AN INTRODUCTION TO p-VARIATION AND YOUNG INTEGRALS.

With emphasis on sample functions of stochastic processes

R. M. Dudley and R. Norvaiša

December 2, 1998

Preface

These lecture notes for a course in Aarhus, January 1999, are a further development of our work on differentiability, p-variation, extended Stieltjes integrals, product integrals, and related topics.

In this survey and exposition, we have not aimed mainly at presenting our own research results, although some are included. More often we have mined the literature, finding under–appreciated gems from up to 85 years ago. We have proved some facts where we saw a need for linking existing results and concepts. We give earlier references where we know them. Our literature searches, although rather extensive in some directions, did not reach out so far in others, and in any case have not (yet) been primarily directed at determining the possible originality of our own results, some of which are recent at this writing. Thus, on any point, the absence of references to previous work, or references only to our own, should not be interpreted as a claim of originality.

We thank Ole Barndorff–Nielsen very much for the opportunity to give these lectures. We also thank Richard Gill and Terry Lyons for stimulating and helpful discussions.

Richard Dudley
Rimas Norvaiša
Cambridge, Mass., November 30, 1998

Contents

Chapter 1

Introduction

Let J be an interval in $\mathbb{R}$, which may be bounded or unbounded, and open or closed at either end. We will be concerned with *regulated* functions f on an interval J, having right limits $f(x+) := \lim_{y\downarrow x} f(y)$ with right jumps $\Delta^+ f(x) := f(x+) - f(x)$ for $x \in J$, not equal to the right endpoint of J, and left limits $f(x-) := \lim_{y\uparrow x} f(y)$ with left jumps $\Delta^-(f)(x) := f(x) - f(x-)$ for $x \in J$, not equal to the left endpoint of J. The functions will at first be real-valued but later may take values in a Banach algebra.

Here are two questions we will address.

1. How should integrals $\int_a^b f dg$ be defined?
2. For differentiability of nonlinear operators on subspaces of the space of regulated functions, what are good modes of differentiability (Fréchet, compact, etc.) and for what norms?

We have found that p-variation norms are useful. They are defined as follows. For any interval $J \subset \mathbb{R}$ let $\mathrm{PP}\,(J)$ denote the set of all *point partitions* of J, namely finite sequences $\{x_j\}_{j=0}^n \subset J$, where $x_0 < x_1 < \cdots < x_n$, and if $J = [a, b]$, a closed bounded interval, then $x_0 = a$ and $x_n = b$. Let ϕ: $[0,\infty) \to [0,\infty)$ be continuous, unbounded, increasing and 0 at 0, and let f: $J \to \mathbb{R}$. For each $\kappa = \{x_i\colon i = 0, 1, \ldots, n\} \in \mathrm{PP}(J)$, let $s_\phi(f;\kappa) := \sum_{i=1}^n \phi(|f(x_i) - f(x_{i-1})|)$. Then the *$\phi$-variation* of f is defined by

$$v_\phi(f; J) := \sup\{s_\phi(f;\kappa)\colon \kappa \in \mathrm{PP}(J)\}.$$

In the special case of $\phi(u) \equiv u^p$, $u \geq 0$, for some $1 \leq p < \infty$, we write $s_p := s_\phi$, $v_p := v_\phi$ and call $v_p(f; J)$ the *p-variation* of f. Let $\mathcal{W}_p(J)$ and $\mathcal{W}_\phi(J)$ be the sets of all functions f: $J \mapsto \mathbb{R}$ for which $v_p(f) < \infty$ and $v_\phi(f) < \infty$, respectively. If ϕ is convex let $\|f\|_{(\phi)} := \inf\{c > 0\colon v_\phi(f/c) \leq 1\}$. Thus for $\phi(u) := u^p$, $1 \leq p < \infty$, $\|f\|_{(p)} := \|f\|_{(\phi)} = v_p(f)^{1/p}$. Then $\|\cdot\|_{(p)}$ is a seminorm on $\mathcal{W}_p(J)$, which we call the p-variation seminorm. For $p = \infty$ let $\mathcal{W}_\infty(J)$ be the set of all regulated functions f on J, with $\|f\|_{(\infty)} := \sup f - \inf f$. Let $\|f\|_\infty := \sup_{x\in J} |f(x)|$. For $1 \leq p \leq \infty$ let $\|f\|_{[p]} := \|f\|_{(p)} + \|f\|_\infty$. Then $\|\cdot\|_{[p]}$ is a norm on $\mathcal{W}_p(J)$, called the p-variation norm.

The p-variation spaces and norms have the following invariance property. Let G be any strictly increasing, continuous function from J onto another interval M. Then $f \in \mathcal{W}_p(M)$ if and only if $f \circ G \in \mathcal{W}_p(J)$, with $\|f\|_{[p]} \equiv \|f \circ G\|_{[p]}$. Of the many norms defined on spaces of functions on intervals (e.g. Triebel [107]) we do not know any beside ϕ-variation norms and the supremum norm which share such wide invariance.

Chapter 2

How should integrals of Stieltjes type be defined?

2.1 Definitions of integrals

We begin with several definitions of integrals $\int_J f\,dg$ we will consider, where J is an interval in the real line $\mathbb{R}$, f is a function called the *integrand* and g is a function called the *integrator*. The integrals will be treated in more detail in Chapter 6.

Let J be a closed interval $[a,b]$ with $-\infty < a < b < \infty$. The *mesh* of a point partition $\xi := \{x_j\}_{j=0}^n$ is $\max_{1\le j\le n}(x_j - x_{j-1})$. A *tagged partition* will be a pair (ξ,η) where $\eta = \{y_j\}_{j=1}^n$ and $x_{j-1} \le y_j \le x_j$ for $j = 1,\ldots,n$. If (ξ,η) is a tagged partition of $[a,b]$ and f, g are two functions: $[a,b] \mapsto \mathbb{R}$, the *Riemann–Stieltjes sum* $S_{RS}(f,g,\xi,\eta)$ is defined as $\sum_{j=1}^n f(y_j)[g(x_j) - g(x_{j-1})]$. The *mesh–Riemann–Stieltjes integral* $(MRS)\int_a^b f\,dg$ is defined and equal to C iff (if and only if) for every $\epsilon > 0$ there is a $\delta > 0$ such that for every tagged partition (ξ,η) with $\text{mesh}(\xi) < \delta$ we have

$$\Big|S_{RS}(f,g,\xi,\eta) - C\Big| < \epsilon. \tag{2.1}$$

A point partition $\{u_i\}_{i=0}^m$ is a *refinement* of $\{x_j\}_{j=0}^n$ iff each $x_j = u_{i(j)}$ for some $i(j)$. The *refinement–Riemann–Stieltjes* integral $(RRS)\int_a^b f\,dg$ is defined and equal to C iff for every $\epsilon > 0$ there exists a partition $\tau \in \mathrm{PP}\,[a,b]$ such that for every tagged partition (ξ,η) with ξ a refinement of τ, (2.1) holds. To give historical credit where it is due, Hildebrandt [50], [52] calls the (RRS) integral the Moore–Pollard–Stieltjes integral. Clearly, if $(MRS)\int_a^b f\,dg$ exists then so does $(RRS)\int_a^b f\,dg$, with the same value. Some texts define the "Riemann–Stieltjes" integral as the (MRS) integral, others as the (RRS) integral.

A *gauge function* $\delta(\cdot)$: $[a,b] \mapsto \mathbb{R}$ is any strictly positive function. The tagged partition (ξ,η) is called $\delta(\cdot)$*–fine* if $y_j - \delta(y_j) \le x_{j-1} \le x_j \le y_j + \delta(y_j)$ for $j = 1,\ldots,n$. A *McShane partition* is a pair (ξ,η) where now $a \le y_1 < y_2 < \cdots < y_n \le b$ and we do not require $y_j \in [x_{j-1},x_j]$.

The *Henstock–Kurzweil* or *gauge* integral $(HK)\int_a^b f\,dg$ is defined to exist and equal C iff for every $\epsilon > 0$ there is a gauge function $\delta(\cdot)$ such that for every $\delta(\cdot)$–fine tagged partition (ξ,η), (2.1) holds. The *McShane integral* is defined in the same way with McShane partitions in place of tagged partitions.

An *interval function* will mean a function defined on intervals which may be open or closed at either end and may be singletons. An interval function ν will be called *finitely* *additive* if whenever $M_1, \cdots, M_k$ are disjoint intervals whose union is an interval M, we have $\nu(M) = \sum_{j=1}^{k} \nu(M_j)$.

Let $\mathcal{R}(J)$ denote the set of all regulated functions on J. If J is a closed interval $[a,b]$ we write $\mathcal{R}[a,b] := \mathcal{R}(J)$. Let g be a regulated function defined on a closed interval $[a,b]$. Let $\nu_g(\{x\}) := g(x+) - g(x-)$ for $a < x < b$, $\nu_g(\{a\}) := g(a+) - g(a)$, $\nu_g(\{b\}) := g(b) - g(b-)$, and $\nu_g((c,d)) := g(d-) - g(c+)$ for $a \le c < d \le b$. Then ν_g extends uniquely to a finitely additive interval function defined on all subintervals of J. If g is of bounded variation, then ν_g extends to a countably additive signed measure on the Borel subsets of J, and the *Lebesgue-Stieltjes integral* $\int_a^b f\,dg$ is defined here as $\int f\,d\nu_g$. (Some other definitions in the literature take different terms at the endpoints.)

An *interval partition* of J will be a collection $\{I_j\}_{j=1}^n$ of disjoint non-empty subintervals of J whose union is J, some of which may be singletons, and such that $x < y$ whenever $x \in I_i$, $y \in I_j$, and $1 \le i < j \le n$. Let IP (J) be the set of all interval partitions of J. For $\kappa, \lambda \in$ IP (J), κ will be said to be a *refinement* of λ if each interval in λ is a union of intervals in κ. Given a real-valued function f on J, a finitely additive interval function ν defined on all subintervals of J and $\lambda = \{I_j\}_{j=0}^n \in \mathrm{IP}(J)$, a *Kolmogorov sum* for f and λ will mean any sum $\sum_{j=0}^n f(y_j)\nu(I_j)$ where $y_j \in I_j$ for each j. The *refinement-Young-Stieltjes integral* (*RYS*) $\int_a^b f\,d\nu$ is defined to exist and equal C if and only if for every $\epsilon > 0$ there is some $\kappa \in$ IP (J) such that for any refinement λ of κ and any Kolmogorov sum S for f and λ we have $|S - C| < \epsilon$. If g is any regulated function on J then we define $(RYS) \int_a^b f\,dg$ as $(RYS) \int_a^b f\,d\nu_g$ if and only if the latter exists. We will also write $⨎_a^b f\,d\nu := (RYS) \int_a^b f\,d\nu$ iff the latter exists, and $⨎_a^b f\,dg := (RYS) \int_a^b f\,d\nu_g$ iff the latter exists and g is right–continuous on (a,b). A *Young partition* will be an interval partition consisting of singletons and open intervals. A *Young sum* will be a Kolmogorov sum for which the partition is a Young partition. Any interval partition has a Young partition as a refinement. On the equivalence of integrals defined from Kolmogorov sums or Young sums, see Theorems 4.35, 6.25 and 6.26 below.

A sum $\sum_{i\in I} x_i$ is said to converge *unconditionally* to S iff for every $\epsilon > 0$ there is a finite set $F \subset I$ such that for every finite $G \supset F$, $|S - \sum_{i\in G} x_i| < \epsilon$. A sum $\sum_{i\in I} x_i$ with values in $\mathbb{R}$ or a finite–dimensional normed space converges unconditionally (to some S) iff the sum converges absolutely, $\sum_{i\in I} |x_i| < \infty$. In infinite–dimensional Banach spaces, absolute convergence of a sum implies unconditional convergence but not conversely (Dvoretzky and Rogers, [31]). We will give some definitions in terms of unconditional convergence with a view to eventual extensions to the Banach–valued case.

For $f \in \mathcal{R}[a,b]$ recall that

$$(\Delta^+ f)(x) := f(x+) - f(x), \quad a \le x < b \quad \text{and} \quad (\Delta^- f)(x) := f(x) - f(x-), \quad a < x \le b.$$

Also let $(\Delta^\pm f)(x) := (\Delta^+ f)(x) + (\Delta^- f)(x) = f(x+) - f(x-)$ if $a < x < b$. Define functions $f_+^{(b)}$ and $f_-^{(a)}$ on $[a,b]$ by

$$\begin{cases} f_+^{(b)}(x) := f_+(x) := f(x+) := \lim_{z\downarrow x} f(z) & \text{if} \quad a \le x < b, \quad \text{and} \quad f_+^{(b)}(b) := f(b), \\ f_-^{(a)}(x) := f_-(x) := f(x-) := \lim_{z\uparrow x} f(z) & \text{if} \quad a < x \le b, \quad \text{and} \quad f_-^{(a)}(a) := f(a). \end{cases} \tag{2.2}$$

Then the *L. C. Young* or *central Young integral* is defined for $f, h \in \mathcal{R}[a,b]$ by

$$(CY) \int_a^b f\,dh := (RRS) \int_a^b f_+^{(b)} dh_-^{(a)} - (\Delta^+ f \Delta^+ h)(a) + (f\Delta^- h)(b) - \sum_{(a,b)} \Delta^+ f \Delta^\pm h$$

if the (RRS) integral exists and the sum converges unconditionally.

2.2 Choices between integrals

Now, which of the integrals is to be preferred in what situations?

The Lebesgue-Stieltjes integral $(LS)\int_a^b f\,dg$ is at hand if g is of bounded variation (and right-continuous except possibly at a), but what if it is not? The other best-known integrals are the mesh–Riemann-Stieltjes integral $(MRS)\int_a^b f dg$ and the refinement–Riemann–Stieltjes integral $(RRS)\int_a^b f dg$. But $(MRS)\int_a^b f dg$ is not defined if f and g have a common discontinuity and $(RRS)\int_a^b f\,dg$ is not defined if f and g have a discontinuity on the same side of the same point. Thus if $(RRS)\int_a^b f\,df$ is defined, f must be continuous.

Consider, specifically, the following, which we will call the *simple step example*. Let $F = 1_{[1,\infty)}$. Then $\int_0^2 F\,dF$ is not defined for the (MRS) or (RRS) integral. It is defined and equals 1 for the Lebesgue-Stieltjes integral, and all the other integrals defined in the previous section. But the classical integration by parts formula then fails: $2 = \int_0^2 F\,dF + \int_0^2 F\,dF \neq F^2(2) - F^2(0) = 1$. An alternate simple step example is given by $H := 1_{(1,\infty)}$ where again $\int_0^2 H\,dH$ is not defined for the (MRS) or (RRS) integral. The integral is defined and equals 0 for all the other integrals in Figure 2.1 below. Again, classical integration by parts fails since $\int_0^2 H\,dH + \int_0^2 H\,dH = 0 \neq H^2(2) - H^2(0) = 1$.

One useful application of the Lebesgue integral is the Hölder inequality. An analogous application of the less classical integrals defined above is the *Love-Young inequality*, as follows.

Theorem 2.1 (Love-Young inequality). *Let $f \in \mathcal{W}_p[a,b]$ and $g \in \mathcal{W}_q[a,b]$ where $p \geq 1$, $q \geq 1$, and $p^{-1} + q^{-1} > 1$. Then*

$$|\int_a^b f\,dg| \;\leq\; C_{p,q}\|f\|_{[p]}\|g\|_{(q)} \tag{2.3}$$

where $C_{p,q}$ is a constant depending only on p and q, and the integral is defined as a Riemann-Stieltjes integral if f and g have no common discontinuities, as an (RRS) integral if they have no common one-sided discontinuities, and as an (RYS) integral always.

A proof of the Love–Young inequality with $C_{p,q} = \zeta(p^{-1} + q^{-1})$, where $\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$ (the Riemann zeta function), is given in Theorem 4.26.

By making seemingly slight changes in the definition of (RRS) integral – requiring intervals to be disjoint and allowing them to be singletons – integrals $(RYS)\int_a^b f\,dg$ can be defined despite common one-sided discontinuities of f and g. (Just taking disjoint intervals, e.g. left open, right closed intervals $(a,b]$, is not sufficient.) Also, for regulated g, whenever $(RRS)\int_a^b f dg$ exists, so does $(RYS)\int_a^b f dg$ and has the same value by Theorem 6.13 below.

Why has the superiority of the (RYS) integral over the (MRS) and (RRS) integrals only been noticed by a small number of analysts? We venture the following historical remarks, each to the best of our knowledge.

W. H. Young (1914), in a paper written in detail by his wife G. C. Young, first defined what we call Young sums but only for step functions. Their daughter R. C. Young (1928) defined the sums more generally but considered only the mesh $\to 0$ case, where the integral obtained does not go beyond the Riemann-Stieltjes integral (Proposition 6.35). Kolmogorov (1930) defined a refinement integral in general spaces. The (RYS) integral as we defined it is the special case of Kolmogorov's integral for interval partitions. The relation with the Youngs' work was noticed only later. Glivenko (1936, pp. 87-117) wrote a book on the Stieltjes integral where he defined and pointed out the superiority of the integral $(RYS)\int_a^b f\,dg$ over the (RRS) integral for g of bounded variation. Glivenko's work seems to have passed almost unnoticed.

L. C. Young (1936), son of W. H. and G. C. Young, first published a proof of the Love–Young inequality, noting that E. R. Love (unpublished) had proved a form for finite sums earlier. In 1936 Young invented and used the (CY) integral. L. C. Young [118, Section 5] adopted the (RYS) integral. But the Love–Young inequality itself, and thus the need for a suitable integral in it, also received relatively little notice. Hildebrandt (1938) formulated the (RYS) integral. Gehring (1954) proved a sharp inequality for convolutions $\int f(x-y)dg(y)$ in terms of the (RYS) integral. Hildebrandt [52, pp. 92-96] devotes several pages to the (RYS) integral. But Hildebrandt does not cite L. C. Young or Gehring, thus the value of the (RYS) integral in analysis was not made evident. Hildebrandt [52, Section II.19.3.13] states and proves the integration by parts formula, when f and g are of bounded variation,

$$\int_a^b f\,dg + \int_a^b g\,df = fg|_a^b - \sum_{a\le x<b} \Delta^+ f \Delta^+ g + \sum_{a<x\le b} \Delta^- f \Delta^- g \tag{2.4}$$

for the (RYS) integral only. Perhaps some analysts have been deterred from accepting the (RYS) integral by this formula. But (2.4) also holds for the Lebesgue-Stieltjes integral when f and g are both right-continuous or both left-continuous. In that sense (2.4) is correct. In the simple step examples, it gives: $2 = \int_0^2 F\,dF + \int_0^2 F\,dF = F^2(2) - F^2(0) + (\Delta^- F(1))^2 = 2$, $0 = \int_0^2 H\,dH + \int_0^2 H\,dH = H^2(2) - H^2(0) - (\Delta^+ H(1))^2 = 0$. The integration by parts formula (2.4) is not so much a helpful device for evaluating $\int_a^b f\,dg$ but a fact of life. For other forms of integration by parts see [28, Theorem 4.8].

Advanced textbooks of real analysis, including one by one of us (Dudley [22]) usually adopt the Lebesgue integral with perhaps some mention of the (MRS) and/or (RRS) integrals. Intermediate-level texts generally treat the three integrals with little if any treatment of others. The (MRS) and (RRS) integrals are treated as if to some degree they were the integrals previously known by the student and their defects are perhaps taken mainly to show the advantages of the Lebesgue integral. Thus, it is not noticed that integrals $\int_a^b f\,dg$ where g is not of bounded variation could be useful.

It seems that in recent years more and more authors are recognizing the inadequacy of the Riemann–Stieltjes integrals, and are presenting in detail other integrals, beside the Lebesgue–Stieltjes integral, at not too advanced levels of study. Specifically Bartle [3] and Gordon [45], [46] propose or give expositions of the gauge integral, while Ross [95], [96] and Love [74] present the (rediscovered) refinement–Young–Stieltjes integral. The form of gauge integral given by McShane [80, pp. 552-553] turns out to be equivalent to the Lebesgue–Stieltjes integral for g of bounded variation.

What are the implications between existence of different integrals and p-variation conditions? In Figure 2.1, $\longrightarrow$ means that existence of the integral to the left of it implies that of the integral to the right of it, with the same value. For "$\longleftarrow$" left and right are interchanged. The marking "$f \in \mathcal{R}$" or "$g \in \mathcal{R}$" means that the implication holds for regulated f or g, respectively. "$\Longrightarrow$" means that the condition to the left of it implies existence of the integral to the right of it. The condition $\frac{1}{p}+\frac{1}{q}=1$ on the right has no arrows from (or to) it, signifying that there exist $f \in \mathcal{W}_p$ and $g \in \mathcal{W}_q$ with $\frac{1}{p}+\frac{1}{q}=1$, $1<p,q<\infty$ such that $\int_a^b f\,dg$ does not exist for any of the definitions given.

It seems to us that the (MRS) and (RRS) integrals should not be preferred because of their weakness in case of common discontinuities. When g is of bounded variation the Lebesgue–Stieltjes integral is available, as is its extension the (HK) integral. For more general g we still have a choice between a few integrals. If f is not necessarily regulated then the gauge (HK)

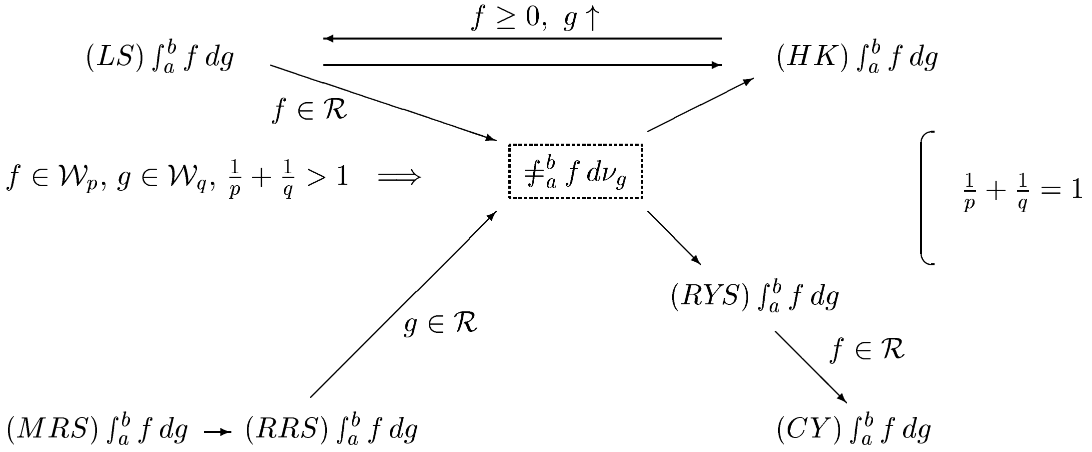

Figure 2.1: Implications for integrals

Table 2.1: References to proofs of the implications shown in Figure 2.1:

$(MRS)\int_a^b f\,dg \longrightarrow (RRS)\int_a^b f\,dg$	Proposition 6.1
$f\in\mathcal{W}_p,\ g\in\mathcal{W}_q,\ \frac{1}{p}+\frac{1}{q}>1 \Longrightarrow \fint_a^b f\,d\nu_g$	Theorem 4.26
$(RRS)\int_a^b f\,dg \longrightarrow \fint_a^b f\,d\nu_g,\quad g\in\mathcal{R}$	Theorem 6.13
$(LS)\int_a^b f\,dg \longrightarrow \fint_a^b f\,d\nu_g,\quad f\in\mathcal{R}$	Lemma 3.3 of [23] and Proposition 6.22
$(LS)\int_a^b f\,dg \longrightarrow (HK)\int_a^b f\,dg$	Theorem VI.8.1 in [98] and Theorem 6.30
$(HK)\int_a^b f\,dg \longrightarrow (LS)\int_a^b f\,dg,\quad f\geq 0,\ g\uparrow$	Theorem 6.32
$\fint_a^b f\,d\nu_g \longrightarrow (HK)\int_a^b f\,dg$	Theorem 6.31
$\fint_a^b f\,d\nu_g \longrightarrow (RYS)\int_a^b f\,dg$	immediate
$(RYS)\int_a^b f\,dg \longrightarrow (CY)\int_a^b f\,dg$	Theorem 6.20
$f\in\mathcal{W}_p,\ g\in\mathcal{W}_q,\ \frac{1}{p}+\frac{1}{q}=1,\ 1<p<\infty$ does not imply any of the above integrals exists	Theorem 4.29

integral seems preferable among the integrals defined here. We are concerned with f and g regulated. Let's compare integrals in that case.

The (RYS) integral, while "stronger", is virtually the same as the $\fint$ integral. In $\fint_a^b f\,dg$, g is assumed right–continuous. The existence and value of $(RYS)\int_a^b f\,dg$ only depend on $g(x-)$ and $g(x+)$ for $a<x<b$, not on $g(x)$ if it happens to differ from both limits. Thus one can add to g a function in $c_0((a,b)) := \{\sum c_n 1_{\{x_n\}} \colon a<x_n<b,\ c_n\to 0 \text{ as } n\to\infty\}$, without changing the value of the integral. Such an addition can cause $(HK)\int_a^b f\,dg$ to be undefined (Example 2.1 in Schwabik [100], attributed to Vrkoč, and Proposition F.6 of [28]), but this in itself does not show any real superiority of the (RYS) over the (HK) integral. Likewise, in our examples where $(CY)\int_a^b f\,dg$ exists but $(RYS)\int_a^b f\,dg$ does not (Proposition 6.21), $f\in c_0((a,b))$, so it is not clear whether the (CY) integral has any real advantage. Its definition as an integral plus a sum is relatively complex.

We are then left with the comparison between the $\fint$ and (HK) integrals. The latter is more inclusive, but in what directions? For $f \geq 0$ and g non–decreasing the (HK) integral is equivalent to the Lebesgue–Stieltjes integral and thus, if f is regulated, to the $\nmid$ integral.

For a set of examples of functions f, g with $f \in \mathcal{W}_p$, $g \in \mathcal{W}_q$, $p^{-1} + q^{-1} = 1$, $1 < p, q < \infty$, all the integrals (including the (HK)) are undefined (Theorem 4.29). For $p^{-1} + q^{-1} > 1$, $1 < p, q < \infty$, the $\nmid$ integral and thus all to the right of it in Figure 2.1 are defined. For almost all paths B_t of Brownian motion, $\int_0^1 B_t \, dB_t$ does not exist for any of the integrals in Figure 2.1 (Proposition 2.2 below).

Perhaps the best–known feature of Perron–type integrals, including the (HK) integral, which goes beyond other integrals, is the following: if f is a continuous function from an interval $[a, b]$ into $\mathbb{R}$, everywhere differentiable on (a, b) except possibly on a countable set, then $(HK) \int_a^x f'(t) \, dt$ exists and equals $f(x) - f(a)$ for $a \leq x \leq b$ (Theorem 6.33). If h is any regulated function on $[a, b]$, it is bounded and measurable and has an indefinite integral $f(x) := \int_a^x h(t) \, dt$. Then $f'(x) = h(x)$ at any x where h is continuous, and so, except on the at most countable set of jumps of h. Since h is continuous almost everywhere for Lebesgue measure and bounded, it satisfies a well known criterion for Riemann integrability (Riemann [94]). So $\int_a^x h(t) \, dt$ can be any of the integrals in Figure 2.1, and the (HK) integral does not go beyond the others for derivatives, if they are regulated. If f' exists everywhere on (a, b) and is regulated (having limits also at a, b), then clearly f' is continuous on (a, b) and extends to a continuous function on $[a, b]$.

Summing up, if f and g are regulated, as they are if f is in $\mathcal{W}_p$ and g is in $\mathcal{W}_q$, we do not know any examples where $(HK) \int_a^b f \, dg$ exists but $\nmid_a^b f \, dg$ does not. Since the definition of $\nmid_a^b f \, dg$ is simpler, we prefer it for regulated functions, at least for the present.

2.3 Stochastic integrals

We recall first a classical stochastic integral.

The Itô integral Let $B = \{B(t) \colon t \geq 0\}$ be a Brownian motion on a probability space $(\Omega, \mathcal{A}, \Pr)$. For $t \geq 0$, let $\mathcal{F}_t$ be the σ–algebra generated by the random variables $\{B(r) \colon 0 \leq r \leq t\}$. Then for $0 \leq t < r < \infty$ we have $\{\emptyset, \Omega\} := \mathcal{F}_0 \subseteq \mathcal{F}_t \subseteq \mathcal{F}_r \subseteq \mathcal{A}$. For $0 < T < \infty$ consider the class $\mathcal{H}[0, T]$ of stochastic processes $H = \{H(t) \colon 0 \leq t \leq T\}$ satisfying the following properties:

(1) $(t, \omega) \mapsto H(t, \omega)$ is $\mathcal{B} \times \mathcal{A}$–measurable, where $\mathcal{B}$ is the Borel σ–algebra on $[0, T]$;

(2) H is $\mathcal{F}_t$–adapted, i.e. $H(t, \cdot)$ is $\mathcal{F}_t$–measurable for all $0 \leq t \leq T$;

(3) $E[\int_0^T H^2(t) \, dt] < \infty$.

We now define the Itô integral $(I) \int_0^T H \, dB$ for each $H \in \mathcal{H}([0, T])$. First suppose that H is elementary, that is

$$H(t, \omega) = H_0(\omega) 1_{\{0\}}(t) + \sum_{j=1}^n H_j(\omega) 1_{(t_{j-1}, t_j]}(t)$$

for some $0 = t_0 < t_1 < \cdots < t_n = T$, where H_0 is $\mathcal{F}_0$–measurable, each H_j, $j = 1, \ldots, n$, is $\mathcal{F}_{t_{j-1}}$–measurable and $EH_j^2 < \infty$. For such H define the Itô integral by

$$(I) \int_0^T H \, dB := \sum_{j=1}^n H_j [B(t_j) - B(t_{j-1})].$$

For an arbitrary $H \in \mathcal{H}[0,T]$, there exists a sequence $\{H_m \colon m \geq 1\}$ of elementary processes H_m such that

$$\lim_{m\to\infty} E\Big[\int_0^T (H(t) - H_m(t))^2\, dt\Big] = 0.$$

Then the Itô integral for H is defined as the limit

$$(I)\int_0^T H\, dB := \lim_{m\to\infty} (I)\int_0^T H_m\, dB \quad \text{in } L_2(\Omega, \mathrm{Pr}),$$

which does not depend on the actual sequence $\{H_m \colon m \geq 1\}$ chosen. The proof of these statements follows from the Itô isometry

$$E\Big[(I)\int_0^T H\, dB\Big]^2 = E\Big[\int_0^T H^2(t)\, dt\Big]$$

valid for each $H \in \mathcal{H}[0,T]$. This construction can be extended in several ways (see e.g. [14, Chapter VIII]). Condition (3) can be relaxed to $\int_0^T H^2(t)\, dt < \infty$ almost surely. Then $(I)\int_0^T H\, dB$ need no longer be square–integrable but can be an arbitrary $\mathcal{F}_T$–measurable random variable (see [21]). Researchers in stochastic analysis have some decades ago defined $\int H(t)\, dX_t$ rather generally, specifically when X_t is a semimartingale (for example, a right-continuous martingale) and $H(t)$ is a locally bounded predictable process (for example, an adapted process with regulated, left-continuous paths): see Doléans–Dade and Meyer [16], Dellacherie and Meyer [14].

To what extent can stochastic integrals be replaced or supplemented by integrals defined pathwise, for fixed ω?

The Love–Young inequality for $f \in \mathcal{W}_p$ and $g \in \mathcal{W}_q$ with $p^{-1} + q^{-1} > 1$, and the Young [118, Section 5] inequality for $f \in \mathcal{W}_\Phi$ and $g \in \mathcal{W}_\Psi$ with $\sum_{n=1}^\infty \Phi^{-1}(n^{-1})\Psi^{-1}(n^{-1}) < \infty$, give conditions for existence of $\fint_a^b f\, dg$ and bounds for it. Since many stochastic processes on intervals have sample paths in $\mathcal{W}_q$ for some q (see Chapter 5), one can then treat integrands f and integrators g in suitable $\mathcal{W}_p$, $\mathcal{W}_q$.

However, we have:

Proposition 2.2. *For almost all paths $B(t) := B_t$ of standard Brownian motion, $(A)\int_0^1 B_t\, dB_t$ is not defined for $(A) = (MRS)$, (RRS), (LS), (RYS), (CY) or (HK).*

Proof. It suffices to consider (CY) and (HK) by the implications shown in Figure 2.1. If the (CY) integral existed it would equal the (MRS) integral by sample continuity and the definition of the integral, so $\fint_0^1 B_t\, dB_t$ would exist. Brownian paths almost surely have unbounded 2-variation on every interval $[a, b]$, $a < b$ (Lévy [71, Théorème 9], Taylor [106]). By continuity, there are arbitrarily large sums

$$\sum_{j=1}^n [B(t_j) - B(t_{j-1})][B(v_j) - B(u_j)]$$

with $a = t_0 < \cdots < t_{j-1} < u_j < v_j < t_j < \cdots < t_n = b$. Thus the sums $\sum_{j=1}^n [B(t_j) - B(t_{j-1})]B(w_j)$ with $t_{j-1} < w_j < t_j$ for each j are unbounded. Since these are Young sums, $\fint_0^1 B_t\, dB_t$ doesn't exist, nor does $(HK)\int_0^1 B_t\, dB_t$ by Proposition F.3 of Dudley and Norvaiša [28]. $\square$

Here is a sense in which at least the particular integral $(I)\int_0^1 B_t\,dB_t$ can be done pathwise. It is a special case of the construction suggested by Föllmer [34] for more general processes. Let $\{\lambda(m)\colon m \geq 1\}$ be a sequence of partitions $\lambda(m) = \{0 = t_0^m < t_1^m < \cdots < t_m^m = 1\}$ of $[0,1]$. Since $B_0 = 0$, for every $m \geq 1$, we have

$$S_L(\lambda(m)) := \sum_{i=1}^m B(t_{i-1}^m)[B(t_i^m) - B(t_{i-1}^m)] = \frac{1}{2}[B(1)^2 - s_2(B;\lambda(m))] \tag{2.5}$$

and

$$\sum_{i=1}^m B(t_i^m)[B(t_i^m) - B(t_{i-1}^m)] = \frac{1}{2}[B(1)^2 + s_2(B;\lambda(m))]. \tag{2.6}$$

The right sides of (2.5) and (2.6) have limits provided the limit $\lim_{m\to\infty} s_2(B;\lambda(m))$, called the quadratic variation, exists. It is so almost surely in the two cases:

(a) the sequence $\{\lambda(m)\colon m \geq 1\}$ is nested and $\cup_m \lambda(m)$ is dense in $[0,1]$ (Lévy, [71, Section 4, Théorème 5]) or

(b) the mesh of $\lambda(m)$ is of order $o(1/\log m)$ as $m \to \infty$ (Dudley, [20, Theorem 4.5]), and this condition is sharp (Fernández de la Vega, [32]).

If (a) or (b) holds then $\lim_{m\to\infty} s_2(B;\lambda(m)) = 1$ almost surely. From the definition of Itô integral, if $S_L(\lambda(m))$ converge almost surely, they converge to $(I)\int_0^1 B_t\,dB_t$, since they converge to it in L^2. One can take $\lambda(m) := \lambda_m = \{j/m\}_{j=0}^m$ by (b) or its subsequence $\lambda(2^m)$ by (a) or (b). On the other hand, for any partition λ, the supremum of $s_2(B;\kappa)$ over refinements κ of λ is $+\infty$ almost surely, so $\int_0^1 B_t\,dB_t$ is not a left Cauchy integral, meaning that Riemann-Stieltjes sums with $y_j = x_{j-1}$ do not converge as the mesh goes to 0 or as partitions are refined. Also, to deal with jumps of a process X_t, in defining an integral $\int_0^1 X_t\,dX_t$, jump points would need to be included in $\cup_m \lambda(m)$.

Recall that "cadlag" means regulated and right–continuous. Dellacherie and Meyer [14, Chapter VIII, Section 15] consider a rather general stochastic integral

$$S(t) := \int_0^t U_{s-}\,dX_s, \tag{2.7}$$

where U and X are cadlag processes and X is a semimartingale. They show that the integral is a limit in probability of left Cauchy sums

$$S_n(t) := \sum_{j=1}^{2^n} U((j-1)t/2^n)[X(jt/2^n) - X((j-1)t/2^n)].$$

Taking a subsequence if necessary, $S_{n_k}(t)$ will converge almost surely to $S(t)$ for all rational $t > 0$. By a different choice of partitions, letting $[x]$ be the largest integer $\leq x$, define

$$T_n(t) := U([2^n t]/2^n)[X(t) - X([2^n t]/2^n)] + \sum_{i=1}^{[2^n t]} U((i-1)/2^n)[X(i/2^n) - X((i-1)/2^n)].$$

Then for some subsequence, $T_{n_k}(t) \to S(t)$ almost surely for all $t \geq 0$. In this sense the stochastic integral can be done pathwise and, more generally, stochastic differential equations can be solved constructively – we claim no expertise in this direction, but give as references Bichteler [6], Jacod and Protter [58].

We will just mention that there are also ways to define stochastic integrals with anticipating integrands: Skorohod [103].

We would also suggest that integrals $\int Y_t \, dX_t$ can be defined pathwise via the Love–Young inequality if the functions $t \mapsto Y_t$ have bounded p–variation of order $p < 2$, or for processes X_t that have bounded q–variation, $q < 2$, as do e.g. suitable stable processes of order $\alpha < 2$ and some Gaussian processes such as fractional Brownian motion with the Hurst index $1/2 < H < 1$ (see sections 5.3 and 5.4). On the other hand, when the Love–Young inequality does apply, it does not require any probabilistic conditions such as (semi)martingale properties of X_t or predictability of Y_t. Still, when we come to consider integral equations and the product integral (Chapter 7) we will in fact use left–continuous integrands.

Paul Lévy [72] took note of L. C. Young's inequalities [117], [118] and the possibility of applying them to "Stieltjes–Young" integrals $\int Y_t \, dX_t$ for stochastic processes X_t, Y_t.

Non-random integrands and the Wiener process Historically, after Wiener (1923) defined the Brownian motion process B_t and showed its sample continuity, Paley, Wiener and Zygmund (1933) defined a stochastic integral $L(f) = \int_0^1 f(t) \, dB_t(\omega)$ for any non-random $f \in H = L^2[0,1]$ by an extension process. It turns out that L is the isonormal process on the Hilbert space H, a Gaussian process with mean 0 and covariance $E(L(x)L(y)) = (x,y)$, the inner product in the Hilbert space. Any Gaussian process has its finite-dimensional joint distributions completely determined by its means and covariances. Thus the isonormal process is determined by the abstract geometric Hilbert space structure, which has a great deal of rotational invariance. For example, a unitary linear transformation of $L^2[0,1]$ onto itself can take any element with norm 1 (say, a very smooth function) into any other such element (which may be quite non-smooth).

The isonormal process is a linear isometry, preserving inner products, of H into L^2 of a probability space (Ω, P). If one takes $L^2(\Omega, P)$ in turn as the Hilbert space and defines an isonormal process M on it, then $M \circ L$ has all the properties of L, i. e. it is also an isonormal process on H. Moreover, if $(X_t, t \in T)$ is any Gaussian process with mean 0 on any parameter space T, then $t \mapsto M(X_t)$ is another version of the process X_t, with the same covariances and joint distributions. Thus sample continuity and boundedness of Gaussian processes can be studied by way of the isonormal process, in other words, by putting the "intrinsic metric" $d(s,t) = [E(X_s - X_t)^2]^{1/2}$ on T. In fact, if (T,d) is a compact metric space and $t \mapsto X_t$ is continuous in probability, then X_t has continuous paths on T if and only if M does on $\{X_t(\cdot) \colon t \in T\} \subset L^2(\Omega, P)$ (Dudley [26, Theorem 2.8.2]). The focus on the isonormal process, as "the" Gaussian process, advocated by one of us (Dudley [19], [20]) has been well accepted, at least by use of the intrinsic metric (e. g. Talagrand [105]). One can first define the isonormal process L, specifically on $L^2[0,\infty)$, then define Brownian motion simply by $B_t = L(1_{[0,t]})$, e. g. Dudley [22, p. 352]. Thus it is unnecessary to define a stochastic integral $\int f(t) \, dB_t$ per se for non-random $f \in L^2$.

Chapter 3

Differentiability of some nonlinear operators

Let T be a function defined on an open set U in a Banach space X with norm $\|\cdot\|$. Let T take values in Y, where $(Y, |\cdot|)$ is a Banach space. Then T is *Fréchet differentiable* at a point $x \in U$ if there is a bounded linear operator D from X into Y such that

$$|T(x+u) - T(x) - D(u)| = o(\|u\|) \qquad \text{as } \|u\| \to 0. \tag{3.1}$$

T will be said to have a *remainder bound* of order $\gamma > 1$ if $o(\|u\|)$ in (3.1) can be replaced by $O(\|u\|^\gamma)$. For a smooth function T we will have $\gamma = 2$, and no larger value is possible except in special cases where second derivatives are 0.

3.1 Bilinear operators

If $(X, \|\cdot\|)$, $(Y, |\cdot|)$ and $(Z, \|\cdot\|)$ are normed spaces, a function B from $X \times Z$ into Y is called *bilinear* if $B(\cdot, z)$ is linear on X for each $z \in Z$ and $B(x, \cdot)$ is linear on Z for each $x \in X$. B is called *bounded* if for some $C < \infty$,

$$|B(x,z)| \le C\|x\|\,\|z\| \qquad \text{for all } x \in X \text{ and } z \in Z.$$

Here $X \times Z$ is a normed space with the norm $\|(x,z)\| := \|x\| + \|z\|$. A bilinear function B is continuous: $X \times Z \mapsto Y$ if and only if it is bounded.

For any $x, u \in X$ and $z, w \in Z$, we have

$$B(x+u, z+w) = B(x,z) + B(u,z) + B(x,w) + B(u,w).$$

If B is bounded, then for fixed x and z,

$$(u,w) \mapsto B(u,z) + B(x,w)$$

is clearly bounded and linear. Since $|B(u,w)| \le C(\|u\| + \|w\|)^2$, B is Fréchet differentiable at (x,z) with remainder bound of order $\gamma = 2$. Indeed B is holomorphic and a polynomial of order 2. Thus for a bilinear operator, boundedness is necessary for continuity but sufficient for holomorphy. So smoothness of B is rather an "all–or–nothing" issue and one will hope to find norms for which "all" smoothness holds.

For a measure μ, the bilinear functional

$$L^p(\mu) \times L^q(\mu) \ni (f, g) \mapsto \int fg\, d\mu \in \mathbb{R}$$

is bounded by the Hölder inequality if $1 \le p \le \infty$, $p^{-1} + q^{-1} = 1$.

The bilinear functional

$$\mathcal{W}_p[a,b] \times \mathcal{W}_q[a,b] \ni (f, g) \mapsto \fint_a^b f\, dg \in \mathbb{R}$$

is bounded by the Love–Young inequality for $p^{-1} + q^{-1} > 1$, $1 \le p < \infty$, $1 \le q < \infty$, or for $(p, q) = (\infty, 1)$ or $(1, \infty)$ (see Theorem 4.26 and Proposition 4.27).

The bilinear indefinite integral operator

$$\mathcal{W}_p[a,b] \times \mathcal{W}_q[a,b] \ni (f, g) \mapsto \fint_a^{(\cdot)} f\, dg \in \mathcal{W}_q[a,b]$$

is bounded by Corollary 4.28 for the same (p, q) as for the bilinear functional above.

Let $F{*}G(x) := \fint_{-\infty}^{\infty} F(x-y)\, dG(y)$ for $x \in \mathbb{R}$. The bilinear convolution operator

$$\mathcal{W}_p(\mathbb{R}) \times \mathcal{W}_q(\mathbb{R}) \ni (F, G) \mapsto F{*}G \in \mathcal{W}_r(\mathbb{R})$$

is bounded for $p^{-1} + q^{-1} - r^{-1} = 1$, $1 \le p < \infty$, $1 \le q < \infty$, $1 \le r < \infty$, as shown by Gehring [39], see also Dudley and Norvaiša [28, Appendix A].

For any interval $J \subset \mathbb{R}$ (possibly unbounded) and $1 \le p \le \infty$ the multiplication operator

$$\mathcal{W}_p(J) \times \mathcal{W}_p(J) \ni (f, g) \mapsto fg \in \mathcal{W}_p(J)$$

is bounded as noted by Krabbe [66], [67], and $(\mathcal{W}_p(J), \|\cdot\|_{[p]})$ is a Banach algebra.

3.2 Compact differentiability

Statisticians have done fundamental, pioneering work on differentiability of non–bilinear operators whose arguments f are functions that may be non–differentiable and even discontinuous, e.g. Reeds [93], Fernholz [33], Gill [41], [42], Gill and Johansen [43], Andersen, Borgan, Gill and Keiding [1]. It seems that analysts had overlooked the possibility of differentiability for such f. The vehicle for most of the pioneering has been compact differentiability (in the supremum norm), where an operator T is not Fréchet differentiable but in place of (3.1),

$$|T(x + tu) - T(x) - tD(u)| = o(\|t\|) \qquad \text{as } t \to 0,$$

uniformly for u in any compact set. In further forms of differentiability, sometimes $x + tu$ is restricted to a (non–open) set A and/or there are other restrictions. Useful theoretical properties of compact differentiability in statistics have been shown, e.g. van der Vaart [108], van der Vaart and Wellner [109]. While honoring the pioneers, we note that one has not only a choice of modes of differentiability but a choice of norms. The supremum norm controls, after all, only the supremum of a function, not its oscillation. Thus it is not surprising that for several operators of interest, in the supremum norm, Fréchet differentiability and compact differentiability both fail. One possibility is to consider still weaker forms of differentiability such as compact differentiability under restriction to subsets A, for example, sets of uniformly bounded variation (that is, 1-variation; Gill [41], Gill and Johansen [43]). Non–normalized empirical processes $F_n - F$, where F is a probability distribution function and F_n an empirical distribution function for it, have 1-variation uniformly bounded (by 2), but other processes of interest do not. Also, use of both supremum ($\mathcal{W}_\infty$) and $\mathcal{W}_1$ norms suggests use of intermediate spaces $\mathcal{W}_p$. As we are indicating, Fréchet differentiability with respect to p-variation norms holds for many (though not all) of the cases previously studied for compact differentiability and the sup norm ([28]).

We next treat three non–bilinear operators: the two–function composition operator, the quantile operator and the product integral.

3.3 The two–function composition operator

For functions $g: X \mapsto Y$ and $f: Y \mapsto Z$, and any sets X, Y and Z, let $(f \circ g)(x) := f(g(x))$. Then $f \circ g$ is called the *composition* of f and g. To treat differentiability of the operator $(F, G) \mapsto F \circ G$, we will take Y and Z to be vector spaces. Clearly $F \mapsto F \circ G$ is linear in F for fixed G, but $G \mapsto F \circ G$ is not linear in G unless F is linear.

We consider the operator in the form $(f, g) \mapsto (F + f) \circ (G + g)$ as $f, g \to 0$ since for differentiability, the properties of the F, G at which one differentiates can differ, as will be seen, from those of f, g.

Let $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$, $(\mathcal{G}, \|\cdot\|_{\mathcal{G}})$ and $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ be normed spaces of functions from Y into Z, X into Y, and X into Z respectively. Our question then is whether

$$\mathcal{F} \times \mathcal{G} \ni (f, g) \mapsto (F + f) \circ (G + g) \in \mathcal{H} \quad \text{is Fréchet differentiable at } f = g = 0.$$

If it is, then $f \mapsto (F + f) \circ G$ must be Fréchet differentiable at $f = 0$ from $\mathcal{F}$ into $\mathcal{H}$. The derivative must be $f \mapsto f \circ G$, which is clearly linear, and must be a bounded operator from $\mathcal{F}$ into $\mathcal{H}$. Also, the operator $g \mapsto F \circ (G + g)$ must be Fréchet differentiable from $\mathcal{G}$ into $\mathcal{H}$. The more general operator $g \mapsto [x \mapsto F(x, (G + g)(x))]$, where F is a function of two variables, is sometimes called a *Nemitskii operator*. It has been much studied, see Appell and Zabrejko [2], Runst and Sickel [97]. Let $(X, \mathcal{S}, \mu)$ be a measure space. We have the spaces $L^p = L^p(X, \mathcal{S}, \mu)$ with their norms $\|\phi\|_p := \left(\int |\phi|^p \, d\mu\right)^{1/p}$, $1 \le p < \infty$. Let $Y = Z = \mathbb{R}$, let $1 \le p < s < \infty$ and let $\mathcal{G} := L^s(X, \mathcal{S}, \mu)$, $\mathcal{H} := L^p(X, \mathcal{S}, \mu)$. It turns out that if F is continuous and $G \in L^s$ then for differentiability of $g \mapsto F \circ (G + g)$, F must be differentiable $(\mu \circ G^{-1})$–almost everywhere and the derivative linear operator is $g \mapsto (F' \circ G) g$ with $F' \circ G \in \mathcal{L}^{sp/(s-p)}$ (Theorems 3.12 and 3.13 and the sentence after (2.56) in Appell and Zabrejko [2]). There will then be a remainder

$$R_2(g, G, F) := F \circ (G + g) - F \circ G - (F' \circ G) g.$$

Now for the two–function operator we will have

$$(F + f) \circ (G + g) - F \circ G - f \circ G - (F' \circ G) g = R_2(g, G, F) + R_1(f, G, g),$$

where $R_1(f, G, g) := f \circ (G + g) - f \circ G$. We will take $f \in \mathcal{F} := \mathcal{W}_p(\mathbb{R})$. Since our interest is in applying p-variation, while R_2 doesn't depend on f, we will focus on the remainder term $R_1(f, G, g)$. This does not involve F. For $1 \le p < \infty$, let $\gamma(p, s) := 1 + s/[p(1 + s)]$ for $1 \le s < \infty$ and $\gamma(p, \infty) := 1 + 1/p$. For R_1 we have:

Theorem 3.3. *Let $(X, \mathcal{S}, \mu)$ be a finite measure space and let G be a bounded measurable real–valued function on X. Let $b := \mathrm{ess.\,sup}\, G$ and $a := \mathrm{ess.\,inf}\, G$. Then the following are equivalent:*
(A) *there is a $C < \infty$ such that for some $p \in [1, \infty)$ and $s \in [1, \infty]$,*

$$\|R_1\|_p = \|f\circ(G+g) - f\circ G\|_p \le C\|f\|_{(p)}\|g\|_s^{\gamma(p,s)-1} \tag{3.2}$$

for $f \in \mathcal{W}_p(\mathbb{R})$ and $g \in L^s(X, \mathcal{S}, \mu)$;
(A′) (A) *holds for all $p \in [1, \infty)$ and $s \in [1, \infty]$;*
(B) *$a < b$ and there is a $K < \infty$ such that*

$$(\mu\circ G^{-1})((c, d]) \le K(d - c)$$

for $a \le c < d \le b$.

Remark. Proposition 3.5 will show that in (3.2) the power $\gamma(p, s) - 1$ on $\|g\|_s$, the power 1 on $\|f\|_{(p)}$ and the norm $\|\cdot\|_{(p)}$ itself are all optimal.

Proof. We can assume $\mu(X) = 1$, changing only the constants C and K.

