STUDIA MATHEMATICA 149 (3) (2002)

Chain rules and p-variation

by

R. NORVAIŠA (Vilnius)

Abstract. The main result is a Young–Stieltjes integral representation of the composition $\phi \circ f$ of two functions f and ϕ such that for some $\alpha \in (0,1]$, ϕ has a derivative satisfying a Lipschitz condition of order α, and f has bounded p-variation for some $p < 1+\alpha$. If given $\alpha \in (0,1]$, the p-variation of f is bounded for some $p < 2+\alpha$, and ϕ has a second derivative satisfying a Lipschitz condition of order α, then a similar result holds with the Young–Stieltjes integral replaced by its extension.

1. Introduction and results. In this paper an integral representation of the composition $\phi \circ f$ of a smooth function ϕ and a rough function f is proved. This representation is analogous to the Itô formula for the composition of a smooth function and a semimartingale. Itô's formula is based on the stochastic integral with respect to a semimartingale. Our main result (Theorem 1.1) is proved using the Young–Stieltjes integral (see Subsection 2.2 for the definition). The integral representation also holds for two other extensions of the Riemann–Stieltjes integral: for the central Young integral, which was suggested by L. C. Young [39], and further modified by Dudley [5], as well as for the variant of the Perron–Stieltjes integral defined by Ward [38], and further developed by Kurzweil [17] and Henstock [12]. The p-variation of f is used to control its roughness. The main result holds for functions f having bounded p-variation with $0 < p < 2$. L. C. Young [39], [40] proved the existence of the Young–Stieltjes integral when the integrand and integrator have bounded p- and q-variations, respectively, with $p, q > 0$ and $p^{-1} + q^{-1} > 1$. This is the best possible condition in terms of p-variation. Theorem 1.1 extends the L. C. Young existence result to integrals with a special form of the integrand when both the integrand and integrator are in a suitable subspace $\mathcal{W}_2^*$ of the space $\mathcal{W}_2$ of all functions of bounded 2-variation. This result is also best possible. For functions having bounded p-variation for some $p < 3$, a similar integral representation is proved (Theorem 1.4) defining an integral analogous to the mean Stieltjes

2000 *Mathematics Subject Classification*: Primary 26A42; Secondary 26A45, 60H05.

integral originating from Smith [35]. The new integral, called the symmetric Young–Stieltjes integral, extends the Young–Stieltjes integral for functions with values at points of discontinuity equal to the average of the left and right limits, and it is equivalent to the Young–Stieltjes integral if in addition the functions are in $\mathcal{W}_2^*$.

1.1. *Preliminary notation.* To be more specific we need some notation. A *partition* of a closed interval $[a,b]$ of real numbers is an ordered set of points $\{x_i : i=0,\dots,n\}$ in $[a,b]$ such that $a=x_0<x_1<\dots<x_n=b$. Let $P([a,b])$ be the set of all partitions of $[a,b]$. For $0<p<\infty$ and a real-valued function f on $[a,b]$, the *p-variation* of f is defined by

$$v_p(f)=v_p(f;[a,b]):=\sup\{s_p(f;\kappa):\kappa\in P([a,b])\},$$

where $s_p(f;\kappa):=\sum_{i=1}^n|f(x_i)-f(x_{i-1})|^p$ for $\kappa=\{x_i : i=0,\dots,n\}$. The set of all functions f of bounded p-variation will be denoted by $\mathcal{W}_p=\mathcal{W}_p([a,b])$. A function f of bounded p-variation is *regulated*; that is, the left-limit $f(x-):=\lim_{y\uparrow x}f(y)$ exists for each $a<x\le b$, and the right-limit $f(x+):=\lim_{y\downarrow x}f(y)$ for each $a\le x<b$. The class of all regulated functions on $[a,b]$ will be denoted by $\mathcal{R}=\mathcal{R}([a,b])$. For $f\in\mathcal{R}([a,b])$ and for any function g on $[a,b]$, define $g\Delta^-f$ on $(a,b]$ and $g\Delta^+f$ on $[a,b)$ respectively by

$$[g\Delta^-f](y):=g(y)[f(y)-f(y-)],\qquad [g\Delta^+f](x):=g(x)[f(x+)-f(x)],$$

whenever $a\le x<y\le b$. Clearly, Δ^-f and Δ^+f are defined as $g\Delta^-f$ and $g\Delta^+f$, respectively, with $g\equiv 1$.

