$$

which yields that g is absolutely continuous. □

4.2 Langevin's equation

Consider the movement of a particle of mass m along a line, with velocity $v(t)$ at time t . Due to friction there is a force $-av(t)$ acting on the particle, where $a \geq 0$ is a constant. Suppose that many small influences in the environment of the particle are acting on the particle, resulting in a force, which is erratic and very irregular. We will call this force “noise”. The velocity of the particle then satisfies

$$mv'(t) = -av(t) + \text{“noise”}.$$

By means of rescaling, we can rewrite the equation as

$$y'(t) = -ay(t) + \text{“noise”} \quad (10)$$

How to model noise? Let us first consider the case $a = 0$. We conceive of the noise as a force which is so erratic that it cannot be modeled deterministically. We think of tossing a coin at each time t and a force hitting the particle in positive direction if the outcome is head and in negative direction if the outcome is tail. For a good mathematical model we consider discretization of the differential equation

$$y'(t) = \text{“noise”} \quad (11)$$

Fix $t > 0$ and $n \in \mathbb{N}$. Let X_1, \dots, X_n be mutually independent random variables such that $X_k = 1$ with probability $1/2$ and $X_k = -1$ with probability $1/2$. We let $t_k = \frac{k}{n}t$ and consider the discretized equation

$$\frac{y(t_k) - y(t_{k-1})}{t_k - t_{k-1}} = bX_k, \quad k = 1, \dots, n,$$

where b is a constant. Then

$$y(t_k) = y(t_{k-1}) + b(t_k - t_{k-1})X_k = y(t_{k-1}) + b\frac{1}{n}X_k,$$

so

$$y(t) = y(t_n) = y(0) + \frac{b_n}{n}t \sum_{k=1}^n X_k.$$

We do this for each n ,

$$y_n(t) = y(0) + \frac{b}{n}t \sum_{k=1}^n X_{n,k}.$$

We are looking for a limit as $n \rightarrow \infty$. If b_n is too small, the random influence may converge to 0. If b_n is too big, the sums may not converge. The Central Limit Theorem says that $\frac{1}{\sqrt{n}}t \sum_{k=1}^n X_{n,k}$ converges in law to a random variable $W(t)$ with normal (Gaussian) law with mean 0 and variance t . Therefore we choose $b_n = \sqrt{n}$. The limiting random variables $W(t)$ can be chosen in such a way that the dependence on t is continuous, as is stated in the next theorem. Its proof is difficult and beyond these notes.

Theorem 4.9. *There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $W(t)$, $t \geq 0$, such that*

- $W_0 = 0$ a.e. on Ω ,
- W_t has normal law with mean 0 and variance t ,
- $W_t - W_s$ is independent of W_r for all $0 \leq r \leq s \leq t$,
- $t \mapsto W_t(\omega)$ is continuous for a.e. $\omega \in \Omega$.

The family of random variables W_t , $t \geq 0$, of the above theorem is called a *Brownian motion* or *Wiener process*.

Recall that a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ is a set Ω , a σ -algebra \mathcal{F} , and a measure \mathbb{P} on \mathcal{F} with $\mathbb{P}(\Omega) = 1$. A *random variable* is a map $X: \Omega \rightarrow \mathbb{R}$ that is measurable with respect to \mathcal{F} and the Borel σ -algebra in \mathbb{R} . The *law* or *distribution* of X is the measure μ_X on \mathbb{R} satisfying

$$\mu_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega: X(\omega) \in A\}),$$

where $A \subseteq \mathbb{R}$ is a Borel set. In words, $\mu_X(A)$ is the probability that X has values in A . There can be different random variables (even defined on different probability spaces) with the same law. If X and Y are random variables defined on Ω , then their *joint law* is the measure $\mu_{(X,Y)}$ on \mathbb{R}^2 defined by

$$\mu_{(X,Y)}(C) = \mathbb{P}(\{\omega \in \Omega: (X(\omega), Y(\omega)) \in C\}),$$