(B) implies (A′). Fix $p \in [1, \infty)$ and $s \in [1, \infty]$. By definition of a and b, $\mu\circ G^{-1}$ is concentrated on $[a, b]$. Let $H(x) := (\mu\circ G^{-1})([a, x])$ for $a \le x \le b$. Then H is a non-decreasing function from $[a, b]$ into $[0, 1]$ with $H(b) = 1$ and $0 \le H(x)-H(u) \le K(x-u)$ for $a \le u \le x \le b$, a Lipschitz property. Let $H^{\leftarrow}(y) := \inf\{x\colon H(x) \ge y\}$. Then $H^{\leftarrow}$ is strictly increasing from $[0, 1]$ into $[a, b]$. Since H is right–continuous, $H^{\leftarrow}(y) \le x$ if and only if $H(x) \ge y$. Thus for $\lambda :=$ Lebesgue measure on $[0, 1]$, and $a \le x \le b$, $\lambda\circ(H^{\leftarrow})^{-1}([a, x]) = H(x) = (\mu\circ G^{-1})([a, x])$. Thus $\lambda\circ(H^{\leftarrow})^{-1} = \mu\circ G^{-1}$ on the Borel sets of $[a, b]$. We have $H^{\leftarrow}(u) - H^{\leftarrow}(v) \ge (u - v)/K$ for $0 \le v \le u \le 1$.

Take the unit square $I^2 := I \times I$ with $I := [0, 1]$ and Lebesgue measure λ^2 on I^2. For any $g \in \mathcal{L}^s(X, \mathcal{S}, \mu)$ there is a measurable function γ on I^2 such that $(H^{\leftarrow}(t), \gamma(t, u))$ have the same joint distribution for λ^2 as (G, g) for μ (Skorohod [102, proof of Theorem 1]; Dudley and Philipp [30, Lemma 2.11]). For $s < \infty$, by Theorem 2.2 of Dudley [24] there is a constant $C < \infty$ depending only on K (and $\mu(X)$) such that for each $u \in I$,

$$\int |f\circ(H^{\leftarrow} + \gamma) - f\circ H^{\leftarrow}|^p(t, u)\, dt \le C^p v_p(f)\Big(\int_0^1 |\gamma(t, u)|^s\, dt\Big)^{1/(1+s)}. \tag{3.3}$$

The statement and proof in Dudley [24] would give $\|f\|_{[p]}^p$ rather than $\|f\|_{(p)}^p = v_p(f)$, but note that we can subtract a constant from f to get $\sup f = -\inf f$ and then $\|f\|_\infty \le \|f\|_{(p)}/2$. Integrating with respect to u, we have

$$\int_0^1 \Big(\int_0^1 |\gamma(t, u)|^s\, dt\Big)^{1/(1+s)} du \le \Big(\int_0^1 \int_0^1 |\gamma(t, u)|^s\, dt\, du\Big)^{1/(1+s)}$$

by Jensen's inequality. Taking pth roots gives

$$\begin{aligned}\|f\circ(G + g) - f\circ G\|_p &= \|f\circ(H^{\leftarrow} + \gamma) - f\circ H^{\leftarrow}\|_p \\ &\le C\|f\|_{(p)}\|\gamma\|_s^{s/(p(1+s))} = C\|f\|_{(p)}\|g\|_s^{s/(p(1+s))}.\end{aligned}$$

For $s = \infty$ we apply instead Theorem 2.5 of Dudley and Norvaiša [28]. So (A′) is proved.

Clearly (A′) implies (A). Lastly, to show (A) implies (B), assume (A). If $a = b$, then $G \equiv a$. Let $f(x) := 1$ for $x \geq a$ and 0 for $x < a$. Let $g := g_n \equiv -1/n$. Then $\|f\|_{(p)} = 1$ and $\|f\circ(G+g_n) - f\circ G\|_p \equiv 1$, but $\|g_n\|_s \to 0$, contradicting (A). So $a < b$. If (B) fails, then for $n = 1, 2, \ldots$, there exist c_n, d_n with $a \leq c_n < d_n \leq b$ and $q_n := (\mu\circ G^{-1})((c_n, d_n]) > n(d_n - c_n)$. Let $g_n := 2(c_n - d_n)$ on $G^{-1}((c_n, d_n])$, $g_n := 0$ elswhere, and let $f_n := 1_{[c_n,\infty)}$. Then for all n, $\|f_n\|_{(p)} = 1$, $\|f_n\circ(G+g_n) - f_n\circ G\|_p = q_n^{1/p}$. For $s < \infty$, $\|g_n\|_s = 2(d_n - c_n)q_n^{1/s}$. Letting $\delta_n := d_n - c_n > 0$ gives $q_n > n\delta_n$ while by (A), $q_n^{1/p} \leq C(2\delta_n)^{s/(p(1+s))} q_n^{1/(p(1+s))}$, giving $q_n^{s/(1+s)} \leq C(2\delta_n)^{s/(1+s)}$, $q_n \leq 2C^{(s+1)/s}\delta_n$, a contradiction. For $s = \infty$, $q_n^{1/p} \leq C \cdot 1 \cdot (2\delta_n)^{1/p}$ so $q_n \leq 2C^p\delta_n$ while again $q_n > n\delta_n$, a contradiction. So (B) holds. □

The bound (3.2) implies the bound

$$\|f\circ(G+g) - f\circ G\|_p \leq C(\|f\|_{[p]}^{\gamma} + \|g\|_s^{\gamma})$$

for $\gamma := \gamma(p, s)$.

If $\mu\circ G^{-1}$ has atoms, the above proof shows that $\|f\circ(G+g) - f\circ G\|_p$ need not even approach 0 as $\|g\|_\infty \to 0$ for fixed $f \in \mathcal{W}_p$. Consider the following

Example 3.4. Let $G(x) = x$, $0 \leq x \leq 1$. Let $f_1(x) := 1_{\{x \leq 1/2\}}$. Let $J(n)$ be the interval $[1/2, 1/2 + 1/n)$, and $g_n := -n^{-1}1_{J(n)}$, $n = 2, 3, \ldots$. Then for $1 \leq p < \infty$, $|f_1\circ(G+g_n) - f_1\circ G|_p = n^{-1/p}$ and $\|f_1\|_{(p)} = 1$. Also $\|g_n\|_s = n^{-(1+s)/s}$ for $1 \leq s < \infty$, $\|g_n\|_\infty = 1/n$ and $\|g_n\|_{[r]} = 2^{1/r}/n$, $1 \leq r < \infty$. We can also take multiples $f = tf_1$ with $t \to 0$.

The example gives the following

Proposition 3.5. *If $(Y, \|\cdot\|)$ is a normed space of functions on $[0,1]$ containing f_1 of Example 3.4, and for some $C < \infty$*

$$\|f\circ(G+g_n) - f\circ G\| \leq C\|f\|^{\lambda}\|g_n\|_s^{\alpha}$$

for f, G, and g_n as in the example, then $\lambda \leq 1$ and $\alpha \leq s/(p(1+s))$ for $1 \leq s < \infty$, $\alpha \leq 1/p$ for $s = \infty$. Thus the powers $\lambda = 1$ in Theorem 3.3 (A) and $\alpha = s/(p(1+s))$ for $s < \infty$, $\alpha = 1/p$ for $s = \infty$, are separately optimal, and no norm on f allows a better exponent than $\|\cdot\|_{(p)}$ does.

On the other hand the exponents for $s = \infty$ are not improved if we replace $\|\cdot\|_\infty$ by a stronger r-variation norm $\|g\|_{[r]}$, $1 \leq r < \infty$.

3.4 The quantile (inverse) operator

Let J be an interval in $\mathbb{R}$ and F a function from J into $\mathbb{R}$. For any $y \in \mathbb{R}$, let $F^{\leftarrow}(y) := F_J^{\leftarrow}(y) := \inf\{x \in J\colon F(x) \geq y\}$, or the right endpoint of J (which may be $+\infty$) if there is no such x. $F^{\leftarrow}$ is called the *quantile function* of F. We use the notation $F^{\leftarrow}$ when F is not necessarily 1-1, while F^{-1} is reserved for the inverse image of sets, or a point function in case F is 1-1.

For any function F from J into $\mathbb{R}$ let $G_F(x) := \sup_{y \leq x,\ y \in J} F(y)$. Then G_F is a nondecreasing function and clearly $G_F^{\leftarrow} \equiv F^{\leftarrow}$ except possibly at some of at most countably many jumps of G_F. For differentiability of $c \mapsto (G_F + c)^{\leftarrow}$ into L^p for $p > 1$ at $c = 0$ along constant functions, G_F must be strictly increasing. In fact we have the

Proposition 3.6. *Let $1 < p < \infty$ and $q := p' := p/(p-1)$. Let F be a real–valued function from an interval $J \subset \mathbb{R}$ into an interval $[a,b]$, $-\infty < a < b < \infty$. Suppose that $c \mapsto (F+c)^{\leftarrow}$ is Fréchet differentiable at $c = 0$ along constant functions c into $L^p[a,b]$. Then for some $\delta > 0$, $G_F(x) - G_F(u) > \delta(x-u)^q$ for any $u < x$ in J.*

Proof. Suppose that for some $u_n < x_n$ in J, and $n = 1, 2, \ldots$, $G_F(x_n) - G_F(u_n) \le (x_n - u_n)1/n$. Then for $c > 0$,

$$(F+c)^{\leftarrow}(G_F(u_n)+v) \le u_n \qquad \text{for } v < c,$$

and

$$(F-c)^{\leftarrow}(G_G(u_n)+v) \ge x_n \qquad \text{for } v > (x_n-u_n)^q n^{-1} - c.$$

Taking $c_n := (x_n - u_n)^q/n$ we get

$$\|(F+c_n)^{\leftarrow} - (F-c_n)^{\leftarrow}\|_p \ge (x_n - u_n)c_n^{1/p} = n^{1/q}c_n,$$

contradicting differentiability. $\square$

Regarding regularity conditions on F in the following theorem note that F can be chosen quite regular (the identity function) in some applications to probability and statistics, see Theorem 3.11 below. Dudley [24, Corollary 2.4], gives the following:

Theorem 3.7. *Let $1 \le p < \infty$ and let F be a continuous function from an interval $[a,b]$ onto an interval $[c,d]$. Suppose that F is differentiable on (a,b) with for some $\kappa > 0$, $F'(x) \ge \kappa$ for $a < x < b$, and such that F' satisfies the Hölder condition $\sup_{a<x<u<b} |F'(u) - F'(x)|/(u-x)^{1/p} < \infty$. Then $f \mapsto (F+f)^{\leftarrow}$ is Fréchet differentiable at $f = 0$ from $\mathcal{W}_p[a,b]$ into $L^p[c,d]$ for Lebesgue measure, with derivative*

$$f \mapsto -(f \circ F)/(F' \circ F^{-1})$$

and such that the remainder

$$R_f := (F+f)^{\leftarrow} - F^{\leftarrow} + (f \circ F')/(F' \circ F^{-1}). \tag{3.4}$$

satisfies, for some constant $K_p < \infty$,

$$\|R_f\|_p \le K_p \|f\|_{[p]}^{(p+1)/p} \qquad \text{for all } f \in \mathcal{W}_p[a,b] \text{ with } \|f\|_{[p]} \le 1. \tag{3.5}$$

The remainder bound $O(\|f\|_{[p]}^{(p+1)/p})$ in (3.5) cannot be replaced by $o(\|f\|^{(p+1)/p})$, for the p-variation norm or any other norm $\|\cdot\|$ on a space containing indicators $1_{[u,v]}$ of intervals, as the following easily checked example (Dudley [24, Proposition 2.5, with $y := Y$]) shows:

Proposition 3.8. *Let $1 \le p < \infty$, $[a,b] = [c,d] = [0,1]$ and $F(x) = x$, $0 \le x \le 1$. Let $h = 1_{[u,v]}$ for any fixed $0 < u < v < 1$. Then there is a constant $C_p > 0$ such that for $0 < t < \min(v-u, 1-v)$, we have $\|R_{th}\|_p = C_p t^{(p+1)/p}$.*

Here is a fact about combining the composition and quantile operators.

Proposition 3.9. (a) *For any two functions F, G from $\mathbb{R}$ into $\mathbb{R}$, and any $y \in \mathbb{R}$,*

$$(F \circ G)^{\leftarrow}(y) \ge (G^{\leftarrow} \circ F^{\leftarrow})(y).$$

(b) *If F is non-decreasing and right-continuous, then for any G,*

$$(F \circ G)^{\leftarrow} \equiv G^{\leftarrow} \circ F^{\leftarrow}. \tag{3.6}$$

Proof. For (a), if $F(G(x)) \geq y$, then $G(x) \geq F^{\leftarrow}(y)$, so $x \geq G^{\leftarrow}(F^{\leftarrow}(y))$ and (a) follows. For (b), we have $F(x_n) \geq y$ for some $x_n \downarrow F^{\leftarrow}(y)$, so since F is right-continuous, $F(F^{\leftarrow}(y)) \geq y$. If $G(x) \geq F^{\leftarrow}(y)$, then since F is non-decreasing, $F(G(x)) \geq y$ and with (a), now (b) follows. □

Examples. If $F(x) \equiv G(x) \equiv -x$ then $(F\circ G)(x) \equiv x$ so $(F\circ G)^{\leftarrow}(y) = y$ for all y, but $F^{\leftarrow}(y) = G^{\leftarrow}(y) = -\infty$ for all y.

Let $F := 1_{(0,\infty)}$ and $G := 1_{[1,\infty)}$. Then F and G are non-decreasing but only G is right-continuous. For $0 < y \leq 1$, we have $(F\circ G)^{\leftarrow}(y) = 1$ while $G^{\leftarrow}(F^{\leftarrow}(y)) = G^{\leftarrow}(0) = -\infty$. Thus the right-continuity of F in Proposition 3.9 cannot be dispensed with.

We also have the following right and left inverse properties, whose proofs are straightforward; (a) is stated e.g. by M. Csörgő [12, p. 1].

Proposition 3.10. (a) *Let F be a non-decreasing, continuous function on a possibly unbounded interval $J \subset \mathbb{R}$. Then for all y in the range of F, $F(F_J^{\leftarrow}(y)) = y$ where if u is the left endpoint of J and $-\infty \leq u \notin J$, we let $F(u) := \lim_{v\downarrow u} F(v)$.*
(b) *If F is strictly increasing: $J \mapsto \mathbb{R}$ then $F^{\leftarrow}(F(x)) = x$ for all $x \in J$.*

For the quantile operator, as shown in Dudley [24, before Theorem 2.6], we get via p-variation, and letting $p \downarrow 2$, for F smooth enough, that the remainder in differentiating $f \mapsto (F+f)^{\leftarrow}$ at $f = F_n - F$ is (in L^2 norm) of order $O_p(n^{\epsilon-3/4})$ for any $\epsilon > 0$, where $-3/4$ is the Bahadur-Kiefer correct exponent. It can be shown by other methods that the ϵ is unnecessary (Dudley [24, Theorem 2.6]). Kiefer [61] shows that for $f := f_n := F_n - F$, for the supremum norm, $\|R_f\|_\infty$ is of exact order $(\log n)^{1/2}/n^{3/4}$ in probability and $(\log n)^{1/2}(\log\log n)^{1/4}/n^{3/4}$ almost surely.

For the composition operator, statisticians' interest has been not so much in $F_n\circ G_m$ but rather in the so-called procentile-procentile or P-P plot $F_n\circ G_m^{\leftarrow}$: $(0,1) \mapsto [0,1]$, where F_n, G_m are empirical distribution functions for F, G, respectively, see e.g. Beirlant and Deheuvels [5]. Let U_m and V_n be empirical distribution functions for the $U[0,1]$ distribution function U. Then we can write $F_m \equiv U_m \circ F$ and $G_n \equiv V_n \circ G$. The following is known:

Theorem 3.11. *If $F = G$ is continuous, then the P-P process $F_m\circ G_n^{\leftarrow}$ for $0 < y < 1$ has the same distribution as when $F = G = U$ for $0 \leq x \leq 1$.*

Proof. By Propositions 3.9 and 3.10, we can write

$$F_m\circ G_n^{\leftarrow} \equiv U_m\circ F\circ(G^{\leftarrow}\circ V_n^{\leftarrow}) \equiv U_m\circ V_n^{\leftarrow},$$

and the conclusion follows. □

So, for $F = G$ continuous, only the case $F = G = U$ needs to be studied: Reeds [93], Fernholz [33].

3.5 The product integral

Let f be a function from an interval $[a,b]$ into a Banach algebra $\mathbb{B}$ with identity $\mathbb{I}$, for example the algebra M_k of real $k \times k$ matrices. For any partition $\kappa := \{x_i\}_{i=0}^n \in \mathrm{PP}([a,b])$ we have a product

$$P(f;\kappa) := \prod_{i=1}^n (\mathbb{I} + f(x_i) - f(x_{i-1})),$$

where $\prod_{i=1}^{n} a_i := a_n \cdots a_2 a_1$. If as partitions are refined, the products converge to a limit, it is called the product integral $\prod_a^b (\mathbb{I} + df)$. In Chapter 7 it will be seen how the product integral solves some differential and integral equations. We can define p-variation of $\mathbb{B}$–valued functions by putting norms in place of absolute values. It turns out that $f \mapsto \prod_a^b (\mathbb{I} + df) \in \mathbb{B}$ is defined for f in the space $\mathcal{W}_p([a,b];\mathbb{B})$ of $\mathbb{B}$–valued functions of bounded p-variation if $p < 2$ (Freedman, [35], for f continuous; in general [29]) and holomorphic into the same space ([29]). Some inequalities of T. Lyons [76] allow a proof that the product integral operator $f \mapsto [x \mapsto \prod_a^x (\mathbb{I} + df)]$ has a Taylor series with infinite radius of uniform convergence from $\mathcal{W}_p$ to $\mathcal{W}_p$, $p < 2$ (Theorem 7.18 below). For a real–valued function to have a non–zero product integral it is necessary, but not sufficient, that $f \in \mathcal{W}_2$ (Theorem 7.6). Another form of product integral, $f \mapsto \Pi_{[a,\cdot]} e^{df}$, is holomorphic from the supremum norm to itself for regulated functions into any *commutative* Banach algebra (Theorem 7.21). But on M_2 (the 2×2 matrices) neither $f \mapsto \prod_0^1 (\mathbb{I} + df)$ nor $f \mapsto \Pi_{[0,1]} e^{df}$ is continuous from $\mathcal{W}_p([0,1];M_2)$ to M_2 for $p > 2$ (Proposition 7.22). Thus, p-variation conditions around $p = 2$ are critical for existence of product integrals and continuity and differentiability of product integral operators. Lyons [76] applied his inequalities to solution of non–linear ordinary differential equations by Picard iteration, also using p-variation for $p < 2$.

Chapter 4

The p-variation

The discovery by Weierstrass and by Cellérier of curves without tangents marks indeed an epoch in the history of Mathematics. We of the twentieth century are bound to recognise it in its full importance. It is not only as Du Bois Reymond and a few thinkers of the 19th century pointed out, a remarkable addition to the philosophy of Mathematics. These curves afford us a means of rendering more veracious the representation of the physical universe by the realm of Mathematics. I cannot but think that these curves will serve as the basis of the geometrical theory of molecular phenomena, in the same sense as the conic sections have served as a first approximation to the movements of the planets.

Nevertheless the apparent trajectory of an element in the ultramicroscope is not a curve without tangents. We cannot follow all the details; a mental effort is required to coordinate the impressions made on the retina by the sparking molecules as they pass. But neither is the path of a planet really an ellipse.

At the present moment the reduction of the trajectory of an atom to a question of analytical geometry is not possible. But who knows whether there will not arise a Newton to illuminate our subject? Let us then patiently attack the study of these remarkable curves, leaving to others the work of making the discoveries which may lead to the application of our results in the world of atoms.

G. C. Young. "On infinite derivatives". An Essay[1] [116, p. 140].

4.1 Elementary properties

If $f \in \mathcal{W}_p[a,b]$ for $0 < p < 1$ then f is a pure jump function as shown in Proposition 4.14. Thus we will concentrate on the range $1 \le p \le \infty$. For such p recall that $\|f\|_{(p)} := v_p(f)^{1/p}$, $\|f\|_\infty := \sup|f|$ and $\|f\|_{[p]} := \|f\|_{(p)} + \|f\|_\infty$.

The next two facts are easy to prove.

Lemma 4.1. *For $p \ge 1$, $\|\cdot\|_{(p)}$ is a seminorm on $\mathcal{W}_p$, 0 only on constants, and $\|\cdot\|_{[p]}$ is a norm.*

Theorem 4.2. *For $p \ge 1$, $(\mathcal{W}_p, \|\cdot\|_{[p]})$ is complete, thus a Banach space.*

Lemma 4.3. *For $1 \le p < \infty$, $f \in \mathcal{W}_p[a,b]$ if and only if $f = g \circ h$ for a bounded nondecreasing nonnegative function h on $[a,b]$ and a function g defined on $[h(a), h(b)]$ satisfying a Hölder condition with exponent $1/p$.*

[1]The Gamble Prize (October, 1915) for Mathematics at Girton College, Cambridge, was awarded for this essay to the author, Dr. Grace Chisholm Young.

Proof. For each $x \in [a,b]$, let $h(x) := v_p(f;[a,x])$. Then h is bounded, nondecreasing and nonnegative function on $[a,b]$. By Lemma 4.6, for each $x, y \in [a,b]$, we have

$$|f(y) - f(x)|^p \le |h(x) - h(y)|.$$

Define g on $\{h(x)\colon x \in [a,b]\}$ by $g(h(x)) := f(x)$ and extend by linearity elsewhere. Thus the "only if" part of the statement holds. The proof is complete because the "if" part is clear. □

The proof of the following is immediate:

Lemma 4.4. *For $0 < p < \infty$ and $f\colon [a,b] \mapsto \mathbb{R}$, $v_p(f) = v_p(f \circ H)$ for any increasing continuous function H from $[a,b]$ onto itself.*

Lemma 4.5. *$\|f\|_{(p)}$ is a non–increasing function of p.*

Proof. This follows from the inequality

$$\Big(\sum_1^n a_i^p\Big)^{1/p} \le \Big(\sum_1^n a_i^q\Big)^{1/q}$$

valid for $0 < q < p < \infty$ and any $\{a_i\}$. □

Lemma 4.6. *Let $a < c < b$ and $1 \le p < \infty$. Then*

$$v_p(f;[a,c]) + v_p(f;[c,b]) \le v_p(f;[a,b]) \le 2^{p-1}[v_p(f;[a,c]) + v_p(f;[c,b])].$$

Proof. The first inequality is immediate. For the second, if $a \le s \le c \le t \le b$ we have

$$|f(t) - f(s)|^p \le (|f(t) - f(c)| + |f(c) - f(s)|)^p, \quad \text{and} \quad \Big(\frac{1}{2}(a+b)\Big)^p \le \frac{1}{2}(a^p + b^p)$$

for $a, b \ge 0$ by Jensen's inequality (e.g. Theorem 10.2.6 in [22]) since $x \mapsto x^p$ is convex, $1 \le p < \infty$, $x \ge 0$. Then conclusion follows. □

Lemma 4.7. *Let $f \in \mathcal{W}_p[a,b]$ and $c > 0$. There exists $\{x_i\colon i = 0, \dots, n\} \in \mathrm{PP}\,([a,b])$ such that $\max_{1 \le i \le n} v_p(f;(x_{i-1}, x_i)) \le c$ and $n \le 1 + v_p(f;[a,b])/c$.*

Proof. Let $\rho(x) := v_p(f;[a,x])$ for $x \in [a,b]$. The lemma holds if $c \ge \rho(b)$. Suppose $c < \rho(b)$. Let M be the minimal integer such that $cM \ge \rho(b)$ and let $x_0 := a$. For each $m = 1, \dots, M$, let $x_1^m := \inf\{x \in (x_0, b]\colon \rho(x) \ge mc \wedge \rho(b)\}$ and let $m_1 := \min\{m = 1, \dots, M\colon x_1^m > x_0\}$. If $m_1 = M$ let $x_1 := b$. Then $\rho(a+) \ge Mc$ and $\rho(b-) - \rho(a+) \le Mc - Mc = 0$. Thus the lemma holds with $n = 1$ in this case. If $m_1 < M$ let $x_1 := x_1^{m_1}$. It then follows that $\rho(x_1-) \le m_1 c$ and $\rho(x_0+) \ge (m_1 - 1)c$.

Suppose $x_i < b$ is chosen such that $m_1 + \dots + m_i < M$, $\rho(x_i-) \le (m_1 + \dots + m_i)c$ and $\rho(x_{i-1}+) \ge (m_1 + \dots + m_i - 1)c$. For each $m = m_1 + \dots + m_i + 1, \dots, M$, let $x_{i+1}^m := \inf\{x \in (x_i, b]\colon \rho(x) \ge mc \wedge \rho(b)\}$ and let $m_{i+1} := \min\{m = m_1 + \dots + m_i + 1, \dots, M\colon x_{i+1}^m > x_i\}$. If $m_{i+1} = M$ let $x_{i+1} := b$. Otherwise let $x_{i+1} := x_{i+1}^{m_{i+1}}$ and proceed further. The recursion stops after at most M steps. Thus $n \le M \le 1 + \rho(b)/c$ and $v_p(f;(x_{i-1}, x_i)) \le \rho(x_i-) - \rho(x_{i-1}+) \le c$ for each $i = 1, \dots, n$ by Lemma 4.6. This completes the proof. □

The following properties are proved in Lemmas 2.18 and 2.19 of [29].

Lemma 4.8. *Let $f \in \mathcal{W}_p[a,b]$. For any $y \in [a,b)$,*

$$\lim_{x \downarrow y} v_p(f;(y,x]) = 0 \qquad \text{and} \qquad \lim_{x \downarrow y} v_p(f;[y,x]) = |(\Delta^+ f)(y)|^p.$$

For any $y \in (a,b]$,

$$\lim_{x \uparrow y} v_p(f;[x,y)) = 0 \qquad \text{and} \qquad \lim_{x \uparrow y} v_p(f;[x,y]) = |(\Delta^- f)(y)|^p.$$

4.2 Regulated functions

Recall that a function $f\colon [a,b] \mapsto \mathbb{R}$ is called *regulated* if, for each $a \leq x < y \leq b$, there exist finite limits $\lim_{z\uparrow y} f(z)$ and $\lim_{z\downarrow y} f(z)$. The class of all regulated functions on $[a,b]$ will be denoted by $\mathcal{R}[a,b]$. Jumps of the regulated function f on the closed interval $[a,b]$ are defined by:

$$\Delta_a^- f(y) := f(y) - f_-^{(a)}(y) = \begin{cases} f(y) - f(y-) & \text{if } a < y \leq b \\ 0 & \text{if } y = a \end{cases} \tag{4.1}$$

and

$$\Delta_b^+ f(x) := f_+^{(b)}(x+) - f(x) = \begin{cases} f(x+) - f(x) & \text{if } a \leq x < b \\ 0 & \text{if } x = b. \end{cases} \tag{4.2}$$

The following fact was noted by Wiener [113, p. 75].

Proposition 4.9. *If $v_p(f;[a,b]) < \infty$ for some $0 < p < \infty$ then $f \in \mathcal{R}[a,b]$.*

Proof. This follows from Lemma 4.3. □

In the following, a *step function* is a linear combination of indicators of intervals, some of which may be singletons.

Theorem 4.10. *The following properties are equivalent for a function $f\colon [a,b] \to \mathbb{R}$:*
1. *$f \in \mathcal{R}([a,b])$;*
2. *for $\epsilon > 0$ there exists $\{x_i\colon i = 0,\dots,n\} \in PP([a,b])$ such that $\mathrm{Osc}(f;(x_{i-1},x_i)) \leq \epsilon$ for each $i = 1,\dots,n$;*
3. *f is a uniform limit of step functions.*

Proof. (1) $\Rightarrow$ (2). Let $\epsilon > 0$. For each $y \in (a,b)$ there exists $\delta_y > 0$ such that $\mathrm{Osc}(f;(y-\delta_y,y)) \leq \epsilon$ and $\mathrm{Osc}(f;(y,y+\delta_y)) \leq \epsilon$. Also, there exist $\delta_a > 0$ and $\delta_b > 0$ such that $\mathrm{Osc}(f;(a,a+\delta_a)) \leq \epsilon$ and $\mathrm{Osc}(f;(b-\delta_b,b)) \leq \epsilon$. The sets

$$[a,a+\delta_a),\dots,(y-\delta_y,y+\delta_y),\dots,(b-\delta_b,b]$$

form an open cover of the compact interval $[a,b]$. Thus, there is a finite subcover

$$[a,a+\delta_a),(y_1-\delta_{y_1},y_1+\delta_{y_1}),\dots,(y_m-\delta_{y_m},y_m+\delta_{y_m}),(b-\delta_b,b].$$

Let $x_0 := a$, $x_1 \in (y_1-\delta_{y_1},a+\delta_a)$, $x_2 := y_1$, …, $x_{2m-1} := y_m$, $x_{2m} \in (b-\delta_b,y_m+\delta_{y_m})$ and $x_{2m+1} := b$. So that (2) holds with $n = 2m+1$.

(2) $\Rightarrow$ (3). Let $\epsilon > 0$. Choose $\{x_i\colon i = 0,\dots,n\} \in Q([a,b])$ such that (2) holds. Let $f_\epsilon(x) := f(y_i)$ with $y_i \in (x_{i-1},x_i)$ if $x \in (x_{i-1},x_i)$ and $f_\epsilon(x_i) := f(x_i)$ for $i = 0,\dots,n$. Then f_ϵ is a step function and $|f(x)-f_\epsilon(x)| \leq \epsilon$ for each $x \in [a,b]$ by construction. Since ϵ is arbitrary (3) holds.

(3) $\Rightarrow$ (1). For each $\epsilon > 0$ choose a step function f_ϵ such that $|f(x)-f_\epsilon(x)| \leq \epsilon$ for each $x \in [a,b]$. Let $x \in (a,b]$ and $x_n \uparrow x$. Then, for all integers n,m, we have

$$|f(x_n)-f(x_m)| \leq |f(x_n)-f_\epsilon(x_n)|+|f_\epsilon(x_n)-f_\epsilon(x_m)|+|f_\epsilon(x_m)-f(x_m)| \leq 2\epsilon+|f_\epsilon(x_n)-f_\epsilon(x_m)|.$$

Since f_ϵ is regulated, the right side of the last inequalities can be made arbitrarily small. Thus $f(x-)$ exists. Similarly it follows that $f(x+)$ exists for each $x \in [a,b)$. This proves (1). The proof of Theorem 4.10 is complete. □

Corollary 4.11. *For each $f \in \mathcal{R}([a,b])$ and $\epsilon > 0$,*

$$\mathit{card}\{x \in [a,b]\colon\ |\Delta_a^- f(x)| > \epsilon \text{ or } |\Delta_b^+ f(x)| > \epsilon\} < \infty.$$

Definitions. Let f be a real-valued function on an interval J and $\varepsilon > 0$. For a subinterval $[c,d] \subset J$ let $|f([c,d])| := |f(d) - f(c)|$. Subintervals of J will be called *nonoverlapping* if they are disjoint except possibly for their endpoints. Let $\sup_{\{J_i\}}$ denote the supremum over all sequences $\{J_i\}$ of nonoverlapping subintervals of J. Then let

$$n(f,\varepsilon) \ := \ \sup_{\{J_i\}}\{n\colon\ |f(J_i)| > \epsilon \quad \text{for } i = 1,\dots,n\}.$$

For $n = 1, 2, \dots$, let

$$v(f,n) \ := \ \sup_{\{J_i\}}\left\{\textstyle\sum_{i=1}^n |f(J_i)|\right\}.$$

For a sequence $\Lambda := \{\lambda_i\}_{i\geq 1}$ of positive numbers, the Λ-*variation* of a function f on J is $\sup_{\{J_i\}} \sum_i\ |f(J_i)|/\lambda_i$. Then f is of *bounded* Λ-*variation*, or $f \in \Lambda BV$, iff its Λ-variation is finite.

An *M-function* will be a strictly increasing, convex function Φ on $[0,\infty[$ with $\Phi(0) = 0$. Recall (from the Introduction) that for a real function f on J, the Φ-*variation* $v_\Phi(f)$ is the supremum of sums $\sum_i \Phi(|f(J_i)|)$ over all interval partitions $\{J_i\}$. Also, as before, a step function on an interval J is a finite linear combination of indicator functions of subintervals of J (which may be open or closed at either end).

Among M-functions are, of course, the functions $\Phi(x) = x^r$ on $0 \leq x < \infty$, where $1 \leq r < \infty$.

Proofs of the following known equivalences are given by Dudley [27, Theorem 2.1].

Theorem 4.12. *For a real-valued function f on a closed, bounded interval $J = [a,b]$, the following are equivalent:*

1. *f is regulated on J ($f \in E(J)$).*
2. *For any $\varepsilon > 0$, $n(f,\varepsilon) < \infty$.*
3. *$v(f,n) = o(n)$ as $n \to \infty$.*
4. *For some $\{\lambda_i\}_{i\geq 1}$ such that $\lambda_i > 0$ for all i and $\sum_i 1/\lambda_i = +\infty$, $f \in \Lambda BV$.*
5. *$v_\Phi(f) < \infty$ for some M-function Φ.*
6. *$f \equiv g \circ h$ where g is continuous on $\mathbb{R}$ and h is nondecreasing and real valued on J.*
7. *$f \equiv g \circ h$ where g is continuous on $\mathbb{R}$ and h is strictly increasing and real valued on J.*
8. *f is a uniform limit of step functions on J.*

4.3 The Wiener class

Let $0 < p < \infty$ and $f\colon [a,b] \mapsto \mathbb{R}$. For each $\kappa = \{x_i\colon i = 0, 1, \dots, n\} \in PP([a,b])$, recall that $s_p(f;\kappa) := \sum_{i=1}^n |f(x_i) - f(x_{i-1})|^p$. Let

$$v_p(f;\kappa) := \sup\{s_p(f;\lambda)\colon\ \lambda \supset \kappa\} = \sum_{i=1}^n v_p(f;[x_{i-1},x_i]). \tag{4.3}$$

Notice that $v_p(f;[a,b]) = v_p(f;\{a,b\}) = \sup\{s_p(f;\kappa)\colon \kappa \in PP([a,b])\}$. Then let

$$\begin{aligned} v_p^*(f) &:= v_p^*(f;[a,b]) := \inf\{v_p(f;\kappa)\colon \kappa \in PP([a,b])\} \\ &= \inf\{\sum_{i=1}^n v_p(f;[x_{i-1},x_i])\colon \{x_i\colon i=0,\ldots,n\} \in PP([a,b])\}. \end{aligned}$$

Lemma 4.13. *$v_p^*(f) \le v_p(f)$, and if either is finite then so is the other.*

Proof. The inequality is clear, so we only need to show that if $v_p^*(f) < \infty$ then $v_p(f) < \infty$. Now $v_p^*(f) < \infty$ implies $v_p(f;\kappa) < \infty$ for some $\kappa = \{x_i\colon i = 0,\ldots,n\} \in \mathrm{PP}([a,b])$. By iterating Lemma 4.6 we get $v_p(f;[a,b]) \le 2^{(p-1)(n-1)} v_p(f;\kappa) < \infty$, proving Lemma 4.13. □

Since a function f with bounded p-variation for some $0 < p < \infty$ is regulated, there are at most countably many discontinuity points of f. One can enumerate the points of discontinuity, say $D(f) := \{\xi_1,\xi_2,\ldots\}$, $\xi_j := \xi_j(f)$, $j = 1,2,\ldots$, so that the sequence $\{w_j\colon j \ge 1\}$ is non-increasing, where

$$w_j^p := |\Delta_a^- f(\xi_j)|^p + |\Delta_b^+ f(\xi_j)|^p \qquad \text{for } j = 1,2,\ldots.$$

Let

$$\mathfrak{S}_p(f) := \mathfrak{S}_p(f;[a,b]) := \Big(\sum_{[a,b]} [|\Delta_a^- f|^p + |\Delta_b^+ f|^p]\Big)^{1/p} = \Big(\sum_{j=1}^\infty w_j^p\Big)^{1/p}.$$

Then the relation

$$\mathfrak{S}_p(f)^p \le v_p^*(f) \tag{4.4}$$

always holds. We will see that functions f for which the equality holds instead of $\le$ constitute a proper subspace of $\mathcal{W}_p$ if $p \ge 1$, and they have nice properties if $p > 1$. First we consider the case when $0 < p < 1$. The proofs of the following two statements can be found in [29] (Propositions 2.12 and 2.13, respectively).

Proposition 4.14. *If $0 < p < 1$ then $f \in \mathcal{W}_p[a,b]$ if and only if there exist at most countably many real numbers v_a, u_b, and u_y, v_y for $a < y < b$, such that, for $a < x \le b$,*

$$f(x) = f(a) + v_a + \sum_{a<y<x} (u_y + v_y) + u_x,$$

where each sum converges absolutely, and

$$\sum_{a<y<b} [|u_y|^p + |v_y|^p] < \infty.$$

Then $\Delta^+ f(x) = v_x$ for $a \le x < b$, $\Delta^- f(x) = u_x$ for $a < x \le b$ and

$$v_p^*(f) = v_p(f) = \mathfrak{S}_p(f)^p = |v_a|^p + \sum_{a<y<b} [|u_y|^p + |v_y|^p] + |u_b|^p.$$

Thus the only continuous functions in $\mathcal{W}_p[a,b]$ if $0 < p < 1$ are constants. Next we look at the case $p = 1$.

Proposition 4.15. *For any function $f\colon [a,b] \mapsto \mathbb{R}$, the following holds:*

1. $v_1^*(f) = v_1(f)$;

2. $v_1^*(f) = \mathfrak{S}_1(f)$ *if and only if for* $a < x \le b$,

$$f(x) = f(a) + \Delta^+ f(a) + \sum_{a<y<x} [\Delta^- f(y) + \Delta^+ f(y)] + \Delta^- f(x),$$

where the sum converges absolutely.

For the rest of this section we consider the case $p > 1$.

Definition 4.16. For $1 < p < \infty$, let $\mathcal{W}_p^* := \mathcal{W}_p^*[a,b] := \{f \in \mathcal{W}_p[a,b]\colon v_p^*(f) = \mathfrak{S}_p(f)^p\}$. We call $\mathcal{W}_p^*$ the *Wiener class*.

The rest of this section is based on the results of Love and Young [75].

Theorem 4.17. *The Wiener class* $\mathcal{W}_p^*$ *is closed, hence complete for* $\|\cdot\|_{[p]}$.

For the proof we need the following:

Lemma 4.18. *Let* $f \in \mathcal{R}[a,b]$ *and* $1 < p < \infty$. *Then the following statements are equivalent:*

1. $f \in \mathcal{W}_p^*[a,b]$;
2. *for every* $\epsilon > 0$ *there is a partition* $\lambda = \{x_i\colon i = 0, \ldots, n\}$ *of* $[a,b]$ *such that*

$$\sum_{i=1}^{n} v_p(f; (x_{i-1}, x_i)) < \epsilon; \tag{4.5}$$

3. $\mathfrak{S}_p(f; [a,b]) < \infty$ *and for every* $\epsilon > 0$ *there is a partition* λ *of* $[a,b]$ *such that*

$$\sum_{i=1}^{n} |f(x_i-) - f(x_{i-1}+)|^p < \epsilon \tag{4.6}$$

for each refinement $\{x_i\colon i = 0, \ldots, n\}$ *of* λ.

Proof. (1) $\Rightarrow$ (2). Let $f \in \mathcal{W}_p^*[a,b]$ and $\epsilon > 0$. It is enough to show that (4.5) holds for some partition of $[a,b]$. By (4.3) and (4.4), there exists $\lambda = \{z_j\colon j = 0, \ldots, m\} \in PP([a,b])$ such that

$$\sum_{j=1}^{m} v_p(f; [z_{j-1}, z_j]) < v_p^*(f) + \epsilon/2$$

and

$$\sum_{j=1}^{m} [|\Delta^+ f(z_{j-1})|^p + |\Delta^- f(z_j)|^p] > \mathfrak{S}_p(f)^p - \epsilon/2.$$

Let $\{u_{j-1}, v_j\colon j = 1, \ldots, m\}$ be a set of points in (a,b) such that $z_{j-1} < u_{j-1} < v_j < z_j$ for $j = 1, \ldots, m$. Then we have

$$\sum_{j=1}^{m} v_p(f; [u_{j-1}, v_j]) \le \sum_{j=1}^{m} v_p(f; [z_{j-1}, z_j]) - \sum_{j=1}^{m} [v_p(f; [z_{j-1}, u_{j-1}]) + v_p(f; [v_j, z_j])].$$

For each $j = 1, \ldots, m$ letting $u_{j-1} \downarrow z_{j-1}$ and $v_j \uparrow z_j$, by Lemma 2.19 in [29], it follows that

$$\begin{aligned} \sum_{j=1}^{m} v_p(f; (z_{j-1}, z_j)) &\le \sum_{j=1}^{m} v_p(f; [z_{j-1}, z_j]) - \sum_{j=1}^{m} [|\Delta^+ f(z_{j-1})|^p + |\Delta^- f(z_j)|^p] \\ &< v_p^*(f) + \epsilon/2 - \mathfrak{S}_p(f)^p + \epsilon/2 = \epsilon. \end{aligned}$$

This proves statement (2).

(2) $\Rightarrow$ (1). Let statement (2) hold. Then $f \in \mathcal{W}_p$ by iterating Lemma 4.6. Assume however that $f \notin \mathcal{W}_p^*$. Therefore, since (4.4) always holds, $v_p^*(f) - \mathfrak{S}_p(f)^p \geq C$ for some positive constant C. Let $\kappa = \{x_i \colon i = 0, \ldots, n\} \in PP([a,b])$. Then we have

$$\sum_{i=1}^{n} [|\Delta^+ f(x_{i-1})|^p + v_p(f; (x_{i-1}, x_i)) + |\Delta^- f(x_i)|^p] \geq v_p^*(f).$$

It then follows that

$$\begin{aligned} \sum_{i=1}^{n} v_p(f; (x_{i-1}, x_i)) &\geq v_p^*(f) - \sum_{i=1}^{n} [\Delta^+ f(x_{i-1})|^p + |\Delta^- f(x_i)|^p] \\ &\geq v_p^*(f) - \mathfrak{S}_p(f)^p \geq C > 0. \end{aligned}$$

Since κ is arbitrary, (4.5) can't hold for every $\epsilon > 0$. This contradiction implies that $f \in \mathcal{W}_p^*$.

(2) $\Rightarrow$ (3). Let statement (2) hold. Then it holds for any refinement of λ by Lemma 4.6. Then it is easy to see that $f \in \mathcal{W}_p$, and hence $\mathfrak{S}_p(f) < \infty$ by Lemma 4.13 and (4.4). Let $\epsilon > 0$ and let (4.5) hold for $\{x_i \colon i = 0, \ldots, n\} \in PP([a,b])$. Then we have

$$\begin{aligned} \sum_{i=1}^{n} |f(x_i-) - f(x_{i-1}+)|^p &= \sum_{i=1}^{n} \lim_{y_{i-1} \downarrow x_{i-1},\, y_i \uparrow x_i} |f(y_i) - f(y_{i-1})|^p \\ &\leq \sum_{i=1}^{n} v_p(f; (x_{i-1}, x_i)) < \epsilon. \end{aligned}$$

Therefore statement (3) holds.

(3) $\Rightarrow$ (2). Let $\epsilon > 0$ and let $\lambda \in PP([a,b])$ be such that (4.6) holds for all its refinements. There exists a finite set $\nu \subset [a,b]$ such that

$$\sum_{x \in \mu} [|\Delta^- f(x)|^p + |\Delta^+ f(x)|^p] < \epsilon$$

for each finite set $\mu \subset (a,b)$ with $\mu \cap \nu = \emptyset$. Let $\{z_j \colon j = 0, \ldots, m\} := \lambda \cup \nu \in PP([a,b])$ and, for each $j = 1, \ldots, m$, let $\kappa(j) := \{x_i^j \colon i = 0, \ldots, n(j)\}$ be a set of points in (a,b) such that $z_{j-1} < x_0^j < \cdots < x_{n(j)}^j < z_j$. Then we have

$$\begin{aligned} \sum_{j=1}^{m} s_p(f; \kappa(j)) &\leq 4^{p-1} \sum_{j=1}^{m} \sum_{i=1}^{n(j)} [|\Delta^- f(x_{i-1}^j)|^p + |\Delta^+ f(x_i^j)|^p] \\ &\quad + 4^{p-1} \sum_{j=1}^{m} \sum_{i=1}^{n(j)} |f(x_i^j-) - f(x_{i-1}^j)|^p < 4^p \epsilon / 2. \end{aligned}$$

Since each partition $\kappa(j)$ of (z_{j-1}, z_j) is arbitrary, it follows that

$$\sum_{j=1}^{m} v_p(f; (z_{j-1}, z_j)) \leq 4^p \epsilon / 2.$$

This proves statement (2). The proof of Lemma 4.18 is complete. $\square$

Proof of Theorem 4.17. Let $\epsilon > 0$ and let $\phi \in \mathcal{W}_p^*$ be such that $\|f - \phi\|_{(p)} \le \epsilon^{1/p}/2$. By Lemma 4.18, there is a partition λ of $[a, b]$ such that $\sum_{i=1}^n v_p(\phi; (x_{i-1}, x_i)) < \epsilon/2^p$ for each refinement $\kappa = \{x_i: i = 0, \dots, n\}$ of λ. Then for any such κ, we have

$$\begin{aligned} \sum_{i=1}^n v_p(f; (x_{i-1}, x_i)) &\le 2^{p-1} \sum_{i=1}^n [v_p(f - \phi; (x_{i-1}, x_i)) + v_p(\phi; (x_{i-1}, x_i))] \\ &\le 2^{p-1}[\|f - \phi\|_{(p)}^p + \sum_{i=1}^n v_p(\phi; (x_{i-1}, x_i))] < \epsilon. \end{aligned}$$

Therefore statement (2) of Lemma 4.18 holds, and hence $f \in \mathcal{W}_p^*$. The proof of Theorem 4.17 is complete. $\square$

A function $\phi: [a, b] \mapsto \mathbb{R}$ is a step function as in Theorem 4.10 if and only if there exists a partition $\{y_j: j = 0, \dots, m\}$ of $[a, b]$ such that ϕ is constant on each open interval (y_{j-1}, y_j), $j = 1, \dots, m$. For such a function ϕ and any $0 < p < \infty$, we have

$$v_p^*(\phi) = \sum_{j=1}^m [|\Delta^+\phi(y_{j-1})|^p + |\Delta^-\phi(y_j)|^p] = \mathfrak{S}_p(\phi)^p.$$

Thus each step function $\phi \in \mathcal{W}_p^*$ for any $p > 1$.

Theorem 4.19. *Step functions based on interval partitions are dense in the Wiener class $\mathcal{W}_p^*$.*

For the proof we need:

Lemma 4.20. *Let $f: [a, b] \mapsto \mathbb{R}$ be such that $f(x_i) = 0$ for $i = 0, \dots, n$, where $a = x_0 < x_1 < \dots < x_n = b$, and let $1 \le p < \infty$. Then*

$$v_p(f; [a, b]) \le c_p \sum_{i=1}^n v_p(f; [x_{i-1}, x_i]), \tag{4.7}$$

where $c_p = 2^p$, and $c_p = 1$ if in addition $f \ge 0$.

Proof. Suppose in addition that $f \ge 0$. Then for $a \le x < z < y \le b$,

$$|f(x) - f(y)|^p \le |f(x)|^p + |f(y)|^p = |f(x) - f(z)|^p + |f(z) - f(y)|^p$$

if $f(z) = 0$. Therefore (4.7) holds in this case. Let $f_+ := \max(f, 0)$ and $f_- := -\min(f, 0)$. Then without the additional hypothesis, we have

$$\begin{aligned} v_p(f) &= v_p(f_+ - f_-) \le 2^{p-1}[v_p(f_+) + v_p(f_-)] \\ &\le 2^{p-1} \sum_{i=1}^n [v_p(f_+; [x_{i-1}, x_i]) + v_p(f_-; [x_{i-1}, x_i])] \\ &\le 2^{p-1} \sum_{i=1}^n [v_p(f; [x_{i-1}, x_i]) + v_p(f; [x_{i-1}, x_i])]. \end{aligned}$$

Thus (4.7) holds with $c_p = 2^p$. $\square$

Proof of Theorem 4.19. Let $f \in \mathcal{W}_p^*$. Let $\epsilon > 0$ and $\theta := \epsilon/2^{2p+4}$. There exists $\kappa_1 \in PP([a, b])$ such that for all $\kappa \supset \kappa_1$,

$$v_p(f; \kappa) < v_p^*(f) + \theta.$$

Choose an integer k so large that both

$$\mathfrak{S}_p(f;\kappa)^p := \sum_{j=1}^{k} w_j^p > \mathfrak{S}_p(f)^p - \theta$$

and $j \leq k$ for each $\xi_j = \xi_j(f) \in \kappa_1 \cap D(f)$. We claim that for all $j > 2k$,

$$w_j^p \leq w_{2k}^p < \theta/k. \tag{4.8}$$

If not then

$$\theta \leq k w_{2k}^p \leq \sum_{k+1}^{2k} w_j^p \leq \mathfrak{S}_p(f)^p - \mathfrak{S}_p(f;k)^p < \theta.$$

Thus (4.8) must hold for all $j > 2k$. By (4.8), it follows that

$$|\Delta_a^- f(\xi_j)| \vee |\Delta_b^+ f(\xi_j)| \leq w_j \leq (\theta/k)^{1/p}$$

for each $j > 2k$. Therefore using Theorem 4.10, one can find a partition κ_2, containing all ξ_j, $j = 1, \dots, 2k$, and such that $\mathrm{Osc}(f;J) \leq 2(\theta/k)^{1/p}$ for each interval J formed by adjacent points of κ_2, being open at either end-point if that end-point is ξ_j with $j \leq 2k$. Let κ_ϵ be the partition consisting of all points of κ_1 and κ_2, and all points $(\eta_{i-1}+\eta_i)/2$ such that $\eta_i = \xi_j$ and $\eta_{i-1} = \xi_r$ are adjacent points of $\kappa_1 \cup \kappa_2$ for some $j, r \leq 2k$, $j \neq r$. Let $\kappa = \{x_i\colon i = 0, \dots, n\} \supset \kappa_\epsilon$ and let $f_{\kappa,\epsilon}$ be the step function defined by

$$f_{\kappa,\epsilon}(x) = \begin{cases} f(y_i) & \text{with } y_i \in [x_{i-1}, x_i] \text{ and } y_i \notin D(f) \cap \kappa_\epsilon \\ & \text{if } x \in (x_{i-1}, x_i) \text{ for some } i = 1, \dots, n \\ f(x_i) & \text{if } x = x_i \text{ for some } i = 0, \dots, n \end{cases} \tag{4.9}$$

Let $\{z_j\colon j = 0, \dots, m\} \in PP([x_{i-1}, x_i])$ with $m > 1$ for some $i = 1, \dots, n$, and let $\psi := f - f_{\kappa,\epsilon}$. Then

$$\begin{aligned} \sum_{j=1}^{m} |\psi(z_j) - \psi(z_{j-1})|^p &= |f(z_1) - f(y_i)|^p + \sum_{j=1}^{m-1} |f(z_j) - f(z_{j-1})|^p + |f(y_i) - f(z_{m-1})|^p \\ &\leq 3 v_p(f;[x_{i-1}, x_i]). \end{aligned}$$

Thus for each $i = 1, \dots, n$,

$$v_p(f - f_{\kappa,\epsilon};[x_{i-1}, x_i]) \leq 3 v_p(f;[x_{i-1}, x_i]).$$

If x_i is a ξ_j with $j \leq 2k$, then x_{i-1} is not, and $y_i < x_i$. So given $\{z_j\colon j = 0, \dots, m\} \in PP([a, b])$ with $m > 1$, we have

$$\begin{aligned} \sum_{j=1}^{m} |\psi(z_j) - \psi(z_{j-1})|^p + |\Delta^- f(x_i)|^p &= |f(z_1) - f(y_i)|^p + \sum_{j=1}^{m-1} |f(z_j) - f(z_{j-1})|^p \\ &+ |\Delta^- f(z_m)|^p + |f(y_i) - f(z_{m-1})|^p \\ &\leq 2^p(\theta/k) + v_p(f;[x_{i-1}, x_i]) + 2^p(\theta/k). \end{aligned}$$

Thus for each $i = 1, \dots, n$,

$$v_p(f - f_{\kappa,\epsilon};[x_{i-1}, x_i]) \leq v_p(f;[x_{i-1}, x_i]) - |\Delta^- f(x_i)|^p + 2^{p+1}(\theta/k).$$

Similarly, if x_{i-1} is a ξ_j with $j \le 2k$, then for each $i = 1, \ldots, n$,

$$v_p(f - f_{\kappa,\epsilon}; [x_{i-1}, x_i]) \le v_p(f; [x_{i-1}, x_i]) - |\Delta^+ f(x_{i-1})|^p + 2^{p+1}(\theta/k).$$

Let I be the set of indices $i = 1, \ldots, n$ such that neither x_{i-1} nor x_i is a ξ_j with $j \le 2k$. Then

$$\mathfrak{S}_p(f; 2k)^p + \sum_{i \in I} v_p(f; [x_{i-1}, x_i]) \le v_p(f; \kappa).$$

Therefore by Lemma 4.20, it follows that

$$\begin{aligned}
\|f - f_{\kappa,\epsilon}\|_{(p)}^p &\le 2^p \sum_{i=1}^n v_p(f - f_{\kappa,\epsilon}; [x_{i-1}, x_i]) \\
&\le 2^p \Big\{ \sum_{i=1}^n v_p(f; [x_{i-1}, x_i]) + 2 \sum_{i \in I} v_p(f; [x_{i-1}, x_i]) - \mathfrak{S}_p(f; 2k)^p + 4k 2^{p+1}\theta/k \Big\} \\
&\le 2^p \Big\{ v_p(f; \kappa) + 2[v_p(f; \kappa) - \mathfrak{S}_p(f; 2k)^p] - \mathfrak{S}_p(f; 2k)^p + 2^{p+3}\theta \Big\} \\
&\le 2^p \Big\{ 3[v_p^*(f) + \theta - \mathfrak{S}_p(f)^p + \theta] + 2^{p+3}\theta \Big\} \\
&= 2^p (6 + 2^{p+3})\theta < \epsilon.
\end{aligned}$$

The proof of Theorem 4.19 is complete. $\square$

The following statement follows from Theorems 4.17 and 4.19.