To deal with the convergence with respect to refinements of partitions we use the theory of limits based on the notion of a directed function as defined by McShane [26, p. 11] (or p. 33 in [27]). The *direction of partitions* of $[a,b]$ is the family $\mathfrak{P}$ of all sets $P(\lambda):=\{\kappa\in P([a,b]):\kappa\supset\lambda\}$, $\lambda\in P([a,b])$. Then by Theorem 4.7 of Dudley and Norvaiša [6, Part II], for $1<p<\infty$ and $f\in\mathcal{W}_p$, we have

$$\liminf_{\kappa,\mathfrak{P}} s_p(f;\kappa):=\sup_{\lambda}\inf_{\kappa\in P(\lambda)}s_p(f;\kappa)=\sum_{(a,b]}|\Delta^-f|^p+\sum_{[a,b)}|\Delta^+f|^p, \tag{1.1}$$

where the two sums on the right side are unconditional sums. Here and throughout the paper an *unconditional sum* is the limit of directed finite partial sums, which exists if and only if the sum converges absolutely. Thus each of the two sums in (1.1) is equal to an absolutely convergent series with at most countably many terms. For $f\in\mathcal{W}_p$, $1<p<\infty$, we write $f\in\mathcal{W}_p^*=\mathcal{W}_p^*([a,b])$ if

$$\liminf_{\kappa,\mathfrak{P}} s_p(f;\kappa)=\limsup_{\kappa,\mathfrak{P}} s_p(f;\kappa):=\inf_{\lambda}\sup_{\kappa\in P(\lambda)}s_p(f;\kappa). \tag{1.2}$$

By Theorem 11.4 of McShane and Botts [27, p. 55], $f\in\mathcal{W}_p^*([a,b])$ if and only if the directed function $(s_p(f;\cdot),\mathfrak{P})$ has a limit $\lim_{\kappa,\mathfrak{P}}s_p(f;\kappa)$, which is then equal to the right side of (1.1). It is clear that $\mathcal{W}_q \subset \mathcal{W}_p^*$ for any $1 \le q < p < \infty$.

1.2. *Results.* Now we are ready to formulate the main results. For a real-valued function ϕ defined on the range of f, the *composition* $\phi \circ f$ is defined on $[a,b]$ by $(\phi \circ f)(x) := \phi(f(x))$, $x \in [a,b]$. The next statement is the Young–Stieltjes integral representation of the composition $\phi \circ f$ which is a special case of Theorem 4.2 proved below.

THEOREM 1.1. *For* $0 < \alpha \le 1$, *let* $f \in \mathcal{W}_{1+\alpha}^*([a,b])$ *and let* ϕ *be differentiable with derivative* ϕ' *satisfying a Lipschitz condition of order* α. *The Young–Stieltjes integral* $(\mathrm{YS})\int_a^b (\phi' \circ f)\,df$ *is defined*, *and its value is determined by the relation*

$$(1.3)\qquad (\phi \circ f)(b) - (\phi \circ f)(a) = (\mathrm{YS})\int_a^b (\phi' \circ f)\,df + \sum_{(a,b]} \{\Delta^-(\phi \circ f) - (\phi' \circ f)\Delta^- f\} + \sum_{[a,b)} \{\Delta^+(\phi \circ f) - (\phi' \circ f)\Delta^+ f\},$$

where the two sums are unconditional.

L. C. Young [39], [40] proved that $(\mathrm{YS})\int g\,df$ exists provided $f \in \mathcal{W}_p$, $g \in \mathcal{W}_q$ and $p^{-1} + q^{-1} > 1$, $p, q > 0$. He also showed that the assumption $p^{-1} + q^{-1} > 1$ cannot be replaced in general by the weaker assumption $p^{-1} + q^{-1} = 1$. The same is true for $f \in \mathcal{W}_p^*$ and $g \in \mathcal{W}_q^*$ with $p^{-1} + q^{-1} = 1$. This follows from Theorem 4.11 of Leśniewicz and Orlicz [20] who further refined L. C. Young's result. Notice that in Theorem 1.1, $\phi' \circ f \in \mathcal{W}_q$ for $q = p/(p-1)$, so that $p^{-1} + q^{-1} = 1$ in this case. Therefore the existence of the integral in (1.3) does not follow from L. C. Young's result. The assumption on the p-variation in Theorem 1.1 cannot be improved further. By Proposition 4.4, if $(\mathrm{YS})\int f\,df$ exists and satisfies (1.3) with $\phi(u) = u^2/2$ then $f \in \mathcal{W}_2^*$.

By Theorems 2.5 and 2.7, the statement of Theorem 1.1 also holds for the central Young and Henstock–Kurzweil integrals in place of the Young–Stieltjes integral. On the other hand, by Proposition 2.1, if $(\mathrm{YS})\int_a^b g\,df$ exists and f is continuous then the Riemann–Stieltjes integral $(\mathrm{RS})\int_a^b g\,df$ also exists, and has the same value. Therefore we have:

COROLLARY 1.2. *If in addition to the assumptions of Theorem* 1.1, f *is continuous then the Riemann–Stieltjes integral* $(\mathrm{RS})\int_a^b (\phi' \circ f)\,df$ *is defined*, *and*

$$(1.4)\qquad (\phi \circ f)(b) - (\phi \circ f)(a) = (\mathrm{RS})\int_a^b (\phi' \circ f)\,df.$$

The next example shows that the smoothness assumption on ϕ is essential in our theorems.