$C \subseteq \mathbb{R}^2$ Borel. The random variables X and Y are *independent* if $\mu_{(X,Y)}$ equals the product measure $\mu_X \otimes \mu_Y$. Equivalently, if

$$\mathbb{P}((X, Y) \in A \times B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

for all Borel sets $A \subseteq \mathbb{R}$, $B \subseteq \mathbb{R}$. A law μ is called *normal* or *Gaussian* with mean $m \in \mathbb{R}$ and variance $\sigma^2 \geq 0$ if for every Borel set $A \subseteq \mathbb{R}$,

$$\mu(A) = \int_A \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}.$$

By considering limits of discretizations we have found a mathematical model for noise. As a solution of (11) we found

$$y(t) = y(0) + W_t$$

so

$$\text{“noise”} = y'(t) = \frac{d}{dt}W_t.$$

Langevin’s equation is therefore

$$y'(t) = -ay(t) + \sigma \frac{d}{dt}W_t, \tag{12}$$

where $\sigma \in \mathbb{R}$ is some auxiliary parameter. More explicitly, W_t depends on the variable $\omega \in \Omega$, so also $y(t)$ will depend on ω . We thus have for each $\omega \in \Omega$ the differential equation

$$y'(t)(\omega) = -ay(t)(\omega) + \sigma \frac{d}{dt}W_t(\omega).$$

As usual in probability theory we often do not write the dependence on ω explicitly.

There is a mathematical problem: the map $t \mapsto W_t(\omega)$ is continuous for almost every ω , but it is *not differentiable* and even *nowhere differentiable*. So $\frac{d}{dt}W_t(\omega)$ does not exist. (The proof of this fact is far from easy and we omit it). There is a simple way to avoid this problem: we formulate Langevin's equation as an *integral equation*,

$$y(T) = y(0) + \int_0^T -ay(t) dt + \sigma W_T, \quad T \geq 0. \quad (13)$$

Let us next solve equation (12) or rather its correct formulation (13). The homogeneous equation

$$y'(t) = -ay(t)$$

has solution

$$y(t) = ce^{-at}, \quad t \geq 0.$$

Due to the variation-of-constants formula, the solution of the inhomogeneous equation

$$y'(t) = -ay(t) + f(t), \quad t \geq 0,$$

is given by the convolution

$$y(t) = ce^{-at} + \int_0^t e^{-a(t-s)} f(s) ds,$$

where $c = y(0)$ and f is a continuous function. Formally, the solution of (12) would be

$$\begin{aligned} y(t) &= y(0)e^{-at} + \int_0^t e^{-a(t-s)} \sigma \frac{d}{ds} W_s ds \\ &= y(0)e^{-at} + \sigma \int_0^t e^{-a(t-s)} dW_s. \end{aligned}$$

Be aware that $\frac{d}{ds}W_s$ doesnot exist. The latter expression, however, makes sense as a Stieltjes integral. Indeed, $s \mapsto e^{-a(t-s)}$ is increasing hence of bounded variation and W_s is continuous. Hence

$$\int_0^t e^{-a(t-s)} dW_s(\omega)$$

exists for almost every $\omega \in \Omega$ as a Stieltjes integral. These formal manipulations do yield the correct solution.

Theorem 4.10. *Let $a, \sigma \in \mathbb{R}$. Let W_t , $t \geq 0$, be a Wiener process defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $y_0 \in \mathbb{R}$. Define*

$$y(t) = e^{-at}y_0 + \sigma \int_0^t e^{-a(t-s)} dW_s$$

(for almost every $\omega \in \Omega$ as Stieltjes integral). Then y satisfies the Langevin equation

$$y(t) = y_0 - a \int_0^t y(s) ds + \sigma W(t).$$

Proof. By integration by parts,

$$\begin{aligned}
y(t) &= e^{-at}y_0 + \sigma \int_0^t e^{-a(t-s)} dW_s \\
&= e^{-at}y_0 + \sigma e^{-at} \left(e^{as}W_s \Big|_0^t - \int_0^t W_s de^{as} \right) \\
&= e^{-a(t-s)} + \sigma e^{-at} \left(e^{at}W_t - \int_0^t aW_s e^{as} ds \right) \\
&= e^{-at}y_0 + \sigma W_t - \sigma \int_0^t a e^{-a(t-s)} W_s ds.
\end{aligned}$$

