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

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

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

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

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

Let $F{*}G(x) := \nplus_{-\infty}^{\infty} F(x - y)\, dG(y)$ 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 $g\colon X \mapsto Y$ and $f\colon Y \mapsto Z$, 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 $b := \operatorname{ess.sup} G$ and $a := \operatorname{ess.inf} G$. 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. □

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. $\square$

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. $\square$

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

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

Lemma 4.3. *For $1 \leq p < \infty$, $f \in \mathcal{W}_p[a,b]$ 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

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

Define g on $\{h(x): x \in [a,b]\}$ 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 $f: [a,b] \mapsto \mathbb{R}$, $v_p(f) = v_p(f \circ H)$ 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

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

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

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

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

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

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

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

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

Proof. Let $\rho(x) := v_p(f;[a,x])$ for $x \in [a,b]$. The lemma holds if $c \geq \rho(b)$. Suppose $c < \rho(b)$. Let M be the minimal integer such that $cM \geq \rho(b)$ and let $x_0 := a$. For each $m = 1, \ldots, M$, let $x_1^m := \inf\{x \in (x_0,b]: \rho(x) \geq mc \wedge \rho(b)\}$ and let $m_1 := \min\{m = 1, \ldots, M: x_1^m > x_0\}$. If $m_1 = M$ let $x_1 := b$. Then $\rho(a+) \geq Mc$ and $\rho(b-) - \rho(a+) \leq Mc - Mc = 0$. 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 $\rho(x_1-) \leq m_1 c$ and $\rho(x_0+) \geq (m_1 - 1)c$.

Suppose $x_i < b$ is chosen such that $m_1 + \cdots + m_i < M$, $\rho(x_i-) \leq (m_1 + \cdots + m_i)c$ and $\rho(x_{i-1}+) \geq (m_1 + \cdots + m_i - 1)c$. For each $m = m_1 + \cdots + m_i + 1, \ldots, M$, let $x_{i+1}^m := \inf\{x \in (x_i,b]: \rho(x) \geq mc \wedge \rho(b)\}$ and let $m_{i+1} := \min\{m = m_1 + \cdots + m_i + 1, \ldots, M: x_{i+1}^m > x_i\}$. 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 $n \leq M \leq 1 + \rho(b)/c$ and $v_p(f;(x_{i-1},x_i)) \leq \rho(x_i-) - \rho(x_{i-1}+) \leq c$ for each $i = 1, \ldots, n$ 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 $a \le x < y \le b$, 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:

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

and

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

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 $\mathrm{Osc}(f;(x_{i-1},x_i)) \le \epsilon$ 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 $\mathrm{Osc}(f;(y-\delta_y,y)) \le \epsilon$ and $\mathrm{Osc}(f;(y,y+\delta_y)) \le \epsilon$. Also, there exist $\delta_a > 0$ and $\delta_b > 0$ such that $\mathrm{Osc}(f;(a,a+\delta_a)) \le \epsilon$ and $\mathrm{Osc}(f;(b-\delta_b,b)) \le \epsilon$. 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 $|f(x)-f_\epsilon(x)| \le \epsilon$ 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 $|f(x)-f_\epsilon(x)| \le \epsilon$ for each $x \in [a,b]$. Let $x \in (a,b]$ and $x_n \uparrow x$. Then, for all integers n,m, we have

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

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 $\epsilon > 0$,*

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

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

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

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

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

For a sequence $\Lambda := \{\lambda_i\}_{i\ge 1}$ 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 $0 \le x < \infty$, where $1 \le r < \infty$.

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 $\{\lambda_i\}_{i\ge1}$ 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 $\kappa = \{x_i\colon i = 0,1,\ldots,n\} \in PP([a,b])$, 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

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

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 $\kappa = \{x_i\colon i=0,\dots,n\} \in \mathrm{PP}([a,b])$. 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. $\square$

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 $D(f) := \{\xi_1,\xi_2,\dots\}$, $\xi_j := \xi_j(f)$, $j=1,2,\dots$, so that the sequence $\{w_j\colon j\ge 1\}$ is non-increasing, where

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

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* $\lambda = \{x_i\colon i = 0,\dots,n\}$ *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 $\{x_i\colon i = 0,\dots,n\}$ *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 $\lambda = \{z_j\colon j = 0,\dots,m\} \in PP([a,b])$ 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 $\{u_{j-1}, v_j\colon j = 1,\dots,m\}$ be a set of points in (a,b) such that $z_{j-1} < u_{j-1} < v_j < z_j$ for $j = 1,\dots,m$. 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 $j = 1,\dots,m$ 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. □

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 $a = x_0 < x_1 < \cdots < x_n = b$, 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 $c_p = 2^p$. □

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, $j = 1, \ldots, 2k$, 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 $\kappa = \{x_i\colon i = 0, \ldots, n\} \supset \kappa_\epsilon$ and let $f_{\kappa,\epsilon}$ be the step function defined by

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

Let $\{z_j\colon j = 0, \ldots, m\} \in PP([x_{i-1}, x_i])$ with $m > 1$ for some $i = 1, \ldots, n$, 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 $i = 1, \ldots, n$,

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

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

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

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

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

Let I be the set of indices $i = 1, \dots, n$ 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. ☐