EXAMPLE 1.3. Let $\phi(y) := \max(0, y)$ for $y \in [-1, 1]$. Then ϕ is a convex function with the left derivative $D^-\phi = \mathbb{1}_{(0,1]}$. Let $f(x) := x \sin(\pi/x)$ for $x \in (0, 1]$ and $f(0) := 0$. Then f is a continuous function of bounded p-variation for each $p > 1$ and has unbounded 1-variation. We show that $g := (D^-\phi) \circ f$ is not Riemann–Stieltjes integrable with respect to f on $[0, 1]$. Let $\kappa = \{x_i : i = 0, \dots, n\}$ be a partition of $[0, 1]$. Take the smallest integer m such that $1/(2m) < x_1$, and let $\lambda_m := \kappa \cup \{u_i := (2i+1/2)^{-1}, v_i := (2i)^{-1} : i = m, \dots, m^2\}$. To estimate the difference between two Riemann–Stieltjes sums based on the same partition λ_m, for the terms corresponding to the intervals $[u_i, v_i]$, $i = m, \dots, m^2$, we evaluate g at the left endpoint for the first sum and at the right endpoint for the second sum, and for all other terms we evaluate g at the same point for the two sums. Then the difference between these two Riemann–Stieltjes sums is equal to

$$\sum_{i=m}^{m^2} [g(v_i) - g(u_i)][f(v_i) - f(u_i)] = \sum_{i=m}^{m^2} u_i > (\log m)/3.$$

Thus the two Riemann–Stieltjes sums for $\int_0^1 g\, df$, both based on refinements of κ, differ by an arbitrarily large amount. Therefore the integral on the right side of (1.4) does not exist for this example.

To extend formula (1.3) to functions $f \in \mathcal{W}_p$ with $p < 3$ we extend the YS integral. The new integral (Definition 5.4), called the symmetric Young–Stieltjes integral, or the SYS integral, coincides with the YS integral in the cases described by Theorem 5.6 and Proposition 5.7. The following statement is a special case of Theorem 5.8.

THEOREM 1.4. *For $0 < \alpha \leq 1$, let $f \in \mathcal{W}^*_{2+\alpha}([a, b])$ and let ϕ be twice differentiable with the second derivative satisfying a Lipschitz condition of order α. The symmetric Young–Stieltjes integral* (SYS) $\int_a^b (\phi' \circ f)\, df$ *is defined, and its value is determined by the relation*

$$(1.5)\qquad \begin{aligned}(\phi \circ f)(b) - (\phi \circ f)(a) &= (\mathrm{SYS})\int_a^b (\phi' \circ f)\, df \\ &\quad + \sum_{(a,b]} \Big\{\Delta^-(\phi \circ f) - (\phi' \circ f)\Delta^- f + \frac{\Delta^-\phi' \circ f}{2}\Delta^- f\Big\} \\ &\quad + \sum_{[a,b)} \Big\{\Delta^+(\phi \circ f) - (\phi' \circ f)\Delta^+ f - \frac{\Delta^+\phi' \circ f}{2}\Delta^+ f\Big\},\end{aligned}$$

where the two sums are unconditional.

If the values of f and $g := \phi' \circ f$ at points of discontinuity satisfy the conditions (5.4), then under the conditions of Theorem 1.1, the conclusion of Theorem 1.4 holds and the formula (1.5) coincides with (1.3). In this sense Theorem 1.4 extends Theorem 1.1.

We call (1.3)–(1.5) chain rule formulas for reasons discussed in Section 6. These formulas are related to the Itô formula which plays an important role in the stochastic calculus based on the stochastic integral with respect to a semimartingale. Several modifications of the stochastic integral have been developed which allow integration with respect to sample functions of continuous stochastic processes with zero quadratic variation (Föllmer [10]), or with respect to sample functions of a fractional Brownian motion B_H with the Hurst exponent $H \in (1/2, 1)$ (Ciesielski, Kerkyacharian and Roynette [3], Zähle [43]). We notice that the results of the present paper apply to sample functions of B_H when the Hurst exponent $H \in (1/3, 1)$, which includes the case of a Brownian motion: $H = 1/2$. Further related comments are given at the end of Section 5 below.

The results of the present paper also apply to discontinuous processes. However, the left Young integrals and the corresponding chain rule formula are better suited to solving integral equations with respect to discontinuous processes (see [29]). We continue with a reminder of several extended Riemann–Stieltjes integrals in the next section. Sections 3, 4 and 5 are devoted to a chain rule formula for the composition $\phi \circ f$ in three separate cases according as the function f is of bounded p-variation with $p = 1$, $1 < p < 2$ and $2 \leq p < 3$, respectively. We finish with a discussion of a relation of the formula (1.3) to a chain rule in the context of classical analysis (Section 6).