Hence the above expression for y and Fubini yield

$$\begin{aligned}
\int_0^T y(t) dt &= \sigma \int_0^T W_t dt - \sigma \int_0^T \int_0^t a e^{-a(t-s)} W_s ds dt + \int_0^T e^{-at} y_0 dt \\
&= \sigma \int_0^T W_t dt - \sigma \int_0^T \int_s^T a e^{-a(t-s)} W_s dt ds + \frac{e^{-aT}}{a} y_0 - \frac{y_0}{a} \\
&= \sigma \int_0^T W_t dt + \sigma \int_0^T \left(e^{-a(t-s)} - 1 \right) W_s ds + \frac{e^{-aT}}{a} y_0 - \frac{y_0}{a} \\
&= \sigma \int_0^T e^{-a(T-s)} W_s ds + \frac{e^{-aT}}{a} y_0 - \frac{y_0}{a} \\
&= \frac{1}{a} \sigma W_T - \frac{1}{a} y(T) + \frac{e^{-aT}}{a} y_0 - \frac{y_0}{a}.
\end{aligned}$$

Thus $y(t)$ satisfies the Langevin equation. \square

Concerning uniqueness we have the following.

Proposition 4.11. *Let $a, \sigma \in \mathbb{R}$. Let $W_t, t \geq 0$, be a Wiener process and let $\Omega_0 \subset \Omega$ be such that $\mathbb{P}(\Omega_0) = 1$ and $t \mapsto W_t(\omega)$ is continuous for all $\omega \in \Omega_0$. If $y_1(t)$ and $y_2(t), t \geq 0$, are two families of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $t \mapsto y_1(t)(\omega)$ and $t \mapsto y_2(t)(\omega)$ are continuous for all $\omega \in \Omega_0$ and such that y_1 and y_2 both satisfy (13), then*

$$|y_1(t) - y_2(t)| \leq e^{-at} |y_1(0) - y_2(0)|, \quad t \geq 0.$$

Proof. Let

$$v(t) := y_1(t) - y_2(t), \quad t \geq 0.$$

Then v satisfies

$$v(t) = -a \int_0^t v(s) ds.$$

On Ω_0 , v is continuous and then from the previous formula v is even continuously differentiable. Differentiating yields

$$v'(t) = -av(t),$$

so

$$v(t) = v(0)e^{-at},$$

which yields the desired formula. \square

Consequently, if $y_1(0) = y_2(0)$, then $y_1 = y_2$.

Law of the solution. So far we have considered $\omega \in \Omega$ as a parameter and solved the Langevin equation for each ω separately. We obtained for almost every ω a function $t \mapsto y(t)(\omega)$. Next, we fix t . Then $y(t)$ is an almost everywhere defined map from Ω to \mathbb{R} . If necessary we can assume it everywhere defined by choosing it zero on those ω where it is not defined. It can be shown that the map $y(t)$ is then measurable. Thus for fixed t , $y(t)$ is a random variable. What is its law? We know that

$$y(t) = e^{-at}y_0 + \sigma \int_0^t e^{-a(t-s)} dW_s.$$

By definition of the Stieltjes integral, $y(t)$ is a limit of sums of the form

$$\mathcal{S}(s_1, \dots, s_n) = e^{-at}y_0 + \sigma \sum_{k=1}^n e^{-a(t-s_k)} (W_{s_k} - W_{s_{k-1}}).$$