Corollary 4.21. *The Wiener class $\mathcal{W}_p^*[a, b]$ is the closure for $\|\cdot\|_{[p]}$ of the set of step functions.*

It then follows from Lemma 4.22 below that, for each $1 < p < \infty$

$$\mathcal{W}_p \supset \mathcal{W}_p^* \supset \mathcal{W}_{p-} := \cup_{q<p} \mathcal{W}_p. \tag{4.10}$$

The following examples of functions on $[0, 1]$ show that the Wiener class $\mathcal{W}_p^*$ is identical with neither $\mathcal{W}_p$ nor $\mathcal{W}_{p-}$:

$$g(x) = \sum_{i=1}^\infty c^{-i/p} \sin(c^i x) \quad \text{and} \quad h(x) = \frac{x^{1/p}}{\log x} \cos^2 \frac{\pi}{x}, \quad x > 0, \quad h(0) := 0.$$

If c is a sufficiently large positive integer then g is in $\mathcal{W}_p$ but not in $\mathcal{W}_p^*$, while h is in $\mathcal{W}_p^*$ but not in $\mathcal{W}_{p-}$.

Lemma 4.22. *Let $f \in \mathcal{W}_q[a, b]$ for some $1 \le q < p < \infty$. Then given $\epsilon > 0$, there is a partition $a = x_0 < z_1 < x_1 < z_2 < \cdots < x_{n-1} < z_n < x_n = b$ such that $\|f - \phi\|_{(p)} < \epsilon$ and $x_i - x_{i-1} < \epsilon$ for $i = 1, \ldots, n$, where ϕ is the step function defined by*

$$\phi(x) := \begin{cases} f(z_i) & \text{if } x_{i-1} < x < x_i \text{ for some } i = 1, \ldots, n \\ f(x_i) & \text{if } x = x_i \text{ for some } i = 0, \ldots, n. \end{cases} \tag{4.11}$$

Proof. Given $\epsilon > 0$, let $\delta > 0$ be such that $\delta^{p-q} 3 \cdot 2^p v_q(f) < \epsilon$. By Proposition 4.9 and Theorem 4.10, there is a partition $\{x_i \colon i = 0, \ldots, n\}$ of $[a, b]$ such that $\max_i(x_i - x_{i-1}) < \epsilon$ and

$\mathrm{Osc}(f;(x_{i-1},x_i)) < \delta$. Let ϕ be the step function defined by (4.11). By Lemma 4.20, it follows that

$$v_p(f-\phi) \le 2^p \sum_{i=1}^n v_p(f-\phi;[x_{i-1},x_i]) \le 2^p \sum_{i=1}^n \delta^{p-q} v_q(f-\phi;[x_{i-1},x_i]),$$

because $f-\phi$ has oscillation less than δ in each closed interval $[x_{i-1},x_i]$. To bound $v_q(f-\phi;[x_{i-1},x_i])$, let $\psi := f-\phi$. Given any partition $\kappa_i = \{y_j : j = 0,\dots,m\} \in PP([x_{i-1},x_i])$ for which $m > 1$, we have

$$\begin{aligned} s_q(\psi;\kappa_i) &= |\psi(y_1)|^q + \sum_{j=2}^{m-1} |\psi(y_j)-\psi(y_{j-1})|^q + |\psi(y_{m-1})|^q \\ &= |f(y_1)-f(z_i)|^q + \sum_{j=2}^{m-1} |f(y_j)-f(y_{j-1})|^p \le 3v_q(f;[x_{i-1},x_i]). \end{aligned}$$

Since κ_i is arbitrary partition, it follows that $v_q(f-\phi;[x_{i-1},x_i]) \le 3v_q(f;[x_{i-1},x_i])$ for $i = 1,\dots,n$. Therefore

$$v_p(f-\phi) \le 2^p\delta^{p-q} \sum_{i=1}^n 3v_q(f;[x_{i-1},x_i]) \le 3\cdot 2^p \delta^{p-q} v_q(f;[a,b]) < \epsilon.$$

This completes the proof of Lemma 4.22. $\square$

4.4 Love–Young inequalities

Lemma 4.23 (Love–Young inequality). *Let $f,h\colon [a,b] \mapsto \mathbb{R}$ and let $\kappa = \{x_i : i = 0,\dots,n\}$ be a partition of $[a,b]$. Then for any $y \in \kappa$,*

$$\Big| \sum_{i=1}^n f(x_i)[h(x_i)-h(x_{i-1})] - f(y)[h(b)-h(a)] \Big| \le C_{p,q} V_p(f) V_q(h) \tag{4.12}$$

provided $p,q > 0$ and $p^{-1}+q^{-1} > 1$, where $C_{p,q} = \zeta(p^{-1}+q^{-1})$ and $\zeta(s) := \sum_{n\ge1} n^{-s}$.

Proof. Let $\kappa = \{x_i : i = 0,\dots,n\} \in PP([a,b])$, and let $\Delta_i g := g(x_i) - g(x_{i-1})$ for $i = 1,\dots,n$ and $g = f$ or h. Also, let $S(f,h;\kappa) := \sum_{i=1}^n f(x_i)[h(x_i)-h(x_{i-1})]$. Then we have

$$S(f,h;\kappa) = f(a)[h(b)-h(a)] + \sum_{1\le i\le j\le n} \Delta_i f \Delta_j h. \tag{4.13}$$

On the other hand, since $f(x_j) = f(b) - \sum_{j<i\le n} \Delta_i f$, we also have

$$S(f,h;\kappa) = f(b)[h(b)-h(a)] - \sum_{1\le j<i\le n} \Delta_i f \Delta_j h. \tag{4.14}$$

Now let $y = x_k$ for some $0 < k < n$. Then (4.13) for $a = y$ gives

$$\sum_{i=k+1}^n f(x_i)\Delta_i h = f(y)[h(b)-h(y)] + \sum_{k<i\le j\le n} \Delta_i f \Delta_j h. \tag{4.15}$$

Also, (4.14) for $b = y$ gives

$$\sum_{i=1}^{k} f(x_i)\Delta_i h = f(y)[h(y) - h(a)] - \sum_{1\le j<i\le k} \Delta_i f \Delta_j h. \tag{4.16}$$

Adding (4.15) and (4.16) gives the representation

$$S(f,h;\kappa) - f(y)[h(b) - h(a)] = -\sum_{1\le j<i\le k} \Delta_i f \Delta_j h + \sum_{k<i\le j\le n} \Delta_i f \Delta_j h. \tag{4.17}$$

of the quantity we are aiming to bound. Therefore it is enough to bound the absolute value of the right side of (4.17). To this aim we replace $\Delta_i f$ by a number ϕ_i, $\Delta_j h$ by a number ψ_j and apply the following lemma. This completes the proof of the Love-Young inequality. $\square$

Let $\phi = (\phi_1, \dots, \phi_n)$ and $\psi = (\psi_1, \dots, \psi_n)$ be finite sequences of real numbers. For $p, q > 0$, let

$$V_p(\phi) := V_p(\phi; 1, n) := \max\Big\{\Big(\sum_{j=1}^{m}\Big|\sum_{i=\theta(j-1)+1}^{\theta(j)} \phi_i\Big|^p\Big)^{1/p} : 0 = \theta(0) < \theta(1) < \dots < \theta(m) = n\Big\},$$

and let $V_q(\psi)$ be defined similarly. Then the following holds:

Lemma 4.24. *Let p, q and $C_{p,q}$ be as in the Love–Young inequality. For finite sequences $\phi = (\phi_1, \dots, \phi_n)$ and $\psi = (\psi_1, \dots, \psi_n)$, and for any integer $k = 0, \dots, n$,*

$$\Big| -\sum_{1\le j<i\le k} \phi_i\psi_j + \sum_{k<i\le j\le n} \phi_i\psi_j\Big| \le C_{p,q} V_p(\phi) V_q(\psi). \tag{4.18}$$

Note. This is an extension of inequality (5.1) of Young (1936, p. 254) which is the case $k = 0$. Also, it improves the same inequality by having the constant factor $C_{p,q} = \zeta(p^{-1} + q^{-1})$ in (4.18) instead of $1 + \zeta(p^{-1} + q^{-1})$. To obtain this improvement we used the argument suggested by Love [74, Theorem 17].

Proof. Let $\phi = (\phi_1, \dots, \phi_n)$ and $\psi = (\psi_1, \dots, \psi_n)$ be two sequences of real numbers, and let $k \in \{0, \dots, n\}$. Then let

$$S_L(\phi, \psi; k) := \sum_{0\le j<i\le k} \phi_i\psi_j = \sum_{1\le j<i\le k} \phi_i\psi_j \quad \text{and} \quad S_R(\phi, \psi; k, n) := \sum_{k<i\le j\le n} \phi_i\psi_j,$$

where $\psi_0 := 0$, $S_L(\phi, \psi; 0) := 0$ and $S_R(\phi, \psi; n, n) := 0$. For $1 \le l \le n$, let

$$\begin{array}{llll} \phi'_j := \phi_j \ \text{ for } \ 1 \le j \le l-1, & \phi'_l := \phi_l + \phi_{l+1}, & \phi'_j := \phi_{j+1} \ \text{ for } \ l < j < n \\ \psi'_j := \psi_j \ \text{ for } \ 0 \le j < l-1, & \psi'_{l-1} := \psi_{l-1} + \psi_l, & \psi'_j := \psi_{j+1} \ \text{ for } \ l \le j < n, \end{array}$$

where $\phi_{n+1} := 0$ for the case when $l = n$. Then $\phi' = (\phi'_1, \dots, \phi'_{n-1})$ and $\psi' = (\psi'_1, \dots, \psi'_{n-1})$ are two sequences of real numbers dependent on l. If $1 \le l < k$ then $S_R(\phi', \psi'; k-1, n-1) = S_R(\phi, \psi; k, n)$ and

$$\begin{aligned} S_L(\phi', \psi'; k-1) &= \sum_{0<i\le k-1} \phi'_i(\psi'_0 + \ldots + \psi'_{i-1}) = \sum_{1<i<l} \phi_i(\psi_1 + \ldots + \psi_{i-1}) \\ &\quad + (\phi_l + \phi_{l+1})(\psi_1 + \ldots + \psi_l) + \sum_{l<i\le k-1} \phi_{i+1}(\psi_1 + \ldots + \psi_i) \\ &= \phi_l\psi_l + \sum_{1<i\le k} \phi_i(\psi_1 + \ldots + \psi_{i-1}) = \phi_l\psi_l + S_L(\phi, \psi; k). \end{aligned}$$

If $k < l \le n$ then $S_L(\phi', \psi'; k) = S_L(\phi, \psi; k)$ and

$$\begin{aligned} S_R(\phi', \psi'; k, n-1) &= \sum_{k<i\le n-1} \phi'_i(\psi'_i + \ldots + \psi'_{n-1}) = \sum_{k<i<l} \phi_i(\psi_i + \ldots + \psi_n) \\ &\quad + (\phi_l + \phi_{l+1})(\psi_{l+1} + \ldots + \psi_n) + \sum_{l<i<n} \phi_{i+1}(\psi_{i+1} + \ldots + \psi_n) \\ &= -\phi_l\psi_l + \sum_{k<i\le n} \phi_i(\psi_i + \ldots + \psi_n) = -\phi_l\psi_l + S_R(\phi, \psi; k, n). \end{aligned}$$

Letting $S_n := S_{n,k} := -S_L(\phi, \psi; k) + S_R(\phi, \psi; k, n)$ and

$$T_{n,l} := T_{n,l,k} := \begin{cases} -S_L(\phi', \psi'; k-1) + S_R(\phi', \psi'; k-1, n-1) & \text{if } l < k \\ -S_L(\phi', \psi'; k) + S_R(\phi', \psi'; k, n-1) & \text{if } k < l \end{cases}$$

we obtain that $S_n - T_{n,l} = \phi_l\psi_l$ for $l \in \{1, \ldots, n\} \setminus \{k\}$. Next, we need the following fact: given two ordered m–tuples of real numbers $\xi = \{\xi_1, \ldots, \xi_m\}$ and $\eta = \{\eta_1, \ldots, \eta_m\}$ there is an index r, $1 \le r \le m$, such that

$$|\xi_r\eta_r| \le \Big(\frac{1}{m}\sum_{j=1}^m |\xi_j|^p\Big)^{1/p}\Big(\frac{1}{m}\sum_{j=1}^m |\eta_j|^q\Big)^{1/q}. \tag{4.19}$$

To see this take the least r such that $|\xi_r\eta_r| = \min\{|\xi_j\eta_j|\colon j = 1, \ldots, m\} =: L$. Indeed, we then have

$$\begin{aligned} L &\le |(\xi_1\eta_1)\ldots(\xi_m\eta_m)|^{1/m} = [(|\xi_1|^p \ldots |\xi_m|^p)^{1/m}]^{1/p}[(|\eta_1|^q \ldots |\eta_m|^q)^{1/m}]^{1/q} \\ &\le \Big(\frac{1}{m}\sum_{j=1}^m |\xi_j|^p\Big)^{1/p}\Big(\frac{1}{m}\sum_{j=1}^m |\eta_j|^q\Big)^{1/q}, \end{aligned}$$

where the last inequality holds by the theorem of the arithmetic and geometric means (Theorem 9 in [47]). Apply (4.19) to $m = n-1$, $\xi = \{\phi_l\colon l \in \{1, \ldots, n\} \setminus \{k\}\}$ and $\eta = \{\psi_l\colon l \in \{1, \ldots, n\} \setminus \{k\}\}$ if $k \ge 1$ or $m = n$, $\xi = \phi$ and $\eta = \psi$ if $k = 0$. This gives an index $l' \in \{1, \ldots, n\} \setminus \{k\}$ such that $|\phi_{l'}\psi_{l'}| \le C_{p,q}(n)V_p(\phi)V_q(\psi)$, where

$$C_{p,q}(n) = \begin{cases} n^{-(1/p)-(1/q)} & \text{if } k = 0 \\ (n-1)^{-(1/p)-(1/q)} & \text{if } k \in \{1, \ldots, n\}. \end{cases} \tag{4.20}$$

Then let $S_{n-1} := T_{n,l'}$. Since $S_n - S_{n-1} = \phi_{l'}\psi_{l'}$, this yields the bound

$$|S_n| \le C_{p,q}(n)V_p(\phi)V_q(\psi) + |S_{n-1}|.$$

Next, instead of ϕ, ψ and k, consider $\phi^{(n-1)} := \phi'$, $\psi^{(n-1)} := \psi'$ and $k^{(n-1)}$ equal to $k-1$ if $k > l'$, or equal to k if $k < l'$. Notice that $k^{(n-1)} = 1$ if $k = 1$ and $k^{(n-1)} = 0$ if $k = 0$. Applying the same argument to $\psi^{(n-1)}$, $\psi^{(n-1)}$ and $k^{(n-1)}$, one gets a similar inequality for the sum S_{n-1} in terms of a sum S_{n-2} of the same kind and a term $C_{p,q}(n-1)V_p(\phi^{(n-1)})V_q(\psi^{(n-1)})$. By the definitions, $V_p(\phi^{(n-1)}) \le V_p(\phi)$ and $V_q(\psi^{(n-1)}) \le V_q(\psi)$. Proceeding in this way we obtain sums S_m for $m = n, n-1, \ldots, 2$ and the bound

$$|S_n| \le \Big\{C_{p,q}(n) + C_{p,q}(n-1) + \cdots + C_{p,q}(3)\Big\}V_p(\phi)V_q(\psi) + |S_2|. \tag{4.21}$$

To bound S_2 consider two cases depending on whether $k = 0$ or $k \geq 1$. In the first case, $k^{(2)} = 0$. Then applying the preceding argument again we obtain for $l \in \{1, 2\}$, $T_{2,l} = S_R((\phi^{(2)})', (\psi^{(2)})'; 0, 1) = (\phi^{(2)})_1'(\psi^{(2)})_1'$ and $S_2 - T_{2,l} = \phi_l^{(2)}\psi_l^{(2)}$. Thus applying (4.19), we get the bound

$$|S_2| \leq |S_2 - T_{2,l'}| + |T_{2,l'}| \leq \{2^{-(1/p)-(1/q)} + 1\}V_p(\phi)V_q(\psi).$$

In the second case, $S_2 = S_R(\phi^{(2)}, \psi^{(2)}; 1, 2) = \phi_2^{(2)}\psi_2^{(2)}$ if $k^{(2)} = 1$ and $S_2 = -S_L(\phi^{(2)}, \psi^{(2)}; 1) = -\phi_2^{(2)}\psi_1^{(2)}$ if $k^{(2)} = 2$. Then the bound

$$|S_2| \leq V_p(\phi)V_q(\psi)$$

holds. Inserting these bounds and (4.20) into (4.21) we obtain the desired inequality (4.18). □

The Love–Young inequality will be used to establish the existence of the (MRS), (RRS) and (RYS) integrals. It will also yield similar inequalities for the corresponding integrals provided they exist. To prove the existence of these integrals we have to consider more general sums than those in the Love–Young inequality. Let $\kappa = \{x_i\colon i = 0, \ldots, n\}$ be a partition of $[a, b]$ and let $\sigma = \{y_i\colon i = 1, \ldots, n\}$ be an *intermediate partition* of κ; that is, (κ, σ) is a tagged partition as defined in Section 2.1. Recall also from there that for functions f and h defined on $[a, b]$, the sum

$$S_{RS}(\kappa, \sigma) := S_{RS}(f, h; \kappa, \sigma) := \sum_{i=1}^{n} f(y_i)[h(x_i) - h(x_{i-1})] \tag{4.22}$$

is called the *Riemann-Stieltjes* sum based on (κ, σ). The following corollary of the Love–Young inequality will be used in conjunction with Cauchy criteria to prove the existence of the three integrals.

Corollary 4.25. *Let $\kappa' = \{x_j'\colon j = 0, \ldots, k\}$, $\kappa'' = \{x_j''\colon l = 0, \ldots, l\}$ be partitions of $[a, b]$ and let σ', σ'' be two intermediate partitions of κ', κ'', respectively. With the hypotheses of Theorem 4.23, the inequality*

$$|S_{RS}(f, h; \kappa', \sigma') - S_{RS}(f, h; \kappa'', \sigma'')|$$

$$\leq C_{p,q}\Big\{\sum_{j=1}^{k} V_p(f; [x_{j-1}', x_j'])V_q(h; [x_{j-1}', x_j']) + \sum_{j=1}^{l} V_p(f; [x_{j-1}'', x_j''])V_q(h; [x_{j-1}'', x_j''])\Big\}$$

holds, where $C_{p,q} = \zeta(p^{-1} + q^{-1})$ and $\zeta(s) := \sum_{n\geq 1} n^{-s}$.

Proof. Let $\kappa = \{x_i\colon i = 0, \ldots, n\}$ be a partition of $[a, b]$ containing each point from κ', σ', κ'' and σ'' and let

$$S(\kappa) := S_{RS}(f, h; \kappa, \{x_i\colon i = 1, \ldots, n\}) = \sum_{i=1}^{n} f(x_i)[h(x_i) - h(x_{i-1})]. \tag{4.23}$$

For each index $j \in \{0, \ldots, k\}$ of the partition κ', let $i(j) \in \{0, \ldots, n\}$ be the index of κ such that $x_{i(j)} = x_j'$. Suppose $\{y_j'\colon j = 1, \ldots, k\}$ are points of the intermediate partition σ' of κ'.

Since $y'_j \in \{x_{i(j-1)}, \dots, x_{i(j)}\}$ for $j = 1, \dots, k$, by Theorem 4.23, we have

$$\begin{aligned} \left|S_{RS}(f, h; \kappa', \sigma') - S(\kappa)\right| &\leq \sum_{j=1}^{k} \left| f(y'_j)[h(x'_j) - h(x'_{j-1})] - \sum_{i=i(j-1)+1}^{i(j)} f(x_i)[h(x_i) - h(x_{i-1})] \right| \\ &\leq C_{p,q} \sum_{j=1}^{k} V_p(f; [x'_{j-1}, x'_j]) V_q(h; [x'_{j-1}, x'_j]). \end{aligned}$$

This and a similar bound with (κ'', σ'') instead of (κ', σ'), yield Corollary 4.25. $\square$

Theorem 4.26. *Let $f \in \mathcal{W}_p[a, b]$ and $h \in \mathcal{W}_q[a, b]$ for some $p, q > 0$, $p^{-1} + q^{-1} > 1$. Then the integral $\int_a^b f\,dh$ exists*

(1) *in the mesh–Riemann–Stieltjes sense if f and h have no discontinuities at the same point;*
(2) *in the refinement–Riemann–Stieltjes sense if f and h have no common discontinuities on the right of the same point and no common discontinuities on the left of the same point;*
(3) *in the refinement–Young–Stieltjes sense always.*

In whichever of the three senses the integral exists, the inequality

$$\Big| \int_a^b f\,dh \Big| \leq C_{p,q} \|f\|_{[p]} \|h\|_{(q)} \tag{4.24}$$

holds, where $C_{p,q} = \zeta(p^{-1} + q^{-1})$ and $\zeta(s) := \sum_{n \geq 1} n^{-s}$.

Proof. To prove (1) we show that given $\epsilon > 0$ there is a positive δ such that a difference between two Riemann–Stieltjes sums is small depending on ϵ whenever the mesh of partitions is less than δ. To this aim we use the bound of Corollary 4.25 for p' and q' such that $p' > p$, $q' > q$ and $(p')^{-1} + (q')^{-1} > 1$. For any subinterval $[u, v]$ of $[a, b]$, we have

$$v_{p'}(f; [u, v]) \leq \mathrm{Osc}(f; [u, v])^{p'-p} v_p(f; [u, v]) \quad \text{and} \quad v_{q'}(h; [u, v]) \leq \mathrm{Osc}(h; [u, v])^{q'-q} v_q(h; [u, v]).$$

By Proposition 4.9, f and h are regulated functions on $[a, b]$. Let $\epsilon > 0$ be given. Then by statement 2 of Theorem 4.10, there is a finite set of open interval (z_{j-1}, z_j), $j = 1, \dots, m$ separated by singletons $a = z_0 < z_1 < \cdots < z_m = b$ and such that

$$\mathrm{Osc}(f; (z_{j-1}, z_j)) < \epsilon \quad \text{and} \quad \mathrm{Osc}(h; (z_{j-1}, z_j)) < \epsilon \quad \text{for } j = 1, \dots, m. \tag{4.25}$$

Choose a positive δ such that whenever a closed subinterval has length less than δ then it contains at most one of the points z_j, $j = 0, \dots, m$. Since f and h have no discontinuities at the same point, either f or h has an oscillation less than ϵ over each closed interval of length less than δ. Let $\kappa' = \{x'_j \colon j = 0, \dots, k\}$ be a partition of $[a, b]$ with mesh less than δ. Using Lemmas 4.6 and 4.5 and Hölder's inequality, we obtain the bound

$$\begin{aligned} \sum_{j=1}^{k} V_{p'}(f; [x'_{j-1}, x'_j]) V_{q'}(h; [x'_{j-1}, x'_j]) &\leq \theta(\epsilon) \sum_{j=1}^{k} v_p(f; [x'_{j-1}, x'_j])^{1/p'} v_q(h; [x'_{j-1}, x'_j])^{1/q'} \\ &\leq \theta(\epsilon) v_p(f; [a, b])^{1/p'} v_q(h; [a, b])^{1/q'}, \end{aligned}$$

where $\theta(\epsilon) := \epsilon^{1-(p/p')} \vee \epsilon^{1-(q/q')}$. A similar bound holds for a partition $\kappa'' = \{x''_j \colon j = 0, \dots, l\}$ with mesh also less than δ. Thus by Corollary 4.25, it follows that

$$|S_{RS}(f, h; \kappa', \sigma') - S_{RS}(f, h; \kappa'', \sigma'')| \leq 2\theta(\epsilon) C_{p',q'} v_p(f; [a, b])^{1/p'} v_q(h; [a, b])^{1/q'}. \tag{4.26}$$

Since ϵ is arbitrary, by the Cauchy criterion for Riemann–Stieltjes integrability (the analogue of Proposition 6.6 below for (RRS)), the integral $(MRS)\int_a^b f\,dh$ exists.

Let $\{\kappa_m\colon m\geq 1\}$ be a sequence of partitions $\kappa_m=\{x_i^m\colon i=0,\dots,n(m)\}\in PP([a,b])$ such that the mesh $|\kappa_m|\to 0$, and let $\sigma_m:=\{x_i^m\colon i=1,\dots,n(m)\}$. By the Love–Young inequality (4.12), we have for each $m\geq 1$,

$$|S_{RS}(f,h;\kappa_m,\sigma_m)|\leq C_{p,q}V_p(f)V_q(h)+\|f\|_\infty V_q(h)=C_{p,q}\|f\|_{[p]}\|h\|_{(q)}.$$

Taking the limit on the left side as $m\to\infty$ we get inequality 4.24 for the (MRS) integral.

To prove (2) suppose that f and h have no common one-sided discontinuities. As before, given $\epsilon>0$ one can choose a partition $\lambda=\{z_j\colon j=0,\dots,m\}$ of $[a,b]$ such that (4.25) holds. By adding additional points if necessary, we can assume that every second point of λ is a continuity point of f and h. On each side of each other point of λ, either f or h is one-sidedly continuous. Therefore either f or h has an oscillation less than ϵ over any closed subinterval of $[z_{j-1},z_j]$, $j=1,\dots,m$. Taking a refinement $\kappa'=\{x_j'\colon j=0,\dots,k\}$ of λ it follows that (4.26) holds. Thus the (RRS) integrability follows by the Cauchy criterion (Proposition 6.6 below). The inequality (4.24) for the (RRS) follows similarly as for the (MRS) integral.

To prove (3) we observe that each Young sum can be approximated by corresponding Riemann–Stieltjes sums. Indeed, for $\kappa=\{x_i\colon i=0,\dots,n\}\in PP([a,b])$ and open intermediate partition $\sigma=\{y_i\colon i=1,\dots,n\}$ of κ, let $\tilde{\kappa}:=\kappa\cup\{u_{i-1},v_i\colon i=1,\dots,n\}$, where $a=x_0<u_0<y_1<v_1<x_1<u_1<\cdots<x_{n-1}<u_{n-1}<y_n<v_n<x_n=b$, and $\tilde{\sigma}:=\sigma\cup\kappa$. Then letting $v_{i-1}\downarrow x_{i-1}$ and $v_i\uparrow x_i$ for $i=1,\dots,n$, it follows that $S_{RS}(\tilde{\kappa},\tilde{\sigma})\to S_Y(\kappa,\sigma)$. To prove the existence of the (RYS) integral, let $\epsilon>0$ and let $q'>q$ be such that $p^{-1}+(q')^{-1}>1$. By (4.10) and Lemma 4.18, there exists $\lambda=\{z_j\colon j=0,\dots,m\}\in PP([a,b])$ such that

$$\sum_{j=1}^m v_{q'}(h;(z_{j-1},z_j))<[\epsilon/(2C_{p,q'}\|f\|_{(p)})]^{q'}.$$

For $j=1,\dots,m$, let h^j be the function on $[z_{j-1},z_j]$ equal to h on (z_{j-1},z_j), $h^j(z_{j-1}):=h(z_{j-1}+)$ and $h^j(z_j):=h(z_j-)$. Let κ', κ'' be two refinements of λ, let σ', σ'' be open intermediate partitions of κ', κ'', respectively, and for $j=1,\dots,m$, let κ_j', σ_j', κ_j'', σ_j'' be the restrictions of κ', σ', κ'', σ'' to $[z_{j-1},z_j]$. Approximating Young sums by corresponding Riemann-Stieltjes sums, then using Corollary (4.25) and Hölder's inequality, we get

$$\Big|S_Y(f,h;\kappa',\sigma')-S_Y(f,h;\kappa'',\sigma'')\Big|\leq\sum_{j=1}^m\Big|S_Y(f,h^j;\kappa_j',\sigma_j')-S_Y(f,h^j;\kappa_j'',\sigma_j'')\Big|$$

$$\leq 2C_{p,q'}\sum_{j=1}^m v_p(f;[x_{j-1},x_j])^{1/p}\|h^j\|_{(q')}\leq 2C_{p,q'}\|f\|_{(p)}\Big(\sum_{j=1}^m v_{q'}(h;(z_{j-1},z_j))\Big)^{1/q'}<\epsilon.$$

Therefore the (RYS) integral exists by the Cauchy criterion (Proposition 6.10 below). The inequality (4.24) for the (RYS) integral follows, approximating Young sums by Riemann–Stieltjes sums, from (4.12). The proof of Theorem 4.26 is complete. □

In the cases $p=1$ or $q=1$ one can improve the preceding theorem as follows.

Proposition 4.27. *Let $f\in\mathcal{W}_p[a,b]$ and $h\in\mathcal{W}_q[a,b]$ for $(p,q)=(1,\infty)$ or $(\infty,1)$. In either case the (RYS) integral exists and has the bound*

$$\Big|(RYS)\int_a^b f\,dh\Big|\leq\|f\|_{[p]}\|h\|_{(q)}.\tag{4.27}$$

Proof. A proof for the case $(p,q)=(\infty,1)$ is simple (see II.19.3.11 in Hildebrandt [52]) and we omit it. Let $(p,q)=(1,\infty)$ and let $\epsilon>0$. By Theorem 4.10, there exists $\lambda=\{z_j\colon j=0,\dots,m\}\in PP([a,b])$ such that $\mathrm{Osc}(h;(z_{j-1},z_j))<\epsilon$ for $j=1,\dots,m$. Fix any $j=1,\dots,m$ and let h^j be the function on $[z_{j-1},z_j]$ equal to h on (z_{j-1},z_j), $h^j(z_{j-1}):=h(z_{j-1}+)$ and $h^j(z_j):=h(z_j-)$. Let $\kappa_j\in PP([z_{j-1},z_j])$ and let σ_j be an open intermediate partition of κ_j. Then exists $\tilde{\kappa}_j\in PP([z_{j-1},z_j])$ and an intermediate partition $\tilde{\sigma}_j$ of $\tilde{\kappa}_j$ such that

$$|S_Y(f,h^j;\kappa_j,\sigma_j)-S_{RS}(f,h^j;\tilde{\kappa}_j,\tilde{\sigma}_j)|<\epsilon/m.$$

For the sum $S(\tilde{\kappa}_j)$ defined by (4.23), we have

$$|S_{RS}(f,h^j;\tilde{\kappa}_j,\tilde{\sigma}_j)-S(\tilde{\kappa}_j)|\le\epsilon v_1(f;[z_{j-1},z_j]).$$

By the representation (4.14), it follows that

$$|S(\tilde{\kappa}_j)-f(z_j)[h^j(z_j)-h^j(z_{j-1})]|\le\epsilon v_1(f;[z_{j-1},z_j]).$$

Therefore for any refinement κ of λ and open intermediate partition σ of κ, we have

$$\Big|S_Y(f,h;\kappa,\sigma)-\sum_{j=1}^m f(z_j)[h^j(z_j)-h^j(z_{j-1})]\Big|\le\epsilon+2\epsilon\|f\|_{(1)}.$$

Thus the (RYS) integral exists by the Cauchy criterion.

To get the bound (4.27) we apply summation by parts to a Young sum similar to representation (4.13). For $\kappa=\{x_i\colon i=0,\dots,n\}\in PP([a,b])$ and open intermediate partition $\sigma=\{y_i\colon i=1,\dots,n\}$ of κ, let $\{x_l',y_k'\colon l,k=0,\dots,2n+1\}$ be $x_0':=x_0$, $x_1':=x_0+$, $x_{2n'}:=x_n-$, $x_{2n+1}':=x_n$,

$$x_l':=\begin{cases}x_i- & \text{for } l=2i\\ x_i+ & \text{for } l=2i+1,\end{cases}\qquad\text{for } i=1,\dots,n-1 \text{ and }\quad y_k':=\begin{cases}y_i & \text{for } k=2i\\ x_i & \text{for } k=2i+1,\end{cases}$$

for $i=1,\dots,n$, $y_0':=y_1':=x_0$. Then we have

$$\begin{aligned}S_Y(f,h;\kappa,\sigma)-f(a)[h(b)-h(a)] &= S_Y(f-f(a),h;\kappa,\sigma)\\ &= \sum_{l=1}^{2n+1}[f(y_l')-f(y_{l-1}')][h(b)-h(x_{l-1}')].\end{aligned}$$

This yields the bound

$$|S_Y(f,h;\kappa,\sigma)|\le\|f\|_\infty\|h\|_{(\infty)}+\|f\|_{(1)}\|h\|_{(\infty)}.$$

The proof of Proposition 4.27) is complete. $\square$

Theorem 4.26 and the preceding statement yield that the indefinite (RYS) integral

$$f,h\mapsto I(f,h),\qquad\text{where}\qquad I(f,h)(x):=(RYS)\int_a^x f\,dh,\qquad x\in[a,b],$$

maps $\mathcal{W}_p\times\mathcal{W}_q$ into $\mathcal{W}_q$ provided $p^{-1}+q^{-1}>1$ if $p\vee q<\infty$ or $p^{-1}+q^{-1}=1$ if $p\vee q=\infty$. Namely, the following is true:

Corollary 4.28. *Let $f \in \mathcal{W}_p[a,b]$ and $h \in \mathcal{W}_q[a,b]$ for (p,q) such that $p^{-1}+q^{-1} > 1$ if $p \vee q < \infty$, or $(p,q) = (1,\infty)$, or $(p,q) = (\infty,1)$. Then*

$$\|I(f,h)\|_{(q)} \le C_{p,q}\|f\|_{[p]}\|h\|_{(q)}, \tag{4.28}$$

where $C_{p,q}$ is the same constant as in (4.24) if $p \vee q < \infty$ and $C_{1,\infty} = C_{\infty,1} = 1$.

Proof. We give a proof only for the case when $p \vee q < \infty$ because proofs for the other two cases are similar. Let $\kappa = \{x_i\colon i = 0,\dots,n\} \in PP([a,b])$. By additivity of the (RYS) integral (Proposition 6.11 below) and (4.24), it follows that

$$\begin{aligned} s_q(I(f,h);\kappa) &= \sum_{i=1}^{n}\Big|(RYS)\int_{x_{i-1}}^{x_i} f\,dh\Big|^q \le C_{p,q}^q \sum_{i=1}^{n}\|f\|_{[p]}^q v_q(h;[x_{i-1},x_i]) \\ &\le C_{p,q}^q\|f\|_{[p]}^q\|h\|_{(q)}^q, \end{aligned}$$

where the last inequality follows from Lemma 4.6. Thus (4.28) holds. □

The following examples, from L. C. Young [117, Section 7] and Leśniewicz and Orlicz [70], will show that the hypothesis $p^{-1}+q^{-1} > 1$ in the Love–Young inequality cannot be weakened to $p^{-1}+q^{-1} = 1$. For $1 < p,q < \infty$ and $0 \le x \le 1$, define the "Weierstrass functions"

$$S_p(x) := \sum_{k=1}^{\infty} 2^{-k/p}\sin 2^{k+1}\pi x \quad \text{and} \quad C_q(x) := \sum_{k=1}^{\infty} 2^{-k/q}\cos 2^{k+1}\pi x.$$

Then $|S_p(x+h) - S_p(x)| = O(h^{1/p})$ and $|C_q(x+h) - C_q(x)| = O(h^{1/q})$ for $0 \le h < 1$ and $0 \le x \le 1-h$ by Theorem II.4.9 in Zygmund [121]. Therefore $S_p \in \mathcal{W}_p$ and $C_q \in \mathcal{W}_q$.

Theorem 4.29. *If $1/p + 1/q = 1$ and $1 < p,q < \infty$ then $(A)\int_0^1 S_p\,dC_q$ doesn't exist for $(A) = (MRS)$, (RRS), (LS), (RYS), (CY) or (HK).*

Proof. For $n = 1,2,\dots$, let $x_i := i2^{-n}$, $i = 0,\dots,2^n$. Then $\{x_i\colon i = 0,1,\dots,2^n\}$ is a partition of $[0,1]$, and

$$\begin{aligned} T_n &:= \sum_{i=1}^{2^n} S_p(x_i)[C_q(x_i) - C_q(x_{i-1})] \\ &= \sum_{i=1}^{2^n}\sum_{k=1}^{n-1} 2^{-k/p}\sin(2^{k-n+1}\pi i)\left\{\sum_{l=1}^{n-1} 2^{-l/q}[\cos(2^{l-n+1}\pi i) - \cos(2^{l-n+1}\pi(i-1))]\right\} \\ &= \sum_{k=1}^{n-1}\sum_{l=1}^{n-1} 2^{-k/p-l/q}\sum_{i=1}^{2^n}\sin(2^{k-n+1}\pi i)[\cos(2^{l-n+1}\pi i) - \cos(2^{l-n+1}\pi(i-1))]. \end{aligned}$$

Recall the following identities (Chapter I, §1 in Zygmund, 1959):

$$\frac{1}{m} + \sum_{j=1}^{m}\cos jt = \frac{\sin(m+\frac{1}{2})t}{2\sin\frac{1}{2}t} \quad \text{and} \quad \sum_{j=1}^{m}\sin jt = \frac{\cos\frac{1}{2}t - \cos(m+\frac{1}{2})t}{2\sin\frac{1}{2}t},$$

valid for any positive integer m and $t \in (0, 2\pi)$. Using trigonometric formulas, for $k \neq l$, we get

$$\begin{aligned}
&\sum_{i=1}^{2^n} \sin(2^{k-n+1}\pi i)[\cos(2^{l-n+1}\pi i) - \cos(2^{l-n+1}\pi(i-1))] \\
&\quad = \sum_{i=1}^{2^n} \Big\{ \frac{1}{2}(1 - \cos 2^{l-n+1}\pi)[\sin 2^{-n+1}\pi(2^k + 2^l)i + \sin 2^{-n+1}\pi(2^k - 2^l)i] \\
&\qquad\qquad + \frac{1}{2}\sin 2^{l-n+1}\pi[\cos 2^{-n+1}\pi(2^k + 2^l)i - \cos 2^{-n+1}\pi(2^k - 2^l)i]\Big\} = 0.
\end{aligned}$$

Similarly, for $k = l$, we get

$$\begin{aligned}
&\sum_{i=1}^{2^n} \sin(2^{k-n+1}\pi i)[\cos(2^{k-n+1}\pi i) - \cos(2^{k-n+1}\pi(i-1))] \\
&\quad = \sum_{i=1}^{2^n} \Big\{ \frac{1}{2}(1 - \cos 2^{k-n+1}\pi) \sin 2^{k-n+1}\pi i + \frac{1}{2}\sin 2^{k-n+1}\pi[\cos 2^{k-n+1}\pi i - 1]\Big\} \\
&\quad = 2^n \cdot \frac{1}{2}\sin 2^{k-n+1}\pi \geq 2^n \cdot \frac{1}{2} \cdot \frac{2}{\pi} \cdot 2^{k-n+1} = 2^{k+1}.
\end{aligned}$$

Thus

$$T_n \geq 2\sum_{k=1}^{n-1} 2^{k(1-1/p-1/q)} = 2(n-1)$$

if $1/p + 1/q = 1$. Therefore the integral $(A)\int_0^1 S_p\,dC_q$ is undefined for $(A) = (MRS)$. For continuous functions, as S_p and C_q are, the (CY) integral by its definition after (2.2) reduces to the (RRS) integral which in turn equals the (MRS) integral by Theorem 6.3 below. Thus the integral is also undefined for $(A) = (RRS)$, (RYS) and (CY) by the implications shown in Figure 2.1. Since S_p and C_q both have unbounded variation the (LS) integral is undefined. The integral $(HK)\int_0^1 S_p\,dC_q$ is undefined by Proposition F.3 of [28] and comments following it. □

4.5 Interval functions

Functions having as arguments either real numbers or certain sets of real numbers give rise to different ways of defining integrals. In some cases it is possible to establish a correspondence between the resulting integrals. For example, let $[a, b]$ be a closed interval of real numbers, and let $\mathcal{B}[a, b]$ be the σ-algebra of Borel subsets of $[a, b]$. Let $\mathcal{A}(\mathcal{B}[a, b])$ be the class of σ-additive set functions on $[a, b]$; that is, real functions defined and countably additive on $\mathcal{B}[a, b]$. There is a one-to-one correspondence between $\mathcal{A}(\mathcal{B}[a, b])$ and the class $\mathcal{W}_1[a, b]$ of real functions h defined and of bounded total variation on $[a, b]$ such that $h(a) = 0$ and h is right-continuous on (a, b). Let $\mu_h((u, v]) := h(v) - h(u)$ for $u, v \in [a, b]$. The correspondence is given by the Lebesgue-Stieltjes measure $\mu_h^\star$ generated using the Carathéodory construction from the interval function μ_h. This correspondence leads to the Lebesgue-Stieltjes integral $(LS)\int_{[a,b]} f\,dh$ for a $|dh|$-integrable function f and a function $h \in \mathcal{W}_1[a, b]$. On the other hand the refinement–Young–Stieltjes integral $(RYS)\int_a^b f\,dh$ exists and has the same value as the Lebesgue-Stieltjes integral provided $h \in \mathcal{W}_1[a,b]$ and f is regulated on $[a,b]$ (see II.19.3.11 in Hildebrandt [52] for the first statement, and Dudley [23, Lemma 3.3] together with Proposition 6.22 below, for the second statement, as in Figure 2.1). However the (RYS) integral may exist when neither the integrand f nor the integrator h is of bounded total variation.

In this section we discuss functions defined on intervals and describe a correspondence between regulated point functions and upper continuous additive interval functions (Proposition 4.37 below). The results of this section are used in Section 6.4 to establish a relation between the Kolmogorov integral defined for interval functions and the (RYS) integral. Also, the results of this section are used in Chapter 7 for the product integral with respect to an additive interval function.

An interval in $\mathbb{R}$ is a set of any of the following four forms: for $-\infty < c \le d < +\infty$, $(c,d) := \{x \in \mathbb{R}\colon c < x < d\}$, $[c,d) := \{x \in \mathbb{R}\colon c \le x < d\}$, $(c,d] := \{x \in \mathbb{R}\colon c < x \le d\}$ and $[c,d] := \{x \in \mathbb{R}\colon c \le x \le d\}$. Thus for $c \in \mathbb{R}$, $[c,c] = \{c\}$ is a singleton and $(c,c] = [c,c) = (c,c) = \emptyset$. Let $-\infty < a < b < +\infty$. Define the following two classes of subintervals of the closed interval $[a,b]$:

1. $\mathfrak{I}_{os}([a,b]) := \{(c,d), \{c\}\colon a \le c \le d \le b\}$ (open intervals and singletons),
2. $\mathfrak{I}([a,b]) := \{[c,d], (c,d], [c,d), (c,d)\colon a \le c \le d \le b\}$ (arbitrary subintervals).

Kolmogorov [63, p. 680] mentioned the class $\mathfrak{I}_{os}([a,b])$. His results apply to both classes. Many texts on measure theory consider the class $\{(c,d]\colon a \le c \le d \le b\}$.

Any function $\mu\colon \mathfrak{I}([a,b]) \mapsto \mathbb{R}$ will be called an *interval function on* $[a,b]$. In Chapter 7 interval functions are considered with values in a Banach algebra $\mathbb{B}$ with identity $\mathbb{1}$. The class of all $\mathbb{B}$-valued interval functions on $[a,b]$ is denoted by $\mathcal{I}([a,b];\mathbb{B})$, and it is denoted by $\mathcal{I}[a,b]$ if $\mathbb{B} = \mathbb{R}$. All the definitions and results in this section are stated for real–valued interval functions. They also hold for $\mathbb{B}$-valued interval functions when absolute value is replaced by the norm and 1 is replaced by the identity $\mathbb{1}$.

Definition 4.30. An interval function μ on $[a,b]$ will be called *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \mathfrak{I}([a,b])$ are disjoint and $A \cup B \in \mathfrak{I}([a,b])$.

If μ is an additive interval function then clearly $\mu(\emptyset) = 0$.

For two disjoint intervals A and B, $A \prec B$ will mean that $x < y$ for all $x \in A$ and $y \in B$.

Definition 4.31. An interval function μ on $[a,b]$ will be called *multiplicative* if for any intervals A, B such that $A \prec B$ and $A \cup B$ is an interval, we have $\mu(A \cup B) = \mu(A)\mu(B)$, and $\mu(\emptyset) = 1$.

If $\mu(\emptyset) = 1$ were not assumed, it would hold if and only if there is $A \in \mathfrak{I}([a,b])$ such that $\mu(A) \ne 0$ ($\mu(A)$ is invertible in the case μ has values in a Banach algebra). Indeed, if such an A exists then

$$1 = \mu(A)[\mu(A)]^{-1} = \mu(\emptyset \cup A)[\mu(A)]^{-1} = \mu(\emptyset) \cdot 1 = \mu(\emptyset).$$

The converse implication holds because $\mu(\emptyset) = 1$ and $\emptyset \in \mathfrak{I}([a,b])$.

For intervals as for other sets, $A_n \uparrow A$ will mean $A_1 \subset A_2 \subset \cdots$ and $\cup_{n=1}^{\infty} A_n = A$, while $A_n \downarrow A$ will mean $A_1 \supset A_2 \supset \cdots$ and $\cap_{n=1}^{\infty} A_n = A$, and $A_n \to A$ will mean $1_{A_n}(x) \to 1_A(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$.

Definition 4.32. 1. An interval function μ on $[a,b]$ will be called *upper continuous* if $\mu(A_n) \to \mu(A)$ for any $A, A_1, A_2, \ldots \in \mathfrak{I}([a,b])$ such that $A_n \downarrow A$.

2. An interval function μ on $[a,b]$ will be called *upper continuous at* $\emptyset$ if $\mu(A_n) \to \mu(\emptyset)$ for any $A_1, A_2, \ldots \in \mathfrak{I}([a,b])$ such that $A_n \downarrow \emptyset$.

An additive or multiplicative interval function μ on $[a,b]$ is uniquely determined by its restriction to the class $\mathfrak{I}_{os}([a,b])$.

For an interval function μ on $[a,b]$, define two functions $R_{\mu,a}$ and $L_{\mu,a}$ on $[a,b]$ by

$$R_{\mu,a}(x) := \begin{cases} \mu(\emptyset) & \text{if } x = a \\ \mu([a,x]) & \text{if } x \in (a,b] \end{cases} \quad \text{and} \quad L_{\mu,a}(x) := \begin{cases} \mu([a,x)) & \text{if } x \in [a,b) \\ \mu([a,b]) & \text{if } x = b. \end{cases} \tag{4.29}$$

If μ is upper continuous then $R_{\mu,a}$ and $L_{\mu,a}$ are both regulated point functions on $[a,b]$. The converse is not true as the following shows:

Example 4.33. Let $\mu([c,d)) := \mu((c,d]) := 0$ and $\mu([c,d]) := 1$ for $a \le c \le d \le b$, and $\mu((c,d)) := -1$ for $c < d$. Then μ is an additive interval function, $R_{\mu,a} = 1_{(a,b]}$ and $L_{\mu,a} = 1_{\{b\}}$ are regulated, but μ is not upper continuous at $\emptyset$.

Example 4.34. Let μ be an additive interval function on $[a,b]$. For $A \in \mathfrak{I}([a,b])$, let $\nu(A) := \exp\{\mu(A)\}$. Then ν is a multiplicative interval function on $[a,b]$. If μ is upper continuous then ν is also upper continuous. Indeed, letting $A_n \downarrow A$ we have

$$\begin{aligned} |\nu(A_n) - \nu(A)| &= |\exp\{\mu(A_n) - \mu(A)\} - 1| \exp\{\mu(A)\} \\ &\le |\mu(A_n) - \mu(A)| \exp\{|\mu(A_n) - \mu(A)|\} \exp\{\mu(A)\}, \end{aligned}$$

and the claim follows.

An interval A will be called *right-open at d* or in symbols $A = \{\cdot, d)$ if $A = [c,d)$ or (c,d) for some $c < d$, otherwise *right-closed at d*, or $A = \{\cdot, d]$, if $J = [c,d]$ for some $c \le d$ or $A = (c,d]$ for some $c < d$. An interval A will be called *left-open at c* or $A = (c, \cdot\}$ if $A = (c,d)$ or $(c,d]$ for some $c < d$, otherwise *left-closed at c*, or $A = [c,\cdot\}$, if $A = [c,d]$ for some $c \le d$ or $A = [c,d)$ for some $c < d$.

We say that an interval function μ on $[a,b]$ is *bounded* if $\sup\{|\mu(A)|\colon A \in \mathfrak{I}([a,b])\} < \infty$ and μ is *non-degenerate* if $\mu([a,b]) \ne 0$ ($\mu(A)$ is invertible for each $A \in \mathfrak{I}([a,b])$ in the case μ has values in a Banach algebra).

Theorem 4.35. *Let $\mu \in \mathcal{I}[a,b]$ be either additive or bounded and multiplicative. Then the following five statements are equivalent*:

(*a*) *μ is upper continuous*;

(*b*) *μ is upper continuous at $\emptyset$*;

(*c*) *$\mu(A_n) \to \mu(\emptyset)$ whenever open intervals $A_n \downarrow \emptyset$ and*

$$\operatorname{card}\{x \in [a,b]\colon |\mu(\{x\}) - \mu(\emptyset)| > \epsilon\} < \infty \quad \textit{for all } \epsilon > 0; \tag{4.30}$$

(*d*) *$R_{\mu,a}$ is regulated, $R_{\mu,a}(x-) = \mu([a,x))$ for $x \in (a,b]$ and $R_{\mu,a}(x+) = \mu([a,x])$ for $x \in [a,b)$*;

(*e*) *$L_{\mu,a}$ is regulated, $L_{\mu,a}(x-) = \mu([a,x))$ for $x \in (a,b]$ and $L_{\mu,a}(x+) = \mu([a,x])$ for $x \in [a,b)$.*

If, in addition, the multiplicative interval function μ is non-degenerate, then any of the above statements are equivalent to the following two statements:

(*f*) *$\mu(A_n) \to \mu(A)$ whenever intervals $A_n \uparrow A$*;

(*g*) *$\mu(A_n) \to \mu(A)$ whenever intervals $A_n \to A \ne \emptyset$.*

Proof. The proofs will be given mainly in the additive case. Those for μ multiplicative are similar.