1.3. *Notation.* The following notation complements the preliminary notation in Subsection 1.1 and will be used throughout the paper. Let $[a, b]$ be a closed interval, and let $\kappa = \{x_i : i = 0, \ldots, n\}$ be a partition of $[a, b]$. For $i = 1, \ldots, n$, a point $y_i \in [x_{i-1}, x_i]$ attached to the subinterval $[x_{i-1}, x_i]$ is called a *tag*, and the set $\tau = \tau(\kappa) = \{([x_{i-1}, x_i], y_i) : i = 1, \ldots, n\}$ is called a *tagged partition* of $[a, b]$ associated to κ. The set of all tagged partitions of $[a, b]$ will be denoted by $TP([a, b])$. For $\lambda \in P([a, b])$, let $TP(\lambda)$ be the set of all tagged partitions which are *refinements* of λ, that is, $\tau \in TP(\lambda)$ if and only if $\tau = \tau(\kappa) = \{([x_{i-1}, x_i], y_i) : i = 1, \ldots, n\}$ and $\lambda \subset \kappa = \{x_i : i = 0, \ldots, n\}$. Then the family $\{TP(\lambda) : \lambda \in P([a, b])\}$ is a direction in the sense of McShane [26, p. 10], denoted by $\mathfrak{R}$. If a tagged partition $\tau = \{([x_{i-1}, x_i], y_i) : i = 1, \ldots, n\}$ is such that $x_{i-1} < y_i < x_i$ for $i = 1, \ldots, n$, then τ will be called a *Young tagged partition* of $[a, b]$ and will be denoted by $\{((x_{i-1}, x_i), y_i) : i = 1, \ldots, n\}$. Similarly we define the set of all Young tagged partitions which are refinements of a given partition, and the resulting direction will be denoted by the same letter $\mathfrak{R}$.

For a tagged partition $\tau = \tau(\kappa)$ associated to $\kappa = \{x_i : i = 0, \ldots, n\}$, $|\kappa| := \max_{1\le i\le n}(x_i - x_{i-1})$ is called the *mesh* of τ. Then the family $\mathfrak{M}$ of all sets $TP(\delta)$, $\delta > 0$, such that $\tau = \tau(\kappa) \in TP(\delta)$ if $|\kappa| < \delta$, is another example of a direction in the sense of McShane [26, p. 10].

Let f be a regulated function on $[a, b]$. In addition to the notation $\Delta^- f$ and $\Delta^+ f$ defined in Subsection 1.1, for $x \in (a, b)$, let $\Delta^{\pm} f(x) := \Delta^+ f(x) + \Delta^- f(x) = f(x+) - f(x-)$, and for $a \le u < v \le b$, let

$$\Delta f(u, v) := f(v-) - f(u+), \quad \Delta f(u, v] := f(v) - f(u+),$$
$$\Delta f[u, v) := f(v-) - f(u), \quad \Delta f[u, v] := f(v) - f(u).$$

If g is another regulated function on $[a, b]$, and if E is a subinterval of $[a, b]$ open or closed at either end, then let $[\Delta f \Delta g]E := [\Delta f E][\Delta g E]$. To $f \in \mathcal{R}([a, b])$, we associate the left-continuous function $f_- = f_-^{(a)}$ on $[a, b]$ defined by $f_-^{(a)}(x) := f(x-)$ for $x \in (a, b]$ and $f_-^{(a)}(a) := f(a)$, and the right-continuous function $f_+ = f_+^{(b)}$ on $[a, b]$ defined by $f_+^{(b)}(x) := f(x+)$ for $x \in [a, b)$ and $f_+^{(b)}(b) := f(b)$. For an arbitrary function h on $[a, b]$ and $E \subset [a, b]$, let $\|h\|_\infty := \sup\{|h(x)| : x \in [a, b]\}$ and $\mathrm{Osc}(h; E) := \sup\{|h(x) - h(y)| : x, y \in E\}$.

Acknowledgements. The author wishes to thank R. M. Dudley, S. Leader and the anonymous referee for remarks on the paper which helped to improve the presentation and uncovered myriad mistakes.

2. Extended Riemann–Stieltjes integrals. In this section we briefly recall several extensions of the Riemann–Stieltjes integral which are used in the composition representation formulas. A more detailed discussion of these integrals can be found in Dudley and Norvaiša [7].