From the properties of the Wiener process it follows that the $W_{s_k} - W_{s_{k-1}}$, $k = 1, \dots, n$ are mutually independent normal random variables with means 0 and variances $s_k - s_{k-1}$. Hence $\mathcal{S}(s_1, \dots, s_n)$ has normal law with mean $e^{-at}y_0$ and variance $\sigma^2 \sum_{k=1}^n e^{-2a(t-s_k)} (s_k - s_{k-1})$. It follows that the random variables converge in law to a normal random variable with mean $e^{-at}y_0$ and variance

$$\sigma^2 \sum_{k=1}^n e^{-2a(t-s_k)} ds = \frac{\sigma^2}{2a} (1 - e^{-2at}).$$

Hence $y(t)$ is normal with mean $e^{-at}y_0$ and variance $\frac{\sigma^2}{2a} (1 - e^{-2at})$. The details of the above arguments need more probability theory and we omit them. We summarize the conclusion.

Proposition 4.12. *The solution $y(t)$ at time t of the Langevin equation*

$$y(t) = y_0 - a \int_0^t y(s) ds + \sigma W_t, \quad t \geq 0,$$

has normal law with mean e^{-at} and variance $\frac{\sigma^2}{2a} (1 - e^{-2at})$.

We could also take the initial value y_0 to be a random variable. If we take it independent of W_t for all t and with normal law with mean 0 and variance $\frac{\sigma^2}{2a}$, then it turns out that $y(t)$ has at each time t mean 0 and variance $\frac{\sigma^2}{2a}$. Thus, the random variables $y(t)$ vary in time, but their laws are constant. Such a solution is called a *stationary solution*. In case of the above laws, the solution is called an *Ornstein-Uhlenbeck process*.

4.3 Wiener integral

Let W_t , $t \geq 0$ be a Wiener process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Since $s \mapsto e^{-a(t-s)}$ is of bounded variation, we can define

$$\int_0^t e^{-a(t-s)} dW_s$$

as a Stieltjes integral, almost everywhere on Ω . In the same way we can define

$$\int_0^t f(s) dW_s$$

for any $f: [0, t] \rightarrow \mathbb{R}$ of bounded variation. As before, the corresponding Riemann-Stieltjes sums

$$\sum_{k=1}^n f(s_k) (W_{s_k} - W_{s_{k-1}})$$

have mean 0 and variance

$$\sum_{k=1}^n f(s_k)^2 (s_k - s_{k-1}).$$

In other words,

$$\mathbb{E} \left(\sum_{k=1}^n f(s_k)^2 (W_{s_k} - W_{s_{k-1}}) \right)^2 = \sum_{k=1}^n f(s_k)^2 (s_k - s_{k-1}),$$

where \mathbb{E} denotes expectation:

$$\mathbb{E}X = \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

It can be shown that taking limits leads to

$$\mathbb{E} \left(\int_0^t f(s) dW_s \right)^2 = \mathbb{E} \int_0^t f(s)^2 ds.$$

This can be restated as

$$\left\| \int_0^t f(s) dW_s \right\|_{L^2(\Omega, \mathbb{P})} = \|f\|_{L^2[0, t]}.$$

This identity is called the *Ito isometry*. The map $J: f \mapsto \int_0^t f(s) dW_s$ from $BV[0, t]$ to $L^2(\Omega, \mathbb{P})$ is hence an isometry with respect to $\|\cdot\|_{L^2[0, t]}$ on $BV[0, t]$. J is also linear. Since $BV[0, t]$ is dense in $(L^2[0, t], \|\cdot\|_{L^2[0, t]})$, there is a unique bounded linear map \hat{J} from $L^2[0, t] \rightarrow L^2(\Omega, \mathbb{P})$ that extends J . Moreover, this map \hat{J} is an isometry. We denote \hat{J} by an integral sign:

$$\int_0^t f(s) dW_s := \hat{J}(f).$$

Thus we can “integrate” any L^2 -function $f: [0, t] \rightarrow \mathbb{R}$ with respect to the Wiener process W_s . This integral is called the *Wiener integral*.

The definition of integration with respect to W_s can still be extended further, for instance, to functions f that are themselves random. Then we arrive at a *stochastic integral* called the *Ito integral*.

5 Convex functions