$(a) \Leftrightarrow (b)$. Clearly, (a) implies (b). For $(b) \Rightarrow (a)$, let intervals $A_n \downarrow A$. Then $A_n = B_n \cup A \cup C_n$ for intervals $B_n \prec A \prec C_n$, with $B_n \downarrow \emptyset$ and $C_n \downarrow \emptyset$. So $\mu(A_n) \to \mu(A)$ by additivity or multiplicativity.

$(b) \Rightarrow (c)$. The first part of statement (c) is clear. For the second part, suppose there exists $\epsilon > 0$ and an infinite sequence $\{u_j : j \geq 1\}$ of different points of (a, b) such that

$$|\mu(\{u_j\}) - \mu(\emptyset)| > \epsilon \quad \text{for all } j \geq 1. \tag{4.31}$$

Then there exists $u \in [a, b]$ and a subsequence $\{u_{j'} : j' \geq 1\}$ such that either $u_{j'} \downarrow u$ or $u_{j'} \uparrow u$ as $j' \to \infty$. In the first case, if μ is additive then

$$\mu(\{u_{j'}\}) = \mu((u, u_{j'}]) - \mu((u, u_{j'})) \to 0 \quad \text{as } j' \to \infty$$

because both $(u, u_{j'}] \downarrow \emptyset$ and $(u, u_{j'}) \downarrow \emptyset$. This contradicts (4.31). Similarly, if μ is multiplicative then we have

$$\mu((u, u_{j'}]) - 1 = \mu((u, u_{j'}))\mu(\{u_{j'}\}) - 1 = [\mu((u, u_{j'})) - 1]\,\mu(\{u_{j'}\}) + \mu(\{u_{j'}\}) - 1.$$

Due to the boundedness assumption, $\sup_j |\mu(\{u_j\})| < \infty$. Thus the left side and the first term on the right side tend to zero. Hence $\mu(\{u_{j'}\}) \to 1$ as $j' \to \infty$, also a contradiction. Therefore (4.30) holds in the first case. The proof in the second case is symmetric.

For $(c) \Rightarrow (b)$, let intervals $A_n \downarrow \emptyset$. Then for some u, either for all sufficiently large n, A_n is left-open at $u \in [a, b)$, or A_n is right-open at $u \in (a, b]$. Using additivity if μ is additive, or continuity of the multiplication if μ is multiplicative, in each of the two cases $\mu(A_n) \to \mu(\emptyset)$ follows by (c).

$(b) \Leftrightarrow (d) \Leftrightarrow (e)$. Taking into account the boundedness assumption when μ is multiplicative, the implications $(b) \Rightarrow (d)$ and $(b) \Rightarrow (e)$ are clear. We prove $(d) \Rightarrow (b)$ only, because the proof of $(e) \Rightarrow (b)$ is similar. Let intervals $A_n \downarrow \emptyset$. Then for some u, either for all sufficiently large n, A_n is left-open at $u \in [a, b)$, or A_n is right-open at $u \in (a, b]$. By assumption, in the first case we have $\lim_n \mu(A_n) = R_{\mu,a}(u+) - R_{\mu,a}(u) = 0$ and in the second case we have $\lim_n \mu(A_n) = R_{\mu,a}(u-) - R_{\mu,a}(u-) = 0$.

For $(b) \Rightarrow (g)$, let intervals $A_n \to A \neq \emptyset$. If $A = (u, v)$ then $\{u\} \prec A_n \prec \{v\}$ for n large enough. For such n, there are intervals C_n and D_n with $\{u\} \prec C_n \prec A_n \prec D_n \prec \{v\}$ and $C_n \cup A_n \cup D_n = A$. Also for such n, we have either $C_n = \emptyset$ or $C_n = (u, \cdot\}$, and either $D_n = \emptyset$ or $D_n = \{\cdot, v)$. Clearly $C_n \to \emptyset$. If $N := \{n \colon C_n \neq \emptyset\}$ is infinite, there is a function $j \mapsto n(j)$ onto N such that $C_{n(j)} \downarrow \emptyset$. Thus $\mu(C_n) \to 0$. Similarly $\mu(D_n) \to 0$. Therefore $\mu(A_n) = \mu(A) - \mu(C_n) - \mu(D_n) \to \mu(A)$. A similar argument works for other cases $A = (u, v]$, $[u, v)$ and $[u, v]$.

Clearly, (g) implies (f). For $(f) \Rightarrow (b)$, let intervals $A_n \downarrow \emptyset$. Then $A_1 = B_n \cup A_n \cup C_n$ for some intervals B_n, C_n with $B_n \prec A_n \prec C_n$, and $B_n \uparrow B$, $C_n \uparrow C$ for some intervals B, C with $A_1 = B \cup C$. If μ is additive then $\mu(A_n) \to 0 = \mu(\emptyset)$. If μ is multiplicative then $\mu(B_n)\mu(A_n)\mu(C_n) = \mu(B)\mu(C)$, so $\mu(A_n) = \mu(B_n)^{-1}\mu(B)\mu(C)\mu(C_n)^{-1} \to 1 = \mu(\emptyset)$. The proof of Theorem 4.35 is complete. $\square$

Let $\mathcal{R}[a, b]$ be the set of all regulated functions on $[a, b]$ considered in Section 4.2. Next we show that there is a one-to-one correspondence between additive upper continuous interval functions on $[a,b]$ and classes of regulated point functions on $[a,b]$. For $h \in \mathcal{R}[a,b]$ define an interval function μ_h on $[a,b]$ corresponding to h by

$$\begin{cases} \mu_h((u,v)) := h(v-) - h(u+) \text{ for } a \le u < v \le b, \\ \mu_h(\{u\}) := h(u+) - h(u-) \text{ for } a < u < b, \\ \mu_h(\{a\}) := h(a+) \text{ and } \mu_h(\{b\}) := h(b) - h(b-). \end{cases} \tag{4.32}$$

Theorem 4.36. *The following statements about an additive interval function μ on $[a,b]$ are equivalent*:
(*a*) *μ is upper continuous*;
(*b*) *there exists $h \in \mathcal{R}[a,b]$ such that $h(a) = 0$ and $\mu_h = \mu$ on $\Im([a,b])$*;
(*c*) *there exists $h \in \mathcal{R}[a,b]$ such that $h(b) = 0$ and $\mu_h = \mu$ on $\Im([a,b])$*.

Proof. (*a*) $\Rightarrow$ (*b*). By statement (*c*) of Theorem 4.35, the set $D_\mu := \{x \in (a,b)\colon \mu(\{x\}) \neq 0\}$ is countable. Let D be a countable subset of (a,b) which includes the set D_μ. Let $\{h_x\colon x \in D\}$ be any set of real numbers such that $\operatorname{card}\{x \in D\colon |h_x - \mu([a,x))| \wedge |h_x - \mu([a,x])| > \epsilon\} < \infty$ for each $\epsilon > 0$. For example, one can take h_x equal either to $\mu([a,x])$ or to $\mu([a,x)$. Define a real-valued function h on $[a,b]$ by

$$\begin{cases} h(x) := \mu([a,x)) = \mu([a,x]) \text{ for } x \in (a,b) \setminus D, \\ h(a) := 0,\ h(b) := \mu([a,b]), \\ h(x) := h_x \text{ for } x \in D. \end{cases} \tag{4.33}$$

To show that $h \in \mathcal{R}([a,b])$ let $x \in (a,b]$ and $\epsilon > 0$. There exists $z \in [a,x)$ such that $|\mu((y,x))| < \epsilon$ and $|h_y - \mu([a,y])| < \epsilon$ for all $y \in (z,x)$ by (*a*) and by the definition of $\{h_x\colon x \in D\}$, respectively. Thus $|h(y) - \mu([a,x))| = |\mu((y,x))| < \epsilon$ for $y \in (z,x) \setminus D$ and $|h(y) - \mu([a,x))| \le |h_y - \mu([a,y])| + |\mu((y,x))| < 2\epsilon$ for $y \in (z,x) \cap D$. Since $\epsilon > 0$ is arbitrary, $h(x-) = \lim_{y\uparrow x} h(y) = \mu([a,x))$ for $x \in (a,b]$. Similarly it follows that $h(x+) = \lim_{y\downarrow x} h(y) = \mu([a,x])$ for $x \in [a,b)$. Thus $h \in \mathcal{R}[a,b]$ and $h(a) = 0$. Concerning the interval function μ_h corresponding to h, it now follows that $\mu_h((u,v)) = h(v-) - h(u+) = \mu((u,v))$ for $a \le u < v \le b$ and $\mu_h(\{u\}) = h(u+) - h(u-) = \mu(\{u\})$ for $a \le u \le b$. Therefore (*b*) holds.

(*b*) $\Rightarrow$ (*a*). To show that μ is upper continuous, by Theorem 4.35, it is enough to show its statement (*c*). Let open intervals $A_n \downarrow \emptyset$. Then there exist $\{u, v_n\colon n \ge 1\} \subset [a,b]$ such that, for all large enough n, either $A_n = (v_n, u)$ with $v_n \uparrow u$ or $A_n = (u, v_n)$ with $v_n \downarrow u$. By (*c*), for some $h \in \mathcal{R}[a,b]$, we have $\mu(A_n) = h(u-) - h(v_n+) \to 0$ as $n \to \infty$ in the first case and $\mu(A_n) = h(v_n-) - h(u+) \to 0$ as $n \to \infty$ in the second case. Similarly, it follows that $\operatorname{card}\{x \in [a,b]\colon |\mu(\{x\})| > \epsilon\} < \infty$ for all $\epsilon > 0$. Thus (*a*) holds by Theorem 4.35.

The proof of the equivalence (*a*) $\Leftrightarrow$ (*c*) is similar to the above proof and therefore is omitted. The proof of Theorem 4.36 is thus complete. ☐

For an additive upper continuous interval function μ on $[a,b]$, let

$$[\mu]_a := \{h \in \mathcal{R}[a,b]\colon\ h(a) = 0 \text{ and } \mu_h = \mu \text{ on } \Im([a,b])\}. \tag{4.34}$$

By property (*b*) of Theorem 4.36, the set $[\mu]_a$ is nonempty. If $h \in [\mu]_a$ then any regulated function equal to h except possibly at jump points of h also belongs to $[\mu]_a$. For an additive upper continuous interval function μ on $[a,b]$ one can single out two elements from $[\mu]_a$. Let $R_{\mu,a}$ and $L_{\mu,a}$ be defined by (4.33) using $h_x = \mu([a,x])$ and $h_x = \mu([a,x))$ for $x \in D = D_\mu$, respectively. Then $R_{\mu,a}$ is right-continuous on (a,b) and $L_{\mu,a}$ is left-continuous on (a,b). Also both functions coincide with the functions in (4.29).

The next fact then follows.

Proposition 4.37. *Let $a < b$. Then one-to-one linear operators between the vector spaces:*
(*a*) *the set of all additive upper continuous interval functions on $[a, b]$;*
(*b*) $\{h \in \mathcal{R}[a, b]\colon\ h(a) = 0$ *and h is right-continuous on $(a, b)\}$;*
(*c*) $\{h \in \mathcal{R}[a, b]\colon\ h(a) = 0$ *and h is left-continuous on $(a, b)\}$,*
are given for (*a*) *and* (*b*) *by $h = R_{\mu,a}$ and $\mu = \mu_h$, and between* (*a*) *and* (*c*) *by $h = L_{\mu,a}$ and $\mu = \mu_h$.*

The p-variation for interval functions is defined similarly to the p-variation for point functions.

Definition 4.38. Let $\mu \in \mathcal{I}[a, b]$ and $1 \le p < \infty$. The *p-variation* of μ will be defined by

$$v_p(\mu) := v_p(\mu; [a, b]) := \sup\Big\{\sum_{i=1}^{n}\Big|\mu\Big(\bigcup_{j=1}^{i} A_j\Big) - \mu\Big(\bigcup_{j=1}^{i-1} A_j\Big)\Big|^p\colon (A_1, \dots, A_n) \in \mathrm{IP}\,([a, b])\Big\},$$

where a union over the empty set of indices is defined as the empty set, and as before $\mathrm{IP}\,([a, b])$ is the set of interval partitions of $[a, b]$. The class of all interval functions on $[a, b]$ with bounded p-variation will be denoted by $\mathcal{I}_p = \mathcal{I}_p[a, b]$.

For an additive interval function μ, clearly

$$v_p(\mu) = \sup\Big\{\sum_{i=1}^{n}|\mu(A_i)|^p\colon (A_1, \dots, A_n) \in \mathrm{IP}\,([a, b])\Big\}.$$

The next statement relates the p-variation for interval functions and point functions.

Theorem 4.39. *Let $\mu \in \mathcal{I}[a, b]$ be upper continuous. Suppose that μ is either additive, or bounded and multiplicative. Then the following three statements are equivalent for μ:*

$$(a)\ \mu \in \mathcal{I}_p[a, b]; \qquad (b)\ R_{\mu,a} \in \mathcal{W}_p[a, b]; \qquad (c)\ L_{\mu,a} \in \mathcal{W}_p[a, b].$$

Also, if any of (*a*), (*b*) *or* (*c*) *holds, then $v_p(\mu) = v_p(L_{\mu,a}) = v_p(R_{\mu,a})$.*

Proof. Let μ be an additive upper continuous interval function on $[a, b]$. Since $R_{\mu,a}(a) = \mu(\emptyset) = 0$, for any $a = x_0 < x_1 < \cdots < x_n = b$, we then have

$$\sum_{i=1}^{n}|R_{\mu,a}(x_i) - R_{\mu,a}(x_{i-1})|^p \;=\; |\mu([a, x_1]) - \mu(\emptyset)|^p + \sum_{i=2}^{n}|\mu([a, x_i]) - \mu([a, x_{i-1}])|^p \;\le\; v_p(\mu).$$

Thus $v_p(R_{\mu,a}) \le v_p(\mu)$. Likewise, $v_p(L_{\mu,a}) \le v_p(\mu)$. Conversely, by statement (*d*) of Theorem 4.35, for an interval $A = [a, \cdot\}$ left–closed at a, we have $\mu(A) = R_{\mu,a}(t)$, where $t = v-$ if $A = [a, v)$, $t = v$ if $A = [a, v]$. Thus for any $\mathfrak{I}$-partition $\{A_i\colon i = 1, \dots, n\}$, with A_i ordered from left to right, there are $a = u_0 < u_1 < \cdots < u_n = b$ (where we write $u- < u < u+$ by definition) such that $\mu(\cup_{j\le i}A_j) = R_{\mu,a}(u_i)$ for each i, where $u_i = v_i-$ for some v_i if $A_i = \{\cdot, v_i)$. Replacing each u_i by $w_{i,m} \uparrow u_i$ if $u_i = v_i-$ and u_1 by $w_{1,m} \downarrow a$ if $u_1 = a+$, letting $w_{i,m} = u_i$ otherwise, we get

$$\sum_{i=1}^{n}\Big|\mu\Big(\bigcup_{j=1}^{i} A_j\Big) - \mu\Big(\bigcup_{j=1}^{i-1} A_j\Big)\Big|^p = \lim_{m\to\infty}\sum_{i=1}^{n}|R_{\mu,a}(w_{i,m}) - R_{\mu,a}(w_{i-1,m})|^p \;\le\; v_p(R_{\mu,a}).$$

Thus $v_p(\mu) \le v_p(R_{\mu,a})$. Likewise, $v_p(\mu) \le v_p(L_{\mu,a})$. So $v_p(\mu) = v_p(R_{\mu,a}) = v_p(R_{\mu,a})$ and (*a*), (*b*), (*c*) are equivalent. □

An interval function need not be upper continuous even if it is additive and has bounded p-variation for some $p < \infty$.

Example 4.40. On $[0, 2]$, let μ be an interval function such that $\mu(J) = 0$ for any interval included in $[0, 1]$ or in $[d, 1]$ for any $d > 1$. Let $\mu((1, c\}) = 1$ for $1 < c \le 2$. Then μ is additive and of bounded p-variation for any p, but not upper continuous at $\emptyset$.

For $1 \le p < \infty$ and $\mu \in \mathcal{I}_p$, let $\|\mu\|_{(p)} := v_p(\mu)^{1/p}$, $\|\mu\|_\infty := \sup\{|\mu(J)|: J \in \mathfrak{I}([a, b])\}$ and $\|\mu\|_{[p]} := \|\mu\|_{(p)} + \|\mu\|_\infty$. Similarly to Theorem 4.2, the next fact is easy to prove.

Theorem 4.41. *For* $1 \le p < \infty$, $(\mathcal{I}_p, \|\cdot\|_{[p]})$ *is a Banach space.*

4.6 On the order of decrease of Fourier coefficients

Much of the early work on Hölder conditions and on p-variation was done with Fourier series as one application in view, e.g. Lipschitz (1864), Wiener (1924), L. C. Young (1936). Let $T^1 := \{z: |z| = 1\} = \{e^{i\theta}: 0 \le \theta < 2\pi\}$ be the unit circle in the complex plane $\mathbb{C}$. On T^1 let $d\mu = d\theta/2\pi$ be the rotationally invariant probability measure. For $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, the functions $z \mapsto z^n$, or $\theta \mapsto e^{in\theta}$ for $n \in \mathbb{Z}$ form an orthonormal basis of complex $L^2(T^1, \mu)$ (e. g. [22], Proposition 7.4.2). For $f \in \mathcal{L}^1(T^1, \mu)$, the Fourier coefficients $c_n := \int f(z) z^{-n}\, d\mu$ are defined, yielding the Fourier series

$$f \sim \sum_{n \in \mathbb{Z}} c_n z^n.$$

Kolmogorov [62] showed that for $f \in \mathcal{L}^1(T^1, \mu)$ the Fourier series may diverge everywhere. Carleson [10] for $p = 2$ and R. A. Hunt [56] for $p > 1$ showed that if $f \in \mathcal{L}^p(T^1, \mu)$, then the Fourier series converges almost everywhere to f.

If f is real–valued then it is usual to rewrite the Fourier series as

$$f \sim a_0 + \sum_{n=1}^{\infty} a_n \cos(n\theta) + b_n \sin(n\theta),$$

where $a_n = (c_n + c_{-n})/2$, $b_n = (c_n - c_{-n})/(2i)$, $n \ge 1$, with a_n and b_n real. For $1 \le p \le 2$, the Hausdorff–(W.H.)Young inequality states that if $f \in L^p(T^1, \mu)$ then the sequences $\{a_n\}$, $\{b_n\}$ and $\{c_n\}$ are in $\ell^q = \{\{d_n\}: \sum_n |d_n|^q < \infty\}$, where $1/p + 1/q = 1$. For $2 < p \le \infty$, stronger integrability conditions do not provide any improvement beyond ℓ^2 on the order of the Fourier coefficients. For f continuous on T^1 (= continuous in θ and periodic of period 2π), Carleman [9] showed that $\{c_n\}$ need not be in $\ell^{2-\epsilon}$ for any $\epsilon > 0$. Orlicz [87] showed that for any $\{d_n: n \in \mathbb{Z}\} \notin \ell^2$ one can have $\sum_n |c_n d_n| = +\infty$. Bary [4, pp. 338-340] gives a proof.

If f is periodic of period 2π and on $[0, 2\pi]$ satisfies the Hölder condition

$$|f(\theta') - f(\theta'')| \le K|\theta' - \theta''|^\alpha, \tag{4.35}$$

where $0 < \alpha \le 1$, then $|c_n| = O(|n|^{-\alpha})$ as $n \to \infty$. Functions Hölder of order α on a bounded interval have bounded $1/\alpha$–variation. So more generally, let $f \in \mathcal{W}_p$ for some $1 < p < \infty$ and let $g_n(\theta) = \cos(n\theta)$ or $\sin(n\theta)$ on $[0, 2\pi]$. It is easily seen that $v_q(g_n) = 2n \cdot 2^q$ for any $q < \infty$, so that one can take q such that $p^{-1} + q^{-1} > 1$. Then by the Love–Young inequality

$$|a_n| = \Big|\frac{1}{\pi}\int_0^{2\pi} f(\theta)\, d\sin(n\theta)/n\Big| \le 2^{2+1/p} C_{p,q} \|f\|_{[p]} n^{-1+1/q} = O(n^{-r})$$

for any $r < 1/p$.

In fact, Marcinkiewicz [78, Theorem 3] proved that this is true for $r = 1/p$. Indeed, due to periodicity

$$c_n = a_n + ib_n = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)e^{-in\theta}\,d\theta = \frac{1}{4\pi}\int_0^{2\pi}\Big[f(\theta) - f(\theta + \frac{\pi}{n})\Big]e^{-in\theta}\,d\theta$$

for $n \geq 1$, so that

$$|c_n| \leq \frac{1}{4\pi}\int_0^{2\pi}\Big|f(\theta) - f(\theta + \frac{\pi}{n})\Big|\,d\theta. \tag{4.36}$$

On the other hand, Marcinkiewicz [78, Theorem 4] proved (see also p. 260 in [117]) that for some $C < \infty$,

$$\Big(\int_0^{2\pi}|f(\theta+\phi) - f(\theta)|^p\,d\theta\Big)^{1/p} \leq C|\phi|^{1/p}\|f\|_{(p)} \tag{4.37}$$

for any ϕ. This together with (4.36) implies $|c_n| = O(n^{-1/p})$.

Marcinkiewicz [78, Theorem 5] also proved that the Fourier series of f converges pointwise everywhere to $[f(x-) + f(x+)]/2$ if the p-variation of f is bounded for some $p < \infty$. Wiener [113] had proved this fact assuming that $p \leq 2$. Recall that the Fourier series of a continuous function can diverge at some points as shown by du Bois–Reymond in 1876, see e.g. [22, Proposition 7.4.3]. A continuous function f has its Fourier series converging to it uniformly if f is of bounded p-variation for some $p < \infty$ or if f is of bounded ϕ-variation for $\phi(u) := \exp\{-u^{-\alpha}\}$, where $0 < \alpha < 1$ (Salem [99], see also p. 310 in Bary [4]).

Chapter 5

Stochastic processes and p-variation

It will be seen that several classes of stochastic processes have sample functions of bounded p-variation for some p. But we did not find any results on p-variation for general Markov processes, which do have regulated and right–continuous paths under mild conditions.

5.1 Stochastic processes with regulated sample functions

As usual, a *stochastic process* $X = \{X(t)\colon t \geq 0\}$ is a family of random variables $X(t) = X(t, \cdot)$ defined on a complete probability space $(\Omega, \mathcal{F}, \Pr)$. For each $\omega \in \Omega$, the function $X(\cdot) = X(\cdot, \omega)$ is called a *sample function* of X. A stochastic process $X = \{X(t)\colon t \geq 0\}$ has regulated sample functions if $X(\cdot, \omega) \in \mathcal{R}[0, \infty)$ for almost all $\omega \in \Omega$. In this case we write $X \in \mathcal{R}[0, \infty)$.

Let $X \in \mathcal{R}[0, \infty)$ and let $\Omega_X := \{\omega \in \Omega\colon X(\cdot, \omega) \in \mathcal{R}[0, \infty)\}$. Due to completeness of the underlying probability space, $\Omega_X \in \mathcal{F}$ and $\Pr(\Omega_X) = 1$. For each $t \in [0, \infty)$, let

$$X_+(t) := X_+(t, \omega) := \begin{cases} X(t+, \omega) & \text{if } \omega \in \Omega_X \\ X(t, \omega) & \text{otherwise.} \end{cases}$$

Similarly define $X_-(t)$ for each $t \in (0, \infty)$ and let $X_-(0) := X(0)$. For each $t \in [0, \infty)$, $X_+(t)$ is a random variable because it is the limit for $\omega \in \Omega_X$ of the random variables $X(r_n)$ for r_n rational, $r_n \downarrow t$, and equals $X(t)$ otherwise. Likewise, $X(t-)$ is a random variable. Therefore $X_+ = \{X_+(t)\colon t \geq 0\}$ and $X_- = \{X_-(t)\colon t \geq 0\}$ are stochastic processes on the same probability space as X.

Recall that $\Delta^+ X(t) := X^+(t) - X(t)$ and $\Delta^- X(t) := X(t) - X^-(t)$. For each $t \in [0, \infty)$, let

$$\Omega_d(t) := \{\omega \in \Omega\colon \text{ either } \Delta^- X(t) \neq 0 \text{ or } \Delta^+ X(t) \neq 0\}.$$

A point $t \in [0, \infty)$ is called a point of *fixed discontinuity* if $\Pr(\Omega_d(t)) > 0$. If t is not a point of fixed discontinuity then

$$\lim_{s \to t} X(s) = X(t) \qquad \text{almost surely.} \tag{5.1}$$

Indeed, for each $t \in (0, \infty)$ and $\omega \in \Omega_X \setminus \Omega_d(t)$, we have

$$\lim_{s \uparrow t} X(s, \omega) = X(t-, \omega) = X(t, \omega) = X(t+, \omega) = \lim_{s \downarrow t} X(s, \omega).$$

Since $\Pr(\Omega_X \setminus \Omega_d(t)) = 1$, (5.1) holds when $t \in (0, \infty)$. The same argument yields (5.1) when $t = 0$. If the stochastic process X has no points of fixed discontinuity then the three processes X_-, X_+ and X are modifications of each other, that is

$$\Pr(\{X(t) = X(t-)\}) = \Pr(\{X(t) = X(t+)\}) = 1$$

for each $t \in [0, \infty)$.

A point $t \in [0, \infty)$ is a point of *stochastic continuity* of X if $X(s) \to X(t)$ in probability as $s \to t$.

Theorem 5.1. *Let $X \in \mathcal{R}[0, \infty)$. The set of points of fixed discontinuity of X is at most countable and coincides with the set of points of stochastic discontinuity of X.*

Proof. Since the limit of a sequence convergent in probability is unique almost surely, one can show that X is not stochastically continuous at some $t \in [0, \infty)$ if t is a point of fixed discontinuity. Also, X is stochastically continuous at t whenever t is not a point of fixed discontinuity. This yields the second part of the claim. The first part follows from Theorem 11.1, Ch. VII, of Doob [18]. $\square$

By the preceding theorem, the set of fixed discontinuities of any stochastic process with regulated sample functions is at most countable. However, almost every sample function of such a process may have a non-fixed discontinuity, e.g. $X_t(\omega) = 1_{t \geq \omega}$, where ω has a uniform distribution in $[0, 1]$.

5.2 Martingales

Let $\psi_1(u) := u^2/\log\log(1/u)$ for $0 < u \leq e^{-e}$. Then $\psi_1(u)$ can be defined for $u > e^{-e}$ so that ψ_1 is a strictly increasing continuous convex function from $[0, \infty)$ onto itself. For Brownian motion $B(t) := B_t$ and $0 < T < \infty$, Taylor [106] showed that $v_{\psi_1}(B; [0, T]) < \infty$ almost surely while if $\psi_1(t) = o(\phi(t))$ as $t \downarrow 0$, $v_\phi(B; [0, T]) = +\infty$ almost surely. Thus almost surely

$$v_p(B; [0, T]) \begin{cases} < +\infty & \text{if } p > 2 \\ = +\infty & \text{if } p \leq 2. \end{cases}$$

Let $(\Omega, \mathcal{A}, \Pr)$ be a probability space. Then a *filtration* $\mathbb{F} = \{\mathcal{F}_t \colon t \geq 0\}$ is a family of sub–σ–algebras of $\mathcal{A}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for $0 \leq s \leq t$. The filtration $\mathbb{F}$ is *right–continuous* iff $\mathcal{F}_{t+} := \cap_{s>t}\mathcal{F}_s = \mathcal{F}_t$ for all $t \geq 0$. A real–valued stochastic process $X = \{X_t \colon t \geq 0\}$, is *adapted* to $\mathbb{F}$ iff X_t is $\mathcal{F}_t$ measurable for each t. A pair $(X, \mathbb{F})$ is a *martingale* iff $E(X_t|\mathcal{F}_s) = X_s$ almost surely for $0 \leq s \leq t$. Thus X is automatically adapted. It is a *supermartingale* if it is adapted and $E(X_t|\mathcal{F}_s) \leq X_s$ almost surely for $0 \leq s \leq t$. A stochastic process $Y = \{Y_t \colon t \geq 0\}$ is a *modification* of X iff for each $t \geq 0$, $X_t = Y_t$ almost surely.

It is known that if $\mathbb{F}$ is right–continuous, then a supermartingale $(X, \mathbb{F})$ has a modification whose paths $t \mapsto Y_t(\omega)$ are regulated and right–continuous if and only if $t \mapsto EX_t$ is right–continuous (Meyer, [82, p. 95]). For a martingale EX_t is constant, so every martingale $(X, \mathbb{F})$ with $\mathbb{F}$ right–continuous has a right–continuous modification $M = \{M_t \colon t \geq 0\}$. For any martingale $(X, \mathbb{F})$ and almost all $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ has left and right limits at each point through the countable dense set of rational numbers. Thus since $t \mapsto M_t(\omega)$ is right–continuous, it is automatically also regulated (has left limits).

Theorem 5.2. *Let $p > 2$, $0 < T < \infty$ and let $M = \{M_t: 0 \le t \le T\}$ be a right–continuous martingale. Then M has almost all paths of bounded p-variation.*

Proof. Monroe [84] showed that in distribution, $M_t = B_{\tau(t)}$, $0 \le t \le T$, where $B = \{B_u: u \ge 0\}$ is a Brownian motion and $\tau(t)$ are stopping times with $\tau(s) \le \tau(t)$, $0 \le s \le t \le T$. It follows that $v_p(M; [0,T]) \le v_p(B; [0, \tau(T)]) < +\infty$ almost surely. □

While a local martingale plus a process locally of bounded variation yields a semimartingale (Doléans–Dade and Meyer [16] and Dellacherie and Meyer [14]) Theorem 5.2 allows one to add a process of bounded p-variation for $p > 2$ and get a sum with the same p-variation property.

If $EM_t = 0$, $0 \le t \le T$, and $EM_T^2 < \infty$ then $E\tau(T) = EM_T^2$ (Monroe, [84, Theorems 5 and 11]). Thus for any set of martingales $M_i = \{M_{i,t}: 0 \le t \le T\}$, if $\sup_i EM_{i,T}^2 < +\infty$ then $v_p(M_i; [0,T])$ is bounded in probability uniformly in i.

5.3 Gaussian stochastic processes

Let $X = \{X(t): t \ge 0\}$ be a Gaussian stochastic process, and let $\sigma_X(s,t) := E|X(s) - X(t)|$. For $0 < T < \infty$, let

$$v_p(\sigma_X; [0,T]) := \sup\Big\{\sum_{i=1}^n \sigma_X(t_{i-1}, t_i)^p: \{t_i: i = 0, \dots, n\} \in \mathrm{PP}\,([0,T])\Big\}$$

For a function f on $[a,b]$, define the *index of p-variation* by

$$\upsilon(f; [a,b]) := \begin{cases} \inf\{p > 0: v_p(f) < \infty\} & \text{if the set is nonempty} \\ +\infty & \text{otherwise.} \end{cases}$$

Using the p-variation index the result of Jain and Monrad [59, Theorem 3.2] can be stated as follows:

Theorem 5.3. *Let X be a separable mean zero Gaussian process. Then with probability 1*

$$\upsilon(X; [0,T]) = \inf\{p > 0: v_p(\sigma_X; [0,T]) < \infty\}.$$

Fractional Brownian motion A fractional Brownian motion $B_H = \{B_H(t): t \ge 0\}$ with the Hurst index $H \in (0,1)$ is a mean zero Gaussian process with the covariance function

$$E[B_H(t)B_H(s)] = \frac{1}{2}\Big[t^{2H} + s^{2H} - |t-s|^{2H}\Big] \qquad \text{for } t, s \ge 0 \tag{5.2}$$

and $B_H(0) = 0$ almost surely. Since the right side of (5.2) is equal to $t \wedge s$ for $H = 1/2$, B_H is a Brownian motion in this case. It follows from (5.2) that the incremental variance σ_H of B_H is given by

$$\sigma_H(u)^2 := E[B_H(t+u) - B_H(t)]^2 = u^{2H} \qquad \text{for } t, u \ge 0,$$

and that B_H has stationary increments. For $0 < T < \infty$ and for each $\epsilon > 0$, let $N_H([0,T]; \epsilon)$ be the smallest integer n such that intervals $\{[a_i, b_i]: i = 1, \dots, n\}$ with $\sigma_H(b_i - a_i) \le 2\epsilon$ for $i = 1, \dots, n$, cover $[0,T]$. Then there is a finite constant C such that $N_H([0,T]; \epsilon) \le C\epsilon^{-1/H}$ for all $\epsilon \in (0,1]$. By Theorem 1.1 of Dudley [20], there is a version of B_H with continuous sample functions on $[0,T]$. We assume throughout that B_H is such a sample-continuous version. A stronger statement follows from Corollary 2.3 of Dudley [20]. Namely, the function $\rho_H(u) := u^H\sqrt{|\log u|}$ for $u \geq 0$ is a sample modulus for B_H. That is for almost all $\omega \in \Omega$ there is a $K_\omega < \infty$ such that

$$|B_H(t,\omega) - B_H(s,\omega)| \leq K_\omega \rho_H(|t-s|) \qquad \text{for } t,s \in [0,T]. \tag{5.3}$$

Let $\kappa = \{t_i\colon i = 0,\dots,n\}$ be a partition of the interval $[0,T]$, and let

$$\varrho_H(u) := \left[u/\sqrt{|\log u|}\right]^{1/H} \quad \text{if } u > 0 \text{ and } \varrho_H(0) := 0.$$

For each $\omega \in \Omega$ satisfying (5.3), we then have

$$\sum_{i=1}^n \varrho_H(|B_H(t_i,\omega) - B_H(t_{i-1},\omega)|) \leq \sum_{i=1}^n \varrho_H(K_\omega \rho_H(t_i - t_{i-1})) \leq C_\omega \sum_{i=1}^n (t_i - t_{i-1}) = TC_\omega$$

for some finite constant C_ω which does not depend on κ. Therefore with probability 1,

$$v_\phi(B_H;[0,T]) = \sup\{v_\phi(B_H;\kappa)\colon \kappa \in \mathrm{PP}([0,T])\} < \infty \tag{5.4}$$

whenever $\phi = \varrho_H$. Since this function ϕ is bigger as $u \downarrow 0$ than u^p for any $p > 1/H$, it follows that the p-variation index $\upsilon(B_H) \leq 1/H$ for almost all sample functions of B_H.

However (5.4) also holds for the function $\phi = \phi_H$ defined on $[0,+\infty)$ by

$$\phi_H(t) := \left[t/\sqrt{2LL\,(1/t)}\right]^{1/H} \quad \text{for } t > 0 \text{ and } \quad \phi_H(0) := 0.$$

Define ψ_H on $[0,+\infty)$ by

$$\psi_H(t) := t^H\sqrt{2LL\,(1/t)} \quad \text{for } t > 0 \text{ and } \quad \psi_H(0) := 0,$$

where $L(u) := 1 \vee \log u$. The function ψ_H is continuous and stritly increasing. Let ψ_H^{-1} be its inverse function. Then ϕ_H is *asymptotic* to ψ_H^{-1} near the origin, that is $\phi_H(u)/\psi_H^{-1}(u) \to 1$ as $u \downarrow 0$. The following statement is a special case of the result of Kawada and Kôno [60]. In the case when $H = 1/2$, that is for a Brownian motion, it was proved by Taylor [106].

Theorem 5.4. *Let $1/2 \leq H < 1$ and let ϕ_H be regularly varying and asymptotic to ψ_H^{-1} near 0. For $0 < T < \infty$, with probability 1,*

$$\lim_{\delta\downarrow 0} \sup\{s_{\phi_H}(B_H;\kappa)\colon \kappa \in \mathrm{PP}([0,T]),\ |\kappa| \leq \delta\} = T. \tag{5.5}$$

By the same result of Kawada and Kôno [60], the relation (5.5) with "=" replaced by "≤" also holds when the Hurst index $H \in (0,1/2)$.

It follows from the preceding theorem that if $1/2 \leq H < 1$ and $\phi_H(u) = o(\phi(u))$ as $u \downarrow 0$, then (5.4) fails. Thus the p-variation index of the fractional Brownian motion B_H, $1/2 \leq H < 1$, is given by

$$\upsilon(B_H) = 1/H \quad \text{with probability 1.}$$

5.4 Lévy processes

Homogeneous Lévy processes Let $X = \{X(t)\colon t \geq 0\}$ be a continuous in probability stochastic process with independent increments. If almost all sample functions of X are right continuous with left limits and $X(0) = 0$ almost surely then X is called a *Lévy process* (cf. Itô [57, Section 1.3]). A Lévy process X is called *homogeneous* if the distribution of $X(t+s)-X(t)$ with $t, s \geq 0$ does not depend on t. Given a real number a, a number $b \geq 0$, and a measure L on $\mathbb{R} \setminus \{0\}$ such $\int_{\mathbb{R}\setminus\{0\}}(1 \wedge x^2)\, L(dx) < \infty$, let

$$\eta(u) := \eta_{a,b,L}(u) := iau - bu^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{iux} - 1 - iuh(x))L(dx) \tag{5.6}$$

for $u \in \mathbb{R}$, where $h(x) := x/(1 + x^2)$. The essential properties of h are that it is bounded on $\mathbb{R}$ and is $x + O(x^2)$ as $x \to 0$. Then η is called a *characteristic exponent* and L is called a *Lévy measure*. The characteristic function of a homogeneous Lévy process X is given by $E \exp\{iuX(t)\} = \exp\{t\eta(u)\}$ for each $t \geq 0$ and $u \in \mathbb{R}$.

It is well–known that sample functions of a homogeneous Lévy process X with the characteristic exponent (5.6) are of bounded variation if and only if $b = 0$ and

$$\int_{\mathbb{R}\setminus\{0\}} (1 \wedge |x|)\, L(dx) < \infty \tag{5.7}$$

(cf. e.g. Gikhman and Skorokhod [40, Theorem 3, p. 279]). The following result of Bretagnolle [8, Théorème III b] is less well–known. It sharpened an earlier result of Blumenthal and Getoor [7] and Monroe [83].

Theorem 5.5 (J. Bretagnolle). *Let $1 < p < 2$ and let $X = \{X(t)\colon t \geq 0\}$ be a mean zero homogeneous Lévy process with the characteristic exponent (5.6) such that $b \equiv 0$. Then $v_p(X; [0,1]) < \infty$ with probability 1 if and only if*

$$\int_{\mathbb{R}\setminus\{0\}} (1 \wedge |x|^p)\, L(dx) < \infty. \tag{5.8}$$

If (5.8) fails then $v_p(X; [0,1]) = +\infty$ almost surely.

α-stable Lévy motion A homogeneous Lévy process X is an *α-stable Lévy motion* of index α if its characteristic exponent (5.6) is given by $b \equiv 0$ and the Lévy measure $L = L_\alpha = L_{\alpha,r,q}$, where

$$L_{\alpha,r,q}(dx) := \begin{cases} rx^{-1-\alpha}dx & \text{if } x > 0 \\ q(-x)^{-1-\alpha}dx & \text{if } x < 0 \end{cases}$$

for $\alpha \in (0, 2)$ and $r, q \geq 0$ with $r + q > 0$. It is easy to see that (5.8) with $L = L_\alpha$ holds if and only if $p > \alpha$. Therefore $v_p(X_\alpha; [0,1]) < \infty$ with probability 1 for each $p > \alpha$. Below we give a more exact result for α-stable Lévy motion.

Recall that $h(x) := x/(1 + x^2)$. If $\alpha < 1$ then (5.7) holds and $\int h\, dL < \infty$. In this case, it is said that an α-stable Lévy motion has no drift if $a + \int h\, dL = 0$. The following result is due to Fristedt and Taylor [38, Theorem 2].

Theorem 5.6 (B. Fristedt and S. J. Taylor). *Let $X_\alpha = \{X_\alpha(t)\colon t \in [0,1]\}$ be an α-stable Lévy motion with $\alpha \in (0,2)$ having no drift if $\alpha < 1$ and with $r = q$ if $\alpha = 1$. For an increasing function $\psi\colon [0,\infty) \to [0,\infty)$, with probability* 1

$$\lim_{|\kappa|\to 0} \sum_{i=1}^{n} \psi(|X_\alpha(t_i) - X_\alpha(t_{i-1})|) = \sum_{(0,1]} \psi(|\Delta^- X_\alpha|). \tag{5.9}$$

By [38, Proposition 3] the right side of (5.9) is finite almost surely if and only if

$$\int_0^1 \psi(u) u^{-1-\alpha}\, du < \infty.$$

Xu [115, Section 3] established necessary and/or sufficient conditions for the boundedness of the p-variation of a symmetric α-stable processes with possibly dependent increments. More information about p-variation of stable processes can be found in Fristedt [37].

5.5 Empirical processes

Let P be a probability distribution on $\mathbb{R}$ with a continuous distribution function F. Let $X_1, X_2, \ldots, X_n$ be independent with law P and let let F_n be the empirical distribution function $F_n(t) := n^{-1}\sum_{j=1}^{n} 1_{\{X_j \le t\}}$ for $t \in \mathbb{R}$. Let U be the $U[0,1]$ distribution function $U(t) := \max(0, \min(1,t))$, and U_n an empirical distribution function for it. Let $\alpha_n := n^{1/2}(F_n - F)$, the classical empirical process. Clearly $U \circ F \equiv F$ and take $F_n \equiv U_n \circ F$. Then $\alpha_n = n^{1/2}(U_n - U)\circ F$. Almost surely U_n is continuous at 0 and at 1. Then the p-variation of α_n on $\mathbb{R}$ equals that of $\beta_n := n^{1/2}(U_n - U)$ on $[0,1]$. In studying the p-variation of α_n we can thus assume $F \equiv U$ and $\alpha_n = \beta_n$. Qian [92, Theorems 3.2 and 4.2] proves the following

Theorem 5.7 (J. Qian). *Let $1 \le p < 2$. Then there are constants $C_p, \lambda_p < \infty$ depending only on p such that $Ev_p(\alpha_n) \le C_p n^{1-p/2}$, and $\limsup_{n\to\infty} v_p(\alpha_n)/n^{1-p/2} \le \lambda_p$ almost surely.*

Conversely, since the X_i are almost surely distinct, we have $v_p(\alpha_n) \ge n^{1-p/2}$, so the above theorem is sharp up to the constants C_p and λ_p.

For $p = 2$ we have:

Theorem 5.8. (a) *$v_2(\alpha_n) = O_P(\log\log n)$ as $n \to \infty$.*
(b) *If $0 < c < 1/12$, then $P(v_2(\alpha_n) > c\log\log n) \to 1$ as $n \to \infty$.*

Proof. Dudley [25] proved (a) and Qian [92, Theorem 3.1] proved (b). □

Let b_t, $0 \le t \le 1$, be the Brownian bridge process, a Gaussian stochastic process with mean 0 and $Eb_t b_u = t(1-u)$ for $0 \le t \le s \le 1$, and such that $t \mapsto b_t(\omega)$ is continuous for almost all ω. As for Brownian motion $v_p(t \mapsto b_t(\omega)) < \infty$ almost surely for $p > 2$ and $= +\infty$ almost surely for $p \le 2$. Y.–C. Huang [54], [55] proved the following:

Theorem 5.9 (Y.–C. Huang). *Let $2 < p < \infty$. On some probability space there exist empirical processes α_n and Brownian bridges $b^{(n)}$ such that $E\|\alpha_n - b^{(n)}\|_{[p]} = O(n^{\frac{1}{p}-\frac{1}{2}})$.*

The rate of convergence is best possible up to a constant multiple since $b^{(n)}$ is continuous: if the X_i are distinct, as they are almost surely, $\|\alpha_n - b^{(n)}\|_{(p)} \ge n^{\frac{1}{p}-\frac{1}{2}}$.

5.6 Differentiability of operators on processes

Some of the operators described in Chapter 3, e.g. the product integral operator (treated in Chapter 7 below), are only Fréchet differentiable with respect to p-variation norms for $p < 2$. In the Love–Young inequality and an inequality for convolutions (Gehring [39], see Appendix A in [28] for a detailed proof) at least one of the two functions must be in $\mathcal{W}_p$ for $p < 2$. On the other hand paths of Brownian motion and certain other processes only have bounded p–variation for $p > 2$, and the empirical process only converges in p–variation norm to the Brownian bridge for $p > 2$. The question arises how the differentiability can still be useful. We have examined this question for the empirical process.

Let T be an operator on distribution functions, Fréchet differentiable with respect to the p–variation norm at a distribution function F with a derivative linear operator DT and a remainder bound of order $\gamma > 1$. Then for empirical distribution functions F_n,

$$T(F_n) = T(F) + (DT)(F_n - F) + O(\|F_n - F\|_{[p]}^{\gamma}).$$

It follows from Theorem 5.7 that

$$\sqrt{n}(T(F_n) - T(F)) = (DT)(\sqrt{n}(F_n - F)) + O_{\mathrm{Pr}}(n^{[1-\gamma(p-1)]/2}).$$

For $\gamma = 1$ and $p \leq 2$, the remainder term would not approach 0 as $n \to \infty$. But if for a fixed $\gamma > 1$ we take $p < 2$ close enough to 2, then $1 - \gamma(p-1) < 0$ and the remainder becomes small as $n \to \infty$. For the bilinear operators and the product integral we have $\gamma = 2$, so that the remainder becomes $O_{\mathrm{Pr}}(n^{\epsilon - 1/2})$ for any $\epsilon > 0$. For the quantile operator, one has $\gamma = (p+1)/p \uparrow 3/2$ as $p \uparrow 2$ by Theorem 3.7, so that the remainder is $O_{\mathrm{Pr}}(n^{\epsilon - 1/4})$ for any $\epsilon > 0$, where $-1/4$ is the best possible exponent by Bahadur–Kiefer theorems, cf. Section 3.4 above. Also for the composition operator (Theorem 3.3), γ does not approach 1 as $p \uparrow 2$.

Whenever $1 - \gamma(p-1) < 0$, if $(DT)(\sqrt{n}(F_n - F))$ converges in distribution to $(DT)(b \circ F)$ for a Brownian bridge b, then so does $\sqrt{n}(T(F_n) - T(F))$.

Chapter 6

Integration

A whole world, created by the Young dynasty, with its own ideas, methods and language.

Transl. from F. A. Medvedev. *Development of the Concept of Integral* (in Russian) [81, p. 378].

6.1 The refinement–Riemann–Stieltjes integral

In this section we consider an extension of the mesh–Riemann-Stieltjes integral suggested by in S. Pollard in 1923. The idea of the extension is to replace the convergence as the mesh tends to zero used to define the (MRS) integral by a type of convergence now called net convergence or, from another angle, filter convergence. Net convergence had previously been introduced by Moore [85], Moore and Smith [86]. In his paper [91] Pollard shows that the new integral extends and improves the mesh–Riemann-Stieltjes integral in some ways.

Throughout the chapter f and h are real-valued functions on a closed interval $[a,b]$. Let $\kappa = \{x_i \colon i = 1,\dots,n\}$ be a partition of $[a,b]$ and let $\sigma = \{y_i \colon i = 1,\dots,n\}$ be an intermediate partition of κ, that is $x_{i-1} \leq y_i \leq x_i$ for $i = 1,\dots,n$. As before the set of all partitions κ of $[a,b]$ is denoted by $\mathrm{PP}([a,b])$ and the Riemann-Stieltjes sum based on (κ,σ) is defined by (4.22). Recall that the refinement–Riemann–Stieltjes integral or (RRS) integral of f with respect to h exists and equals the number A, if for each $\epsilon > 0$ there is a partition λ of $[a,b]$ such that

$$\Big|S_{RS}(f,h;\kappa,\sigma) - A\Big| < \epsilon \tag{6.1}$$

holds for all refinements κ of λ and all intermediate partitions σ of κ. If the refinement–Riemann–Stieltjes integral is defined then we write $(RRS)\int_a^b f\,dh := A$.

Proposition 6.1. *The (RRS) integral is an extension of the (MRS) integral.*

Proof. Suppose $(MRS)\int_a^b f\,dh$ exists with a value A. Then given any positive number $\epsilon > 0$, one can find a $\delta > 0$ such that (6.1) holds whenever the mesh $|\kappa| < \delta$. Let $\lambda := \{a, a+\delta/2, a+\delta, a+3\delta/2, \dots, a+n\delta, b\}$ for the minimal integer n such that $a + n\delta + \delta/2 > b$. Then (6.1) also holds for each refinement κ of λ. Thus $(RRS)\int_a^b f\,dh$ exists and has the same value A. □

Example 6.2. Let f be a function on $[a,b]$ and let $\ell_c := 1_{[c,b]}$ be the indicator function of $[c,b]$ for some $a < c < b$. For any $\kappa = \{x_i: i = 1,\dots,n\}$ and intermediate partition $\sigma = \{y_i: i = 1,\dots,n\}$ of κ, we have $S_{RS}(f,\ell_c;\kappa,\sigma) = f(y_i)$ for $x_{i-1} < c \le x_i$ and $y_i \in [x_{i-1},x_i]$. The integral $(MRS)\int_a^b f\,d\ell_c$ exists and equals $f(c)$ if and only if f is continuous at c. Taking $c \in \kappa$ it follows that the integral $(RRS)\int_a^b f\,d\ell_c$ exists and equals $f(c)$ if and only if f is left-continuous at c. Therefore $(RRS)\int_a^b 1_{[a,c]}\,d\ell_c$ exists and equals 1 while the same integral in the mesh–Riemann-Stieltjes sense doesn't exist. However the other integral $(RRS)\int_a^b \ell_c\,d\ell_c$ also doesn't exist.

The following is proved in Hildebrandt (1963, Theorem II.10.9).

Theorem 6.3. *Let f and h be bounded functions on $[a,b]$. The integral $(MRS)\int_a^b f\,dh$ exists if and only if both $(RRS)\int_a^b f\,dh$ exists and the two functions f,h have no common discontinuities on $[a,b]$.*

Properties 6.4. *Consider the following properties of an integral $\int_a^b f\,dh$:*

I. $\int_a^b 1_{[a,b]}\,dh = h(b) - h(a)$.

II. For $u_1,u_2 \in \mathbb{R}$ and $f_1,f_2\colon [a,b] \mapsto \mathbb{R}$,

$$\int_a^b (u_1 f_1 + u_2 f_2)\,dh = u_1\int_a^b f_1\,dh + u_2\int_a^b f_2\,dh$$

where the left side exists provided the right side does.

III. For $v_1,v_2 \in \mathbb{R}$ and $h_1,h_2\colon [a,b] \mapsto \mathbb{R}$,

$$\int_a^b f\,d(v_1 h_1 + v_2 h_2) = v_1\int_a^b f\,dh_1 + v_2\int_a^b f\,dh_2$$

where again the left side exists provided the right side does.