2.1. *The Moore–Pollard–Stieltjes integral.* Let g and f be real-valued functions defined on $[a, b]$. For a tagged partition $\tau = \{([x_{i-1}, x_i], y_i) : i = 1, \ldots, n\} \in TP([a, b])$, the sum

$$S_{\mathrm{RS}}(\tau) := S_{\mathrm{RS}}(g, f; \tau) := \sum_{i=1}^{n} g(y_i)[f(x_i) - f(x_{i-1})]$$

is called the *Riemann–Stieltjes sum* based on τ. The *Moore–Pollard–Stieltjes integral*, or the *MPS integral*, (MPS) $\int_a^b g\,df$, is defined to be the limit of the directed function $(S_{\mathrm{RS}}(g, f; \cdot), \mathfrak{R})$ if it exists. That is, let

$$\text{(MPS)}\int_a^b g\,df := \lim_{\tau,\mathfrak{R}} S_{\mathrm{RS}}(g, f; \tau) \tag{2.1}$$

provided there exists a number I such that for each $\varepsilon > 0$ there exists $\lambda \in P([a, b])$ such that $|I - S_{\mathrm{RS}}(\tau)| < \varepsilon$ for each refinement τ of λ. The

Riemann–Stieltjes integral (RS) $\int_a^b g\,df$ is defined similarly except that the direction $\mathfrak{R}$ is replaced by the direction $\mathfrak{M}$ based on the mesh of a tagged partition. Since $\mathfrak{R}$ is a subdirection of $\mathfrak{M}$, the MPS integral extends the RS integral (see p. 24 of McShane [26]). For bounded functions g and f, existence of the two integrals is equivalent if and only if g and f have no common discontinuities on $[a,b]$ (e.g. Theorem II.10.9 of Hildebrandt [14]).

2.2. *The Young–Stieltjes integral.* If the MPS integral (2.1) is defined then g and f cannot have a common discontinuity on the same side at the same point of $[a,b]$. In the case when there is such a discontinuity, an integral can be defined in the sense suggested by W. H. Young [42] provided the integrator is a regulated function. The Riemann–Stieltjes sum depends on increments of the integrator at consecutive points of a partition $\kappa = \{x_i : i = 0,\dots,n\}$ of $[a,b]$, so that S_{RS} can be viewed as the sum based on the partition into the collection of adjacent closed intervals $[x_0,x_1],[x_1,x_2],\dots,[x_{n-1},x_n]$. The idea of W. H. Young is to use a sum based on the partition $\{x_0\},(x_0,x_1),\{x_1\},\dots,(x_{n-1},x_n),\{x_n\}$ formed by alternating singletons and open intervals instead. The corresponding tagged partition $(\{x_0\},x_0),((x_0,x_1),y_1),(\{x_1\},x_1),\dots,((x_{n-1},x_n),y_n),(\{x_n\},x_n)$ with tags $x_{i-1} < y_i < x_i$, $i = 1,\dots,n$, is called a *Young tagged partition* $\{((x_{i-1},x_i),y_i) : i = 1,\dots,n\}$ (the singletons $\{x_i\}$ with their uniquely determined tags x_i are omitted from the notation). This idea looks more natural when used to define an integral with respect to an interval function as in Kolmogorov [16]. However the Kolmogorov integral is an *additive interval function* and in this sense the resulting theory (see Dudley and Norvaiša [8]) is different from the Riemann–Stieltjes type integration theory adopted in the present paper. Let f be a regulated function on $[a,b]$, and let g be an arbitrary function on $[a,b]$. For a Young tagged partition $\tau = \{((x_{i-1},x_i),y_i) : i = 1,\dots,n\}$ of $[a,b]$, the *Young–Stieltjes sum* $S_{\mathrm{YS}}(\tau) = S_{\mathrm{YS}}(g,f;\tau)$ based on τ is defined by

$$(2.2)\qquad S_{\mathrm{YS}}(\tau) := \sum_{i=1}^{n}\{[g\Delta^{+}f](x_{i-1}) + g(y_i)\Delta f(x_{i-1},x_i) + [g\Delta^{-}f](x_i)\}.$$

Then the *Young–Stieltjes integral*, or the *YS integral*, (YS) $\int_a^b g\,df$, is defined to be the limit (if it exists) of the directed function $(S_{\mathrm{YS}}(g,f;\cdot),\mathfrak{R})$. That is, let

$$(2.3)\qquad (\mathrm{YS})\int_a^b g\,df := \lim_{\tau,\mathfrak{R}} S_{\mathrm{YS}}(g,f;\tau)$$

provided the limit exists. The YS integral extends the MPS integral. Indeed, for each Young tagged partition τ as above, if we take points $\{u_{i-1},v_i : i = 1,\dots,n\}$ so that $u_{i-1}\downarrow x_{i-1}$ and $v_i\uparrow x_i$ for each $i = 1,\dots,n$, then the

Riemann–Stieltjes sums $S_{\mathrm{RS}}(\tau')$ based on the tagged partitions

$$\tau' = \{([x_{i-1}, u_{i-1}], x_{i-1}), ([u_{i-1}, v_i], y_i), ([v_i, x_i], x_i) : i = 1, \ldots, n\}$$

approximate the Young sum $S_{\mathrm{YS}}(\tau)$ arbitrarily closely. The following statement yields that the YS integral in the chain rule formula (1.3) exists in the Riemann–Stieltjes sense with the same value provided f is continuous, that is, (1.4) holds.