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 $i = 1, \dots, n$, where ϕ is the step function defined by*

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

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 $\{x_i \colon i = 0, \dots, n\}$ 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 $\kappa_i = \{y_j\colon j=0,\ldots,m\} \in PP([x_{i-1},x_i])$ 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 $i=1,\ldots,n$. 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 $\kappa=\{x_i\colon i=0,\ldots,n\}$ 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 $\kappa=\{x_i\colon i=0,\ldots,n\}\in PP([a,b])$, and let $\Delta_i g := g(x_i)-g(x_{i-1})$ for $i=1,\ldots,n$ 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

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

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

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

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

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

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

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

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

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

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), $j = 1, \ldots, m$ separated by singletons $a = z_0 < z_1 < \cdots < z_m = b$ and such that

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

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, $j = 0, \ldots, m$. 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 $\kappa' = \{x'_j \colon j = 0, \ldots, k\}$ 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

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

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

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

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 $\{\kappa_m : m \geq 1\}$ be a sequence of partitions $\kappa_m = \{x_i^m : i = 0, \ldots, n(m)\} \in PP([a,b])$ such that the mesh $|\kappa_m| \to 0$, and let $\sigma_m := \{x_i^m : i = 1, \ldots, n(m)\}$. 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 $\lambda = \{z_j : j = 0, \ldots, m\}$ 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 $[z_{j-1}, z_j]$, $j = 1, \ldots, m$. Taking a refinement $\kappa' = \{x_j' : j = 0, \ldots, k\}$ 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 $\kappa = \{x_i : i = 0, \ldots, n\} \in PP([a,b])$ and open intermediate partition $\sigma = \{y_i : i = 1, \ldots, n\}$ of κ, let $\tilde{\kappa} := \kappa \cup \{u_{i-1}, v_i : i = 1, \ldots, n\}$, 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 $i = 1, \ldots, n$, 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 $\lambda = \{z_j : j = 0, \ldots, m\} \in PP([a,b])$ such that

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

For $j = 1, \ldots, m$, 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 $j = 1, \ldots, m$, 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. $\square$

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. □

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):

$$\frac{1}{m}+\sum_{j=1}^m \cos jt = \frac{\sin(m+\frac12)t}{2\sin\frac12 t} \quad\text{and}\quad \sum_{j=1}^m \sin jt = \frac{\cos\frac12 t-\cos(m+\frac12)t}{2\sin\frac12 t},$$

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

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

Similarly, for $k = l$, we get

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

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 $-\infty < c \leq d < +\infty$, $(c,d) := \{x \in \mathbb{R}\colon c < x < d\}$, $[c,d) := \{x \in \mathbb{R}\colon c \leq x < d\}$, $(c,d] := \{x \in \mathbb{R}\colon c < x \leq d\}$ and $[c,d] := \{x \in \mathbb{R}\colon c \leq x \leq d\}$. 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. $\mathfrak{I}_{os}([a,b]) := \{(c,d), \{c\}\colon a \leq c \leq d \leq b\}$ (open intervals and singletons),
2. $\mathfrak{I}([a,b]) := \{[c,d], (c,d], [c,d), (c,d)\colon a \leq c \leq d \leq b\}$ (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 $\{(c,d]\colon a \leq c \leq d \leq b\}$.

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 $\mu(A) \neq 0$ ($\mu(A)$ 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 $\mathfrak{I}([a,b])$*;
(*c*) *there exists $h\in\mathcal{R}[a,b]$ such that $h(b)=0$ and $\mu_h=\mu$ on $\mathfrak{I}([a,b])$*.

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 $\mathrm{card}\{x\in D\colon |h_x-\mu([a,x))|\wedge|h_x-\mu([a,x])|>\epsilon\}<\infty$ 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 $\{u,v_n\colon n\ge1\}\subset[a,b]$ 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 $\mu(A_n)=h(u-)-h(v_n+)\to0$ as $n\to\infty$ in the first case and $\mu(A_n)=h(v_n-)-h(u+)\to0$ as $n\to\infty$ in the second case. Similarly, it follows that $\mathrm{card}\{x\in[a,b]\colon|\mu(\{x\})|>\epsilon\}<\infty$ 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

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

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*) $\{h \in \mathcal{R}[a, b]$: $h(a) = 0$ *and* h *is right-continuous on* $(a, b)\}$;
(*c*) $\{h \in \mathcal{R}[a, b]$: $h(a) = 0$ *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

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

where a union over the empty set of indices is defined as the empty set, and as before 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

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

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 $\{A_i\colon i = 1, \ldots, n\}$, 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 $T^1:=\{z\colon |z|=1\}=\{e^{i\theta}\colon 0\le\theta<2\pi\}$ 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 $\mathbb{Z}=\{0,\pm1,\pm2,\ldots\}$, 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

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

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 $a_n=(c_n+c_{-n})/2$, $b_n=(c_n-c_{-n})/(2i)$, $n\ge1$, 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 $\ell^q=\{\{d_n\}\colon\sum_n|d_n|^q<\infty\}$, 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 $\{d_n\colon n\in\mathbb{Z}\}\notin\ell^2$ one can have $\sum_n|c_nd_n|=+\infty$. 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 $0<\alpha\le1$, 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 $v_q(g_n)=2n\cdot2^q$ 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]).