IV. For $a < c < b$, $\int_a^b f\,dh$ exists if and only if both $\int_a^c f\,dh$ and $\int_c^b f\,dh$ exist, and then

$$\int_a^b f\,dh = \int_a^c f\,dh + \int_c^b f\,dh.$$

V. Let $\int_a^b f\,dh$ exists and let $F(y) := \int_a^y f\,dh$ for $y \in [a,b]$. Then for each $x \in [a,b]$,

$$\lim_{y\to x}\Big\{F(y) - F(x) - f(x)[h(y) - h(x)]\Big\} = 0. \tag{6.2}$$

Pollard [91, p. 80] showed that the "if" part of IV does not hold for the (MRS) integral. Indeed let $\bar f := 1_{[1,2]}$ and $\bar h := 1_{(1,2]}$ be indicator functions on $[0,2]$. Then $(MRS)\int_0^1 \bar f\,d\bar h = 0$ and $(MRS)\int_1^2 \bar f\,d\bar h = h(2) - h(1) = 1$. For any $\kappa = \{x_i\} \in \mathrm{PP}([a,b])$ such that $1 \notin \kappa$ and an intermediate partition $\sigma = \{y_i\}$ of κ, we have

$$S_{RS}(\bar f,\bar h;\kappa,\sigma) = \bar f(y_i)[\bar h(x_i) - \bar h(x_{i-1})] = \begin{cases} 1 & \text{if } y_i = x_i \\ 0 & \text{if } y_i = x_{i-1}, \end{cases}$$

where i is such that $x_{i-1} < 1 < x_i$. Clearly κ can have in addition arbitrarily small mesh. Thus $(MRS)\int_0^2 \bar f\,d\bar h$ doesn't exist. A sufficient condition for the (MRS) integral to satisfy condition IV is given in Theorem II.4.4N of Hildebrandt [52, p. 34].

Theorem 6.5. *The* (RRS) *integral satisfies the above properties* I–V.

Proof. Pollard [91] proved the first four properties. One can also find proofs in Hildebrandt [52, Section II.11]. Here we prove the fifth property for the (RRS) integral.

Suppose the integral $(RRS)\int_a^b f\,dh$ exists. Let $\epsilon > 0$ and $x \in (a, b]$. There exists $\lambda \in \mathrm{PP}\,([a, b])$ such that (6.1) holds for all refinements κ of λ and all intermediate partitions σ of κ. We can and do assume that $x = x_i \in \lambda$ for some $0 \le i < \mathrm{card}\lambda$. Let $y \in (x, x_{i+1}]$. By property IV for the (RRS) integral, there exists a partition $\kappa_{x,y}$ of $[x, y]$ such that the Riemann–Stieltjes sum $S_{RS}(\kappa_{x,y}, \sigma_{x,y})$ differs from $F(y) - F(x)$ by less than ϵ. Let κ_1 and κ_2 be two refinements of λ which coincide with $\kappa_{x,y}$ and $\{x, y\}$ when restricted to $[x, y]$. By (6.1), it then follows that

$$\begin{aligned}\Big|F(y) - F(x) - f(x)[h(y) - h(x)]\Big| &\le \Big|(RRS)\int_x^y f\,dh - S_{RS}(\kappa_{x,y}, \sigma_{x,y})\Big| \\ &\quad +|S_{RS}(\kappa_1, \sigma_1) - S_{RS}(\kappa_2, \sigma_2)| \le 3\epsilon,\end{aligned}$$

where the intermediate partition σ_1 of κ_1, when restricted to $[x, y]$, coincides with the intermediate partition $\sigma_{x,y}$ of $\kappa_{x,y}$, and the intermediate partition σ_2 of κ_2 has an element x corresponding to the interval $[x, y]$. Thus (6.2) holds for $y \downarrow x$ instead of $y \to x$. Similarly, given $x \in (a, b]$, one can show that (6.2) also holds when $y \uparrow x$. The proof of property V for the (RRS) integral is complete. $\square$

Proposition 6.6 (Cauchy criterion). *The integral* $(RRS)\int_a^b f\,dh$ *exists if and only if, for each* $\epsilon > 0$, *there exists* $\lambda \in \mathrm{PP}\,([a, b])$ *such that*

$$|S_{RS}(f, h; \kappa_1, \sigma_1) - S_{RS}(f, h; \kappa_2, \sigma_2)| < \epsilon \tag{6.3}$$

for any refinements κ_1, κ_2 *of* λ *and any intermediate partitions* σ_1, σ_2.

Proof. The "only if" implication is clear. To prove the "if" part let $\epsilon_m \downarrow 0$. For each $m \ge 1$, let λ_m be a partition λ existing by the hypothesis for $\epsilon = \epsilon_m$. One can and does assume that $\lambda_m \subset \lambda_{m+1}$ for each m. Let $\{\sigma_m \colon m \ge 1\}$ be any sequence of intermediate partitions of λ_m, $m \ge 1$. Then $\{S_{RS}(\lambda_m; \sigma_m) \colon m \ge 1\}$ is a Cauchy sequence of real numbers. Let A be its limit. Given $m \ge 1$, let κ be a refinement of λ_m and let σ be an intermediate partition of κ. By (6.3), we then have

$$|S_{RS}(\kappa, \sigma) - A| \le |S_{RS}(\kappa, \sigma) - S_{RS}(\lambda_{m+n}, \sigma_{m+n})| + |S_{RS}(\lambda_{m+n}, \sigma_{m+n}) - A| < 2\epsilon_m$$

for large enough integers $n \ge 1$. Therefore $(RRS)\int_a^b f\,dh$ exists and has value A. $\square$

Proposition 6.7. *If* $(RRS)\int_a^b f\,dh$ *exists then both* f *and* h *can't have discontinuities on the same side at same point; that is, for* $x \in (a, b]$, *either*

$$\lim_{y \uparrow x} f(y) = f(x) \qquad \text{or} \qquad \lim_{y \uparrow x} h(y) = h(x), \tag{6.4}$$

and the same for $x \in [a, b)$ *with* $y \uparrow x$ *replaced by* $y \downarrow x$.

Proof. Let $x \in (a, b]$ and $\epsilon > 0$. Then there exists $\lambda \in \mathrm{PP}\,([a, b])$ such that (6.1) holds for each refinement κ of λ. We can and do suppose that $x \in \lambda$. Subtract two Riemann-Stieltjes sums based on the same κ and two intermediate partitions of κ having all points equal except two of them in the interval with the right endpoint x. This gives the bound

$$|f(y') - f(y'')||h(x') - h(x)| < 2\epsilon$$

for any $x' \le y' < y'' \le x$ and x' close enough to x. Thus (6.4) holds. A symmetric argument yields the second part of the Proposition 6.7. $\square$

Naturally the preceding necessary condition on discontinuities is stronger in the case when the integral exists in the mesh–Riemann-Stieltjes sense.

Proposition 6.8. *If $(MRS)\int_a^b f\,dh$ exists then both f and h can't have a discontinuity at the same point; that is, for $x \in [a,b]$, either $\lim_{y\to x} f(y) = f(x)$ or $\lim_{y\to x} h(y) = h(x)$.*

Proof. Let $x \in [a,b]$ and $\epsilon > 0$. Then there exists $\delta > 0$ such that (6.1) holds for each partition κ with mesh $|\kappa| < \delta$. Subtract two Riemann-Stieltjes sums based on the same κ and two intermediate partitions of κ having all points equal except two of them in the interval containing the point x. This gives the bound

$$|f(y') - f(y'')||h(x') - h(x'')| < 2\epsilon$$

for any $x', x'' \in [a \vee (x-\delta), b \wedge (x+\delta)]$ and $y', y'' \in [x', x'']$. Thus Proposition 6.8 holds. $\square$

6.2 The refinement–Young–Stieltjes integral

Recall the (RYS) integral as defined in Section 2.1. Let f be any function on $[a,b]$, and let $h \in \mathcal{R}[a,b]$. Given a partition $\kappa = \{x_i\colon i = 0,\dots,n\} \in \mathrm{PP}\,([a,b])$, we say that an intermediate partition $\sigma = \{y_i\colon i = 1,\dots,n\}$ of κ is *open* if $x_{i-1} < y_i < x_i$ for $i = 1,\dots,n$, or *left–open* if $x_{i-1} < y_i \le x_i$ for $i = 1,\dots,n$. Then the Young sum $S_Y(f,h;\kappa,\sigma)$ based on (κ,σ) is defined by

$$\begin{aligned} S_Y(f,h;\kappa,\sigma) \; &:= \; \sum_{i=1}^n \{[f\Delta^+h](x_{i-1}) + f(y_i)[h(x_i-) - h(x_{i-1}+)] + [f\Delta^-h](x_i)\} \qquad (6.5) \\ &= \; [f\Delta^+h](a) + [f\Delta^-h](b) + \sum_{i=1}^n f(y_i)[h(x_i-) - h(x_{i-1}+)] + \sum_{i=1}^{n-1}[f\Delta^{\pm}h](x_i). \end{aligned}$$

Example 6.9. As in Example 6.2, for $a < c < b$, let ℓ_c be the indicator function of $[c,b]$ and f be any function, both on $[a,b]$. For each partition κ containing c and an open intermediate partition σ of κ, we have $S_Y(f,\ell_c;\kappa,\sigma) = f(c)$. Therefore the integral $(RYS)\int_a^b f\,d\ell_c$ exists and has value $f(c)$. In particular, $(RYS)\int_a^b \ell_c\,d\ell_c$ exists and has value 1, while the (RRS) integral doesn't exist as shown in Example 6.2. More generally, for any real number r, let $\ell_c^r(x) := \ell_c(x)$ for $x \neq c$ and $\ell_c^r(c) := r$, so that $\ell_c^1 \equiv \ell_c$. Then $(RYS)\int_a^b f\,d\ell_c^r$ exists and has value $f(c)$ for any r.

In fact the (RYS) integral $\int_a^b f\,dh$ does not depend on values of the integrator h at jump points in (a,b). This is because the value of the Young sum (6.5) does not depend on values of h at jump points of the open interval (a,b). The same is not true for the (RRS) integral. Indeed, by changing the value of the integrator h at a point x one can make h either right-continuous or left-continuous at x and so destroy the necessary condition of Proposition 6.7. Unlike the (RRS) integral, the (MRS) integral shares the same property of independence of values of the integrator at discontinuity points with the Young integral. Indeed, by Proposition 6.8, if $(MRS)\int_a^b f\,dh$ exists and h has a jump at $x \in (a,b)$ then f must be continuous at x. For $\xi \in \mathbb{R}$, let $h_\xi := h$ on $[a,x) \cup (x,b]$ and $h_\xi(x) := \xi$. Then for the two Riemann-Stieltjes sums we have

$$S_{RS}(f,h;\kappa,\sigma) - S_{RS}(f,h_\xi;\kappa,\sigma) = \begin{cases} [f(y') - f(y'')][h(x) - \xi] & \text{if } x \in \kappa \\ 0 & \text{if } x \notin \kappa, \end{cases}$$

where $y' \in [x-\delta, x]$ and $y'' \in [x, x+\delta]$ for $\delta := |\kappa|$. Due to continuity of f at x the preceding difference can be made arbitrarily small if the mesh $|\kappa|$ is small enough.

Properties of the (RYS) integral.

Proposition 6.10 (Cauchy criterion). *The integral $(RYS)\int_a^b f\,dh$ exists if and only if, for each $\epsilon > 0$, there exists $\lambda \in \mathrm{PP}([a,b])$ such that*

$$|S_Y(f,h;\kappa_1,\sigma_1) - S_Y(f,h;\kappa_2,\sigma_2)| < \epsilon$$

for any refinements κ_1, κ_2 of λ and any open intermediate partitions σ_1, σ_2.

The proof of the Cauchy criterion for the (RYS) integral is the same as for the (RRS) integral given by Proposition 6.6 above.

Theorem 6.11. *The (RYS) integral satisfies statement IV among Properties 6.4, that is, for $a < c < b$, $(RYS)\int_a^b f\,dh$ exists if and only if both $(RYS)\int_a^c f\,dh$ and $(RYS)\int_c^b f\,dh$ exist, and then*

$$(RYS)\int_a^b f\,dh = (RYS)\int_a^c f\,dh + (RYS)\int_c^b f\,dh. \tag{6.6}$$

Proof. To see that the "if" part holds notice that the additivity relation

$$S_Y(f,h;\kappa,\sigma) = S_Y(f,h;\kappa_1,\sigma_1) + S_Y(f,h;\kappa_2,\sigma_2) \tag{6.7}$$

holds for each $\kappa \in \mathrm{PP}([a,b])$ containing c, $\kappa_1 \in \mathrm{PP}([a,c])$, $\kappa_2 \in \mathrm{PP}([c,b])$, $\kappa_1 \cup \kappa_2 = \kappa$ (as a set), and for corresponding intermediate partitions such that $\sigma_1 \cup \sigma_2 = \sigma$.

For the converse suppose that the integral exists over $[a,b]$. Then the Cauchy condition of Proposition 6.10 holds. We can and do assume that $c \in \lambda \in \mathrm{PP}([a,b])$. Then using this condition when intermediate partitions have the same points from the interval $[c,b]$, one gets that the Cauchy condition holds for the integral over $[a,c]$. A symmetric argument yields that the integral over $[c,b]$ also exists. The additivity relation (6.6) then follows from (6.7). The proof of Theorem 6.11 is complete. $\square$

Let F be an indefinite (RYS) integral defined for a constant C by

$$F(x) := C + (RYS)\int_a^x f\,dh \qquad \text{for } x \in [a,b].$$

According to the following theorem, F is a regulated function, and for $a \le y < x \le b$,

$$(\Delta^- F)(x) = [f\Delta^- h](x) \quad \text{and} \quad (\Delta^+ F)(y) = [f\Delta^+ h](y). \tag{6.8}$$

Theorem 6.12. *The* (RYS) *integral satisfies statement V among Properties 6.4, that is, if* $(RYS)\int_a^b f\,dh$ *exists then, for each* $x \in [a,b]$,

$$\lim_{y\to x}\Big\{F(x)-F(y)-f(x)[h(x)-h(y)]\Big\}=0.$$

Proof. Suppose $(RYS)\int_a^b f\,dh$ exists. Then the indefinite integral F exists by Theorem 6.11. We prove that, for $a<x\le b$,

$$\lim_{y\uparrow x}\Big\{F(x)-F(y)-f(x)[h(x)-h(y)]\Big\}=0. \tag{6.9}$$

The proof when $y\downarrow x$ for $a\le x<b$ is symmetric and is omited. Taking $u_y\in(y,x)$, we have

$$\lim_{y\uparrow x}\Big\{S_Y(f,h;\{y,x\},\{u_y\})-f(x)[h(x)-h(y)]\Big\}$$

$$=\lim_{y\uparrow x}\Big\{[f(u_y)-f(x)][h(x-)-h(y+)]+[f(y)-f(x)]\Delta^+h(y)\Big\}=0.$$

Therefore it is enough to prove

$$\lim_{y\uparrow x}\Big\{F(x)-F(y)-S_Y(f,h;\{y,x\},\{u_y\})\Big\}=0$$

for $u_y\in(y,x)$. Let $a<x\le b$ and $\epsilon>0$. By definition of the (RYS) integral, there exists $\lambda=\{z_j\colon j=0,\dots,m\}\in \mathrm{PP}([a,x])$ such that

$$|S_Y(f,h;\kappa,\sigma)-(RYS)\int_a^x f dh|<\epsilon$$

for each refinement κ of λ and intermediate partition σ of κ. Let $y\in[z_{m-1},x)$. Within ϵ of $(RYS)\int_y^x f\,dh$, choose a Riemann–Stieltjes sum $S_{y,x}$ based on a partition $\kappa_{y,x}$ of $[y,x]$. Let κ_1 and κ_2 be two refinements of λ which coincide with $\kappa_{y,x}$ and $\{y,x\}$, respectively, when restricted to $[y,x]$. Then taking $u_y\in(y,x)$, we have

$$|F(x)-F(y)-S_Y(f,h;\{y,x\},\{u_y\}]|\le\Big|(RYS)\int_y^x f\,dh-S_{y,x}\Big|$$

$$+|S_Y(f,h;\kappa_1,\sigma_1)-S_Y(f,h;\kappa_2,\sigma_2)|\le 3\epsilon$$

Therefore (6.9) holds, and the proof is complete. □

Connection with the (RRS) integral.

Theorem 6.13. *Let f be a function on $[a,b]$, and let h be regulated on $[a,b]$. If $(RRS)\int_a^b f\,dh$ exists then so does $(RYS)\int_a^b f\,dh$, and both have the same value.*

Proof. Let $\kappa=\{x_i\colon i=0,\dots,n\}$ be a partition of $[a,b]$ and let $\sigma=\{y_i\colon i=1,\dots,n\}$ be an open intermediate partition of κ. Let $\mu=\{u_{i-1},v_i\colon i=1,\dots,n\}$ be a set of points in (a,b) such that $x_0<u_0<y_1<v_1<x_1<\cdots<x_{n-1}<u_{n-1}<y_n<v_n<x_n$. Then $\kappa':=\kappa\cup\mu$

is a partition of $[a,b]$ and $\sigma' := \{x_0, y_1, x_1, x_1, y_2, \dots, x_{n-1}, x_{n-1}, y_n, x_n\}$ is an intermediate partition of κ'. Letting $u_{i-1} \downarrow x_{i-1}$ and $v_i \uparrow x_i$ for $i = 1, \dots, n$, we get

$$\begin{aligned} S_{RS}(f,h;\kappa',\sigma') &= \sum_{i=1}^{n} f(y_i)[h(v_i) - h(u_{i-1})] + f(x_0)[h(u_0) - h(x_0)] \\ &+ \sum_{i=1}^{n-1} f(x_i)[h(u_i) - h(v_i)] + f(x_n)[h(x_n) - h(v_n)] \to S_Y(f,h;\kappa,\sigma). \end{aligned}$$

Therefore for each $\kappa \in \mathrm{PP}([a,b])$ and open intermediate partition σ of κ, the Young sums based on κ and σ can be approximated arbitrarily closely by Riemann–Stieltjes sums based on refinements of κ, and the theorem follows. $\square$

Lemma 6.14. *Let $\phi, \psi \in \mathcal{R}[a,b]$ be such that ϕ is left–continuous on $(a,b]$ and ψ is right–continuous on $[a,b)$, or vice versa. If (RYS) $\int_a^b \phi\, d\psi$ exists then so does (RRS) $\int_a^b \phi\, d\psi$, and the two integrals have the same value.*

Proof. Let $f, h \in \mathcal{R}[a,b]$. The proof is given only when $\phi = f_-^{(a)}$ and $\psi = h_+^{(b)}$. A proof for the other case is similar. Let $\kappa = \{x_i \colon i = 0, \dots, n\} \in \mathrm{PP}([a,b])$ and let $\sigma = \{y_i \colon i = 1, \dots, n\}$ be an open intermediate partition of κ, that is $x_{i-1} < y_i < x_i$ for $i = 1, \dots, n$. Then we have

$$S_{RS}(f_-^{(a)}, h_+^{(b)}; \kappa, \sigma) - S_Y(f_-^{(a)}, h_+^{(b)}; \kappa, \sigma) = \sum_{i=1}^{n} f_-^{(a)}(y_i)[h_+^{(b)}(x_i) - h_+^{(b)}(x_{i-1})]$$

$$-\sum_{i=1}^{n} \Big\{ f(y_i-)[h(x_i-) - h(x_{i-1}+)] + f(x_i-)[h_+^{(b)}(x_i) - h(x_i-)] \Big\}$$

$$= \sum_{i=1}^{n} [f(y_i-) - f(x_i-)][h_+^{(b)}(x_i) - h(x_{i-1}+)] - \sum_{i=1}^{n} [f(y_i-) - f(x_i-)][h(x_i-) - h(x_{i-1}+)]. \quad (6.10)$$

Let $\epsilon > 0$. There exists $\lambda = \{z_j \colon j = 0, \dots, m\} \in \mathrm{PP}([a,b])$ such that

$$\Big| S_Y(f_-^{(a)}, h_+^{(b)}; \kappa, \sigma) - (Y) \int_a^b f_-^{(a)}\, dh_+^{(b)} \Big| < \epsilon \quad (6.11)$$

for all refinements $\kappa \in \mathrm{PP}([a,b])$ of λ and all open intermediate partitions σ of κ. Choose $\nu = \{v_{j-1}, u_j \colon j = 1, \dots, m\} \subset [a,b]$ such that $z_{j-1} < v_{j-1} < u_j < z_j$ for each $j = 1, \dots, m$,

$$\mathrm{Osc}(f; [u_j, z_j-]) \le \epsilon/(2m\|h\|_\infty) \quad \text{and} \quad \mathrm{Osc}(h; [z_{j-1}+, v_{j-1}+]) \le \epsilon/(2m\|f\|_\infty). \quad (6.12)$$

Let $\kappa = \{x_i \colon i = 0, \dots, n\} \in \mathrm{PP}([a,b])$ be a refinement of $\lambda \cup \nu \subset$ and let $\sigma = \{y_i \colon i = 1, \dots, n\}$ be an open intermediate partition of κ. We show that the absolute value of the right side of (6.10) is small. Let $\sigma' = \{y_i' \colon i = 1, \dots, n\}$ be another open intermediate partition of κ. Using (6.11) for two Young sums based on (κ, σ) and (κ, σ'), we get

$$\Big| \sum_{i=1}^{n} [f(y_i-) - f(y_i'-)][h(x_i-) - h(x_{i-1}+)] \Big| < 2\epsilon.$$

Then letting $y_i' \uparrow x_i$ for $i = 1, \ldots, n$, it follows that the second sum on the right side of (6.10) has the bound

$$\Big|\sum_{i=1}^{n}[f(y_i-) - f(x_i-)][h(x_i-) - h(x_{i-1}+)]\Big| \le 2\epsilon. \tag{6.13}$$

To bound the first sum on the right side of (6.10) we invoke an auxiliary partition $\kappa' = \{x_i'\colon i = 0, \ldots, n\} \in \mathrm{PP}([a,b])$. For each $j = 0, \ldots, m$, let $i(j) \in \{0, \ldots, n\}$ be such that $x_{i(j)} = z_j$. Let $J := \{i(j-1)+1, i(j)\colon j = 1, \ldots, m\}$ and let $I := \{0, \ldots, n\} \setminus J$. Define κ' by $x_i' := x_i$ for $i \in J$ and $x_i' \in (x_i, y_{i+1})$ for $i \in I$. Notice that $\kappa' \supset \lambda$ and $\sigma = \{y_i\colon i = 1, \ldots, n\}$ is an open intermediate partition of κ'. Let $\sigma' = \{y_i'\colon i = 1, \ldots, n\}$ be another open intermediate partition for κ' defined by $y_i' := y_i$ for $i \in J$ and $y_i' := x_i$ for $i \in I$. Then using (6.11) for two Young sums corresponding to (κ', σ) and (κ', σ') we get the bound

$$\Big|\sum_{i\in I}[f(y_i-) - f(x_i-)][h(x_i'-) - h(x_{i-1}'+)]\Big| < 2\epsilon.$$

Letting $x_i' \downarrow x_i$ for $i \in I$ it follows that

$$\Big|\sum_{i\in I}[f(y_i-) - f(x_i-)][h(x_i+) - h(x_{i-1}+)]\Big| \le 2\epsilon.$$

Since $u_j \le x_{i(j)-1} < y_{i(j)} < z_j$ and $x_{i(j-1)+1} \le v_{j-1}$ for each $j = 1, \ldots, m$, by (6.12), we get

$$\Big|\sum_{i=1}^{n}[f(y_i-) - f(x_i-)][h_+^{(b)}(x_i) - h(x_{i-1}+)]\Big| \le 2\|h\|_\infty \sum_{j=1}^{m} \mathrm{Osc}(f; [u_j, z_j-])$$

$$+2\|f\|_\infty \sum_{j=0}^{m-1} \mathrm{Osc}(h; [z_j+, v_j+]) + \Big|\sum_{i\in I}[f(y_i-) - f(x_i-)][h(x_i+) - h(x_{i-1}+)]\Big| \le 4\epsilon.$$

This in conjunction with (6.13) yields that the absolute value of the right side of (6.10) does not exceed 6ϵ. Therefore, by (6.11), we have the bound

$$\Big|S_{RS}(f_-^{(a)}, h_+^{(b)}; \kappa, \sigma) - (Y)\int_a^b f_-^{(a)}\, dh_+^{(b)}\Big| \le 7\epsilon \tag{6.14}$$

for all $\kappa \in \mathrm{PP}([a,b])$ with $\kappa \supset \lambda \cup \nu$ and all open intermediate partitions σ of κ. For any such $\kappa = \{x_i\colon i = 0, \ldots, n\}$, since $f_-^{(a)}$ is left–continuous, (6.14) also holds for all *left-open* intermediate partitions $\sigma = \{y_i\colon i = 1, \ldots, n\}$ of κ, that is $x_{i-1} < y_i \le x_i$ for $i = 1, \ldots, n$.

To finish the proof we need to extend the above bound to an arbitrary intermediate partition σ of κ. Given $\epsilon > 0$, let λ be the partition $\{z_j\colon j = 0, \ldots, m\}$ of $[a,b]$ satisfying (6.11), and let ν be the set of points $\{v_{j-1}, u_j\colon j = 1, \ldots, m\}$ such that $z_{j-1} < v_{j-1} < u_j < z_j$ for $j = 1, \ldots, m$ and (6.12) holds. By the above proof, (6.14) then holds for any refinement κ of $\lambda \cup \nu'$ for some $\nu' = \{v_{j-1}', u_j'\colon j = 1, \ldots, m\}$ with $z_{j-1} < v_{j-1}' \le v_{j-1} < u_j \le u_j' < z_j$, $j = 1, \ldots, m$, and left-open intermediate partition σ of κ. Let $\kappa = \{x_i\colon i = 0, \ldots, n\} \in \mathrm{PP}([a,b])$ be such that $\kappa \supset \lambda \cup \nu \cup \{(z_j + v_j)/2\colon j = 0, \ldots, m-1\}$ and let $\sigma = \{y_i\colon i = 1, \ldots, n\}$ be an intermediate partition of κ, that is $x_{i-1} \le y_i \le x_i$ for $i = 1, \ldots, n$. Let σ_e be the same as σ except that each $y_i \in \sigma$ with i even is replaced by x_i. Similarly, let σ_o be the same as σ except that each $y_i \in \sigma$ with i odd is replaced by x_i. Also, let $\sigma_m := \{x_i\colon i = 1, \ldots, n\}$ and $S(\sigma) := S_{RS}(f_-^{(a)}, h_+^{(b)}; \kappa, \sigma)$. Then we have $S(\sigma) + S(\sigma_m) = S(\sigma_e) + S(\sigma_o)$. Hence

$$|S(\sigma) - A| \le |S(\sigma_m) - A| + |S(\sigma_e) - A| + |S(\sigma_o) - A| \tag{6.15}$$

with $A = (Y)\int_a^b f_-^{(a)}\,dh_+^{(b)}$. The first term on the right side of (6.15) does not exceed 7ϵ by (6.14) because σ_m is a left-open intermediate partition of κ. To bound the second term on the right side of (6.15), let I_σ be the set of indices $i \in \{1,\dots,n\} \setminus J$ such that $y_i \in \sigma_e$ and $y_i = x_{i-1}$, where $J = \{i(j-1)+1, i(j)\colon j = 1,\dots,m\}$. So, if $i \in I_\sigma$ then $i \in \{2,\dots,n-1\}$ is odd and

$$f(y_i-)[h(x_i+) - h(x_{i-1}+)] + f(y_{i-1}-)[h(x_{i-1}+) - h(x_{i-2}+)] = f(x_{i-1}-)[h(x_i+) - h(x_{i-2}+)]$$

because $y_i = y_{i-1} = x_{i-1}$ by the construction of σ_e. Let $\kappa' := \kappa \setminus \{x_{i-1}\colon i \in I_\sigma\}$. Notice that $\kappa' \supset \lambda \cup \{v'_{j-1}, u_j\colon j = 1,\dots,m\}$ with $v'_{j-1} = x_{i(j-1)+1} < v_{j-1}$ or $v'_{j-1} = x_{i(j-1)+2} \le v_{j-1}$ for $j = 1,\dots,m$. Also, let σ'_e be the same as σ_e except that each y_i, $i \in J$, such that $y_i = x_{i-1}$, is replaced by x_i, and one of the two points x_{i-1}, $i \in I_\sigma$, is removed. Then σ'_e is a left-open intermediate partition of κ'. By (6.14) and (6.12), it then follows that

$$|S(\sigma_e) - A| \le |S_{RS}(f_-^{(a)}, h_+^{(b)}; \kappa', \sigma'_e) - A| + 2\|f\|_\infty \sum_{j=0}^{m-1} \mathrm{Osc}(h; [z_j+, v_j+])$$

$$+2\|h\|_\infty \sum_{j=1}^{m} \mathrm{Osc}(f; [u_j, z_j-]) \le 9\epsilon.$$

The same reasoning yields the bound $|S(\sigma_o) - A| \le 9\epsilon$. Therefore, by (6.15), $|S(\sigma) - A| \le 25\epsilon$ for an arbitrary intermediate partition σ of κ. The proof of Lemma 6.14 is complete. ☐

6.3 The central Young integral

L. C. Young [117] defined (and Dudley [23] specified the endpoint terms of) the integral which we call the Central Young integral, or the (CY) integral. The idea of the (CY) integral is to use the (RRS) integral, avoiding its lack of definition when f and h have common one–sided discontinuities by taking a right–continuous version of f and left–continuous version of h, or vice versa, and adding sums of jump terms to restore the desired value of $\int f\,dh$. In general the (CY) integral extends the (RYS) integral. However, if the integrand and integrator have bounded p– and p'–variation, respectively, with $1/p + 1/p' = 1$, then the (CY) integral is defined only if the (RYS) integral is (see Proposition 6.22 below).

Recall the definitions (2.2) of functions $f_+^{(b)}$ and $f_-^{(a)}$, as well as unconditional convergence as defined in Section 2.1. The (CY) integral has two equivalent forms defined next:

Definition 6.15. Let $f, h \in \mathcal{R}([a,b])$. If $a < b$ define the (Y_1) integral

$$(Y_1)\int_a^b f\,dh := (RRS)\int_a^b f_+^{(b)}\,dh_-^{(a)} - [\Delta^+ f \Delta^+ h](a) + [f\Delta^- h](b) - \sum_{(a,b)} \Delta^+ f \Delta^\pm h \qquad (6.16)$$

if the (RRS) integral exists and the sum converges unconditionally. If $a = b$ define the (Y_1) integral as 0. Similarly, if $a < b$ define the (Y_2) integral

$$(Y_2)\int_a^b f\,dh := (RRS)\int_a^b f_-^{(a)}\,dh_+^{(b)} + [f\Delta^+ h](a) + [\Delta^- f \Delta^- h](b) + \sum_{(a,b)} \Delta^- f \Delta^\pm h \qquad (6.17)$$

if the (RRS) integral exists and the sum converges unconditionally. If $a = b$ define the (Y_2) integral as 0.

By Theorem 6.19 below, the two integrals both exist and have the same value when either exists. To prove this theorem we replace the (RRS) integral in (6.16) and (6.17) by the (RYS) integral. Let

$$f_-^{(a,b)}(x) := \begin{cases} f(a) & \text{for } x = a \\ f(x-) & \text{for } a < x < b \\ f(b) & \text{for } x = b \end{cases} \qquad \text{and} \qquad f_+^{(a,b)}(x) := \begin{cases} f(a) & \text{for } x = a \\ f(x+) & \text{for } a < x < b \\ f(b) & \text{for } x = b \end{cases}$$

Lemma 6.16. *Let $f, h \in \mathcal{R}([a,b])$. The integral $(RRS)\int_a^b f_+^{(b)}\,dh_-^{(a)}$ exists if and only if $(RYS)\int_a^b f_+^{(a,b)}\,dh$ does, and if they exist, they satisfy*

$$(RYS)\int_a^b f_+^{(a,b)}\,dh = (RRS)\int_a^b f_+^{(b)}\,dh_-^{(a)} - [\Delta^+ f\Delta^+ h](a) + [f\Delta^- h](b). \tag{6.18}$$

The integral $(RRS)\int_a^b f_-^{(a)}\,dh_+^{(b)}$ exists if and only if $(RYS)\int_a^b f_-^{(a,b)}\,dh$ does, and if they exist, then

$$(RYS)\int_a^b f_-^{(a,b)}\,dh = (RRS)\int_a^b f_-^{(a)}\,dh_+^{(b)} + [f\Delta^+ h](a) + [\Delta^- f\Delta^- h](b). \tag{6.19}$$

Proof. By Theorem 6.13 and Lemma 6.14, the integral $(RRS)\int_a^b f_+^{(b)}\,dh_-^{(a)}$ exists if and only if $(RYS)\int_a^b f_+^{(b)}\,dh_-^{(a)}$ does, and then both have the same value. Let $\kappa = \{x_i : i = 0, \dots, n\} \in$ PP $([a,b])$ and let $\sigma = \{y_i : i = 1, \dots, n\}$ be an open intermediate partition of κ. Then

$$\begin{aligned} S_Y(f_+^{(b)}, h_-^{(a)}; \kappa, \sigma) &= \sum_{i=1}^{n} f(y_i+)[h(x_i-) - h(x_{i-1}+)] + [f_+\Delta^+ h](a) + \sum_{i=1}^{n-1}[f_+\Delta^{\pm} h](x_i) \\ &= S_Y(f_+^{(a,b)}, h; \kappa, \sigma) + [\Delta^+ f\Delta^+ h](a) - [f\Delta^- h](b). \end{aligned}$$

Therefore, $(RYS)\int_a^b f_+^{(b)}\,dh_-^{(a)}$ exists if and only if $(RYS)\int_a^b f_+^{(a,b)}\,dh$ does, and (6.18) holds. This proves the first statement. The proof of the second one is similar and is omitted. □

The following is an easy consequence of the preceding lemma.

Proposition 6.17. *For $f, h \in \mathcal{R}([a,b])$, the relations*

$$(\mathrm{Y}_1)\int_a^b f\,dh = (RYS)\int_a^b f_+^{(a,b)}\,dh - \sum_{(a,b)} \Delta^+ f\Delta^{\pm} h \tag{6.20}$$

and

$$(\mathrm{Y}_2)\int_a^b f\,dh = (RYS)\int_a^b f_-^{(a,b)}\,dh + \sum_{(a,b)} \Delta^- f\Delta^{\pm} h \tag{6.21}$$

hold; that is, the integral on the left side exists if and only if both the integral on the right side exists and the sum converges unconditionally.

For the proofs of Theorems 6.19 and 6.20, as well as for later use, we prove the following fact:

Lemma 6.18. *Let $f, h \in \mathcal{R}([a,b])$. The following three statements are equivalent:*
(1). $(RYS) \int_a^b f\,dh$ *exists;*
(2). $(RYS) \int_a^b f_+^{(a,b)}\,dh$ *and* $(RYS) \int_a^b \Delta_{a,b}^+ f\,dh$ *exist;*
(3). $(RYS) \int_a^b f_-^{(a,b)}\,dh$ *and* $(RYS) \int_a^b \Delta_{a,b}^- f\,dh$ *exist.*
If any one of the preceding three statements holds then

$$(RYS)\int_a^b f\,dh = (RYS)\int_a^b f_+^{(a,b)}\,dh - \sum_{(a,b)} \Delta^+ f \Delta^\pm h = (RYS)\int_a^b f_-^{(a,b)}\,dh + \sum_{(a,b)} \Delta^- f \Delta^\pm h,$$

where the two sums converge unconditionally.

Proof. We prove only the implication (1) $\Rightarrow$ (2), because the converse follows from linearity of the (RYS) integral and the proof of (2) $\Leftrightarrow$ (3) is similar. Suppose (1) holds. First we show that the sum

$$\sum_{(a,b)} \Delta^+ f \Delta^\pm h \qquad \text{converges unconditionally.} \tag{6.22}$$

Let $\epsilon > 0$. By hypothesis, there exists $\lambda \in \mathrm{PP}([a,b])$ such that

$$\Big| S_Y(f,h;\kappa,\sigma) - (RYS)\int_a^b f\,dh \Big| < \epsilon \tag{6.23}$$

for any refinement κ of λ and open intermediate partition σ of κ. Let $\nu = \{u_l : l = 1, \dots, k\}$ be a finite set of points in (a,b) disjoint from λ. Choose a refinement $\kappa = \{x_i : i = 0, \dots, n\}$ of λ such that for each $l = 1, \dots, k-1$, and some $i(l)$, $l = 1, \dots, k$, $x_{i(l)-1} < u_l < x_{i(l)} < \dots < x_{i(l+1)-1} < u_{l+1} < x_{i(l+1)}$, where $x_{i(l)-1}, x_{i(l)} \notin \lambda$. Let $\sigma' = \{y_i' : i = 1, \dots, n\}$ and $\sigma'' = \{y_i'' : i = 1, \dots, n\}$ be two open intermediate partitions of κ such that $y_{i(l)}' > y_{i(l)}'' = u_l$ for $l = 1, \dots, k$ and $y_i' = y_i'' \in (x_{i-1}, x_i)$ for each $i \notin \{i(1), \dots, i(k)\}$. By (6.23), it follows that

$$|S_Y(f,h;\kappa,\sigma') - S_Y, h;\kappa,\sigma'')| = \Big| \sum_{l=1}^k [f(y_{i(l)}') - f(u_l)][h(x_{i(l)}-) - h(x_{i(l)-1}+)] \Big| < 2\epsilon.$$

Letting $y_{i(l)}' \downarrow u_l$, $x_{i(l)-1} \uparrow u_l$ and $x_{i(l)} \downarrow u_l$, as is possible, one gets $|\sum_{l=1}^k \Delta^+ f(u_l) \Delta^\pm h(u_l)| \le 2\epsilon$ for any finite set $\nu = \{u_l : l = 1, \dots, k\} \subset (a,b)$ which is disjoint from λ. Since $\epsilon > 0$ is arbitrary one can conclude that (6.22) holds.

Next we show that $\Delta_{a,b}^+ f$ is (RYS) integrable with respect to h and

$$(RYS)\int_a^b \Delta_{a,b}^+ f\,dh = \sum_{(a,b)} \Delta^+ f \Delta^\pm h. \tag{6.24}$$

For each partition $\kappa = \{x_i : i = 0, 1, \dots, n\} \in \mathrm{PP}([a,b])$ and open intermediate partition $\sigma = \{y_i : i = 1, \dots, n\}$ of κ, we have

$$\begin{aligned} \Big| S_Y(\Delta_{a,b}^+ f, h;\kappa,\sigma) - \sum_{(a,b)} \Delta^+ f \Delta^\pm h \Big| &\le \Big| \sum_{i=1}^n \Delta^+ f(y_i)[h(x_i-) - h(x_{i-1}+)] \Big| \\ + \quad \Big| \sum_{(a,b)} \Delta^+ f \Delta^\pm h - \sum_{i=1}^{n-1} [\Delta^+ f \Delta^\pm h](x_i) \Big| &=: T_1(\kappa,\sigma) + T_2(\kappa). \end{aligned} \tag{6.25}$$

Given $\epsilon > 0$ choose $\lambda \in \mathrm{PP}([a,b])$ such that (6.23) holds and $T_2(\kappa) < \epsilon$ for all refinements $\kappa = \{x_i\colon i = 0,1,\dots,n\}$ of λ and all open intermediate partitions $\sigma = \{y_i\colon i = 1,\dots,n\}$ of κ. Then

$$\Big|\sum_{i=1}^{n}[f(y_i') - f(y_i)][h(x_i-) - h(x_{i-1}+)]\Big| < 2\epsilon$$

whenever $y_i' \in (y_i, x_i)$ for $i = 1,\dots,n$. Letting $y_i' \downarrow y_i$ for each $i = 1,\dots,n$, yields the bound $T_1(\kappa,\sigma) \le 2\epsilon$. Therefore the right side of (6.25) does not exceed 3ϵ for each refinement κ of λ and each open intermediate partition σ of κ. Since $\epsilon > 0$ is arbitrary (6.24) holds. By linearity of the (RYS) integral, (2) follows and the last conclusion holds. □

Theorem 6.19. *Let $f, h \in \mathcal{R}([a,b])$. The integral $(Y_1)\int_a^b f\,dh$ exists if and only if $(Y_2)\int_a^b f\,dh$ does, and then they are equal.*

Proof. Suppose the (Y_2) integral exists. By Proposition 6.17, $(RYS)\int_a^b f_-^{(a,b)}\,dh$ exists and (6.21) holds. By Lemma 6.18 applied to $f = f_-^{(a,b)}$, $(RYS)\int_a^b f_+^{(a,b)}\,dh$ exists and

$$(RYS)\int_a^b f_+^{(a,b)}\,dh = (RYS)\int_a^b f_-^{(a,b)}\,dh - \sum_{(a,b)} \Delta^{\pm} f \Delta^{\pm} h,$$

where the sum converges unconditionally. Also by linearity, the sum

$$\sum_{(a,b)} \Delta^{+} f \Delta^{\pm} h = \sum_{(a,b)} \Delta^{\pm} f \Delta^{\pm} h - \sum_{(a,b)} \Delta^{-} f \Delta^{\pm} h$$

converges unconditionally. Since the right side of (6.20) is defined, the (Y_1) integral exists and

$$\begin{aligned}
(Y_2)\int_a^b f\,dh &= (RYS)\int_a^b f_-^{(a,b)}\,dh + \sum_{(a,b)} \Delta^{-} f \Delta^{\pm} h \\
&= (RYS)\int_a^b f_+^{(a,b)}\,dh - \sum_{(a,b)} \Delta^{\pm} f \Delta^{\pm} h + \sum_{(a,b)} \Delta^{-} f \Delta^{\pm} h \\
&= (RYS)\int_a^b f_+^{(a,b)}\,dh - \sum_{(a,b)} \Delta^{+} f \Delta^{\pm} h = (Y_1)\int_a^b f\,dh.
\end{aligned}$$

The proof of the converse implication is symmetric and we omit it. □

Theorem 6.20. *For $f, h \in \mathcal{R}([a,b])$, the integral $(RYS)\int_a^b f\,dh$ exists if and only if the two integrals $(Y_2)\int_a^b f\,dh$ and $(RYS)\int_a^b \Delta_{a,b}^{-} f\,dh$ exist. Also, the relations*

$$(RYS)\int_a^b f\,dh = (Y_2)\int_a^b f\,dh \quad \textit{and} \quad (RYS)\int_a^b \Delta_{a,b}^{-} f\,dh = \sum_{(a,b)} \Delta^{-} f \Delta^{\pm} h \tag{6.26}$$

hold with the sum converging unconditionally, whenever the three integrals exist.

Proof. Suppose that $(Y_2)\int_a^b f\,dh$ and $(RYS)\int_a^b \Delta_{a,b}^{-} f\,dh$ exist. Then by Proposition 6.17, $(RYS)\int_a^b f_-^{(a,b)}\,dh$ exists. Thus $(RYS)\int_a^b f\,dh$ exists by linearity of the (RYS) integral. The relations (6.26) then follow from (6.21).

Suppose that $(RYS)\int_a^b f\,dh$ exists. By statement (3) of Lemma 6.18, $(RYS)\int_a^b \Delta_{a,b}^- f\,dh$ and the right side of (6.21) exist. Thus $(Y_2)\int_a^b f\,dh$ exists and equals $(RYS)\int_a^b f\,dh$ by Proposition 6.17. The proof of Theorem 6.20 is complete. □

Notation. For a set E of real numbers, let $c_0(E)$ be the set of all functions $f\colon E \mapsto \mathbb{R}$ for which there exists a set $\{\xi_i\colon i \geq 1\} \subset E$ of distinct numbers such that $f(\xi_i) \to 0$ as $i \to \infty$ and $f(x) = 0$ if $x \in E \setminus \{\xi_i\colon i \geq 1\}$. Then

$$c_0(E) = \Big\{\sum_i f_i 1_{\{\xi_i\}}\colon\ \xi_i \neq \xi_j \text{ for } i \neq j,\ \xi_i \in E,\ f_i \to 0 \text{ as } i \to \infty\Big\}. \tag{6.27}$$

If $f \in c_0(J)$ for an open interval J then $f \in \mathcal{R}(J)$ with $f_+ \equiv f_- \equiv 0$. If $f \in c_0((a,b))$ then $f_+(a) = f_-(b) = 0$. Thus defining $f(a) = f(b) := 0$, we have $f_-^{(a)} \equiv f_+^{(b)} \equiv 0$ on $[a,b]$.

Proposition 6.21. *There exist functions f and h on $[0,1]$, where $f \in c_0((0,1))$ and h is continuous, for which $(Y_2)\int_0^1 f\,dh$ exists, while $(RYS)\int_0^1 \Delta_{0,1}^- f\,dh$ and hence $(RYS)\int_a^b f\,dh$ don't exist.*

Proof. Let $f(x) := k^{-1/2}$ if $x = 1/(3k)$ for $k = 1,2,\ldots$, and $f(x) := 0$ otherwise. Let $h(1/(3k+1)) := 0$ and $h(1/(3k-1)) := k^{-1/2}$ for $k = 1,2,\ldots$. Let $h(0) = h(1) := 0$ and let h be "linear in between", i.e. on each closed interval where h is so far defined only at the endpoints. Since h is continuous and $f_-^{(0)} \equiv f_+^{(1)} \equiv 0$, $(Y_2)\int_0^1 f\,dh$ exists and is 0. Since $f_-^{(0,1)} \equiv 0$ it is enough to show the non-existence of the integral $(RYS)\int_0^1 f\,dh$. Let $\lambda = \{z_j\colon j = 0,\ldots,m\}$ be any partition of $[0,1]$. Take the smallest m such that $x_m := 1/(3m+1) \leq z_1$. If $\kappa_1 := \lambda \cup \{x_m\}$ then the contribution to any Young sum based on κ_1 coming from $[0,x_m]$ is 0. For $n > m$, consider partitions

$$\kappa_n := \lambda \cup \{1/(3k+1), 1/(3k-1)\colon k = m+1,\ldots,n\}.$$

We form a Young sum based on κ_n by letting it be the same as one for κ_1 on $[x_m,1]$, and by evaluating f at $1/(3k)$ for $k = m+1,\ldots,n$. Thus the part of our Young sum for κ_n coming from $[0,x_m]$ is

$$\sum_{k=m+1}^{n} f\Big(\frac{1}{3k}\Big)\Big[h\Big(\frac{1}{3k-1}\Big) - h\Big(\frac{1}{3k+1}\Big)\Big] + 0\cdot\Big[h\Big(\frac{1}{3k-2}\Big) - h\Big(\frac{1}{3k+1}\Big)\Big] = \sum_{k=m+1}^{n} k^{-1/2}k^{-1/2} \to \infty$$

as $n \to \infty$. Thus two Young sums for $(RYS)\int_0^1 f\,dh$, both based on refinements of λ, differ by an arbitrarily large amount. So $(RYS)\int_0^1 f\,dh$ doesn't exist. □

For a set E and $0 < p < \infty$, let

$$\ell_p(E) := \Big\{f = \sum_i f_i 1_{\{\xi_i\}} \in c_0(E)\colon \sum_i |f_i|^p < \infty\Big\}.$$

Proposition 6.22. *Let $f,h \in \mathcal{R}([a,b])$ be such that $\Delta^- f \in \ell_p((a,b))$ and $h \in \mathcal{W}_{p'}([a,b])$ for some p, p' with $1/p+1/p' = 1$, $1 \leq p \leq \infty$. Then $(RYS)\int_a^b f\,dh$ exists if and only if $(Y_2)\int_a^b f\,dh$ does, and the two integrals are equal.*

Proof. The proof is given only for the case $p < \infty$ because a proof for the case $p = \infty$ is similar. By Theorem 6.20, it is enough to prove that $(RYS)\int_a^b \Delta_{a,b}^- f\,dh$ exists. By Hölder's inequality, the sum $\sum_{(a,b)} \Delta^- f \Delta^{\pm} h$ converges unconditionally. Let $\epsilon > 0$. Then there exists a finite set $\lambda_1 \subset (a,b)$ such that

$$\Big| \sum_{x\in\mu} [\Delta^- f \Delta^{\pm} h](x) - \sum_{(a,b)} \Delta^- f \Delta^{\pm} h \Big| < \epsilon$$

for any set $\mu \subset (a,b)$ containing λ_1. Since $\Delta^- f \in \ell_p((a,b))$ and $h \in \mathcal{W}_{p'}([a,b])$, there exists a finite set $\lambda_2 \subset (a,b)$ such that

$$\sum_{y\in\nu} |\Delta^- f(y)|^p < \Big(\epsilon / V_{p'}(h;[a,b])\Big)^p$$

for any set $\nu \subset (a,b)$ disjoint from λ_2. Let $\lambda := \{a,b\} \cup \lambda_1 \cup \lambda_2$. Then $\lambda \in \mathrm{PP}\,([a,b])$. For any refinement $\kappa = \{x_i \colon i = 0,\ldots,n\}$ of λ and open intermediate partition $\sigma = \{y_i \colon i = 1,\ldots,n\}$ of κ, we then have

$$\Big| S_Y(\Delta^-_{a,b} f, h; \kappa, \sigma) - \sum_{(a,b)} \Delta^- f \Delta^{\pm} h \Big| \le \Big| \sum_{i=1}^{n-1} [\Delta^- f \Delta^{\pm} h](x_i) - \sum_{(a,b)} \Delta^- f \Delta^{\pm} h \Big|$$

$$+ \Big| \sum_{i=1}^{n} \Delta^- f(y_i)[h(x_i-) - h(x_{i-1}+)] \Big| \le \epsilon + \Big(\sum_{i=1}^{n} |\Delta^- f(y_i)|^p \Big)^{1/p} V_{p'}(h;[a,b]) < 2\epsilon.$$

Therefore the integral $(RYS) \int_a^b \Delta^-_{a,b} f\, dh$ exists, and the proof is complete. □

6.4 The $(\mathfrak{I})$–integral of A. N. Kolmogorov

Kolmogorov [63] suggested a unified approach to a class of integrals with respect to interval or set functions. The Kolmogorov integral includes the Lebesgue, Lebesgue-Stieltjes and several other integrals. Some of them are considered below. We recall and use Kolmogorov's definition (Definition 6.24 below) in the case when only finite partitions are used.

Let $\mathfrak{M}$ be a multiplicative system of sets; that is, $J_1 \cap J_2 \in \mathfrak{M}$ whenever $J_1, J_2 \in \mathfrak{M}$. For $J \in \mathfrak{M}$, a vector $\mathfrak{D} = (J_1,\ldots,J_N)$ of pairwise disjoint sets in $\mathfrak{M}$ will be called an *$\mathfrak{M}$-partition of J* if $J = \cup_{i=1}^N J_i$. The sets $J_1,\ldots,J_N$ are called *elements of the $\mathfrak{M}$-partition* $\mathfrak{D}$ and written $J_i \in \mathfrak{D}$, $i = 1,\ldots,N$. Let $\mathrm{IP}\,(\mathfrak{M})$ be the set of all $\mathfrak{M}$-partitions $\mathfrak{D}$ of J. For example, if $\mathfrak{M}$ consists of a single set $J \neq \emptyset$ then $\mathrm{IP}\,(\mathfrak{M})$ consists of the single partition of J into itself. An $\mathfrak{M}$-partition $\mathfrak{D}''$ is an *$\mathfrak{M}$-refinement* of an $\mathfrak{M}$-partition $\mathfrak{D}'$ if, for each element J'' of $\mathfrak{D}''$, $J'' \subset J'$ for some element J' of $\mathfrak{D}'$. In this case we write $\mathfrak{D}' \sqsubseteq \mathfrak{D}''$. Then $(\mathrm{IP}\,(\mathfrak{M}), \sqsubseteq)$ is a partially ordered set (a poset). For $\mathfrak{D}', \mathfrak{D}'' \in \mathrm{IP}\,(\mathfrak{M})$ let $\mathfrak{D} := \{J' \cap J'' \colon J' \in \mathfrak{D}', J'' \in \mathfrak{D}''\}$. Then $\mathfrak{D}' \sqsubseteq \mathfrak{D}$, $\mathfrak{D}'' \sqsubseteq \mathfrak{D}$ and $\mathfrak{D} \in \mathrm{IP}\,(\mathfrak{M})$ because $\mathfrak{M}$ is multiplicative. Therefore the poset $(\mathrm{IP}\,(\mathfrak{M}), \sqsubseteq)$ is upwards directed.