Proposition 2.1. *Let g and f be respectively bounded and continuous functions on $[a, b]$. If* $(\mathrm{YS}) \int_a^b g\,df$ *exists then* $(\mathrm{RS}) \int_a^b g\,df$ *also does, and the two integrals are equal.*

Proof. By Theorem II.10.9 of Hildebrandt [14] already cited above, it is enough to prove that $(\mathrm{MPS}) \int_a^b g\,df$ exists with the same value as $(\mathrm{YS}) \int_a^b g\,df$. Let $\varepsilon > 0$. Since f is continuous there exists a partition $\lambda = \{z_j : j = 0, \ldots, m\}$ of $[a, b]$ such that

$$\Big| \sum_{i=1}^{n} g(y_i)[f(x_i) - f(x_{i-1})] - (\mathrm{YS}) \int_a^b g\,df \Big| < \varepsilon \tag{2.4}$$

for any Young tagged partition $\{((x_{i-1}, x_i), y_i) : i = 1, \ldots, n\}$ provided $\lambda \subset \{x_i : i = 0, \ldots, n\}$. Thus we have to extend (2.4) to the case when the tags y_i can be taken to be the endpoints of (x_{i-1}, x_i) for any i. Again since f is continuous, one can choose a set $\{v_{j-1}, u_j : j = 1, \ldots, m\}$ of points in (a, b) so that $z_{j-1} < v_{j-1} < u_j < z_j$ for $j = 1, \ldots, m$, and

$$\max_{1 \le j \le m} \{\mathrm{Osc}(f; [z_{j-1}, v_{j-1}]) \vee \mathrm{Osc}(f; [u_j, z_j])\} < \varepsilon/(2m\|g\|_\infty). \tag{2.5}$$

Let $\eta := \lambda \cup \{(z_{j-1} + v_{j-1})/2, v_{j-1}, u_j, (u_j + z_j)/2 : j = 1, \ldots, m\}$, and let $\tau = \{([x_{i-1}, x_i], y_i) : i = 1, \ldots, n\}$ be a tagged partition such that $\eta \subset \{x_i : i = 0, \ldots, n\}$. We will show that (2.4) holds for τ with ε replaced by 3ε. Since $\lambda \subset \eta \subset \{x_i : i = 0, \ldots, n\}$, for each $j = 0, \ldots, m$, there is an index $i(j) \in \{0, 1, \ldots, n\}$ such that $x_{i(j)} = z_j$. By joining adjacent intervals if necessary, we can assume that for each $i \neq i(j)$, $j = 0, \ldots, m$, $y_{i+1} > x_i$ if $y_i = x_i$ and $y_{i-1} < x_{i-1}$ if $y_i = x_{i-1}$. Since $\{(z_{j-1} + v_{j-1})/2, (z_j + v_j)/2 : j = 1, \ldots, m\} \subset \{x_i : i = 0, \ldots, n\}$, we thus have

$$z_{j-1} < x_{i(j-1)+1} < v_{j-1} < u_j < x_{i(j)-1} < z_j, \qquad j = 1, \ldots, m. \tag{2.6}$$

Define another partition $\kappa' = \{x_i' : i = 0, \ldots, n\}$ of $[a, b]$ so that $x_{i(j)}' := x_{i(j)}$ for $j = 0, \ldots, m$, $x_i' := x_i$ if $y_i \in (x_{i-1}, x_i)$, $x_i' \in (x_i, y_{i+1})$ if $y_i = x_i$ and $x_i' \in (y_i, x_i)$ if $y_{i+1} = x_i$. Let $I := \{i(j-1) + 1, i(j) : j = 1, \ldots, m\}$ and $J := \{0, \ldots, n\} \setminus I$. Since $x_{i(j-1)+1}' \le v_{j-1}$ and $u_j \le x_{i(j)-1}'$ for $j = 1, \ldots, m$, by (2.5) and (2.6) the absolute value of the sum

$$\sum_{i\in I} g(y_i)[f(x_i)-f(x_{i-1})]-\sum_{i\in I} g(y_i)[f(x_i')-f(x_{i-1}')]$$
$$=\sum_{j=0}^{m-1} g(y_{i(j)+1})[f(x_{i(j)+1})-f(x'_{i(j)+1})]$$
$$+\sum_{j=1}^{m} g(y_{i(j)})[f(x'_{i(j)-1})-f(x_{i(j)-1})]$$

is less than ε. For all $i\notin\{i(j):j=0,\dots,m\}$, taking x_i' close enough to x_i, which is possible, we have

$$\Big|\sum_{i\in J} g(y_i)[f(x_i)-f(x_{i-1})]-\sum_{i\in J} g(y_i)[f(x_i')-f(x_{i-1}')]\Big|<\varepsilon.$$

Since κ' is a refinement of λ, (2.4) holds with all x_i replaced by x_i'. Therefore (2.4) holds with 3ε instead of ε for any tagged partition $\tau=\{([x_{i-1},x_i],y_i): i=1,\dots,n\}$ such that $\eta\subset\{x_i:i=0,\dots,n\}$. ■

2.3. *The* Y_1, Y_2 *and* CY *integrals.* In some cases it is more useful to work with an integral defined by means of left- and/or right-continuous modifications of the integrand and integrator. The following definition originated from the work of L. C. Young [39]. The present form is due to Dudley [5].

Definition 2.2. Let $g,f\in\mathcal{R}([a,b])$, and let $a\le c\le d\le b$. Define the Y_1 *integral* on $[c,d]$ by

$$(\mathrm{Y}_1)\int_c^d g\,df:=(\mathrm{MPS})\int_c^d g_+^{(d)}\,df_-^{(c)}-\sum_{[c,d)}\Delta^+g[f_+-f_-^{(c)}]+[g\Delta^-f](d)$$

provided the MPS integral exists, the sum converges unconditionally and $c<d$. Similarly, define the Y_2 *integral* on $[c,d]$ by

$$(\mathrm{Y}_2)\int_c^d g\,df:=(\mathrm{MPS})\int_c^d g_-^{(c)}\,df_+^{(d)}+[g\Delta^+f](c)+\sum_{(c,d]}\Delta^-g[f_+^{(d)}-f_-]$$

provided the MPS integral exists, the sum converges unconditionally and $c<d$. If $c=d$ both integrals are defined to be 0.

The Y_1 and Y_2 integrals satisfy standard properties of integrals, such as linearity and additivity over adjacent intervals. The next statement follows from Theorems 3.6 and 3.7 of Dudley and Norvaiša [6, Part II] (see also Theorem 6.19 of Dudley and Norvaiša [7] for a simpler proof).

THEOREM 2.3. *For $g, f \in \mathcal{R}([a,b])$, if either of the two integrals $(\mathrm{Y}_1)\int_a^b g\,df$ and $(\mathrm{Y}_2)\int_a^b g\,df$ exists then so does the other, and the two integrals are equal.*

The preceding theorem justifies the following definition. Under the assumptions of Definition 2.2 define the *central Young integral*, or the *CY integral*, on $[c,d]$ by

$$(\mathrm{CY})\int_c^d g\,df := (\mathrm{Y}_1)\int_c^d g\,df = (\mathrm{Y}_2)\int_c^d g\,df$$

if either the Y_1 or the Y_2 integral exists on $[c,d]$. The CY integral extends the YS integral on the class of regulated functions. Namely, the following statement is a special case of Proposition 3.17 of Dudley and Norvaiša [6, Part II] (see also Theorem 6.20 of Dudley and Norvaiša [7] for a different proof).

THEOREM 2.4. *For $g, f \in \mathcal{R}([a,b])$, if $(\mathrm{YS})\int_a^b g\,df$ exists then so does $(\mathrm{CY})\int_a^b g\,df$, and the two integrals are equal.*

By Proposition 3.16 of Dudley and Norvaiša [6, Part II], there are regulated functions g and f such that $(\mathrm{CY})\int_a^b g\,df$ exists while $(\mathrm{YS})\int_a^b g\,df$ does not. However, under the conditions stated in the following statement, both integrals coincide.

THEOREM 2.5. *For $1/p + 1/p' \geq 1$ with $1 < p < \infty$, let $f \in \mathcal{W}_p^*([a,b])$ and $g \in \mathcal{W}_{p'}([a,b])$. If either of the two integrals $(\mathrm{CY})\int_a^b g\,df$ and $(\mathrm{YS})\int_a^b g\,df$ exists then so does the other, and the two integrals are equal.*

Theorem 2.5 is a special case of Corollary 3.20 of Dudley and Norvaiša [6, Part II]. By the preceding statement, Theorem 1.1 also holds with the YS integral replaced by the CY integral. In addition to the CY integral one can define the left and right Young integrals which make solutions of corresponding linear integral equations driven by discontinuous functions particularly simple (see Subsection 5.4 in [6, Part II]). The chain rule formulas in this case are similar to but different from Theorem 1.1 (see [29]).

2.4. *The Ward–Perron–Stieltjes and Henstock–Kurzweil integrals.* Ward [38] defined a Perron–Stieltjes type integral which includes both the Lebesgue–Stieltjes and Moore–Pollard–Stieltjes integrals. Given two real-valued functions f, g on $[a,b]$, say M is a *major function* of g with respect to f if $M(a) = 0$, M has finite values on $[a,b]$, and for each $x \in [a,b]$ there exists $\delta(x) > 0$ such that

$$\begin{aligned} M(z) &\geq M(x) + g(x)[f(z) - f(x)] \quad \text{if } x \leq z \leq \min\{b, x+\delta(x)\}, \\ M(z) &\leq M(x) + g(x)[f(z) - f(x)] \quad \text{if } \max\{a, x-\delta(x)\} \leq z \leq x. \end{aligned}$$