Definition 6.23. Let S be a function on the set of $\mathfrak{M}$-partitions of $J \in \mathfrak{M}$ whose values are sets of real numbers. We say that a real number A is the limit of S *under refinements of $\mathfrak{M}$-partitions of* J and write $A = \lim_{\mathfrak{D}\uparrow} S(\mathfrak{D})$, if for each $\epsilon > 0$ there is an $\mathfrak{M}$-partition $\mathfrak{D}$ of J such that for all $\mathfrak{D}' \sqsupseteq \mathfrak{D}$, we have $\sup\{|x - A| \colon x \in S(\mathfrak{D}')\} < \epsilon$.

For $J \in \mathfrak{M}$, let $\mathfrak{M}(J)$ be the collection of elements of all partitions $\mathfrak{D}$ of J. Then $\mathfrak{M}(J)$ is a multiplicative system. For a function ν defined on all elements of an $\mathfrak{M}$-partition $\mathfrak{D} = (J_1, \ldots, J_n)$, let

$$S_{\mathfrak{M}}(\nu; \mathfrak{D}) := \sum_{i=1}^{n} \nu(J_i).$$

Definition 6.24. Let ν be a function on $\mathfrak{M}(J)$ for some $J \in \mathfrak{M}$ whose values are sets of real numbers. The $(\mathfrak{M})$*-integral of* ν *on* J will be defined by

$$(\mathfrak{M}) \int_J d\nu := \lim_{\mathfrak{D}\uparrow} S_{\mathfrak{M}}(\nu; \mathfrak{D})$$

provided the limit of $S_{\mathfrak{M}}(\nu; \cdot)$ exists under refinements of $\mathfrak{M}$-partitions of J.

We refer to Kolmogorov [63] for elementary properties of the $(\mathfrak{M})$-integral.

Recall that in Section 4.5 we considered two classes of subintervals of $[a,b]$: $\mathfrak{I}_{os}([a,b]) := \{(c,d), \{c\}\colon a \le c \le d \le b\}$ and $\mathfrak{I}([a,b]) := \{[c,d], (c,d], [c,d), (c,d)\colon a \le c \le d \le b\}$. Next we shall consider the $(\mathfrak{M})$-integral when $\mathfrak{M}(J) = \mathfrak{I}_{os}([a,b])$ or $\mathfrak{M}(J) = \mathfrak{I}([a,b])$, and ν is a point function f multiplied by an additive interval function μ.

Let $\mathfrak{I}$ be either of the two classes $\mathfrak{I}_{os}([a,b])$ or $\mathfrak{I}([a,b])$. These classes possess the following decomposability property (used in Chapter 3, §2 by Kolmogorov [63]): if $J', J \in \mathfrak{I}$ and $J' \subset J$ then there is an $\mathfrak{I}$-partition $\mathfrak{D}$ of J such that J' is an element of $\mathfrak{D}$. Let $f\colon [a,b] \mapsto \mathbb{R}$ be a point function and let $\mu \in \mathcal{I}[a,b]$ be additive. For each $\mathfrak{D} = (J_1, \ldots, J_N) \in \mathrm{IP}\,(\mathfrak{I})$ and $\theta = (y_1, \ldots, y_N) \in J_1 \times \cdots \times J_N$, let

$$S_{\mathfrak{I}}(f, \mu; \mathfrak{D})(\theta) := \sum_{j=1}^{N} f(y_j)\mu(J_j).$$

According to Definition 6.24, the $(\mathfrak{I})$–integral of f with respect to an additive interval function μ on $[a,b]$ is defined by

$$(\mathfrak{I}) \int_{[a,b]} f\, d\mu := \lim_{\mathfrak{D}\uparrow} S_{\mathfrak{I}}(f, \mu; \mathfrak{D})$$

provided the limit exists under refinements of $\mathfrak{I}$–partitions of $[a,b]$. Thus the $(\mathfrak{I})$-integral with $\mathfrak{I} = \mathfrak{I}([a,b])$ is the same as the (RYS) integral and the class of partitions $\mathrm{IP}\,(\mathfrak{I}([a,b]))$ is the same as $\mathrm{IP}\,([a,b])$, both defined in Section 2.1.

Next we describe a relation between $\mathfrak{I}_{os}$–partitions from $\mathrm{IP}_{os}([a,b]) := \mathrm{IP}\,(\mathfrak{I}_{os}([a,b]))$ and corresponding point partitions of $[a,b]$. Let $a = x_0 < x_1 < \cdots < x_n = b$. Then $\kappa = \{x_i\colon i = 0, 1, \ldots, n\}$ is a partition of $[a,b]$ by points and $\mathrm{PP}\,([a,b])$ is the set of all point partitions κ of $[a,b]$. For each $\mathfrak{D} = (J_1, \ldots, J_N) \in \mathrm{IP}_{os}([a,b])$, we have $J_1 = \{a\}$, $J_2 = (a, x_1)$, $J_3 = \{x_1\}$, $J_4 = (x_1, x_2)$, $\ldots$, $J_{2n} = (x_{n-1}, b)$, $J_{2n+1} = \{b\}$ and $N = 2n+1$. Let $\wp$ be the mapping from $\mathrm{IP}_{os}([a,b])$ into $\mathrm{PP}\,([a,b])$ defined by $\wp(\mathfrak{D}) := \{x_0, x_1, \ldots, x_n\} \in \mathrm{PP}\,([a,b])$. Then $\wp$ is an order isomorphism between $(\mathrm{IP}_{os}([a,b]), \sqsubseteq\,)$ and $(\mathrm{PP}\,([a,b]), \subseteq\,)$, that is, for each $\mathfrak{D}_1, \mathfrak{D}_2 \in \mathrm{IP}_{os}([a,b])$, $\mathfrak{D}_1 \sqsubseteq \mathfrak{D}_2$ if and only if $\wp(\mathfrak{D}_1) \subseteq \wp(\mathfrak{D}_2)$.

Let $h \in \mathcal{R}[a,b]$, $\kappa = \{x_i\colon i = 0, \ldots, n\} \in \mathrm{PP}\,([a,b])$ and let $\sigma = \{y_i\colon i = 1, \ldots, n\}$ be an open intermediate partition of κ, meaning that $x_{i-1} < y_i < x_i$ for $i = 1, \ldots, n$. Recall that the Young sum is defined by

$$S_Y(f, h; \kappa, \sigma) := \sum_{i=1}^{n} \{[f\Delta^+ h](x_{i-1}) + f(y_i)[h(x_i-) - h(x_{i-1}+)] + [f\Delta^- h](x_i)\}.$$

Notice that the value of the Young sum is the same for each function $h \in [\mu_h]_a$ (cf. (4.34) above); that is, the value of the integral $(RYS) \int_a^b f\,dh$ does not depend on the values of h at its jumps. Also, the existence of $(RYS) \int_a^b f\,dh$ implies the existence of $(\mathfrak{I}_{os}) \int_{[a,b]} f\,d\mu_h$. This follows by comparing the Young sum $S_Y(f,h;\kappa,\sigma)$ with the sum $S_{\mathfrak{I}_{os}}(f,\mu_h;\mathfrak{D})(\theta)$, where $\wp(\mathfrak{D}) = \kappa$ and the coordinates of θ corresponding to the open intervals of $\mathfrak{D}$ are the coordinates of σ. We show next that the converse implication holds whenever an additive interval function μ is upper continuous (see Definition 4.32).

Theorem 6.25. *Let $\mu \in \mathcal{I}[a,b]$ be additive. The following statements about μ are equivalent:*
- (a) *μ is upper continuous;*
- (b) *for any $f\colon [a,b] \to \mathbb{R}$, if $(\mathfrak{I}_{os}) \int_{[a,b]} f\,d\mu$ exists then so does $(RYS) \int_a^b f\,dh$ for each $h \in [\mu]_a$, and the two integrals are equal;*
- (c) *there exists $h \in \mathcal{R}[a,b]$ with $h(a) = 0$ such that, for any $f\colon [a,b] \to \mathbb{R}$, $(RYS) \int_a^b f\,dh$ exists whenever $(\mathfrak{I}_{os}) \int_{[a,b]} f\,d\mu$ does, and the two integrals are equal.*

Proof. (a) $\Rightarrow$ (b). Let $\mathfrak{D} = (J_1,\dots,J_N) \in \mathrm{IP}_1([a,b])$ and $J_{2i} = (x_{i-1},x_i)$ for $i = 1,\dots,n$ with $N = 2n+1$. Then $\wp\colon \mathfrak{D} \mapsto \wp(\mathfrak{D}) = \{x_i\colon i = 0,1,\dots,n\} =: \kappa \in \mathrm{PP}([a,b])$ is an order isomorphism between $\mathrm{IP}_{os}([a,b])$ and $\mathrm{PP}([a,b])$. Let $\theta = \{x_0,y_1,x_1,\dots,y_n,x_n\}$, where $y_i \in (x_{i-1},x_i)$ for $i = 1,\dots,n$. Since μ is upper continuous and $h \in [\mu]_a$, by Theorem 4.36, we have $\mu((x_{i-1},x_i)) = \mu_h((x_{i-1},x_i)) = h(x_i-) - h(x_{i-1}+)$ for $i = 1,\dots,n$, $\mu(\{a\}) = h(a+)$, $\mu(\{x_i\}) = \mu_h(\{x_i\}) = h(x_i+) - h(x_i-)$ for $i = 1,\dots,n-1$ and $\mu(\{b\}) = h(b) - h(b-)$. Then we have

$$S_{\mathfrak{I}_1}(f,\mu;\mathfrak{D})(\theta) = \sum_{i=1}^n f(y_i)\mu((x_{i-1},x_i)) + \sum_{i=0}^n f(x_i)\mu(\{x_i\}) = S_Y(f,h;\kappa,\sigma),$$

where $\sigma = (y_1,\dots,y_n)$. Therefore $(RYS) \int_a^b f\,dh$ exists whenever $(\mathfrak{I}_{os}) \int_{[a,b]} f\,d\mu$ does, and the two integrals are equal.

(b) $\Rightarrow$ (c). This implication is clear because $h \in \mathcal{R}[a,b]$ and $h(a) = 0$ for each $h \in [\mu]_a$ by definition (4.34).

(c) $\Rightarrow$ (a). Let $a \le u < v \le b$. Then we have

$$\mu_h((u,v)) = h(v-) - h(u+) = (RYS) \int_a^b 1_{(u,v)}\,dh = (\mathfrak{I}_{os}) \int_{[a,b]} 1_{(u,v)}\,d\mu = \mu((u,v)).$$

Similarly, it follows that $\mu_h(\{w\}) = \mu(\{w\})$ for $w \in [a,b]$. The implication then follows from Theorem 4.36. The proof is complete. $\square$

Since $\mathrm{IP}_{os}([a,b]) \subset \mathrm{IP}([a,b])$ and each $\mathfrak{I}$-partition can be $\mathfrak{I}_{os}$-refined, if $(\mathfrak{I}) \int_{[a,b]} f\,d\mu$ exists then so does $(\mathfrak{I}_{os}) \int_{[a,b]} f\,d\mu$, and the two integrals are equal. We show next that the converse holds for upper continuous additive interval functions.

Theorem 6.26. *Let $\mu \in \mathcal{I}[a,b]$ be additive and upper continuous, and let $f\colon [a,b] \to \mathbb{R}$. If $(\mathfrak{I}_{os}) \int_{[a,b]} f\,d\mu$ exists then so does $(\mathfrak{I}) \int_{[a,b]} f\,d\mu$, and the two are equal.*

Proof. Let $\epsilon > 0$. There exists $\mathfrak{D}_0 \in \mathrm{IP}_{os}([a,b]) \subset \mathrm{IP}([a,b])$ such that

$$|S_{\mathfrak{I}_{os}}(f,\mu;\mathfrak{D})(\theta) - (\mathfrak{I}_{os}) \int_{[a,b]} f\,d\mu| \le \epsilon \tag{6.28}$$

for all $\mathfrak{I}_{os}$-refinements $\mathfrak{D} = \{J_1, \dots, J_N\}$ of $\mathfrak{D}_0$ and all $\theta \in J_1 \times \cdots \times J_N$. It will be shown that (6.28) also holds for $\mathfrak{I}$ in place of $\mathfrak{I}_{os}$, with the integral having the same value.

Let $\mathfrak{D}_1 = (J_1, \dots, J_N)$ be an $\mathfrak{I}$-refinement of $\mathfrak{D}_0$, $\theta = (y_1, \dots, y_N) \in J_1 \times \cdots \times J_N$ and let $\lambda := \wp(\mathfrak{D}_0) = \{z_0, \dots, z_L\} \in \mathrm{PP}\,([a,b])$. Let $\kappa := \{x_i : i = 0, \dots, m\}$ be the set of endpoints of intervals in $\mathfrak{D}_1$. So $\kappa \in \mathrm{PP}\,([a,b])$. For each $i = 0, 1, \dots, m$, define a sequence $\{x_{in} : n \geq 1\}$ as follows. If $\{x_i\}$ is a singleton in $\mathfrak{D}_1$ then let $x_{in} := x_i$ for all $n \geq 1$. If an interval $\{\cdot, x_i] \in \mathfrak{D}_1$ let $x_{in} \downarrow x_i$ or if some $[x_i, \cdot\} \in \mathfrak{D}_1$ let $x_{in} \uparrow x_i$, where in either case x_{in} are not atoms of μ. This can be done by statement (e) of Theorem 4.35. For n large enough, $x_{0n} = a < x_{1n} < \cdots < x_{mn} = b$. Then there is a unique $\mathfrak{I}_{os}$-partition $\mathfrak{D}_{2,n}$ having $\wp(\mathfrak{D}_{2,n}) = \{x_{in} : i = 0, \dots, m\} \in \mathrm{PP}\,([a,b])$. Since each singleton $\{z_j\} = \{x_i\} \in \mathfrak{D}_1$ for some i, $\mathfrak{D}_{2,n}$ is a refinement of $\mathfrak{D}_0$. The intervals in $\mathfrak{D}_{2,n}$ converge to the intervals in $\mathfrak{D}_1$ as $n \to \infty$, except for singletons in $\mathfrak{D}_{2,n}$ with $\mu = 0$ which do not contribute to sums. Each point y_i is eventually in the interval of $\mathfrak{D}_{2,n}$ corresponding to the interval of $\mathfrak{D}_1$. Thus there are sums $S_{\mathfrak{I}_{os}}(f, \mu; \mathfrak{D}_{2,n})(\theta_n)$ converging to $S_{\mathfrak{I}}(f, \mu; \mathfrak{D}_1)(\theta)$, where the coordinates of θ_n are those of θ, and possibly some y_{in} with $\{y_{in}\} \in \mathfrak{D}_{2,n}$ and $\mu(\{y_{in}\}) = 0$. Theorem 6.26 then follows. $\square$

6.5 The Ward–Perron–Stieltjes and gauge integrals

> The Henstock integral is designed to integrate highly oscillatory functions which the Lebesgue integral fails to do. It is known as nonabsolute integration and is a powerful tool.
>
> – Lee Peng–Yee [89, p. vii].

> If we think of the Lebesgue integral as God sent, then differentiable functions whose derivatives are not Lebesgue integrable may appear evil. The challenge is to resolve the conflict ...
>
> – unattributed aphorism in Pfeffer [90, p. xii].

Ward (1936) defined a Perron–Stieltjes integral which includes both the Lebesgue–Stieltjes and Moore–Pollard–Stieltjes (RRS) integrals. Given two real-valued functions f, g on $[a,b]$, say M is a *major function* of f with respect to g if $M(a) = 0$, M has finite values on $[a,b]$, and for each $x \in [a,b]$ there exists $\delta(x) > 0$ such that

$$\begin{cases} M(z) \geq M(x) + f(x)[g(z) - g(x)] & \text{if} \quad x \leq z \leq \min\{b, x + \delta(x)\}, \\ M(z) \leq M(x) + f(x)[g(z) - g(x)] & \text{if} \quad \max\{a, x - \delta(x)\} \leq z \leq x. \end{cases} \tag{6.29}$$

Thus for each major function M there exists a positive function $\delta_M(\cdot) = \delta(\cdot)$ on $[a,b]$ which satisfies (6.29). Let $\mathcal{U}(f,g)$ be the class of all major functions of f with respect to g, and let

$$U(f,g) := \begin{cases} \inf\{M(b) : M \in \mathcal{U}(f,g)\} & \text{if } \mathcal{U}(f,g) \neq \emptyset, \\ +\infty & \text{if } \mathcal{U}(f,g) = \emptyset. \end{cases}$$

A function m is a *minor function* of f with respect to g if $-m \in \mathcal{U}(-f, g)$. Let $\mathcal{L}(f,g)$ be the class of all minor functions of f with respect to g. Thus $m \in \mathcal{L}(f,g)$ provided $m(a) = 0$, m has finite values on $[a,b]$, and there exists a positive function $\delta(\cdot) = \delta_m(\cdot)$ on $[a,b]$ such that

$$\begin{cases} m(z) \leq m(x) + f(x)[g(z) - g(x)] & \text{if} \quad x \leq z \leq \min\{b, x + \delta(x)\}, \\ m(z) \geq m(x) + f(x)[g(z) - g(x)] & \text{if} \quad \max\{a, x - \delta(x)\} \leq z \leq x. \end{cases} \tag{6.30}$$

Let

$$L(f,g) := \begin{cases} \sup\{m(b) \colon m \in \mathcal{L}(f,g)\} & \text{if } \mathcal{L}(f,g) \neq \emptyset, \\ -\infty & \text{if } \mathcal{L}(f,g) = \emptyset. \end{cases}$$

Then the following is true:

Lemma 6.27. *$L(f,g) \le U(f,g)$.*

Proof. The statement is true if at least one of the sets $\mathcal{L}(f,g)$ or $\mathcal{U}(f,g)$ is empty. So suppose that $\mathcal{L}(f,g)$ and $\mathcal{U}(f,g)$ are not empty. Let $m \in \mathcal{L}(f,g)$ and $M \in \mathcal{U}(f,g)$. Also, let $\delta := \delta_m \wedge \delta_M$ and $w(x) := M(x) - m(x)$ for $x \in [a,b]$. By (6.29) and (6.30), for each $x \in (a,b]$, it follows that

$$\begin{cases} w(z) \ge w(x) & \text{if} \quad x \le z \le \min\{b, x+\delta(x)\}, \\ w(z) \le w(x) & \text{if} \quad \max\{a, x-\delta(x)\} \le z \le x. \end{cases}$$

Therefore $\inf\{M(x) \colon M \in \mathcal{U}(f,g)\} - \sup\{m(x) \colon m \in \mathcal{L}(f,g)\}$ is a nondecreasing function of x. Then the statement of the lemma holds because $m(a) = M(a) = 0$. $\square$

If $U(f,g) = L(f,g)$ is finite then denote the common value by $(WPS) \int_a^b f\,dg$ and call it the *Ward–Perron–Stieltjes integral*, or the (WPS) integral. The following is Theorem 5 of Ward [111]. Note that in it, f and g are not necessarily regulated.

Theorem 6.28. *If $(RRS) \int_a^b f\,dg$ exists then so does $(WPS) \int_a^b f\,dg$, and has the same value.*

Proof. For $a \le u < v \le b$, let $M(u,v)$ be the least upper bound for Riemann–Stieltjes sums $S_{RS}(f,g;\kappa,\sigma)$ based on partitions $\kappa \in \mathrm{PP}([u,v])$ and intermediate partitions σ of κ. Then

$$\begin{cases} M(u,y) \ge M(u,x) + f(x)[g(y) - g(x)] & \text{if} \quad u < x < y, \\ M(u,y) \le M(u,x) + f(x)[g(y) - g(x)] & \text{if} \quad u < y < x. \end{cases} \tag{6.31}$$

Suppose $(RRS) \int_a^b f\,dg$ exists and $\epsilon > 0$. Then exists $\lambda = \{z_j \colon j = 0, \dots, m\} \in \mathrm{PP}([a,b])$ such that

$$(RRS) \int_a^b f\,dg - \epsilon < S_{RS}(f,g;\kappa,\sigma) < (RRS) \int_a^b f\,dg + \epsilon \tag{6.32}$$

for each refinement κ of λ and intermediate partition σ of κ. For each $x \in (a,b]$, let $j(x)$ be the greatest integer such that $z_{j(x)} \le x$. Let $M(a) := 0$, and for $x \in (a,b]$ let

$$M(x) := \sum_{j=1}^{j(x)} M(z_{j-1}, z_j) + M(z_{j(x)}, x),$$

where the sum over the empty set is 0. Define a positive function $\delta(\cdot)$ on $[a,b]$ by $\delta(x) := \min\{x - z_{j(x)}, z_{j(x)+1} - x\}$ if $x \in (z_{j(x)}, z_{j(x)+1})$, $\delta(a) := z_1 - a$, $\delta(b) := b - z_{m-1}$ and $\delta(z_j) := \min\{z_j - z_{j-1}, z_{j+1} - z_j\}$ for $j = 1, \dots, m-1$. By (6.31), the function M so defined satisfies (6.29). Hence it is a major function of f with respect to g. By (6.32), it follows that $M(b) \le (RRS) \int_a^b f\,dg + \epsilon$. Similarly one can define a minor function m such that $m(b) \ge (RRS) \int_a^b f\,dg - \epsilon$. Since ϵ is arbitrarily small this proves the theorem. $\square$

Kurzweil [68, Section 1.2] suggested an equivalent definition of the (WPS) integral based on an extension of the limit as the mesh of partitions tends to zero in the definition of the RS integral. For a partition $a = x_0 < x_1 < \cdots < x_n = b$, points $y_i \in [x_{i-1}, x_i]$, $i = 1, \dots, n$, are called *tags* and the pair of sets $\tau = (\{x_i \colon i = 0, \dots, n\}, \{y_i \colon i = 1, \dots, n\})$ is called a *tagged partition* of $[a, b]$. A *gauge function* is any function with strictly positive values. Given a gauge function $\delta(\cdot)$ on $[a, b]$, a tagged partition $\tau = (\{x_i \colon i = 0, \dots, n\}, \{y_i \colon i = 1, \dots, n\})$ is *δ-fine* if $y_i - \delta(y_i) \le x_{i-1} \le y_i \le x_i \le y_i + \delta(y_i)$ for $i = 1, \dots, n$. Let $TP(\delta, [a, b])$ be the set of all δ–fine tagged partitions of $[a, b]$.

Lemma 6.29. *For any gauge function δ, $TP(\delta, [a, b])$ is nonempty.*

Proof. Let $\delta(\cdot)$ be a gauge function on $[a, b]$. Then the system of open intervals $\{(y - \delta(y), y + \delta(y)) \colon y \in [a, b]\}$ is an open cover of $[a, b]$. Since $[a, b]$ is compact there is a finite subcover $\{J_i := (y_i - \delta(y_i), y_i + \delta(y_i) \colon i = 1, \dots, n\}$ of $[a, b]$ such that any $n - 1$ of J_i do not cover $[a, b]$. Since y_i, $i = 1, \dots, n$ are different we can assume that $y_1 < \cdots < y_n$. Also, $J_i \cap J_{i+1} \neq \emptyset$ for $i = 1, \dots, n - 1$, $J_i \cap J_j = \emptyset$ for $|i - j| > 1$, $i, j = 1, \dots, n$, $a \in J_1$ and $b \in J_n$. Hence one can find numbers $a = x_0 < x_1 < \ldots < x_n = b$ such that $x_i \in J_i \cap J_{i+1}$ for $i = 1, \dots, n - 1$ and $y_i \in [x_{i-1}, x_i]$ for $i = 1, \dots, n$. Thus $(\{x_i \colon i = 0, \dots, n\}, \{y_i \colon i = 1, \dots, n\}) \in TP(\delta, [a, b])$. $\square$

The sum

$$S_\delta(\tau) := S_\delta(f, g; \tau) := \sum_{i=1}^{n} f(y_i)[g(x_i) - g(x_{i-1})]$$

based on a δ–fine tagged partition $\tau = (\{x_i \colon i = 0, \dots, n\}, \{y_i \colon i = 1, \dots, n\})$ will be called the *gauge sum* based on τ. The *gauge* or *Henstock-Kurzweil* integral $(HK) \int_a^b f\, dg$ is defined as the number A, whenever it exists, such that for each $\epsilon > 0$ there is a gauge function $\delta(\cdot)$ on $[a, b]$ such that $|S_\delta(f, g; \tau) - A| < \epsilon$ for each δ-fine tagged partition τ of $[a, b]$. Kurzweil [68, Theorem 1.2.1] proved the following statement:

Theorem 6.30. *The integral $(WPS) \int_a^b f\, dg$ exists if and only if $(HK) \int_a^b f\, dg$ exists, and then their values are equal.*

Proof. Suppose $(WPS) \int_a^b f\, dg$ is defined. Then given $\epsilon > 0$, there exist a major function $M \in \mathcal{U}(f, g)$ and a minor function $m \in \mathcal{L}(f, g)$ such that

$$M(b) - \epsilon < (WPS) \int_a^b f\, dg < m(b) + \epsilon. \tag{6.33}$$

Let $\delta := \delta_M \wedge \delta_m$ and let $\tau = (\{x_i \colon i = 0, \dots, n\}, \{y_i \colon i = 1, \dots, n\})$ be any δ–fine tagged partition of $[a, b]$, which exists by Lemma 6.29. By (6.29), we have

$$M(x_i) - M(y_i) \ge f(y_i)[g(x_i) - g(y_i)] \quad \text{and} \quad M(y_i) - M(x_{i-1}) \ge f(y_i)[g(y_i) - g(x_{i-1})] \tag{6.34}$$

for each $i = 1, \dots, n$. Adding up $2n$ inequalities (6.34) we get the upper bound $M(b) \ge S_\delta(\tau)$ for the gauge sum $S_\delta(\tau)$. Similarly, using (6.30), we get the lower bound $S_\delta(\tau) \ge m(b)$. By (6.33), it then follows that

$$(WPS) \int_a^b f\, dg - \epsilon < S_\delta(\tau) < (WPS) \int_a^b f\, dg + \epsilon.$$

Thus $(HK) \int_a^b f\, dg$ exists with the same value as the (WPS) integral.

Now suppose $(HK) \int_a^b f\, dg$ is defined. Given $\epsilon > 0$ there exists a gauge function $\delta(\cdot)$ such that for each δ–fine tagged partition τ,

$$(HK) \int_a^b f\, dg - \epsilon < S_\delta(\tau) < (HK) \int_a^b f\, dg + \epsilon. \tag{6.35}$$

For each $x \in (a, b]$, let δ_x be the gauge function δ restricted to the interval $[a, x]$. Then for each $x \in (a, b]$, let

$$m(x) := \inf\{S_{\delta_x}(\tau) \colon \tau \in TP(\delta_x, [a, x])\}, \qquad M(x) := \sup\{S_{\delta_x}(\tau) \colon \tau \in TP(\delta_x, [a, x])\}$$

and $m(a) = M(a) := 0$. Then $m(x)$ and $M(x)$ have finite values for each $x \in (a, b]$. Let $x \in (a, b)$ and $z \in (x, x + \delta(x) \wedge b]$. Then for each $\tau = (\{a = x_0, \dots, x_m = x\}, \{y_1, \dots, y_m\}) \in TP(\delta_x, [a, x])$, $(\{x_0, \dots, x_m, z\}, \{y_1, \dots, y_m, x\}) \in TP(\delta_z, [a, z])$. Thus

$$S_{\delta_x}(\tau) + f(x)[g(z) - g(x)] \le M(z)$$

for any $\tau \in TP(\delta_x, [a, x])$. Thus the first inequality in (6.29) holds. Similarly one can show that the second one and (6.30) also hold. Therefore m and M are minor and major functions respectively. By Lemma 6.27 and (6.35), it then follows that

$$(WPS)\int_a^b f\,dg - \epsilon \le m(b) \le L(f, g) \le U(f, g) \le M(b) \le (WPS)\int_a^b f\,dg + \epsilon.$$

Since ϵ is arbitrary, $L(f, g) = U(f, g)$ and equals the (HK) integral. The proof of Theorem 6.30 is complete. $\square$

Schwabik [100, Theorem 3.1] proved the following for g of bounded variation, where f is not necessarily regulated, if (a) f is bounded, or (b) f is arbitrary and for any t such that $g(t-) = g(t+)$, we have $g(t) = g(t-)$ also.

Theorem 6.31. *Suppose f and g are regulated on $[a, b]$ and g is right–continuous on $(a, b]$. If $\ddagger_a^b f\,dg$ exists then $(HK)\int_a^b f\,dg$ exists and has the same value.*

Proof. Given $\epsilon > 0$, take a partition $\lambda = \{z_j \colon j = 0, \dots, m\}$ of $[a, b]$ such that for any refinement κ of λ and open intermediate partition σ of κ,

$$\Big|S_Y(f, g; \kappa, \sigma) - \ddagger_a^b f\,dg\Big| < \epsilon. \tag{6.36}$$

By Lemma 6.18 we can assume that (6.36) holds also for f replaced by $f_+^{(a,b)}$ or by $\Delta_{a,b}^+ f$, taking λ as a common refinement $\lambda = \lambda' \cup \lambda''$ of partitions λ' for $f_+^{(a,b)}$ and $\epsilon/2$, and λ'' for $\Delta_{a,b}^+ f$ and $\epsilon/2$. It suffices to prove the Theorem for $f_+^{(a,b)}$ and for $\Delta_{a,b}^+ f$.

Define a gauge function $\delta(\cdot)$ on $[a, b]$ by: if $x \notin \lambda$, let $\delta(x) := \min_j |x - z_j|/2$. Then any $\delta(\cdot)$–fine tagged partition must contain each z_j as a tag. For $j = 0, \dots, m$, define $\delta(z_j)$ such that: $\delta(z_j) \le \min_{i \ne j} |z_i - z_j|/3$ and

$$\|f\|_\infty \sum_{j=1}^m \Big\{\mathrm{Osc}(g; [z_j - \delta(z_j), z_j)) + \mathrm{Osc}(g; (z_{j-1}, z_{j-1} + \delta(z_{j-1})])\Big\} < \epsilon. \tag{6.37}$$

Let $\tau = (\{u_i \colon i = 0, \dots, n\}, \{v_i \colon i = 1, \dots, n\})$ be any $\delta(\cdot)$–fine tagged partition of $[a, b]$ and let

$$S_\delta(\tau) := \sum_{i=1}^n f(v_i)[g(u_i) - g(u_{i-1})] \tag{6.38}$$

be the gauge sum based on τ. We can assume that $v_i = u_i = v_{i+1}$ never occurs since if it did, the ith and $(i+1)$st term in S could be replaced by $f(v_i)[g(u_{i+1}) - g(u_{i-1})]$, and we would still have a gauge sum based on a $\delta(\cdot)$–fine tagged partition equal to $S_\delta(\tau)$. Since the tags $\{v_i: i = 1,\dots,n\}$ must contain the points of λ, for each index $j \in \{0,\dots,m\}$ there is an index $i(j) \in \{1,\dots,n\}$ such that $z_j = v_{i(j)}$. Let $\mu := \{i(j): j = 0,\dots,m\} \subset \{1,\dots,n\}$. For $j = 1,\dots,m-1$, by definition of $\delta(\cdot)$, we must have

$$u_{i(j)-1} < v_{i(j)} < u_{i(j)}. \tag{6.39}$$

We also have $i(0) = 1$, $v_1 = z_0 = u_0 = a$ and $i(m) = n$, $v_n = z_m = u_n = b$.

First suppose that f is right–continuous on (a,b); that is the case when $f = f_+^{(a,b)}$. We will show that one can find a Young sum based on a refinement of λ, which is arbitrarily close to the gauge sum S. To this aim we will replace the values $g(u_i)$ in S, $i = 1,\dots,n-1$, by values $g(x_i)$ at continuity points x_i of g, close to u_i. Then the Young sum terms $f(x_i)\Delta^{\pm}g(x_i)$ will be 0. For each $i = 1,\dots,n-1$ such that $u_i < v_{i+1}$, replacing u_i in S by a slightly larger $x_i > u_i$ which is a continuity point of g gives a sum T with $|T - S| < \epsilon$. If $v_{i+1} = u_i$, then let U be a sum equal to T except that for each such i, we replace v_{i+1} by y_{i+1} and u_i by x_i with $u_i < x_i < y_{i+1} < u_{i+1}$, where x_i is a continuity point of g and let $y_{i+1} - u_i$ be as small as desired. Since f is right–continuous on (a,b), we can make $|U - T| < \epsilon$. In either case $v_i \le u_i < x_i$. Let $x_0 := a$ and $x_n := b$. Let $y_i := v_i$ for each $i = 2,\dots,n-1$ for which y_i was not previously defined. Thus $y_{i(j)} = v_{i(j)}$ for $j = 1,\dots,m-1$ by (6.39). Choose any $y_1 \in (a,u_1)$ and $y_n \in (x_{n-1},b)$. Let $V := S_Y(f,g;\zeta,\rho)$ be the Young sum based on $\zeta := \{x_i: i = 0,\dots,n\}$ and $\rho := \{y_i: i = 1,\dots,n\}$. Then

$$\begin{aligned}
|V - U| &= |f(a)\Delta^+g(a) + f(y_1)[g(x_1) - g(a+)] - f(a)[g(x_1) - g(a)] \\
&\quad + f(b)\Delta^-g(b) + f(y_n)[g(b-) - g(x_{n-1})] - f(b)[g(b) - g(x_{n-1})]| \\
&= |[f(y_1) - f(a)][g(x_1) - g(a+)] + [f(y_n) - f(b)][g(b-) - g(x_{n-1})]| < 4\epsilon
\end{aligned}$$

by (6.37). Now let $\kappa := \lambda \cup \zeta = \{s_l: l = 0,\dots,k\}$, a partition of $[a,b]$ with $k = m+n-2$. Each interval (x_{i-1},x_i) for $i = 2,\dots,n-1$ contains z_j if $i = i(j) \in \mu$, otherwise contains no z_j and becomes an interval (s_{l-1},s_l) for some l. In the latter case let $t_l := y_i$. In the former case we have for some l, $s_{l-2} = x_{i-1} < z_j = s_{l-1} < x_i = s_l$. Choose any $t_{l-1} := w_j$ with $s_{l-2} < t_{l-1} < s_{l-1}$ and $t_l := r_j$ with $s_{l-1} < t_l < s_l$. Let $W := S_Y(f,g;\kappa,\sigma)$ be the Young sum based on κ and its open intermediate partition $\sigma := \{t_l: l = 1,\dots,k\}$. Then

$$\begin{aligned}
|W - V| &= \Big|\sum_{j=1}^{m-1}\Big\{f(w_j)[g(z_j-) - g(x_{i(j)-1})] + f(z_j)\Delta^-g(z_j) \\
&\quad + f(r_j)[g(x_{i(j)}-) - g(z_j)] - f(z_j)[g(x_{i(j)}-) - g(x_{i(j)-1})]\Big\}\Big| \\
&\le \sum_{j=1}^{m-1}|[f(w_j) - f(z_j)][g(z_j-) - g(x_{i(j)-1})]| \\
&\quad + \sum_{j=1}^{m-1}|[f(r_j) - f(z_j)][g(x_{i(j)}-) - g(z_j)]| < 2\epsilon
\end{aligned}$$

by (6.37). Here W is a Young sum over a refinement of λ, so $|W - {\not\!\!\int}_a^b f\,dg| < \epsilon$. Thus $|S - {\not\!\!\int}_a^b f\,dg| < 9\epsilon$. The conclusion follows when f is right–continuous on (a,b).

Now suppose that $f \in c_0((a,b))$ (cf. notation (6.27)); that is the case when $f = \Delta_{a,b}^+ f$. There is an open intermediate partition $\{w_j: j = 1,\dots,m\}$ of $\lambda = \{z_j: j = 0,\dots,m\}$ such that $f(w_j) = 0$ for all j. Thus by (6.36)

$$\Big|\sum_{j=1}^{m-1} f(z_j)\Delta^- g(z_j) - \underset{a}{\overset{b}{\rlap{\,=}\int}} f\,dg\Big| < \epsilon. \tag{6.40}$$

For any set $\nu \subset \{1, \ldots, n-1\}$, consider partitions $\kappa = \{s_l \colon l = 0, \ldots, k\}$, $k = m + n - 2$, consisting of λ, u_i for $i \in \nu$, and for each $i \in \{1, \ldots, n-1\} \setminus \nu$, a continuity point of g close to u_i. Let $\sigma = \{t_l \colon l = 1, \ldots, k\}$ be an open intermediate partition of κ with $f(t_l) = 0$ for all l. Then from (6.40) and (6.36) we obtain

$$\Big|\sum_{i\in\nu} f(u_i)\Delta^- g(u_i)\Big| < 2\epsilon. \tag{6.41}$$

Consider also partitions (κ, σ) defined in the same way except that for $\xi \subset \{1, \ldots, n-1\}$, if $i \in \xi \setminus \mu$ and i is even, we take $s_l = s_{l(i)}$ to be, instead of u_i, a continuity point of g a little larger than u_i, while $s_{l(i)-1}$ is a continuity point of g a little smaller than u_{i-1} and $t_{l(i)} = u_{i-1}$. For each i there will be an $s_{l(i)}$, a continuity point of g near u_i. Let $f(t_l) = 0$ for other l as before. Letting $s_{l(i)} \downarrow u_i$ and $s_{l(i)-1} \uparrow u_{i-1}$ it follows that

$$\Big|\sum \{f(u_{i-1})[g(u_i) - g(u_{i-1}-)] \colon i \in \xi \setminus \mu,\ i \text{ even}\}\Big| \le 2\epsilon.$$

The same holds likewise for i odd. Thus

$$\Big|\sum_{i\in\xi\setminus\mu} f(u_{i-1})[g(u_i) - g(u_{i-1}-)]\Big| \le 4\epsilon, \quad \text{or} \quad \Big|\sum_{i+1\in\xi\setminus\mu} f(u_i)[g(u_{i+1}) - g(u_i-)]\Big| \le 4\epsilon.$$

With $\nu = \{i \colon i+1 \in \xi \setminus \mu\}$ in (6.41) this gives

$$\Big|\sum_{i+1\in\xi\setminus\mu} f(u_i)[g(u_{i+1}) - g(u_i)]\Big| \le 6\epsilon.$$

If $v_{i+1} = u_i$ then $i+1 \notin \mu$ by (6.39). Thus

$$\Big|\sum \{f(u_i)[g(u_{i+1}) - g(u_i)] \colon u_i = v_{i+1},\ \ i = 1, \ldots, n-2\}\Big| \le 6\epsilon.$$

Here $f(u_i)$ could be replaced by $f(u_i) - f(y_{i+1})$, where $u_i < y_{i+1} < u_{i+1}$ and $f(y_{i+1}) = 0$. So the gauge sum (6.38) differs by at most 6ϵ from the sum

$$S' := \sum_{i=1}^{n} f(w_i)[g(u_i) - g(u_{i-1})],$$

where $w_i := y_i$ if $u_{i-1} = v_i$ and $i = 2, \ldots, n-1$, $w_i := v_i$ otherwise. Then $u_{i-1} < w_i \le u_i$ for $i = 2, \ldots, n-1$. The rest of the proof follows as in the case where f is right–continuous except that now there is no need to consider the sum U. The proof of Theorem 6.31 is complete. □

Ward [111] stated and Saks [98, Theorem VI.8.1] gave a proof of the fact that $(WPS) \int_a^b f\,dg$ is defined provided the corresponding Lebesgue–Stieltjes integral $(LS) \int_a^b f\,dg$ is defined, and then they are equal. By Theorem 6.30, the gauge integral is in the same relation with the (LS) integral. The following gives conditions when the converse holds.

Theorem 6.32. *Let f be non-negative on $[a,b]$, and let g be non-decreasing on $[a,b]$. Then $(HK)\int_a^b f\,dg$ exists if and only if $(LS)\int_a^b f\,dg$ does, and then the two are equal.*

Proof. Recall that the McShane integral is defined as the (HK) integral except that in a tagged partition $(\{x_i\colon i=0,\ldots,n\},\{y_i\colon i=1,\ldots,n\})$ the tags y_i need not be in the corresponding intervals $[x_{i-1},x_i]$. By Corollary 6.3.5 of Pfeffer [90, p. 113], under the present hypotheses $(HK)\int_a^b f\,dg$ exists if and only if the corresponding McShane integral exists, and then the two are equal. The equivalence between the McShane and Lebesgue–Stieltjes integral was proved by McShane [80, pp. 552,553]. Also, it follows from Theorem 4.4.7, Proposition 3.6.14 and Theorem 2.3.4 of Pfeffer [90]. The proof of the theorem is complete. $\square$

Suppose a continuous function $f\colon [a,b]\mapsto\mathbb{R}$ is differentiable everywhere on (a,b). It may be that f' is not Lebesgue integrable, e.g. if $[a,b]=[0,1]$, $f(x)=x^2\sin(\pi/x^2)$, $0<x<1$, $f(x)=0$ elsewhere. The Denjoy and Perron integrals were defined so that for every such f', $\int_a^x f'(t)\,dt=f(x)-f(a)$, $a\le x\le b$. These integrals and the later–invented Henstock–Kurzweil integral turned out to be equivalent for integrals $\int_a^b g(x)\,dx$: Gordon [45, Chapter 11]. The following holds:

Theorem 6.33. *Let f be a continuous function: $[a,b)\mapsto\mathbb{R}$ having a derivative $f'(x)$ for $a<x<b$ except for at most countably many x. Then $(HK)\int_a^x f'(t)\,dt$ exists and equals $f(x)-f(a)$ for $a\le x<b$.*

Proof. Apply Theorems 7.2 and 11.1 of Gordon [45]. $\square$

For another example, $f(x):=x^2\sin(e^{1/x})$, $x\neq 0$, $f(0):=0$, satisfies the conditions of Theorem 6.33 although f' has very wild oscillations.

While there has been relatively little literature about the Young (RYS) and (CY) integrals, there has been much more about Henstock–Kurzweil (gauge) integrals, e.g. Lee Peng–Yee [89]. A 1991 book by Henstock [48] has a reference list of more than 1200 papers and books, mainly on the theme of "non–absolute integration", if not necessarily about the (HK) integral itself. In fact, Henstock [48] treats integration over general spaces ("division spaces").

6.6 Comments and related results

Integration by parts In stochastic analysis, and for some integral equations as will be seen in Section 7.4, there are advantages in integrating left-continuous integrands with respect to right-continuous functions. In that case we have the following:

Theorem 6.34 (integration by parts). *Let f and g be right-continuous and of bounded variation on $[a,b]$. Then for Lebesgue-Stieltjes integrals,*

$$\int_a^b f_-^{(a)}\,dg+\int_a^b g_-^{(a)}\,df \;=\; fg|_a^b-\sum_{a<x\le b}(\Delta^- f\Delta^- g)(x). \tag{6.42}$$

Proof. By (2.4), $\int_a^b f_-^{(a)}dg=f_-^{(a)}g|_a^b-\int_a^b g\,df_-^{(a)}$. By formulas for the (CY) integral, $\int_a^b g\,df_-^{(a)}=\int_a^b g_-^{(a)}\,df_-^{(a)}+\sum_{(a,b)}\Delta^- g\Delta^- f$, and $\int_a^b g_-^{(a)}df_-^{(a)}=\int_a^b g_-^{(a)}\,df-(g_-\Delta^- f)(b)$. Collecting terms, the result follows. $\square$

The integration by parts formula (6.42) also holds for any of the integrals following from the (LS) integral in Diagram 2.1 (the (RYS), (HK), or (CY) integral). Note that the sum $\sum \Delta^- f \Delta^- g$ appears in (6.42) and in (2.4) with opposite signs. Naturally, Δ^+ terms as in (2.4) don't occur in (6.42) because of right continuity. In stochastic analysis, there are integration by parts formulas with the terms of (6.42) and/or further continuous terms (see e.g. pp. 227-235 and 325-330 in [14]).

W. H. Young's integrals Let h be a monotonic function on $[a, b]$. Suppose f is a step function on $[a, b]$, that is f is constant on each open interval (x_{i-1}, x_i) of a partition $\kappa = \{x_i\colon i = 0, \dots, n\} \in \mathrm{PP}\,([a, b])$. Then let

$$\int_a^b f\,dh := \sum_{i=1}^n \Big\{[f\Delta^+ h](x_{i-1}) + f(x_{i-1}+)[h(x_i-) - h(x_{i-1}+)] + [f\Delta^- h](x_i)\Big\}. \qquad (6.43)$$

Starting with such integrals for step functions f, W. H. Young (1914) defined integrals for more general functions f by his method of monotone sequences, later adopted by Daniell [13] in defining the "Daniell integral." Thus Young obtained an integral equivalent to the Lebesgue–Stieltjes integral, (see Hildebrandt [49, p. 193]), which Lebesgue had previously defined by a change of variables. If $h(x) = x$, then the right side of (6.43) is a Riemann sum for f because $f(x_{i-1}+) = f(y_i)$ for each $y_i \in (x_{i-1}, x_i)$, $i = 1, \dots, n$. Also, the right side of (6.43) is equal to the sum $S_Y(f, h; \kappa, \sigma)$ defined by (6.5) for any open intermediate partition σ of κ. Therefore it was natural to define an integral following the definition of the Riemann-Stieltjes integral except that Riemann-Stieltjes sums are replaced by Young sums. However it was known as stated by R. C. Young (1929, p. 221) that such an integral would coincide with the Riemann-Stieltjes integral for monotonic functions h. Namely the following is true:

Proposition 6.35. *Let f be a bounded function on $[a, b]$, let h be of bounded variation on $[a, b]$, and let h be continuous at each $x \in (a, b)$ such that $h(x-) = h(x+)$. Suppose that a number A exists with the property: for each $\epsilon > 0$ there exists $\delta > 0$ such that $|S_Y(f, h; \kappa, \sigma) - A| < \epsilon$ for each $\kappa \in \mathrm{PP}\,([a, b])$ with mesh $|\kappa| < \delta$ and for each open intermediate partition σ of κ. Then $(MRS)\int_a^b f\,dh$ exists and has value A.*

Remark. Any monotonic function h satisfies the above hypothesis. A different proof of the preceding statement is indicated by Hildebrandt ([52, II.19.3.12], cf. also [50, 8.11]). The following proof also extends to the case when $h \in \mathcal{W}_p$, $f \in \mathcal{W}_q$, $p > 1$, $q > 1$ and $1/p + 1/q = 1$ in place of the assumptions f bounded and $h \in \mathcal{W}_1$.

Proof. Let $\epsilon > 0$ and let $x \in (a, b)$. By hypothesis there exists $\delta > 0$ satisfying the stated property. Let $u < x < v$, $v - u < \delta$ and $y \in (u, v)$. Consider a partition $\kappa_{u,v}$ of $[a, b]$ containing u, v and no other points in between. Let σ_x and σ_y be two open intermediate partitions of κ with the same points except the points x and y. Subtracting two Young sums based on the same partition $\kappa_{u,v}$ and on two open intermediate partitions σ_x and σ_y, we get the bound

$$|[f(x) - f(y)][h(v-) - h(u+)]| < 2\epsilon.$$

Since $\epsilon > 0$ and y are arbitrary, it follows that f is continuous at x whenever $\Delta^{\pm} h(x) = h(x+) - h(x-) \neq 0$. Therefore at each $x \in (a, b)$ either f or h must be continuous. A similar argument applied for $x = a$ and $x = b$ yields that at each endpoint one of the two functions must be one-sidedly continuous. Due to this continuity implication, to show the Riemann-Stieltjes integrability of f with respect to h it is enough to consider Riemann-Stieltjes sums for open intermediate partitions. Let $\kappa = \{x_i\colon i = 0, \ldots, n\} \in \mathrm{PP}\,([a,b])$ and let $\sigma = \{y_i\colon i = 1, \ldots, n\}$ be an open intermediate partition of κ. Then we have

$$\begin{aligned} R(\kappa, \sigma) &:= S_{RS}(f, h; \kappa, \sigma) - S_Y(f, h; \kappa, \sigma) \\ &= \sum_{i=1}^{n} \{[f(y_i) - f(x_{i-1})]\Delta^+ h(x_{i-1}) + [f(y_i) - f(x_i)]\Delta^- h(x_i)\}. \end{aligned} \tag{6.44}$$

Again fix $\epsilon > 0$. Since h has bounded variation, there exists a finite set μ of jump points of h of cardinality m such that the sum of the absolute values of jumps at other points is less than $\epsilon/(2\|f\|_\infty)$. Then choose $\delta > 0$ such that the oscillation of f around each point of μ is less than $\epsilon/(2m\|h\|_\infty)$. By (6.44), it follows that $|R(\kappa, \sigma)| < 2\epsilon$ whenever the mesh $|\kappa| < \delta$. The statement of the proposition follows at once. $\square$

The preceding statement shows that replacing Riemann–Stieltjes sums by Young sums in the definition of the mesh–Riemann–Stieltjes integral essentially gives nothing new. In Section 4.3 we saw that this is not so if the same change is made in the definition of the (RRS) integral. It led to the new integral called the refinement–Young–Stieltjes integral, or just the (RYS) integral. Full-fledged uses of this integral appeared in the works of L. C. Young [118] and Gehring [39]. Hildebrandt discussed the (RYS) integral in his survey paper [50, p. 275] and in his book [52, pp. 88-96].

The early history of Stieltjes integrals Medvedev [81] gives a history of concepts of integration. He describes how Stieltjes (1894) defined an integral $\int_a^b f(u)\,dg(u)$ for f continuous and g nondecreasing. König (1897), in Hungarian, gave the full definition of mesh Riemann-Stieltjes integral, and showed that the classical integration by parts formula holds with endpoint corrections

$$\int_a^b f\,dg + \int_a^b g\,df - (fg)|_a^b = (\Delta^+ f \Delta^+ g)(a) - (\Delta^- f \Delta^- g)(b),$$

which actually must be 0 if the integrals are defined (in the mesh or the later refinement Riemann–Stieltjes sense). Kowalewski (1905) rediscovered the definition and now proved that

$$\int_a^b f\,dg + \int_a^b g\,df = (fg)|_a^b$$

provided that both integrals exist. According to Medvedev, F. Riesz in 1909 made known the existence of König's paper to the world outside of Hungary, but without characterizing it.

Chapter 7

Product integration

7.1 Ordinary differential equations

The product integral first arose as a way of representing the solutions of initial value problems for linear ordinary differential equations. Consider a linear kth order ordinary differential equation

$$\frac{d^k f(t)}{dt^k} + A_{k-1}(t)\frac{d^{k-1} f(t)}{dt^{k-1}} + \ldots + A_0(t) f(t) = B(t), \qquad a < t < b, \tag{7.1}$$

where for the present f and the coefficients A_j and B are real–valued. As is often done in differential equations one can write an equivalent first order linear vector and matrix differential equation

$$dY/dt = M \cdot Y, \tag{7.2}$$

where Y is the $(k+1) \times 1$ column vector $Y = (y_0, y_1, \ldots, y_k)'$ with $y_0 \equiv 1$, $y_1 \equiv f(t)$, $y_j = d^{j-1} f(t)/dt^{j-1}$ for $j = 2, \ldots, k$ and

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & & \ldots & 0 \\ 0 & 0 & 1 & 0 & & \ldots & 0 \\ 0 & 0 & 0 & 1 & & \ldots & 0 \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & . & . & . & & \ldots\ 0 & 1 \\ B & -A_0 & . & . & & \ldots & -A_{k-1} \end{pmatrix}$$

with $B := B(t)$, $A_j := A_j(t)$, $j = 0, 1, \ldots, k-1$. If (7.2) holds at a point t, then

$$Y(t+h) - Y(t) = hM(t) \cdot Y(y) + o(h) \qquad \text{as } h \downarrow 0, \qquad \text{or}$$

$$Y(t+h) = (I + hM(t)) \cdot Y(t) + o(h) \qquad \text{as } h \downarrow 0, \tag{7.3}$$

where I is the $(k+1) \times (k+1)$ identity matrix. Let $N(t) := \int_a^t M(u)\, du$. If (7.2) holds for $a < t < b$ and Y is continuous on $[a, b]$ then informally, one sees by iterating (7.3) and letting $h \downarrow 0$ that for the product integral as defined in Section 3.5,

$$Y(u) = \Big(\prod_t^u (I + dN)\Big)\, Y(t) \qquad \text{for } a \le t \le u \le b. \tag{7.4}$$

More formally, making a stronger regularity assumptions than necessary for simplicity, if M is continuous, then (7.2) implies the integral equation v

$$Y(u) = Y(t) + \underset{t}{\overset{u}{\not\!\!\int}} dN(x) \cdot Y(x), \qquad a \le t \le u \le b, \tag{7.5}$$

and (7.4) will indeed give the unique solution of (7.5) and thus of (7.1) by Theorem 7.13 below, or by Theorem 5.21 in [29], as follows. These facts give that $H(u) := \prod_t^u (\mathbb{I} + dN)$ is the unique matrix–valued solution of the matrix integral equation

$$H(u) = I + \int_t^u dN(x) \cdot H(x), \qquad t \le u \le b. \tag{7.6}$$

Suppose that $Z(\cdot)$ were another vector–valued solution of

$$Z(u) = Y(t) + \int_t^u dN(x) \cdot Z(x).$$

Let $A(u) = (A_{ij}(u))$ be defined by

$$A_{ij}(u) := \begin{cases} (Z_j - Y_j)(u) & \text{for } i = 1 \\ 0 & \text{for } i > 1. \end{cases}$$

Then $H + A$ would give a solution of (7.6) different from H, a contradiction.

The above considerations also apply to systems of first–order linear ordinary differential equations, where M can be a general (continuous) square matrix–valued function instead of having the specific form coming from one kth order equation.

Equation (7.4) gives the solution of an initial–value problem for linear ordinary differential equations in terms of an operator which is clearly linear in the initial values $Y(t)$. It is non–linear in the coefficients. For a kth order linear ordinary differential equation (7.1) the solution will be $y_1(t)$. The operator will be shown to be holomorphic in the coefficients in Section 7.5.

It's easy to check that for two 3×3 matrices A, B of the form of M, so that $k = 2$, we have $AB = BA$ if and only if $A = B$. The same is true for any $k \ge 2$: consider the next–to–last row of the products. Thus commuting matrices will be obtained only for differential equations with constant coefficients if $k \ge 2$.

7.2 Product integrals

We were introduced to probabilistic and statistical uses of the product integral by Gill and Johansen [43], who use an interval function formulation. In our work [29] we used point functions. Now we have come to prefer interval functions, but perhaps there is some advantage in seeing and comparing both formulations.

The book of Dollard and Friedman [17] contains a survey and examples of applications of several types of product integrals. In the case when a function h is real valued, the product integral with respect to h usually can be written explicitly as in Theorem 7.8 below. Therefore we consider product integrals for functions with values in a Banach algebra $\mathbb{B}$ with an identity $\mathbb{I}$. Several results stated in earlier sections for real–valued functions will be used for $\mathbb{B}$–valued functions. Their proofs in the more general case are the same.

In Section 7.5 we apply some ingenious inequalities of T. Lyons [76] to bounding Peano series expansions of product integrals. It has not escaped our attention that these inequalities can have other applications. We present what we could write up in the time available so far.

Let μ be an additive interval function on an interval $[a,b]$ with values in $\mathbb{B}$. As before the class of all such functions is denoted by $\mathcal{I}([a,b];\mathbb{B})$. Also, as in Section 4.5 we use two classes of subintervals of $[a,b]$: $\mathfrak{I}_{os}([a,b]) := \{(c,d),\{c\}\colon a \le c \le d \le b\}$ and $\mathfrak{I}([a,b]) := \{[c,d],(c,d],[c,d),(c,d)\colon a \le c \le d \le b\}$. Let $\mathrm{IP}_{os}([a,b])$ be the set of all partitions of $[a,b]$ by intervals from $\mathfrak{I}_{os}([a,b])$. Recall that the set $\mathrm{IP}([a,b])$ of interval partitions was introduced in Section 2.1. For $J \in \mathfrak{I}([a,b])$ and $\mathfrak{D} = (J_1,\dots,J_N) \in \mathrm{IP}(J)$, let

$$P(\mu;\mathfrak{D}) := P_{\mathfrak{I}(J)}(\mu;\mathfrak{D}) := \prod_{j=1}^{N}[\mathbb{I}+\mu(J_j)] := [\mathbb{I}+\mu(J_N)]\cdot\ldots\cdot[\mathbb{I}+\mu(J_1)].$$

Similarly define $P_{\mathfrak{I}_{os}(J)}(\mu;\mathfrak{D})$. Since $\mathbb{B}$ may not be commutative, the product sign is used below with the prescribed order. Some authors use the reverse order.

Definition 7.1. Let $\mu \in \mathcal{I}([a,b];\mathbb{B})$ be additive and let $J \in \mathfrak{I}([a.b])$. The *product integral with respect to μ over J* is defined to be the limit

$$\prod_J(1+d\mu) := \lim_{\mathfrak{D}\uparrow} P_{\mathfrak{I}(J)}(\mu;\mathfrak{D}) \tag{7.7}$$

if it exists under refinements of $\mathfrak{I}$–partitions of J (cf. Definition 6.23).

Lemma 7.2. *Let $\mu \in \mathcal{I}([a,b];\mathbb{B})$ be additive and upper continuous. Then the limit in (7.7) exists if and only if it exists under refinements of $\mathfrak{I}_{os}$–partitions of J, and the two are equal.*

Proof. Since $\mathrm{IP}_{os}(J) \subset \mathrm{IP}(J)$ and each $\mathfrak{I}$–partition can be $\mathfrak{I}_{os}$–refined, if (7.7) exists then so does the limit under refinements of $\mathfrak{I}_{os}$-partitions of J, and the two are equal. The proof of the converse implication is similar to the proof of Theorem 6.26 and is omitted. □

As in Dudley and Norvaiša [29, Definition 4.20] given a $\mathbb{B}$-valued function h on $[a,b]$, the product integral

$$\prod_a^b(\mathbb{I}+dh) \tag{7.8}$$

is defined to exist and equal A iff for every $\epsilon > 0$ there is a $\lambda \in \mathrm{PP}([a,b])$ such that

$$\|P(h;\kappa) - A\| < \epsilon \tag{7.9}$$

for each refinement κ of λ, where $P(h;\kappa) := \prod_{i=1}^{n}[1+h(x_i)-h(x_{i-1})]$ for $\kappa = \{x_i\colon i = 0,\dots,n\} \in \mathrm{PP}([a,b])$.

To relate the product integral (7.7) with (7.8) recall the definition (4.32) of the interval function μ_h corresponding to a regulated function h. Then we have:

Theorem 7.3. *Let $h \in \mathcal{R}([a,b];\mathbb{B})$ be either right-continuous or left-continuous at each point of (a,b) and let $h(a)=0$.*

(1) *The existence of any of the three product integrals*:

$$\prod_a^b(\mathbb{I}+dh),\qquad \prod_{a+}^{b}(\mathbb{I}+dh),\qquad \prod_a^{b-}(\mathbb{I}+dh) \tag{7.10}$$

implies the existence of the corresponding product integral with the same value:

$$\prod_{[a,b]}(\mathbb{I}+d\mu_h),\qquad \prod_{(a,b]}(\mathbb{I}+d\mu_h),\qquad \prod_{[a,b)}(\mathbb{I}+d\mu_h). \tag{7.11}$$

(2) *If, in addition, there is a finite constant $C(h)$ such that*

$$\sup_{\mathfrak{D}\in \mathrm{IP}\,(J)} \|P(\mu_h;\mathfrak{D})\| \le C(h) \tag{7.12}$$

for each subinterval $J\subset [a,b]$, then the converse to (1) *holds.*

Remark. By Lemma 4.19 of [29], condition (7.12) is satisfied provided $v_p(h;[a,b]) < \infty$ for some $0<p<2$.

Proof. We prove the stated relation only for the product integrals $\prod_a^b(\mathbb{I}+dh)$ and $\prod_{[a,b]}(\mathbb{I}+d\mu_h)$ because a proof for the other two pairs in (7.10) and (7.11) is similar. To prove (1) let $\epsilon>0$. By definition of the product integral with respect to h, there exists $\lambda\in \mathrm{PP}\,([a,b])$ such that (7.9) holds for each refinement κ of λ. Let $\mathfrak{D}_\lambda$ be the $\mathfrak{I}_{os}$-partition of $[a,b]$ such that the corresponding point partition $\wp(\mathfrak{D}_\lambda)=\lambda$ and let $\mathfrak{D}$ be any $\mathfrak{I}_{os}$-refinement of $\mathfrak{D}_\lambda$. Let $\kappa:=\{x_i\colon i=0,\dots,n\}=\wp(\mathfrak{D})$ be the corresponding point partition of $[a,b]$. Let $I_r:=\{i=1,\dots,n-1\colon \Delta^+h(x_i)=0\}$ and $I_l:=\{i=1,\dots,n-1\}\setminus I_r$. Then $\Delta^-h(x_i)=0$ for each $i\in I_l$ by assumption. Choose points x_i', $i=0,\dots,n$ recursively as follows: let $x_0'\in(x_0,x_1)$ and supposing x_{i-1}' to be chosen, then, if $i<n$, take either $x_i'\in(x_{i-1}\vee x_{i-1}',x_i)$ if $i\in I_r$, or $x_i'\in(x_i,x_{i+1})$ if $i\in I_l$, while if $i=n$ take $x_n'\in(x_{n-1}\vee x_{n-1}',x_n)$. Let $D_0':=[\mathbb{I}+h(x_1'\wedge x_1)-h(x_0')][\mathbb{I}+h(x_0')]$, for each $i=1,\dots,n-1$, let

$$D_i' := \begin{cases} [\mathbb{I}+h(x_{i+1}'\wedge x_{i+1})-h(x_i)][\mathbb{I}+h(x_i)-h(x_i')] & \text{if } i\in I_r\\ [\mathbb{I}+h(x_{i+1}'\wedge x_{i+1})-h(x_i')][\mathbb{I}+h(x_i')-h(x_i)] & \text{if } i\in I_l\end{cases}$$

and let $D_n':=\mathbb{I}+h(x_n)-h(x_n')$. Letting $x_i'\downarrow x_i$ if $i\in\{0\}\cup I_l$ and $x_i'\uparrow x_i$ if $i\in I_r\cup\{n\}$, we get that $D_0'\to[\mathbb{I}+h(x_1-)-h(x_0+)][\mathbb{I}+h(x_0+)]=[\mathbb{I}+\mu_h((x_0,x_1))][\mathbb{I}+\mu_h(\{x_0\})]$,

$$D_i'\to\left\{\begin{matrix}[\mathbb{I}+h(x_{i+1}-)-h(x_i)][\mathbb{I}+\Delta^-h(x_i)] & \text{if } i\in I_r\\ [\mathbb{I}+h(x_{i+1}-)-h(x_i+)][\mathbb{I}+\Delta^+h(x_i)] & \text{if } i\in I_l\end{matrix}\right\} = [\mathbb{I}+\mu_h((x_i,x_{i+1}))][\mathbb{I}+\mu_h(\{x_i\})]$$

for each $i=1,\dots,n-1$, and $D_n'\to\mathbb{I}+\mu_h(\{x_n\})$. Therefore

$$P(h;\kappa')=D_n'\cdots D_1'D_0'\to P_{\mathfrak{I}_{os}}(\mu_h;\mathfrak{D})\qquad\text{as } \kappa'\to\kappa.$$

Since $\lambda\subset\kappa\subset\kappa'$, by (7.9), it follows that $\|\prod_a^b(\mathbb{I}+dh)-P_{\mathfrak{I}_{os}}(\mu_h;\mathfrak{D})\|\le\epsilon$ for any $\mathfrak{I}_{os}$-refinement $\mathfrak{D}$ of $\mathfrak{D}_\lambda$. Statement (1) now follows from Lemma 7.2.

To prove (2), take $C(h)\ge1$ and again let $\epsilon>0$. Since each $\mathfrak{I}$-partition can be $\mathfrak{I}_{os}$-refined, there is an $\mathfrak{I}_{os}$-partition $\mathfrak{D}_\epsilon$ of $[a,b]$ such that

$$\Big\|\prod_{[a,b]}(\mathbb{I}+d\mu_h)-P_{\mathfrak{I}_2}(\mu_h;\mathfrak{D})\Big\|<\epsilon \tag{7.13}$$

for each $\mathfrak{I}$-partition $\mathfrak{D}\sqsupseteq\mathfrak{D}_\epsilon$. Let $\{z_j\colon j=0,\dots,m\}:=\wp(\mathfrak{D}_\epsilon)$ be the corresponding point partition, and let $\{v_{j-1},u_j\colon j=1,\dots,m\}$ be a set of points such that $z_{j-1}<v_{j-1}<u_j<z_j$ for $j=1,\dots,m$ and

$$\max_{1\le j\le m}[\mathrm{Osc}(h;(z_{j-1},v_{j-1}])\vee\mathrm{Osc}(h;[u_j,z_j))]<\epsilon/[2\|h\|_\infty C(h)^2(m+1)]. \tag{7.14}$$

Let $\lambda:=\{z_j,v_{j-1},u_j\colon j=0,\dots,m\}$ and let $\kappa=\{x_i\colon i=0,\dots,n\}\in\mathrm{PP}\,([a,b])$ be a refinement of λ. Define the sets I_r, I_l as in the above proof of statement (1), and for each $j=0,\dots,m$, let $i(j)\in\{0,1,\dots,n\}$ be such that $x_{i(j)}=z_j$. Let $\mathfrak{D}_\kappa$ be the $\mathfrak{I}$-partition of $[a,b]$ consisting of the intervals: $\{z_0\},J_1,\dots,J_{i(j)},\{z_j\},J_{i(j)+1},\dots,J_n,\{z_m\}$, where, for $j=0,\dots,m-1$,

$$J_{i(j+1)} := \begin{cases} (x_{i(j+1)-1}, z_{j+1}) & \text{if } i(j+1) \in I_r \\ [x_{i(j+1)-1}, z_{j+1}) & \text{if } i(j+1) \in I_l, \end{cases} \quad \text{and for } i(j)+1 < i < i(j+1)$$

$$J_{i(j)+1} := \begin{cases} (z_j, x_{i(j)+1}] & \text{if } i(j)+1 \in I_r \\ (z_j, x_{i(j)+1}) & \text{if } i(j)+1 \in I_l \end{cases} \qquad J_i := \begin{cases} [x_{i-1}, x_i] & \text{if } i-1 \in I_l,\, i \in I_r \\ (x_{i-1}, x_i] & \text{if } i-1, i \in I_r \\ [x_{i-1}, x_i) & \text{if } i-1, i \in I_l \\ (x_{i-1}, x_i) & \text{if } i-1 \in I_r,\, i \in I_l. \end{cases}$$

Then $\mathfrak{D}_\kappa \in \mathrm{IP}_2([a,b])$ and $\mathfrak{D}_\kappa \sqsupseteq \mathfrak{D}_\epsilon$. Due to the assumption of left- or right-continuity we have $\mu_h(J_{i(j+1)}) = h(z_{j+1}-) - h(x_{i(j+1)-1})$, $\mu_h(J_{i(j)+1}) = h(x_{i(j)+1}) - h(z_j+)$ and $\mu_h(J_i) = h(x_i) - h(x_{i-1})$. Let $B_n := [\mathbb{I} + \mu_h(\{z_m\})][\mathbb{I} + \mu_h(J_n)]$,

$$B_i := \begin{cases} [\mathbb{I} + \mu_h(\{z_j\})][\mathbb{I} + \mu_h(J_{i(j)})] & \text{if } i(j) \in I_r, \\ [\mathbb{I} + \mu_h(J_{i(j)+1})][\mathbb{I} + \mu_h(\{z_j\})] & \text{if } i(j) \in I_l, \\ \mathbb{I} + \mu_h(J_i) & \text{if } i \in \{1, \dots, n-1\} \setminus (I_l \cup I_r) \end{cases}$$

and $B_1 := (\mathbb{I} + \mu_h(J_1))(\mathbb{I} + \mu_h(\{z_0\}))$. Using a telescoping sum representation, we get

$$T := P(h;\kappa) - P_{\mathfrak{J}}(\mu_h; \mathfrak{D}_\kappa) = \prod_{i=1}^n (\mathbb{I} + \Delta^i h) - \prod_{i=1}^n B_i = [\mathbb{I} + \Delta^n h - B_n] \prod_{i=1}^{n-1} B_i$$

$$+ \sum_{j=1}^{m-1} \Big(\prod_{i=i(j)+1}^{n} (\mathbb{I} + \Delta^i h) \Big) [\mathbb{I} + \Delta^{i(j)} h - B_{i(j)}] \Big(\prod_{i=1}^{i(j)-1} B_i \Big) + \prod_{i=2}^n (\mathbb{I} + \Delta^i h)[\mathbb{I} + h(x_1) - B_1].$$

Then using (7.12) and (7.14), we get the bound

$$\|T\| \le C(h) \|\Delta^- h(b)[h(b-) - h(x_{n-1})]\| + \sum_{j=1}^{m-1} C(h)^2 H_j + C(h) \|h(a+)[h(x_1) - h(a+)]\| < \epsilon,$$

where

$$H_j := \begin{cases} \|\Delta^- h(z_j)[h(z_j) - h(x_{i(j)-1})]\| & \text{if } i(j) \in I_r \\ \|[h(x_{i(j)+1}) - h(z_j+)]\Delta^+ h(z_j)\| & \text{if } i(j) \in I_l. \end{cases}$$

By (7.13), it then follows that $\| \prod_{[a,b]} (\mathbb{I} + d\mu_h) - P(h;\kappa)\| < 2\epsilon$ for any point partition $\kappa \supset \lambda$. Since $\epsilon > 0$ is arbitrary, (2) holds. The proof of Theorem 7.3 is complete. $\square$

As the following example shows, the one–sided continuity assumptions of the preceding theorem cannot be dropped.

Example 7.4. For $\xi \in [0,1]$, let $h_\xi(x) := 0$ for $x \in [0,1)$, $h_\xi(1) := \xi$ and $h_\xi(x) := 1$ for $x \in (1,2]$. Then the product integrals with respect to h_ξ and μ_{h_ξ} exist, and

$$\prod_0^2 (1 + dh_\xi) = (1+\xi)(2-\xi) \quad \text{and} \quad \prod_{[0,2]} (1 + d\mu_{h_\xi}) = 2.$$

Therefore the two product integrals have the same value if and only if either $\xi = 0$ or $\xi = 1$; that is if and only if h_ξ is either left-continuous or right-continuous.

Recall that $\mathcal{I}_p([a,b], \mathbb{B})$ is the set of all $\mu \in \mathcal{I}([a,b]; \mathbb{B})$ with bounded p-variation as defined by (4.38). Next are the main facts about the product integral to be used in the subsequent sections.

Corollary 7.5. *Let $\mu \in \mathcal{I}_p([a,b];\mathbb{B})$, $1 \le p < 2$, be additive and upper continuous. Then the product integral with respect to μ is defined over any subinterval of $[a,b]$ and the mapping*

$$\mathfrak{I}([a,b]) \ni J \mapsto \prod_J (\mathbb{I} + d\mu) \in \mathbb{B} \tag{7.15}$$

is a multiplicative and upper continuous interval function on $[a,b]$ with bounded p-variation.

Proof. The functions $R_{\mu,a}$ and $L_{\mu,a}$ defined by (4.29), are respectively right–continuous and left–continuous on (a,b), and both are 0 at a. By Proposition 4.37 and Theorem 4.39, $\mu = \mu_h$ with $h = R_{\mu,a}$, or $h = L_{\mu,a}$, and h has bounded p-variation. Therefore the existence of the product integral with respect to μ over any subinterval of $[a,b]$ follows from Theorem 7.3 above and Theorem 4.23 of [29]. To see that (7.15) is multiplicative let $a < c < b$. Then by Theorem 7.3 above, Corollary 4.24 and Lemma 5.1 of [29], we have

$$\begin{aligned} \prod_{[a,b]} (\mathbb{I} + d\mu) &= \prod_a^c (\mathbb{I} + dR_{\mu,a}) \prod_{c+}^b (\mathbb{I} + dR_{\mu,a}) = \prod_{[a,c]} (\mathbb{I} + d\mu) \prod_{(c,b]} (\mathbb{I} + d\mu) \\ &= \prod_a^{c-} (\mathbb{I} + dL_{\mu,a}) \prod_c^b (\mathbb{I} + dL_{\mu,a}) = \prod_{[a,c)} (\mathbb{I} + d\mu) \prod_{[c,b]} (\mathbb{I} + d\mu). \end{aligned}$$

Similarly using Theorem 7.3 above and Lemma 5.2 of [29] one can show that (7.15) is upper continuous at $\emptyset$. This yields upper continuity by Theorem 4.35. The interval function (7.15) has bounded p-variation by an analogous result for the indefinite product integral with respect to point functions proved in Proposition 5.3 of [29] and by Theorem 4.39 and Lemma 4.13. Thus the conclusion of Corollary 7.5 holds. $\square$

The condition on p-variation in the preceding statement is sufficient but not necessary. In the case of real–valued functions we have necessary and sufficient conditions for existence of a non–zero product integral. The following statement is Theorem 4.4 of Dudley and Norvaiša [29].

Theorem 7.6. *For a real–valued function h on $[a,b]$, the product integral* (7.8) *exists and is non zero if and only if the following two conditions hold:*

(1) $h \in \mathcal{W}_2^*[a,b]$;
(2) $\Delta^- h(x) \neq -1 \neq \Delta^+ h(y)$ *for* $a < x \le b$, $a \le y < b$.

Let g be a $k \times k$ matrix-valued function defined on an interval $[a,b]$ and let $\kappa = \{x_i : i = 0, \dots, n\}$ be a partition of $[a,b]$. Volterra [110, Opere, I, p. 235] defined the product integral of the function g over $[a,b]$ to be the limit

$$\lim_{\kappa} \prod_{i=1}^{n} [\mathbb{I} + g(y_i)(x_i - x_{i-1})] \tag{7.16}$$

if it exists, where $\mathbb{I}$ is the identity matrix, y_i are any points from $[x_{i-1}, x_i]$ for $i = 1, \dots, n$ and the limit is taken as the mesh $|\kappa|$ of the partition κ tends to zero. He proved the existence of the limit under conditions analogous to those under which Riemann [94] proved the existence of his integral. Masani [79, Sec. V] proved among other things that for g with values in a Banach algebra the limit (7.16) exists if and only if the function g is Riemann integrable. Next we prove a similar result for the (RYS) integral. Let f be a real-valued function on $[a,b]$

and let $h \in \mathcal{R}[a,b]$. For $\kappa = \{x_i: i = 0,\dots,n\} \in \mathrm{PP}\,([a,b])$ and open intermediate partition $\sigma = \{y_i: i = 1,\dots,n\}$ of κ, let

$$P_Y(f,h;\kappa,\sigma) := \prod_{i=1}^{n} \{1 + (f\Delta^+h)(x_{i-1})\}\{1 + f(y_i)[h(x_i-) - h(x_{i-1}+)]\}\{1 + (f\Delta^-h)(x_i)\}.$$

The subscript Y for P_Y refers to the similarity between this product and the corresponding Young sum.

Definition 7.7. Let f be a function on $[a,b]$ and let $h \in \mathcal{R}[a,b]$. The *product integral* of f with respect to h over $[a,b]$ exists and has value A, if for each $\epsilon > 0$ there exists a partition $\lambda \in \mathrm{PP}\,([a,b])$ such that

$$|P_Y(f,h;\kappa,\sigma) - A| < \epsilon \tag{7.17}$$

for each refinement κ of λ and open intermediate partition σ of κ. If the product integral is defined then we write

$$\prod_a^b(1 + f\,dh) := A. \tag{7.18}$$

A product $\prod_i A_i$ will be said to *converge absolutely* if $\sum_i |A_i - 1| < \infty$ and $A_i \neq -1$ for all i. Let

$$\chi[z] := (1+z)e^{-z}, \qquad z \in \mathbb{R}. \tag{7.19}$$

For $J \subset [a,b]$ and $\delta > 0$, let $I^\sharp(\delta;J) := \{x \in J: |\Delta^\sharp_{a,b}(f)| > \delta\}$, where $\sharp = -$ and $\sharp = +$. Let

$$\prod_J \chi[f\Delta^\pm h] := \lim_{\delta\downarrow 0} \prod_{x\in I^-(\delta;J)} \chi[f\Delta^-_{a,b}h](x) \prod_{x\in I^+(\delta;J)} \chi[f\Delta^+_{a,b}h](x) \tag{7.20}$$

provided the two products converge absolutely.

Theorem 7.8. *Let f be a bounded function on $[a,b]$, $h \in \mathcal{W}_2^*[a,b]$, both real–valued, and*

$$f(y)\Delta^-h(y) \neq -1 \neq f(x)\Delta^+h(x) \tag{7.21}$$

for $a \le x < y \le b$. Then the integral $(RYS)\int_a^b f\,dh$ exists if and only if the product integral (7.18) *does, and if so they satisfy the relation*

$$\prod_a^b(1 + f\,dh) = \exp\Big\{(RYS)\int_a^b f\,dh\Big\} \prod_{[a,b]} \chi(f\Delta^\pm h). \tag{7.22}$$

Remarks. The product over jumps is as in the stochastic case, e.g. when $f \equiv 1$ ([15]). Condition (7.21) does not restrict small jumps of h, that is, (7.21) holds if $|\Delta^-h(y)| \vee |\Delta^+h(x)| < 1/\|f\|_\infty$ for $a \le x < y \le b$.

Proof. A Taylor series expansion with remainder for the function (7.19) yields

$$\chi(u) = 1 - \theta(u)u^2, \quad \text{where } 1/(4\sqrt{e}) \le \theta(u) \le 3\sqrt{e}/4 \text{ for } |u| \le 1/2.$$

Thus for any finite set $\nu \subset (a,b)$ of points x such that $|[f\Delta^-h](x)| \vee |[f\Delta^+h](x)| \le 1/2$, we have the bound

$$\sum_\nu |1 - \chi(f\Delta^\pm h)| \le (3\sqrt{e}/4)\|f\|_\infty^2 \mathfrak{S}_2(h)^2 < \infty.$$

That is, the product $\prod_{[a,b]} \chi(f\Delta^{\pm}h)$ in (7.22) converges absolutely.

By another Taylor series expansion with remainder, it follows that

$$\log(1+u) = u - \theta(u)u^2, \qquad \text{where } 2/9 \le \theta(u) \le 2 \text{ for } |u| \le 1/2. \tag{7.23}$$

Let $\epsilon > 0$. There exists $\lambda = \{z_j\colon j = 0,\dots,m\} \in \mathrm{PP}([a,b])$ satisfying the following four properties:

(1) $D := I^-(1/2;[a,b]) \cup I^+(1/2;[a,b]) \subset \lambda$;

(2) $\mathrm{Osc}(h;(z_{j-1},z_j)) < 1/(2\|f\|_\infty)$ for $j = 1,\dots,m$;

(3) for each finite set $\nu \subset (a,b)$ disjoint from λ,

$$\sum_{\nu} [(\Delta^- h)^2 + (\Delta^+ h)^2] < \epsilon/(4\|f\|_\infty);$$

(4) for each refinement $\kappa = \{x_i\colon i = 0,\dots,n\}$ of λ,

$$\sum_{i=1}^{n} [h(x_i-) - h(x_{i-1}+)]^2 < \epsilon/(4\|f\|_\infty).$$

The existence of λ follows from Corollary 4.11, Theorem 4.10, $\mathfrak{S}_2(f) < \infty$ and Lemma 4.18, applied in that order. Let $\kappa = \{x_i\colon i = 0,\dots,n\}$ be a refinement of λ and let $\sigma = \{y_i\colon i = 1,\dots,n\}$ be an open intermediate partition of κ. Then let

$$\Delta_i^\sharp := \Delta_i^\sharp(\kappa) := [f\Delta^\sharp h](x_i) \quad \text{and} \quad \Delta_i := \Delta_i(\kappa,\sigma) := f(y_i)[h(x_i-) - h(x_{i-1}+)],$$

where $\sharp = -$ and $\sharp = +$. Also, for each index $j = 0,\dots,m$ of λ, let $i(j) \in \{0,\dots,n\}$ be the index of κ such that $x_{i(j)} = z_j$ and let $I := \{0,\dots,n\} \setminus \{i(j)\colon j = 0,\dots,m\}$. By the above conditions (1) and (2), we have

$$\begin{aligned}
\log\{P_Y(f,h;\kappa,\sigma)\} &= \log\Big\{\prod_{j=1}^{m}(1+\Delta_{i(j-1)}^+)(1+\Delta_{i(j)}^-)\Big\} + \sum_{i\in I}[\log(1+\Delta_i^-) + \log(1+\Delta_i^+)] \\
&\quad + \sum_{i=1}^{n}\log(1+\Delta_i) \qquad \text{by (7.23) and (7.20)} \\
&= \log\Big\{\prod_{\lambda}\chi[f\Delta^{\pm}h]\Big\} + S_Y(f,h;\kappa,\sigma) - R(\kappa,\sigma),
\end{aligned} \tag{7.24}$$

where

$$R(\kappa,\sigma) := \sum_{i=1}^{n}\theta(\Delta_i)(\Delta_i)^2 + \sum_{i\in I}[\theta(\Delta_i^-)(\Delta_i^-)^2 + \theta(\Delta_i^+)(\Delta_i^+)^2].$$

By the above conditions (3) and (4), it then follows that

$$R(\kappa,\sigma) \le 2\|f\|_\infty \sum_{i=1}^{n}[h(x_i-) - h(x_{i-1}+)]^2 + 2\|f\|_\infty \sum_{i\in I}[(\Delta^+ h(x_i))^2 + (\Delta^- h(x_i))^2] < \epsilon$$

for any refinement κ of λ and open intermediate partition σ of κ. Since $\log\{\prod_\lambda \chi[f\Delta^{\pm}h]\}$ in (7.24) does not depend on κ and σ, the (RYS) integral exists if and only if the corresponding product integral (7.18) does, and relation (7.7) follows. The proof of the theorem is complete. □

7.3 The Duhamel formula

We prove the Duhamel formula, which is an integral representation of a difference between two product integrals. Its discrete analog is a simple algebraic identity:

$$\prod_{i=1}^{n} a_j - \prod_{j=1}^{n} b_j = \sum_{i=1}^{n} \Big(\prod_{j=i+1}^{n+1} a_j \Big)(a_i - b_i) \Big(\prod_{j=0}^{i-1} b_j \Big) \tag{7.25}$$

valid by a telescoping sum for any elements $a_1, \dots, a_n$ and $b_1, \dots, b_n$ of a Banach algebra, with $b_0 := a_{n+1} := \mathbb{I}$. An integral $\ne\!\!\!\!\int_{[a,b]} f(y)\,\rho(dy)\,g(y)$ is defined as a limit under refinement of sums $\sum_{j=1}^{n} f(y_j)\rho(I_j)g(y_j)$ over interval partitions $\{I_j\}_{j=1}^{n}$ of $[a,b]$ and $y_j \in I_j$, $j = 1, \dots, n$. This reduces to the previous definition if $\mathbb{B}$ is commutative.

Theorem 7.9 (Duhamel formula). *Let $\mu, \nu \in \mathcal{I}_p([a,b], \mathbb{B})$, $1 \le p < 2$, be additive and upper continuous. Then*

$$\prod_{[a,b]} (\mathbb{I} + d\mu) - \prod_{[a,b]} (\mathbb{I} + d\nu) = \ne\!\!\!\!\int_{[a,b]} \prod_{(y,b]} (\mathbb{I} + d\mu)(\mu - \nu)(dy) \prod_{[a,y)} (\mathbb{I} + d\nu), \tag{7.26}$$

where a product integral over the empty set is defined as $\mathbb{I}$.

First proof. Let $R := R_{\mu,a}$ and $L := L_{\nu,a}$ be the functions defined by (4.29). By Theorem 4.39, R and L have bounded p-variation. Thus by Theorem 7.3 and Corollary 7.5, the following product integrals are defined and satisfy the relations:

$$\prod_{[a,b]} (\mathbb{I} + d\mu) - \prod_{[a,b]} (\mathbb{I} + d\nu) = \prod_{a}^{b} (\mathbb{I} + dR) - \prod_{a}^{b} (\mathbb{I} + dL), \qquad \prod_{[a,y)} (\mathbb{I} + d\nu) = \prod_{a}^{y-} (\mathbb{I} + dL)$$

for $a < y \le b$, and

$$\prod_{(y,b]} (\mathbb{I} + d\mu) = \prod_{y+}^{b} (\mathbb{I} + dR) = \Big(\prod_{y}^{b} (\mathbb{I} + dR) \Big)^{(a+)}$$

for $a \le y < b$, because R is right–continuous only on (a,b). Here $f^{(a+)}$ is a function on $[a,b]$ defined by $f^{(a+)}(x) := f(x)$ for $a < x \le b$ and $f^{(a+)}(a) := f(a+)$. Since R is right–continuous on (a,b) and L is left–continuous on (a,b), the Duhamel formula of [29, Theorem 5.5] can be written in terms of the (CY) integral as follows:

$$\begin{aligned} \prod_{a}^{b} (\mathbb{I} + dR) - \prod_{a}^{b} (\mathbb{I} + dL) &= (CY) \int_a^b \Big(\prod_{y}^{b} (\mathbb{I} + dR) \Big)^{(a+)} d(R - L) \Big(\prod_{a}^{y} (\mathbb{I} + dL) \Big)_{-}^{(a)} \\ &= (RYS) \int_a^b \prod_{(y,b]} (\mathbb{I} + d\mu)\, d(R - L) \prod_{[a,y)} (\mathbb{I} + d\nu), \end{aligned}$$

where the last equality holds by Proposition 6.22. Since the (RYS) integral does not depend on values of the integrator at jump points, and since $(\mu - \nu)([a,x]) = R(x) - L(x)$ for all $x \in [a,b]$ except at atoms of $\mu - \nu$, the Duhamel formula (7.26) holds. $\square$

We give a second proof of the Duhamel formula which in view of what is used from [29] would give a shorter total proof. It shows how the (RYS) integral appears in (7.26) as compared with the Left Young and Right Young integrals used in the Duhamel formula of [29, Theorem 5.5].

Second proof. Let $\mathfrak{D} = (J_1, \dots, J_N)$ be a Young interval partition of $[a,b]$ and let $\wp(\mathfrak{D}) = \{x_0, x_1, \dots, x_n\}$ be the corresponding point partition. For any $I \subset \{1, \dots, N\}$ and an additive interval function γ, let

$$P_\gamma^{\mathfrak{D}}(\cup_{i \in I} J_i) := \prod_{i \in I} [\mathbb{I} + \gamma(J_i)] \quad \text{and} \quad P_\gamma^{\mathfrak{D}}(\emptyset) := \mathbb{I}.$$

Define step functions $G^{\mathfrak{D}}$ and $F^{\mathfrak{D}}$ on $[a,b]$ by

$$G^{\mathfrak{D}}(y) := \begin{cases} P_\mu^{\mathfrak{D}}([x_i, b]) & \text{if } y \in (x_{i-1}, x_i) \text{ for some } i = 1, \dots, n \\ P_\mu^{\mathfrak{D}}((x_i, b]) & \text{if } y = x_i \text{ for some } i = 0, \dots, n \end{cases}$$

and

$$F^{\mathfrak{D}}(y) := \begin{cases} P_\nu^{\mathfrak{D}}([a, x_{i-1}]) & \text{if } y \in (x_{i-1}, x_i) \text{ for some } i = 1, \dots, n \\ P_\nu^{\mathfrak{D}}([a, x_i)) & \text{if } y = x_i \text{ for some } i = 0, \dots, n. \end{cases}$$

Letting $\gamma := \mu - \nu$, by (7.25), we have

$$\begin{aligned} \oint_a^b G^{\mathfrak{D}}\, d\gamma\, F^{\mathfrak{D}} &= \sum_{i=1}^n P_\mu^{\mathfrak{D}}([x_i, b])\gamma((x_{i-1}, x_i))P_\nu^{\mathfrak{D}}([a, x_{i-1}]) + \sum_{i=0}^n P_\mu^{\mathfrak{D}}((x_i, b])\gamma(\{x_i\})P_\nu^{\mathfrak{D}}([a, x_i)) \\ &= P_\mu^{\mathfrak{D}}([a,b]) - P_\nu^{\mathfrak{D}}([a,b]). \end{aligned} \tag{7.27}$$

In the next lemma we construct an increasing sequence $\{\mathfrak{D}_m : m \geq 1\}$ of Young partitions of $[a,b]$ such that $D := \cup_m \wp(\mathfrak{D}_m)$ is dense in $[a,b]$, contains all the atoms of γ and, for each $y \in D$,

$$\lim_{m \to \infty} G^{\mathfrak{D}_m}(y) = \prod_{(y,b]} (\mathbb{I} + d\mu), \qquad \lim_{m \to \infty} F^{\mathfrak{D}_m}(y) = \prod_{[a,y)} (\mathbb{I} + d\nu), \tag{7.28}$$

$$\lim_{m \to \infty} P_\mu^{\mathfrak{D}_m}([a,b]) = \prod_{[a,b]} (\mathbb{I} + d\mu) \quad \text{and} \quad \lim_{m \to \infty} P_\nu^{\mathfrak{D}_m}([a,b]) = \prod_{[a,b]} (\mathbb{I} + d\nu).$$

Then (7.26) will follow provided

$$\lim_{m \to \infty} \oint_a^b G^{\mathfrak{D}_m}\, d\gamma\, F^{\mathfrak{D}_m} = \oint_a^b \prod_{(y,b]} (\mathbb{I} + d\mu)\gamma(dy) \prod_{[a,y)} (\mathbb{I} + d\nu). \tag{7.29}$$

Lemma 7.10. *There exists an increasing sequence $\{\mathfrak{D}_m : m \geq 1\}$ of Young interval partitions of $[a,b]$ such that $D := \cup_m \wp(\mathfrak{D}_m)$ is dense in $[a,b]$, contains all the atoms of γ,*

$$\lim_{m \to \infty} P_\mu^{\mathfrak{D}_m}([y,b]) = \prod_{[y,b]} (\mathbb{I} + d\mu), \qquad \lim_{m \to \infty} P_\nu^{\mathfrak{D}_m}([a,y]) = \prod_{[a,y]} (\mathbb{I} + d\nu), \tag{7.30}$$

and (7.28) *holds for $y \in D$.*

Proof. Let $\{\xi_m : m \geq 1\}$ and $\{\zeta_m : m \geq 1\}$ denote the atoms of μ and ν, respectively. For each $m \geq 1$, put $r_m := \{a + l2^{-m} : l = 1, \dots, [2^m(b-a)]\}$. We will proceed recursively. By Corollary 7.5, there exists a Young partition $\mathfrak{D}_1$ such that, for each refinement $\mathfrak{D}$ of $\mathfrak{D}_1$,

$$\|P_\mu^{\mathfrak{D}}([a,b]) - \prod_{[a,b]} (\mathbb{I} + d\mu)\| \vee \|P_\nu^{\mathfrak{D}}([a,b]) - \prod_{[a,b]} (\mathbb{I} + d\nu)\| < 1.$$

We may and do assume that ξ_1, ζ_1 and the points of r_1 are endpoints of some elements of $\mathfrak{D}_1$. Assume that $m \geq 2$ and we already have Young partitions $\mathfrak{D}_1, \dots, \mathfrak{D}_{m-1}$ with $\wp(\mathfrak{D}_{m-1}) =$

$\{x_0^{m-1}, \ldots, x_{n(m-1)}^{m-1}\}$. To construct $\mathfrak{D}_m$ using Corollary 7.5, for each $i = 0, \ldots, n(m-1) - 1$, one can find a Young partition $\mathfrak{D}_m^{i,\mu}$ of $[x_i^{m-1}, b]$ such that

$$\Big\| P_\mu^{\mathfrak{D}}([x_i^{m-1}, b]) - \prod_{[x_i^{m-1}, b]} (\mathbb{I} + d\mu) \Big\| \vee \Big\| P_\mu^{\mathfrak{D}}((x_i^{m-1}, b]) - \prod_{(x_i^{m-1}, b]} (\mathbb{I} + d\mu) \Big\| < 1/m$$

for all refinements $\mathfrak{D}$ of $\mathfrak{D}_m^{i,\mu}$. Similarly for each $i = 1, \ldots, n(m-1)$, one can find a Young partition $\mathfrak{D}_m^{i,\nu}$ of $[a, x_i^{m-1}]$ such that

$$\Big\| P_\nu^{\mathfrak{D}}([a, x_i^{m-1}]) - \prod_{[a, x_i^{m-1}]} (\mathbb{I} + d\mu) \Big\| \vee \Big\| P_\nu^{\mathfrak{D}}([a, x_i^{m-1})) - \prod_{[a, x_i^{m-1})} (\mathbb{I} + d\mu) \Big\| < 1/m$$

for all refinements $\mathfrak{D}$ of $\mathfrak{D}_m^{i,\nu}$. We may and do assume that among the endpoints of elements of the new partitions $\mathfrak{D}_m^{i,\mu}$ and $\mathfrak{D}_m^{i,\nu}$ are all points $r_m \cap [x_i^{m-1}, b]$ and $r_m \cap [a, x_i^{m-1}]$, respectively. Let $\mathfrak{D}_m$ be the Young partition of $[a, b]$ such that

$$\wp(\mathfrak{D}_m) = \{\xi_m, \zeta_m\} \cup \Big(\cup_{i=0}^{n(m-1)-1} \wp(\mathfrak{D}_m^{i,\mu}) \Big) \cup \Big(\cup_{i=1}^{n(m-1)} \wp(\mathfrak{D}_m^{i,\nu}) \Big).$$

We arrive recursively at a sequence of Young partitions $\{\mathfrak{D}_m \colon m \geq 1\}$. By construction the set $D := \cup_m \wp(\mathfrak{D}_m)$ is dense in $[a, b]$, contains all atoms of γ and relations (7.28), (7.30) hold. □

To prove (7.29) we will use a convergence theorem for (RYS) integrals. The theorem applies to sequences $\{f_m\} = \{f_m \colon m \geq 1\}$ of functions which have uniformly bounded p-variation and converge locally uniformly. That is, we say that $\{f_m\}$ is a *$\mathcal{W}_p$–sequence* if $\sup_m \|f_m\|_{[p]} < \infty$ and $\{f_m\}$ *converges locally uniformly to f at y*, if given $\epsilon > 0$, there is an m_0 and a $\delta > 0$ such that

$$\|f_m(z) - f(z)\| < \epsilon \tag{7.31}$$

for all $m \geq m_0$ and all z distant less than δ from y. To show local uniform convergence in the proof below we first show *uniform convergence on the left at y* meaning that (7.31) holds for $z < y$, and the other conditions are the same. Then we show *uniform convergence on the right at y* meaning that (7.31) holds for $z \geq y$.

Proposition 7.11. *Let $h \in \mathcal{W}_p([a, b]; \mathbb{B})$ and let $\{g_m\}$, $\{f_m\}$ be $\mathcal{W}_q$–sequences of $\mathbb{B}$–valued functions on $[a, b]$, where $p^{-1} + q^{-1} > 1$, $p, q > 0$. Suppose that $\{g_m\}$ and $\{f_m\}$ converge pointwise on a dense set to functions $g, f \in \mathcal{W}_q([a, b]; \mathbb{B})$, respectively, and that both sequences converge locally uniformly to their limits at each discontinuity of h. Then*

$$\lim_{m \to \infty} (RYS) \int_a^b g_m \, dh \, f_m = (RYS) \int_a^b g \, dh \, f. \tag{7.32}$$

Proof. By Proposition 6.22, the (RYS) integrals exist and are equal to the corresponding (CY) integrals. Thus the stated convergence theorem follows from Proposition 3.33 of [29]. □

Continuation of the second proof of Theorem 7.9. We check the hypotheses of Proposition 7.11 with $q = p$ only for $g_m = G^m := G^{\mathfrak{D}_m}$, $m \geq 1$. The hypotheses for $\{F^{\mathfrak{D}_m}\}$ can be verified similarly. For the rest of the proof we use without special mention Theorems 4.39 and 7.3 relating the p-variation and the product integrals for μ, $R_{\mu,a}$ and $L_{\mu,a}$. Let

$$C(\mu) := \sup \{ \|P_\mu^{\mathfrak{D}}(J)\| \colon \mathfrak{D} \in \mathrm{IP}_1(J),\ J \in \mathfrak{I}_{os}([a, b]) \}.$$

Since $v_p(\mu) < \infty$, by Lemma 4.19 of [29], $C(\mu) < \infty$. Suppose for a moment that $\rho := C_p\|\mu\|_{(p)} < 1$ for a suitable constant C_p from Lemma 4.18 of [29]. Then for $\kappa \in \mathrm{PP}([a,b])$, by inequality (4.52) of [29], we have

$$\begin{aligned} s_p(G^m;\kappa) &\le C(\mu)^p \sup\Big\{\sum_j \|P_\mu^{\mathfrak{D}_m}(J_j) - \mathbb{I}\|^p \colon \{J_j\} \in \mathrm{IP}(\mathfrak{I}_{os}([a,b])),\ \{J_j\} \subset \mathfrak{D}_m\Big\} \\ &\le [C(\mu)/(1-\rho)]^p \sup\Big\{\sum_j v_p(\mu; J_j) \colon \{J_j\} \in \mathrm{IP}(\mathfrak{I}_{os}([a,b])),\ \{J_j\} \subset \mathfrak{D}_m\Big\} \\ &\le [C(\mu)/(1-\rho)]^p \|\mu\|_{(p)}^p. \end{aligned}$$

Since κ and m are arbitrary, $\{G^m\}$ is a $\mathcal{W}_p$–sequence in the case when $\|\mu\|_{(p)} < 1/C_p$. In the general case we use Lemma 4.7 above to split the interval $[a,b]$ by a Young partition $\mathfrak{D}$ so that the above bound holds on each open subinterval of $\mathfrak{D}$. This yields that $\{G^m\}$ is a $\mathcal{W}_p$–sequence.

Let y be an atom of $\gamma = \mu - \nu$. Then $y \in D = \cup_m \wp(\mathfrak{D}_m)$. First, we show that if $y > a$,

$$G^m \to \prod_{(\cdot,b]} (\mathbb{I} + d\mu) \quad \text{as } m \to \infty, \text{ uniformly on the left at } y. \tag{7.33}$$

Let $\epsilon > 0$. By Lemma 7.10, there exists m_0 so large that, for all $m \ge m_0$, $y \in \wp(\mathfrak{D}_m)$ and

$$\Big\| G^m(y-) - \prod_{[y,b]} (\mathbb{I} + d\mu) \Big\| < \epsilon. \tag{7.34}$$

By Lemmas 2.18 and 5.2 of [29], there exists $\delta \in (0, y-a)$ such that

$$v_p(\mu; [y-\delta, y)) < \{[\epsilon/(2C(\mu))] \wedge (2C_p)^{-1}\}^p, \tag{7.35}$$

where C_p is as before a suitable constant from Lemma 4.18 of [29], and for all $z \in [y-\delta, y)$,

$$\Big\| \prod_{(z,b]} (\mathbb{I} + d\mu) - \prod_{[y,b]} (\mathbb{I} + d\mu) \Big\| = \Big\| \prod_{z+}^{b} (\mathbb{I} + dL_{\mu,a}) - \prod_{y}^{b} (\mathbb{I} + dL_{\mu,a}) \Big\| \le \epsilon \tag{7.36}$$

because $L_{\mu,a}$ is left–continuous at y. Thus by inequality (4.52) of [29] and (7.35) above, we have for $z \in [y-\delta, y)$,

$$\begin{aligned} \|G^m(z) - G^m(y-)\| &\le C(\mu) \max\{\|P_\mu^{\mathfrak{D}_m}(\{x,y)) - \mathbb{I}\| \colon x \in \wp(\mathfrak{D}_m) \cap [y-\delta, y)\} \\ &\le 2C(\mu) v_p(\mu; [y-\delta, y))^{1/p} < \epsilon, \end{aligned}$$

where $\{\cdot, y)$ is an interval right open at y. Thus by (7.34) and (7.36), for each $m \ge m_0$ and $z \in [y-\delta, y)$, we get

$$\begin{aligned} \Big\| G^m(z) - \prod_{(z,b]} (\mathbb{I} + d\mu) \Big\| &\le \|G^m(z) - G^m(y-)\| + \Big\| G^m(y-) - \prod_{[y,b]} (\mathbb{I} + d\mu) \Big\| \\ &\quad + \Big\| \prod_{[y,b]} (\mathbb{I} + d\mu) - \prod_{(z,b]} (\mathbb{I} + d\mu) \Big\| < 3\epsilon. \end{aligned}$$

Therefore (7.33) holds.

Second, we show that if $y < b$,

$$G^m \to \prod_{(\cdot,b]} (\mathbb{I} + d\mu) \quad \text{as } m \to \infty, \text{ uniformly on the right at } y. \tag{7.37}$$

Let $\epsilon > 0$. By Lemma 7.10, there exists m_0 so large that, for all $m \geq m_0$, $y \in \wp(\mathfrak{D}_m)$ and

$$\Big\| G^m(y) - \prod_{(y,b]} (\mathbb{I} + d\mu) \Big\| < \epsilon. \tag{7.38}$$

By Lemmas 2.18 and 5.2 of [29], there exists $\delta \in (0, b-y)$ such that

$$v_p(\mu; (y, y+\delta]) < \{[\epsilon/(2C(\mu))] \wedge (2C_p)^{-1}\}^p, \tag{7.39}$$

where C_p is again a suitable constant from Lemma 4.18 of [29], and for all $z \in (y, y+\delta]$,

$$\Big\| \prod_{(z,b]} (\mathbb{I} + d\mu) - \prod_{(y,b]} (\mathbb{I} + d\mu) \Big\| = \Big\| \prod_{z+}^{b} (\mathbb{I} + dL_{\mu,a}) - \prod_{y+}^{b} (\mathbb{I} + dL_{\mu,a}) \Big\| \leq \epsilon. \tag{7.40}$$

We can and do assume that $y + \delta \in \wp(\mathfrak{D}_m)$ for $m \geq m_0$. Then by inequality (4.52) of [29] and (7.39) above, we have for $z \in (y, y+\delta]$,

$$\begin{aligned} \|G^m(z) - G^m(y)\| &\leq C(\mu) \max\{\|P_\mu^{\mathfrak{D}_m}((y,x\}) - \mathbb{I}\| \colon x \in \wp(\mathfrak{D}_m) \cap (y, y+\delta)\} \\ &\leq 2C(\mu) v_p(\mu; (y, y+\delta])^{1/p} < \epsilon, \end{aligned}$$

where $(y, \cdot\}$ is an interval left open at y. Thus by (7.38) and (7.40), for each $m \geq m_0$ and $z \in (y, y+\delta]$, we get

$$\begin{aligned} \Big\| G^m(z) - \prod_{(z,b]} (\mathbb{I} + d\mu) \Big\| &\leq \|G^m(z) - G^m(y)\| + \Big\| G^m(y) - \prod_{(y,b]} (\mathbb{I} + d\mu) \Big\| \\ &\quad + \Big\| \prod_{(y,b]} (\mathbb{I} + d\mu) - \prod_{(z,b]} (\mathbb{I} + d\mu) \Big\| < 3\epsilon. \end{aligned}$$

Therefore (7.37) holds. This together with (7.33) yields that $\{G^m\}$ converges locally uniformly at each atom of γ. By Lemma 7.10 it also converges densely. Hence all the hypotheses of Theorem 7.11 are satisfied, and hence (7.29) holds. The second proof of the Duhamel formula (7.26) is complete. $\square$

7.4 Integral equations

First, as an example, let's consider the integral equation

$$G(x) = C + \int_0^x G(y)\, dF(y).$$

Let $0 \leq x \leq 2$ and $F := 1_{[1,\infty)}$. Then for the integrals we are considering other than the (MRS) and (RRS) integrals (which may be undefined), the equation gives

$$G(x) = \begin{cases} C, & \text{if } 0 \leq x < 1; \\ C + G(1), & \text{if } 1 \leq x \leq 2. \end{cases}$$

For $x = 1$, this yields a contradiction (unless $C = 0$). In accordance with some developments in analysis (see below) and stochastic analysis (e.g. Doléans-Dade [15]) it seems desirable that in such an equation the integrand should be left–continuous while the solution is right–continuous. Writing

$$G(x) = G_+^{(2)}(x) = C + \int_0^x G_-^{(0)}(y)\, dF(y)$$

we get

$$G(x) = \begin{cases} C, & \text{if } 0 \le x < 1; \\ 2C, & \text{if } 1 \le x \le 2 \end{cases}$$

and now there is no contradiction. The integral can be any of those we are considering (in Figure 2.1) except for the (MRS) integral but now including the (RRS) integral.

In analysis several approaches have been suggested to solving a linear Stieltjes type integral equation with respect to a possibly discontinuous function F of bounded variation. Using an integral equivalent to the (RYS) integral, Hildebrandt [51, p. 359] proved that if a matrix–valued function F with bounded variation has a finite number of discontinuities on $[a, b]$ then the integral equation

$$G(x) = G(a) + \int_a^x dF(y)\, G(y), \qquad a \le x \le b, \tag{7.41}$$

has a unique solution if and only if the matrices $\mathbb{I} - \Delta^- F(x)$ have inverses for all points of discontinuity of F. Thus a situation similar to the above example is excluded. Hildebrandt considers cases where F may have non–zero jumps $(\Delta^+ F)(u)$ and $(\Delta^- F)(u)$ on the two sides of the same point and seeks an actual solution of (7.41) with $G(y)$ (not $G(y-)$). The solution is a kind of product integral where the factor at a point u where F has jumps on one side or the other is $[\mathbb{I} + \Delta^+ F(u)][\mathbb{I} - \Delta^- F(u)]^{-1}$. MacNerney [77] used the left Cauchy and right Cauchy integrals defined as (RRS) integrals except that tagged partitions $(\{x_i\}, \{y_i\})$ are used such that $y_i = x_{i-1}$ and $y_i = x_i$, respectively. Hönig [53] used the interior (RRS) integral defined as the (RRS) integral except that the tagged partitions $(\{x_i\}, \{y_i\})$ defining the integral satisfy the relation: $x_{i-1} < y_i < x_i$. The book of Schwabik, Tvrdý and Vejvoda [101] treats linear integral equations with respect to functions of bounded variation using the (HK) integral.

The form of the Duhamel formula from the preceding section suggests the following:

Definition 7.12. Let ν be a $\mathbb{B}$–valued additive and upper continuous interval function on $[a, b]$. We say that a $\mathbb{B}$–valued interval function α satisfies the *forward linear integral equation with respect to* ν, if the integral $⨎_{[a,b]} \nu(dx)\, \alpha([a, x))$ is defined, and for all $y \in [a, b]$,

$$\alpha([a, y]) = \mathbb{I} + ⨎_{[a,y]} \nu(dx)\, \alpha([a, x)). \tag{7.42}$$

Similarly, we say that a $\mathbb{B}$–valued interval function α satisfies the *backward linear integral equation with respect to* ν, if the integral $⨎_{[a,b]} \alpha((x, b])\, \nu(dx)$ is defined, and for all $y \in [a, b]$,

$$\alpha([y, b]) = \mathbb{I} + ⨎_{[y,b]} \alpha((x, b])\, \nu(dx). \tag{7.43}$$

Theorem 7.13. *Let $\nu \in \mathcal{I}_p([a, b]; \mathbb{B})$, $1 \le p < 2$ be additive and upper continuous. Then the product integral $\prod(\mathbb{I} + d\nu)$ is defined, is in $\mathcal{I}_p([a, b]; \mathbb{B})$ and is the unique solution in $\mathcal{I}_r([a, b]; \mathbb{B})$, for any $r \ge p$ such that $1/p + 1/r > 1$, of the forward integral equation* (7.42).

Proof. The product integral $\prod(\mathbb{I}+d\nu)$ exists and is in $\mathcal{I}_p([a,b];\mathbb{B})$ by Corollary 7.5. Taking $\mu=0$ in the Duhamel formula (7.26) we get that $\prod(\mathbb{I}+d\nu)$ satisfies the forward linear integral equation with respect to ν. Let μ be another solution in $\mathcal{I}_p([a,b];\mathbb{B})$ and let

$$F(y) := \prod_{[a,y]}(\mathbb{I}+d\nu)-\mu([a,y]) \quad \text{for } a<y\le b \text{ and } F(a):=\mathbb{I}-\mu(\emptyset).$$

Then for some $h\in[\nu]_a$ (cf. (4.34)) and all $y\in[a,b]$,

$$F(y)=(RYS)\int_a^y dh\,F_-^{(a)}.$$

So $F(a)=0$. By (6.8), $\Delta^+F(a)=(\Delta^+h)(a)F(a)=0$. Thus $F(a+)=0$. By the Love–Young inequality for the indefinite (RYS) integral (Corollary 4.28), and since $\|F\|_{[p]}$ is a non-increasing function of p (Lemma 4.5), for $r\ge p$ such that $1/p+1/r>1$, it follows that

$$\|F\|_{[r]}\le\|F\|_{[p]}\le D_{p,r}\|h\|_{(p)}\|F\|_{[r]}, \tag{7.44}$$

where $D_{p,r}=2\zeta(p^{-1}+r^{-1})$. Then (7.44) holds for norms over any subinterval $(a,y]$ of $[a,b]$. Taking y close enough to a we will have $\|h\|_{(p)}<1/D_{p,r}$ by Lemma 4.8, and thus $F\equiv0$ on $[a,y]$. There is a largest y for which this holds, with $y\le b$. If $y=b$ then we are done. If $y<b$ then we can repeat the argument with y in place of a and contradict maximality of y. So uniqueness is proved. □

7.5 Lyons' inequalities and series expansions

In this section we prove that the product integral with respect to a regulated additive interval function μ on $[a,b]$ with bounded p-variation $1\le p<2$, has the series expansion:

$$\prod_{[a,b]}(\mathbb{I}+d\mu)=\mathbb{I}+\mu([a,b])+\sum_{k\ge2}\,⨎_{[a,b]}\mu(dx_1)\,⨎_{[a,x_1)}\mu(dx_2)\cdots⨎_{[a,x_{k-1})}\mu(dx_k).$$

Proposition 7.14. *For $1\le p<2$, let $\mu\in\mathcal{I}_p([a,b];\mathbb{B})$ be additive and upper continuous, and let $f\in\mathcal{W}_p([a,b];\mathbb{B})$. Suppose that f is right–continuous on $[a,b)$, $f(a)=0$ and $v_p(f;[u,v])\le\rho^k(v)-\rho^k(u)$ for $a\le u<v\le b$ and $k\ge1$, where $\rho(x):=v_p(\mu;[a,x])$ for $x\in[a,b]$. Then for the function H defined on $[a,b]$ by $H(x):=⨎_{[a,x]}d\mu\,f_-^{(a)}$, and any $a\le u<v\le b$,*

$$V_p(H;[u,v])\le C_p\Big(\frac{\rho^{k+1}(v)-\rho^{k+1}(u)}{k+1}\Big)^{1/p}, \tag{7.45}$$

where $C_p:=1+4^{1/p}\zeta(2/p)$, and H is right–continuous on $[a,b)$.

The proof of this proposition will be based on two facts. The first one is a part of Theorem 1.1 of Lyons [76], up to the constant 4.

Lemma 7.15. *Let ρ be a non-decreasing function on $[a,b]$, k be a positive integer and $\kappa=\{x_i\colon i=0,\dots,n\}\in\mathrm{PP}([a,b])$. Then there exists an index $i\in\{1,\dots,n-1\}$ such that*

$$[\rho^k(x_i)-\rho^k(x_{i-1})][\rho(x_{i+1})-\rho(x_i)]\le\frac{4}{(n-1)^2}\Big(\frac{\rho^{k+1}(b)-\rho^{k+1}(a)}{k+1}\Big).$$

Proof. Denoting $\Delta^i\phi := \phi(x_i) - \phi(x_{i-1})$ for a function ϕ, for each $i = 1, \dots, n-1$, we have

$$\begin{aligned}
\Delta^i\rho^k\Delta^{i+1}\rho &= k\int_{\rho(x_{i-1})}^{\rho(x_i)} u^{k-1}\,du\,\Delta^{i+1}\rho \le k\int_{\rho(x_{i-1})}^{\rho(x_i)} u^{\frac{k-1}{2}}\,du\,[\rho(x_i)]^{\frac{k-1}{2}}[\rho(x_{i+1}) - \rho(x_i)] \\
&\le k\int_{\rho(x_{i-1})}^{\rho(x_i)} u^{(k-1)/2}\,du\int_{\rho(x_i)}^{\rho(x_{i+1})} u^{(k-1)/2}\,du. \qquad (7.46)
\end{aligned}$$

Then using the inequality $2\sqrt{xy} \le x + y$ for $x, y \ge 0$, we get

$$\begin{aligned}
\min_{1\le i\le k-1}\{\Delta^i\rho^k\Delta^{i+1}\rho\} &\le \Big\{\frac{1}{n-1}\sum_{i=1}^{n-1}\sqrt{\Delta^i\rho^k\Delta^{i+1}\rho}\Big\}^2 \\
\text{by (7.46)} \qquad &\le \Big\{\frac{1}{n-1}\frac{\sqrt{k}}{2}\Big(\int_{\rho(a)}^{\rho(x_{n-1})} u^{\frac{k-1}{2}}\,du + \int_{\rho(x_1)}^{\rho(b)} u^{\frac{k-1}{2}}\,du\Big)\Big\}^2 \\
&\le \Big\{\frac{\sqrt{k}}{n-1}\int_{\rho(a)}^{\rho(b)} u^{\frac{k-1}{2}}\,du\Big\}^2 = \Big\{\frac{2\sqrt{k}}{n-1}\Big(\frac{\rho^{(k+1)/2}(b) - \rho^{(k+1)/2}(a)}{k+1}\Big)\Big\}^2 \\
&\le \frac{4}{(n-1)^2}\Big(\frac{\rho^{k+1}(b) - \rho^{k+1}(a)}{k+1}\Big),
\end{aligned}$$

where in the last step, the inequality $(d-c)^2 < d^2 - c^2$ for $0 < c < d$ was used. The proof is complete. $\square$

The second fact to be used in proving Proposition 7.14 is the following:

Lemma 7.16. *Let $f, h \in \mathcal{W}_p([a,b];\mathbb{B})$ for $1 \le p < 2$, let f be right–continuous on $[a,b)$ and let h be right–continuous on (a,b). Then for any $\epsilon > 0$ there is a partition $\{z_j\}_{j=0}^m \in \mathrm{PP}\,([a,b])$ such that*

$$\Big\|(RYS)\int_a^b dh\,f_-^{(a)} - \sum_{j=1}^m [h(z_j) - h(z_{j-1})]f(z_{j-1})\Big\| < \epsilon.$$

Proof. Since $\|f_-^{(a)}\|_{(p)} \le \|f\|_{(p)}$, and $f_-^{(a)}$, h have no common one–sided discontinuities, the (RYS) integral exists and equals an (RRS) integral by Theorems 4.26 and 6.13. Thus there exists a partition $\{z_j\}_{j=0}^m$ such that

$$\Big\|(RRS)\int_a^b dh\,f_-^{(a)} - \sum_{j=1}^m [h(z_j) - h(z_{j-1})]f_-^{(a)}(y_j)\Big\| < \epsilon$$

whenever $z_{j-1} \le y_j \le z_j$ for $j = 1, \dots, m$. We can let $y_j \downarrow z_{j-1}$ for $j = 1, \dots, m$, giving $f(z_{j-1})$ by right–continuity of f on $[a,b)$. The conclusion follows. $\square$

Proof of Proposition 7.14. Let $a \le u < v \le b$. Since $R_{\mu,a}$ is right–continuous on (a,b), by Proposition 3.25 and Corollary 3.20 of [29], we have

$$H(u) - H(v) = (RYS)\int_u^v dR_{\mu,a}\,f_-^{(u)}.$$

Let $\epsilon > 0$. By Lemma 7.16, there exists $\lambda = \{x_i \colon i = 0, \dots, n\} \in \mathrm{PP}\,[u,v]$ such that

$$\Big\|(RYS)\int_u^v dR_{\mu,a}\,f_-^{(u)} - S(\lambda)\Big\| < \epsilon,$$

where $S(\lambda) := \sum_{i=1}^{n}[R_{\mu,a}(x_i)-R_{\mu,a}(x_{i-1})]f(x_{i-1})$. Let $\lambda^i := \lambda\backslash\{x_i\}$ for some $i \in \{1,\dots,n-1\}$. Then we have

$$\begin{aligned}\|S(\lambda)-S(\lambda^i)\| &\le \|R_{\mu,a}(x_{i+1})-R_{\mu,a}(x_i)\|\,\|f(x_i)-f(x_{i-1})\|\\ &\le \{[\rho(x_{i+1})-\rho(x_i)]\,[\rho^k(x_i)-\rho^k(x_{i-1})]\}^{1/p}.\end{aligned}$$

By Lemma 7.15, there exists an index $i \in \{1,\dots,n-1\}$ such that

$$\|S(\lambda)-S(\lambda^i)\| \le \Big\{\frac{4}{(n-1)^2}\Big(\frac{\rho^{k+1}(v)-\rho^{k+1}(u)}{k+1}\Big)\Big\}^{1/p}.$$

Deleting further points according to Lemma 7.15, we obtain the bound

$$\begin{aligned}\Big\|(RYS)\int_u^v dR_{\mu,a}\, f_-^{(u)}\Big\| &\le 4^{1/p}\sum_{n=2}^{\infty}\frac{1}{(n-1)^{2/p}}\Big(\frac{\rho^{k+1}(v)-\rho^{k+1}(u)}{k+1}\Big)^{1/p}\\ &\quad+\|[R_{\mu,a}(v)-R_{\mu,a}(u)]f(u)\|+\epsilon.\end{aligned}$$

Since $f(a)=0$, we have

$$\|[R_{\mu,a}(v)-R_{\mu,a}(u)]f(u)\|^p \le [\rho(v)-\rho(u)]\rho^k(u) \le \int_{\rho(u)}^{\rho(v)} t^k\,dt = \frac{\rho^{k+1}(v)-\rho^{k+1}(u)}{k+1}.$$

Letting $\epsilon \downarrow 0$, it follows that

$$\Big\|(RYS)\int_u^v dR_{\mu,a}\, f_-^{(u)}\Big\| \le [1+4^{1/p}\zeta(2/p)]\Big(\frac{\rho^{k+1}(v)-\rho^{k+1}(u)}{k+1}\Big)^{1/p}.$$

Using this inequality on subintervals, (7.45) follows. $\square$

Proposition 7.17. *Let $\mu \in \mathcal{I}_p([a,b];\mathbb{B})$, $1 \le p < 2$, be additive and upper continuous. For each $k = 2,3,\dots$ the function $I_k(\mu)$ on $[a,b]$ defined by*

$$I_k(\mu)(y) := \oint_{[a,y]}\mu(dx_1)\oint_{[a,x_1)}\mu(dx_2)\cdots\oint_{[a,x_{k-1})}\mu(dx_k),\quad a<y\le b,\qquad I_k(\mu)(a):=0$$

exists, and satisfies the bound

$$\|I_k(\mu)\|_{[p]} \le C_p^{k-1}\Big(\frac{\rho^k(b)-\rho^k(a)}{k!}\Big)^{1/p}, \tag{7.47}$$

where $\rho(y) := v_p(\mu;[a,y])$ and $C_p = 2(1+4^{1/p}\zeta(2/p))$.

Proof. Let $I_1(\mu)(x) := \mu((a,x])$, $x \in [a,b]$. Then $I_k(\mu)(y) = \oint_{[a,y]}\mu(dx)(I_{k-1}(\mu))_-^{(a)}$ for each $k = 2,3,\dots$. It is clear that $I_1(\mu)$ is right–continuous on $[a,b)$ and 0 at a. Because of the relations $a \le x_k < x_{k-1} < \dots < x_1 \le y$, we have

$$I_k(\mu)(y) = \oint_{(a,y]}\mu(dx_1)\oint_{(a,x_1)}\mu(dx_2)\cdots\oint_{[a,x_{k-1})}\mu(dx_k),$$

so that each $I_k(\mu)$ is right–continuous on $[a,b)$ and 0 at a by the definition. Thus Proposition 7.17 follows from Proposition 7.14 by induction. $\square$

Theorem 7.18. *Let $\mu \in \mathcal{I}_p([a,b];\mathbb{B})$, $1 \le p < 2$, be additive and upper continuous. Then the series $\sum_{k\ge 2} I_k(\mu)$ converges in $\mathcal{W}_p([a,b];\mathbb{B})$, the product integral with respect to μ exists, and they satisfy the relation:*

$$\prod_{[a,y]} (\mathbb{I} + d\mu) = \mathbb{I} + \mu([a,y]) + \sum_{k\ge 2} I_k(\mu)(y), \quad \textit{for } y \in [a,b]. \tag{7.48}$$

Proof. The series $\sum_{k\ge 2} I_k(\mu)$ converges in $\mathcal{W}_p$ by Proposition 7.17. The product integral with respect to μ exists by Corollary 7.5. To prove (7.48) let α be the interval function on $[a,b]$ defined by

$$\alpha(J) := \mathbb{I} + \mu(J) + \sum_{k\ge 2} =\!\!\!\!\!\!\int_J d\mu\, (I_{k-1}(\mu))_-^{(u)},$$

where u is the left endpoint of $J \subset [a,b]$ and $I_1(\mu)(x) := \mu((a,x])$, $x \in [a,b]$. By Theorem 4.39, α has bounded p-variation. Also for each $y \in [a,b]$, it follows that

$$\begin{aligned} =\!\!\!\!\!\!\int_{[a,y]} \mu(dx)\alpha([a,x)) &= \mu([a,y]) + =\!\!\!\!\!\!\int_{[a,y]} \mu(dx)\mu([a,x)) + =\!\!\!\!\!\!\int_{[a,y]} \mu(dx) \sum_{k\ge 2} =\!\!\!\!\!\!\int_{[a,x)} d\mu\, (I_{k-1}(\mu))_-^{(a)} \\ &= \mu([a,y]) + =\!\!\!\!\!\!\int_{[a,y]} \mu(dx_1) =\!\!\!\!\!\!\int_{[a,x_1)} \mu(dx_2) + \sum_{k\ge 3} =\!\!\!\!\!\!\int_{[a,y]} d\mu\, (I_k(\mu))_-^{(a)} \\ &= \alpha([a,y]) - \mathbb{I}. \end{aligned}$$

To show that the integrals in the series all exist it is enough to apply the Love–Young inequality. Thus α satisfies the forward linear integral equation with respect to μ. Since the unique solution in $\mathcal{I}_p([a,b];\mathbb{B})$ of this equation is the product integral with respect to μ by Theorem 7.13, (7.48) holds. The proof is complete. $\square$

Corollary 7.19. *Let $\mu \in \mathcal{I}_p([a,b];\mathbb{B})$, $1 \le p < 2$, be additive and upper continuous. Then the mapping*

$$\mathcal{I}_p \ni \mu \mapsto \prod(\mathbb{I} + d\mu) \in \mathcal{I}_p \tag{7.49}$$

is analytic with infinite radius of uniform convergence of its Taylor series around any point.

Proof. By Corollary 7.5, the mapping (7.49) is defined. (7.48) gives a power series expansion of the product integral with an infinite radius of uniform convergence at 0 by (7.47). Thus by Theorem 11.11 of [11, p. 163], where we suggest that $\|A_m\|$ be replaced by $\|\widehat{A}_m\|$ in the proof, the same is true at any other element of $\mathcal{I}_p$. $\square$

7.6 Comments and related results

Survival and hazard functions Let T be a positive and bounded random variable on a probability space $(\Omega, \Pr)$. The *survival function* S on $[0,\infty)$ is defined by $S(t) := \Pr(\{T > t\})$ for $0 \le t < \infty$. Let $\tau_S := \sup\{t \ge 0\colon S(t) > 0\}$ which is finite because T is bounded. A nonnegative measure Λ on $[0,\infty)$ is a *hazard measure* if, letting $\tau_\Lambda := \inf\{t \ge 0\colon \Lambda((t,\infty)) = 0\}$, we have

(*a*) $0 < \tau_\Lambda < \infty$;
(*b*) for $0 < s < \tau_\Lambda$, $\Lambda(\{s\}) < 1$ and $\Lambda([0,s]) < \infty$;
(*c*) either $\Lambda([0,\tau_\Lambda)) < \infty$ and $\Lambda(\{\tau_\Lambda\}) = 1$ or $\Lambda([0,\tau_\Lambda)) = \infty$ and $\Lambda(\{\tau_\Lambda\}) = 0$.

Theorem 7.20. *There is a* $1-1$ *correspondence between hazard measures* Λ *on* $[0,\infty)$ *and survival functions* S *on* $[0,\infty)$ *given by*

$$S(t) = \prod_{[0,t]} (1 - d\Lambda), \quad 0 \le t < \tau_\Lambda, \qquad S(\tau_\Lambda) = 0 \qquad \textit{and}$$

$$\Lambda([0,t]) = -\int_{[0,t]} \frac{dS(u)}{S(u-)}, \quad 0 \le t < \tau_S, \quad \Lambda(\{\tau_S\}) := \begin{cases} 1 & \textit{if } S(\tau_S-) > 0 \\ 0 & \textit{otherwise} \end{cases} \quad \Lambda((\tau_S,\infty)) := 0,$$

with $\tau_\Lambda = \tau_S$.

This is approximately Theorem 11 of Gill and Johansen [43] (see also [42, Section 5], [1, Theorem II.6.6]).

In survival analysis there is a sample of n individuals each of whom has a lifetime T_i and a "censoring time" C_i. Let T_i be independent identically distributed positive random variables with an unknown law having a survival function S to be estimated, and let $C_1, \dots, C_n$ be independent identically distributed positive random variables with a law having a survival function H. For each i one observes $X_i := \min(T_i, C_i)$ and $D_i = 1_{\{X_i = T_i\}}$, that is whether X_i actually equals T_i or C_i. We assume that for some $\tau_0 < \infty$, $C_i \le \tau_0$ for all i. Let $\tau_n := \max_{1 \le i \le n} X_i \le \tau_0$. For $0 \le t < \tau_n$, let

$$N_n(t) := \frac{1}{n}\#\{i\colon X_i \le t,\, D_i = 1\}, \quad Y_n(t) := \frac{1}{n}\#\{i\colon X_i \ge t\}, \quad \text{and} \quad \Lambda_n([0,t]) := \int_{[0,t]} \frac{dN_n}{Y_n}.$$

The *empirical hazard measure* Λ_n so defined is called the Nelson–Aalen estimator of the hazard measure. Since $\Lambda_n([0,\tau_n))$ is always finite in this case, $\Lambda_n(\{\tau_n\}) := 1$ is needed to fit the above definition of hazard measure. Unlike the point masses at $X_i = T_i < \tau_n$, representing deaths, or other endpoints of studies, $\Lambda_n(\{\tau_n\}) = 1$ simply means that only the distribution of $\min(X_i, \tau_n)$ can be estimated at all. The *empirical survival function* S_n defined by

$$S_n(t) := \prod_{[0,t]} (1 - d\Lambda_n), \qquad 0 \le t \le \tau_n,$$

is called the Kaplan-Meier estimator of the unknown survival function $S(t) = \Pr(\{T_i > t\})$. S_n is the nonparametric maximum likelihood estimator of S on $[0,\tau_S)$ from the given data set, and has other good properties. However, no estimator can be reliable for t large enough so that $Y_n(t)$ is small.

An alternative definition of the product integral Let M_k be the set of $k \times k$ matrices with complex entries, let μ be a nonnegative measure on an interval $[a,b]$ and $f \in \mathcal{L}^1([a,b],\mu,M_k)$. Dollard and Friedman [17, pp. 155-156] define a form of product integral

$$\prod_{(a,b]} \exp\{f\,d\mu\} \quad \text{as a limit of products} \quad \prod_{i=1}^{n} \exp\{\nu((x_{i-1},x_i])\} \tag{7.50}$$

over partitions $a = x_0 < x_1 < \cdots < x_n = b$ as the mesh of partitions goes to 0, where $\nu((c,d]) := \int_{(c,d]} f\,d\mu$. The exponential is defined by its usual power series. If the signed measure ν has no atoms, then

$$\prod_{(a,b]} \exp\{f\,d\mu\} = \prod_{(a,b]} (\mathbb{1} + d\nu) \tag{7.51}$$

(Dollard and Friedman [17, Theorem I.7.1, p. 52]; the continuity assumption there can be avoided because of our different formulation). The function $g(z) = e^z$ or $1+z$ can be replaced by any function analytic in a neighborhood of 0 with $g(0) = g'(0) = 1$. Thus $g(z) = 1+z+O(z^2)$ as $z \to 0$. Beside $1+z$ and e^z, the function $1/(1-z)$ also has been applied, cf. Hildebrandt [51] and Freedman [36]. If ν has atoms the product integral depends on the choice of g and corresponding definitions give different results.

If the values of ν on different intervals commute, e.g. if ν is real–valued, then simply

$$\prod_{(a,b]} \exp\{f\,d\mu\} = \exp\{\nu((a,b])\}, \tag{7.52}$$

which is never 0 since ν is a finite signed measure. If ν is taken as a hazard measure and has atoms, (7.50) does not give the Kaplan–Meier estimator but a different, unwanted estimator.

While the product integral $\prod(\mathbb{I}+d\nu)$ gives solutions of integral equations (7.42) and (7.43), the form

$$P(a,b) := \prod_{(a,b]} \exp\{f\,d\mu\}$$

gives solutions of integral equations

$$P(a,b) = \mathbb{I} + \int_{(a,b]} f_\phi(s)P(a,s)\,\mu(ds),$$

where $f_\phi(s) := \phi(f(s)\mu(\{s\}))f(s)$ and $\phi(x) := (e^z-1)/z$, $z \neq 0$, $\phi(0) := 1$ (Dollard and Friedman [17, pp. 160-161]). For natural integral equations with atoms (jumps), then, it seems that $\prod(\mathbb{I}+d\nu)$ might be preferred. If ν has no atoms, then in light of (7.52) one can see how the form $\prod_{(a,b]}\exp\{f d\mu\}$ might have been preferred, although it is equivalent to $\prod(\mathbb{I}+d\nu)$ by (7.51).

We next extend the definition of the product integral $\Pi\, e^{df}$. Let $\kappa := \{I_j\}_{j=1}^n$ be a partition of an interval J into disjoint subintervals (some of which may be singletons), i.e. $\kappa \in \mathrm{IP}\,(J)$. Then $x < y$ whenever $x \in I_i$, $y \in I_j$, and $1 \le i < j \le n$. Let ν be an additive interval function defined on all subintervals of J, with values in a Banach algebra $\mathbb{B}$ with identity $\mathbb{I}$. Recall that for $A \in \mathbb{B}$, $e^A := \exp(A) := \sum_{j=0}^\infty A^j/j!$, where $A^0 := \mathbb{I}$. We define the product $\Pi_\kappa e^{\Delta\nu} := \Pi_{j=1}^n \exp\{\nu(I_j)\}$. We say that the product integral $\Pi_J e^{d\nu}$ exists and equals $A \in \mathbb{B}$ if for every $\epsilon > 0$ there is a $\kappa \in \mathrm{IP}\,(J)$ such that for every refinement λ of κ, $\|A - \Pi_\lambda e^{\Delta\nu}\| < \epsilon$.

If $\mathbb{B}$ is commutative then $e^{A+B} = e^A e^B$ for any $A, B \in \mathbb{B}$ and the following is immediate:

Theorem 7.21. *If $\mathbb{B}$ is a commutative Banach algebra with identity then for an arbitrary finitely additive interval function ν into $\mathbb{B}$, defined on all subintervals I of an interval J, $\Pi_I e^{d\nu}$ exists and equals $e^{\nu(I)}$. In fact, $\Pi_\kappa e^{\Delta\nu} = e^{\nu(I)}$ for any partition κ of I. If I ranges over subintervals of J then the map $\nu \mapsto (I \mapsto \Pi_I e^{d\nu})$ is holomorphic from the supremum norm $\|\nu\|_\infty := \sup_{I\subset J}\|\nu(I)\|$ to itself.*

In the noncommutative case, it is not hard to show that $\Pi_J e^{d\nu}$ is defined in $\mathbb{B}$ at least for $\nu \in \mathcal{W}_1(J,\mathbb{B})$. Note that such a ν is countably additive. But we have:

Proposition 7.22. *Let M_2 be the Banach algebra of 2×2 real matrices with $I := \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$. Then the maps $\nu \mapsto \Pi_{[0,1]}e^{d\nu}$ and $\nu \mapsto \prod_{[0,1]}(I+d\nu)$ on $\mathcal{W}_1([0,1],M_2)$ are not continuous in the p-variation norm for any $p > 2$.*

Proof. Let $B := \left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right)$, $C := \left(\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right)$, and for $0 < \delta < 1/2$ let $X := I + \delta B$ and $Y := I + \delta C$,. Then $B^2 = C^2 = 0$, so $X^{-1} = I - \delta B$, $Y^{-1} = I - \delta C$, $e^{\delta B} = X$ and $e^{\delta C} = Y$. Let $T := YXY^{-1}X^{-1}Y^{-1}X^{-1}YX$. Then straightforward algebra gives $T_{11} = 1 - 2\delta^2 + \delta^4 + \delta^6$, $T_{12} = T_{21} = 2\delta^5 + \delta^7$, and $T_{22} = 1 + 2\delta^2 + 3\delta^4 + 3\delta^6 + \delta^8$. Here T is symmetric (self-adjoint), as can be seen directly from its definition since $X = Y^t$, etc. Thus T can be diagonalized by some rotation. The characteristic equation for its eigenvalues is $\lambda^2 - t\lambda + 1 = 0$ where $t = 2 + 4\delta^4 + 4\delta^6 + \delta^8$, and T has determinant 1 since each factor does. The roots are thus $\lambda = [t \pm (t^2 - 4)^{1/2}]/2$. It follows easily that the larger eigenvalue is $\geq 1 + 2\delta^2 \geq \exp\{\delta^2\}$.

Now for each $n = 1, 2, \cdots$, let κ be the partition of $(0, 1]$ into the $8n$ subintervals $((j - 1)/(8n), j/(8n)]$, $j = 1, \cdots, 8n$. Let the values of ν on the first eight subintervals be successively $\delta C, \delta B, -\delta C, -\delta B, -\delta C, -\delta B, \delta C, \delta B$; and then the cycle repeats n times. If I is one of the $8n$ intervals in κ and $J \subset I$ let $\nu(J) := 8n\lambda(J)\nu(I)$ where λ is Lebesgue measure. The sum of the 8 values in a cycle is 0. It follows by Lemma 4.20 that the p-variation of ν is bounded above by $Kn\delta^p$ for some constant K. Since $p > 2$ we can choose $\delta_n := n^{-2/(2+p)}$ so that $n\delta_n^p \to 0$ but $n\delta_n^2 \to \infty$ as $n \to \infty$. Then $\Pi_{[0,1]}e^{d\nu} = T^n$, via the partition κ and any refinement of it, and $\prod_{(0,1]} \mathbb{I} + d\nu = T^n$ also. Since T^n has an eigenvalue $\geq \exp(n\delta_n^2)$, the conclusion follows. $\square$

In Proposition 7.22, p-variation for $p > 2$ can be replaced by ϕ-variation for any ϕ-function such that $\phi(x) = o(x^2)$ as $x \downarrow 0$.

Comparing product integrals and integral equations There are different possibilities for defining product integrals and integral equations they satisfy, which are actually different in case there are jumps. As we have seen,

Dollard and Friedman [17] defined a product integral Πe^{df} and obtained an unusual integral equation.

We, in [29], defined a product integral $\prod(\mathbb{I} + df)$ and obtained an unusual integral equation in terms of "Left Young" or "Right Young" integrals.

Hildebrandt [51] solved a natural integral equation and found an unusual product integral with factors $(\mathbb{I} + \Delta^+ f)(\mathbb{I} - \Delta^- f)^{-1}$.

Stochastic analysts, some analysts, Gill and Johansen [43], and Gill [42], defined integral equations with left-continuous integrands and right-continuous solutions. Also, results were formulated in terms of interval functions in [43], [42], where the product integral is $\prod(\mathbb{I} + d\mu)$. These formulations, as we have seen in Sections 7.4 and 7.5, seem to fit together well.

Bibliography

[1] P. K. Andersen, Ø. Borgan, R. D. Gill and N. Keiding. *Statistical Models based on Counting Processes.* Springer, New York, 1993.

[2] J. Appell and P. P. Zabrejko. *Nonlinear superposition operators.* Cambridge University Press, 1990.

[3] * R. G. Bartle. Return to the Riemann integral. *Amer. Math. Monthly*, **103** (1996), 625-632.

[4] N. K. Bary. *A treatise on trigonometric series.* Vol. 1. Transl. by M. F. Mullins. MacMillan, New York, 1964.

[5] J. Beirlant and P. Deheuvels. On the approximation of P-P and Q-Q plot processes by Brownian bridges. *Statist. Probab. Letters*, **9** (1990), 241-251.

[6] K. Bichteler. Stochastic integration and L^p-theory of semimartingales. *Ann. Probab.*, **9** (1981), 49-89.

[7] R. M. Blumenthal and R. K. Getoor. Sample functions of stochastic processes with stationary independent increments. *J. Math. and Mech.*, **10** (1961), 493-516.

[8] J. Bretagnolle. p-variation de fonctions aléatoires. In: *Séminaire de Probabilités VI*; Lect. Notes in Math. **258**, 51-71. Springer, Berlin, 1972.

[9] T. Carleman. Über die Fourierkoeffizienten einer stetigen Funktion. *Acta Math.* (Sweden), **41** (1918), 377-384.

[10] L. Carleson. On convergence and growth of partial sums of Fourier series. *Acta Math.* (Sweden) **116** (1966), 135-157.

[11] S. B. Chae. *Holomorphy and calculus in normed spaces.* Dekker, New York, 1985.

[12] M. Csörgő. *Quantile Processes with Statistical Applications.* SIAM, Philadelphia, 1983.

[13] P. J. Daniell. Stieltjes derivatives. *Bull. Amer. Math. Soc.* **26**, 444-448.

[14] C. Dellacherie and P.-A. Meyer. *Probabilities and potential B.* Transl. by J. P. Wilson. North–Holland, Amsterdam, 1982.

[15] C. Doléans–Dade. Quelques applications de la formule de changement de variables pour les semimartingales. *Z. Wahrsch. Verw. Gebiete*, **16** (1970), 181-194.

[16] C. Doléans–Dade and P. A. Meyer. Intégrales stochastiques par rapport aux martingales locales. *Séminaire de Probabilités IV*, Univ. Strasbourg, *Lecture Notes in Math.* (Springer) **124** (1970), 77-107.

[17] J. D. Dollard and C. N. Friedman. *Product Integration with Applications to Differential Equations. Encyclopedia of Math. and its Appl.*, Vol. 10, Addison-Wesley, Reading, Massachusetts, 1979.

[18] J. L. Doob. *Stochastic Processes.* Wiley, New York, 1953.

[19] R. M. Dudley. The size of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, **1** (1967), 290-330.

[20] R. M. Dudley. Sample functions of the Gaussian process. *Ann. Probab.*, **1** (1973), 66-103.

[21] R. M. Dudley. Wiener functionals as Itô integrals. *Ann. Probab.*, **5** (1977), 140-141.
[22] R. M. Dudley. *Real analysis and probability.* 2d printing, Chapman & Hall, New York, 1993.
[23] R. M. Dudley. Fréchet differentiability, p-variation and uniform Donsker classes. *Ann. Probab.*, **20** (1992), 1968-1982.
[24] R. M. Dudley. The order of the remainder in derivatives of composition and inverse operators for p-variation norms. *Ann. Statist.*, **22** (1994), 1-20.
[25] R. M. Dudley. Empirical processes and p-variation. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. L. Yang, eds.) 219-233. Springer, New York, 1997.
[26] R. M. Dudley. *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge, to appear 1999.
[27] R. M. Dudley. Differentiability of the composition and quantile operators for regulated and a.e. continuous functions. Preprint.
[28] R. M. Dudley and R. Norvaiša. A survey on differentiability of six operators in relation to probability and statistics. Preprint.
[29] R. M. Dudley and R. Norvaiša. Product integrals, Young integrals and p-variation. Preprint.
[30] R. M. Dudley and W. Philipp. Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrscheinlichkeitsth. verw. Geb.*, **62** (1983), 509-552.
[31] A. Dvoretzky and C. A. Rogers. Absolute and unconditional convergence in normed linear spaces. *Proc. Nat. Acad. Sci. U.S.A.*, **36** (1950), 192-197.
[32] W. Fernández de la Vega. On almost sure convergence of quadratic Brownian variation. *Ann. Probab.*, **2** (1974), 551-552.
[33] L. T. Fernholz. *von Mises calculus for statistical functionals. Lect. Notes in Statist.*, **19**. Springer, New York, 1983.
[34] H. Föllmer. Calcul d'Itô sans probabilités. *Séminaire de Probabilités XV*, Eds. J. Azéma and M. Yor; *Lect. Notes in Math.* **850** (1981), 143-150.
[35] M. A. Freedman. Operators of p-variation and the evolution representation problem. *Trans. Amer. Math. Soc.*, **279** (1983), 95-112.
[36] M. A. Freedman. Necessary and sufficient conditions for discontinuous evolutions with applications to Stieltjes integral equations. *J. Integral Equations*, **5** (1983), 237-270.
[37] B. Fristedt. Sample functions of stochastic processes with stationary, independent increments. In: *Advances in probability and related topics*, Vol. 3. Dekker, New York, 1974, pp. 241-396.
[38] B. Fristedt and S. J. Taylor. Strong variation for the sample functions of a stable process. *Duke Math. J.*, **40** (1973), 259-278.
[39] F. W. Gehring. A study of α-variation. I. *Trans. Amer. Math. Soc.*, **76** (1954), 420-443.
[40] I. I. Gikhman and A. V. Skorokhod. *Introduction to the theory of random processes.* W.B. Saunders, Philadelphia, 1969.
[41] R. D. Gill. Non– and semi–parametric maximum likelihood estimators and the von Mises method (Part 1). *Scand. J. Statist.*, **16** (1989), 97-128.
[42] R. D. Gill. Lectures on Survival Analysis. In *Ecole d'été de Probabilités de Saint-Flour* (P. Bernard, ed.). *Lecture Notes in Math.*, **1581** (1994), 115-241. Springer, Berlin.
[43] R. D. Gill and S. Johansen. A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.*, **18** (1990), 1501-1555.
[44] V. I. Glivenko. *The Stieltjes integral.* (In Russian) ONTI, Leningrad, 1936.

[45] R. A. Gordon. *The integrals of Lebesgue, Denjoy, Perron, and Henstock.* Amer. Math. Soc., Providence, 1994.

[46] R. A. Gordon. The use of tagged partitions in elementary real analysis. *Amer. Math. Monthly*, **105** (1998), 107-117.

[47] G. H. Hardy, J. E. Littlewood and G. Pólya. *Inequalities.* Cambridge University Press, 2d ed. 1952, repr. 1967.

[48] R. Henstock. *The General Theory of Integration.* Clarendon Press, Oxford, 1991.

[49] T. H. Hildebrandt. On integrals related to and extensions of the Lebesgue integrals. *Bull. Amer. Math. Soc.*, **24** (1917), 113-144 and 177-202.

[50] T. H. Hildebrandt. Definitions of Stieltjes integrals of the Riemann type. *Amer. Math. Monthly*, **45** (1938), 265-278.

[51] T. H. Hildebrandt. On systems of linear differentio–Stieltjes–integral equations. *Illinois J. Math.*, **3** (1959), 352-373.

[52] T. H. Hildebrandt. *Introduction to the theory of integration.* Academic Press, New York, 1963.

[53] C. S. Hönig. *Volterra Stieltjes-Integral Equations.* In: *Functional Differential Equations and Bifurcation*, (A. F. Izé, ed.). Lect. Notes in Math. (Springer), **799** (1980), 173-216.

[54] Yen-Chin Huang. *Empirical distribution function statistics, speed of convergence, and p-variation.* Ph. D. thesis, Massachusetts Institute of Technology, 1994.

[55] Yen-Chin Huang. Speed of convergence of classical empirical processes in p-variation norm. Preprint, 1998.

[56] R. A. Hunt. On the convergence of Fourier series. In: *Orthogonal Expansions and their Continuous Analogues*, ed. D. T. Haimo, Southern Illinois Univ. Press [Carbondale], (1968), 235-255.

[57] K. Itô. Stochastic processes. *Lecture Notes Series*, **16**. Aarhus Universitet, 1969.

[58] J. Jacod and P. Protter. Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** (1998), 267-307.

[59] N. C. Jain and D. Monrad. Gaussian measures in B_p. *Ann. Probab.*, **11** (1983), 46-57.

[60] T. Kawada and N. Kôno. On the variation of Gaussian processes. In: *Proc. Second Japan-USSR Symp. on Probab. Theory*; (G. Maruyama and Yu. V. Prokhorov, eds.). *Lect. Notes Math.* (Springer), **330** (1973), 176-192.

[61] J. Kiefer. Deviations between the sample quantile process and the sample df. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.), 299-319. Cambridge University Press, 1970.

[62] A. N. Kolmogorov. Une série de Fourier-Lebesgue divergente partout. *C. R. Acad. Sci. Paris*, **183** (1926), 1327-1328.

[63] A. N. Kolmogorov. Untersuchungen über den Integralbegriff. *Math. Ann.*, **103** (1930), pp. 654-696. Russian transl. by P. Osvald in *A. N. Kolmogorov. Collected works. Mathematics and mechanics*, Moskow, Nauka, 1985, pp. 96-134.

[64] * J. König. A határozott intégralok elmeletéz. *Mat. és. Természettudmányi értesitö*, 1897, 380-384.

[65] G. Kowalewski. Über den zweiten Mittelwertsatz der Infinitesimalrechnung. *Math. Annalen*, **60** (1905), 151-156.

[66] G. L. Krabbe. Integration with respect to operator-valued functions. *Bull. Amer. Math. Soc.*, **67** (1961), 214-218.

[67] G. L. Krabbe. Integration with respect to operator-valued functions. *Acta Sci. Math.* (Szeged), **22** (1961), 301-319.

[68] J. Kurzweil. Generalized ordinary differential equations and continuous dependence on a parameter. *Czechoslovak Math. J.*, **7** (1957), 418-446.
[69] M. Ledoux and M. Talagrand. *Probability in Banach spaces.* Springer, Berlin, 1991.
[70] R. Leśniewicz and W. Orlicz. On generalized variations (II). *Studia Math.* **45** (1973), 71-109.
[71] P. Lévy. Le mouvement brownien plan. *Amer. J. Math.*, **62** (1940), 487-550.
[72] P. Lévy. Random functions: general theory with special reference to Laplacian random functions. *Univ. of Calif. Publs. in Statist.*, **1** no. 12, 1953, 331-388; repr. in *Oeuvres de Paul Lévy* IV, *Processus Stochastiques*, Gauthier–Villars, Paris, 1980, 187-286.
[73] R. Lipschitz. De explicatione per series trigonometricas instituenda functiorum unius variabilis arbitrariarum ... *J. reine angew. Math.*, **63** (1864), 296-308. French. transl. by P. Montel in *Acta Math.*, **36** (1912-13), 281-295.
[74] E. R. Love. The refinement–Ross–Riemann–Stieltjes (R^3S) integral. In: *Analysis, Geometry and Groups: A Riemann Legacy Volume*, (H. M. Srivastava and Th. M. Rassias, eds.), pp. 289-312; Hadronic Press, Palm Harbor, Florida, 1993.
[75] E. R. Love and L. C. Young. On fractional integration by parts. *Proc. London Math. Soc.*, Ser. 2, **44** (1938), 1-35.
[76] T. Lyons. Differential equations driven by rough signals (I): an extension of an inequality of L. C. Young. *Math. Res. Letters*, **1** (1994), 451-464.
[77] J. S. MacNerney. Integral equations and semigroups. *Illinois J. Math.*, **7** (1963), 148-173.
[78] J. Marcinkiewicz. On a class of functions and their Fourier series. *C. R. Soc. Sci. Varsovie*, **26** (1934), 71-77. Reprinted in: *J. Marcinkiewicz. Collected Papers.* (A. Zygmund, ed.) Państwowe Wydawnictwo Naukowe, Warsaw, 1964.
[79] P. R. Masani. Multiplicative Riemann integration in normed rings. *Trans. Amer. Math. Soc.*, **61** (1947), 147-192.
[80] E. J. McShane. *Unified integration.* Academic Press, Orlando, 1983.
[81] F. A. Medvedev. *Development of the Concept of Integral* (in Russian). Moscow, Nauka, 1974.
[82] P.-A. Meyer. *Probability and Potentials.* Blaisdell, Waltham, MA, 1966.
[83] I. Monroe. On the γ-variation of processes with stationary independent increments. *Ann. Math. Statist.*, **43** (1972), 1213-1220.
[84] I. Monroe. On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.*, **43** (1972), 1293-1311.
[85] E. H. Moore. Definition of limit in general integral analysis. *Proc. Nat. Acad. Sci. U.S.A.*, **1** (1915), 628-632.
[86] E. H. Moore and H. L. Smith. A general theory of limits. *Amer. J. Math.*, **44** (1922), 102-121.
[87] W. Orlicz. Beiträge zur Theorie der Orthogonalentwicklungen (III). *Bull. Acad. Pol. Sci.*, **1932**, 229-238.
[88] R. E. A. C. Paley, N. Wiener and A. Zygmund. Notes on random functions. *Math. Z.*, **37** (1933) 647-668. Reprinted in: *Norbert Wiener. Collected Works with commentaries*, Vol. I (P. Masani, ed.) pp. 536-557; MIT Press, Cambridge, 1976.
[89] Lee Peng–Yee. *Lanzhou Lectures on Henstock integration.* World Scientific, Singapore, 1989.
[90] W. F. Pfeffer. *The Riemann approach to integration.* Cambridge University Press, Cambridge, 1993.

[91] S. Pollard. The Stieltjes' integral and its generalisations. *Quart. J. Pure and Appl. Math.*, **49** (1923), 73-138.

[92] Jinghua Qian. The p-variation of partial sum processes and the empirical process. *Ann. Probab.*, **26** (1998), 1370-1383.

[93] J. A. Reeds III. *On the definition of von Mises functionals.* Ph. D. thesis, Statistics, Harvard University, 1976.

[94] G. F. B. Riemann. Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe. *Abh. Gesell. Wiss. Göttingen*, **13** (1866-67: publ. 1868), math. Kl., 87-132. Reprinted in: *Oeuvres mathématiques de Riemann*, (L. Laugel, ed.), 1968, Paris, 225-279.

[95] K. A. Ross. Another approach to Riemann–Stieltjes integrals. *Amer. Math. Monthly*, **87** (1980), 660-662.

[96] K. A. Ross. *Elementary analysis: The theory of calculus.* Springer, New York, 1980.

[97] T. Runst and W. Sickel. *Sobolev Spaces of Fractional Order, Nemytskij Operators, and Nonlinear Partial Differential Equations.* de Gruyter, Berlin, 1996.

[98] S. Saks. *Theory of the Integral.* 2d ed. English transl. by L. C. Young. Hafner, New York. Repr. corrected Dover, New York, 1964.

[99] R. Salem. La convergence des series de Fourier. *C. R. Acad. Sci. Paris*, **202** (1938), 662-66. Repr. as: Essais sur les séries trigonometriques. Chap. VI, in *Actual. Sci. et Industr.*, **862**. Paris, 1940, and in: *Œuvres Mathématiques*, pp. 156-161. Hermann, Paris, 1967.

[100] Š. Schwabik. On the relation between Young's and Kurzweil's concept of Stieltjes integral. *Časopis Pěst. Mat.*, **98** (1973), 237-251.

[101] Š. Schwabik, M. Tvrdý and O. Vejvoda. *Differential and integral equations.* Reidel, Dordrecht, 1979.

[102] A. V. Skorohod. On a representation of random variables. *Theory Probab. Appl.*, **21** (1976), 628-632.

[103] A. V. Skorohod. On a generalization of the stochastic integral. *Theory Probab. Appl.*, **20** (1975), 219-233.

[104] $*$ T. J. Stieltjes. Recherches sur les fractions continues. *Ann. Fac. Sci. Toulouse*, Sér. 1 **8** (1894), 1-122.

[105] M. Talagrand. Regularity of Gaussian processes. *Acta Math.* (Sweden), **159** (1987), 99-149.

[106] S. J. Taylor. Exact asymptotic estimates of Brownian path variation. *Duke Math.*, **39** (1972), 219-241.

[107] H. Triebel. *Theory of function spaces.* Birkhäuser, Basel, 1983.

[108] A. van der Vaart. Efficiency and Hadamard differentiability. *Scand. J. Statist.*, **18** (1991), 63-75.

[109] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes, with Applications to Statistics.* Springer, New York, 1996.

[110] V. Volterra. Sui fondamenti della teoria delle equazioni differenziali lineari. *Memorie della Società Italiana delle Scienze (detta dei XL)*, serie III, vol. VI, n. 8 (1887). In: "Opere Matematiche" I, 1954, pp. 209-290. Accad. Nazionale dei Lincei, Roma.

[111] A. J. Ward. The Perron–Stieltjes integral. *Math. Z.*, **41** (1936), 578-604.

[112] N. Wiener. Differential space. *J. Math. Phys.* M.I.T. **2** (1923), 131-174. Also in Wiener (1976), I, pp. 455-498.

[113] N. Wiener. The quadratic variation of a function and its Fourier coefficients. *J. Math. and Phys.* (MIT, Cambridge, Mass), **3** (1924), 72-94. Also in Wiener (1976), II, pp. 36-58.

[114] N. Wiener (1976-1986). *Norbert Wiener: Collected Works with Commentaries.* (P. Masani, ed.) 4 vols. M.I.T. Press, Cambridge, Mass.

[115] Quanhua Xu. Espaces d'interpolation réels entre les espaces V_p: propriétés géométriques et applications probabilistes. In: *Séminaire d'Analyse Fonctionnelle* 1985/ 1986/1987, Paris VI-VII. Paris: Publ. Math. Univ. Paris VII **28** 1988, pp. 77-123.

[116] G. C. Young. On infinite derivatives. An Essay. *Quarterly J. Pure Appl. Math.*, **47** (1916), 127-175.

[117] L. C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* (Sweden), **67** (1936), 251-282.

[118] L. C. Young. General inequalities for Stieltjes integrals and the convergence of Fourier series. *Math. Ann.*, **115** (1938), 581-612.

[119] R. C. Young. On Riemann integration with respect to a continuous increment. *Math. Z.*, **29** (1929), 217-233.

[120] W. H. Young. On integration with respect to a function of bounded variation. *Proc. London Math. Soc.* (Ser. 2), **13** (1914), 109-150.

[121] A. Zygmund. *Trigonometric series.* Cambridge University Press, 2d ed. 1959, repr. with corrections and some additions 1968, repr. 1988.

$*$ We learned of these references from secondary sources and have not seen them in the original.