Let $\mathcal{U}(g,f)$ be the class of all major functions of g with respect to f, and let

$$(\mathrm{UPS})\int_a^b g\,df := \begin{cases} \inf\{M(b) : M \in \mathcal{U}(g,f)\} & \text{if } \mathcal{U}(g,f) \neq \emptyset,\\ +\infty & \text{if } \mathcal{U}(g,f) = \emptyset.\end{cases}$$

A function m is a *minor function* of g with respect to f if $-m \in \mathcal{U}(-g,f)$. Let $\mathcal{L}(g,f)$ be the class of all minor functions of g with respect to f, and let

$$(\mathrm{LPS})\int_a^b g\,df := \begin{cases} \sup\{m(b) : m \in \mathcal{L}(g,f)\} & \text{if } \mathcal{L}(g,f) \neq \emptyset,\\ -\infty & \text{if } \mathcal{L}(g,f) = \emptyset.\end{cases}$$

If $(\mathrm{UPS})\int_a^b g\,df = (\mathrm{LPS})\int_a^b g\,df$ is finite then denote the common value by $(\mathrm{WPS})\int_a^b g\,df$ and call it the *Ward–Perron–Stieltjes integral*, or the *WPS integral*. By Theorem 5 of Ward [38], if $(\mathrm{MPS})\int_a^b g\,df$ exists then so does $(\mathrm{WPS})\int_a^b g\,df$, and has the same value. Ward [38] stated and Saks [33, Theorem VI.8.1] gave a proof of the fact that $(\mathrm{WPS})\int_a^b g\,df$ is defined provided the corresponding Lebesgue–Stieltjes integral is defined. Then $(\mathrm{WPS})\int_a^b g\,df = (\mathrm{LS})\int_a^b g\,df$.

Kurzweil [17, Section 1.2] suggested an equivalent definition of the WPS integral based on an extension of the limit as the mesh of partitions tends to zero in the definition of the RS integral. A *gauge function* is any function with strictly positive values. Given a gauge function $\delta(\cdot)$ on $[a,b]$, a tagged partition $\{([x_{i-1},x_i],y_i) : i=1,\dots,n\}$ is *δ-fine* if $y_i - \delta(y_i) \le x_{i-1} \le y_i \le x_i \le y_i + \delta(y_i)$ for $i=1,\dots,n$. The *Henstock–Kurzweil* integral $(\mathrm{HK})\int_a^b g\,df$ is defined as the number I, whenever it exists, such that for each $\varepsilon > 0$ there is a gauge function $\delta(\cdot)$ on $[a,b]$ such that $|S_{\mathrm{RS}}(\tau) - I| < \varepsilon$ for each δ-fine tagged partition τ of $[a,b]$. Kurzweil [17, Theorem 1.2.1] proved the following statement:

THEOREM 2.6. *The integrals* $(\mathrm{WPS})\int_a^b g\,df$ *and* $(\mathrm{HK})\int_a^b g\,df$ *either both exist with the same values, or both do not exist.*

The following theorem is a special case of Theorem F.2 of Dudley and Norvaiša [6, Part I].

THEOREM 2.7. *For* $1/p + 1/p' \ge 1$ *with* $1 < p < \infty$, *let* $f \in \mathcal{W}_p^*([a,b])$ *and* $g \in \mathcal{W}_{p'}([a,b])$. *If* $(\mathrm{YS})\int_a^b g\,df$ *exists then so does* $(\mathrm{HK})\int_a^b g\,df$, *and the two integrals are equal.*

By these statements, Theorem 1.1 also holds with the YS integral replaced either by the WPS integral or by the HK integral.

2.5. *Existence of integrals.* To give a sufficient condition for the existence of the YS integral we need the notion of the Φ-variation, where Φ is a continuous function on $[0,\infty)$ strictly increasing from 0 to ∞. For a function f, its *Φ-variation* $v_\Phi(f)$ is the least upper bound of the sums $\sum_{i=1}^n \Phi(|f(x_i)-f(x_{i-1})|)$ over all partitions $\{x_i : i=0,\dots,n\} \in P([a,b])$. The Φ-variation reduces to the p-variation when $\Phi(u)=u^p$. The following result is due to L. C. Young [40, Theorem 5.1].

THEOREM 2.8. *Let Φ, Ψ be continuous functions on $[0,\infty)$, strictly increasing from 0 to ∞ and such that*

$$\sum_{n=1}^{\infty} \Phi^{-1}(1/n)\Psi^{-1}(1/n) < \infty, \tag{2.7}$$

where Φ^{-1} and Ψ^{-1} are the inverse functions. If $v_\Phi(f)<\infty$ and $v_\Psi(g)<\infty$, then g is Young–Stieltjes integrable on $[a,b]$ with respect to f, and for any $x\in[a,b]$,

$$\Big|(\mathrm{YS})\int_a^b [g-g(x)]\,df\Big| \le K\sum_{n=1}^{\infty}\Phi^{-1}(v_\Phi(f)/n)\Psi^{-1}(v_\Psi(g)/n) < \infty,$$

where K is an absolute constant.

Leśniewicz and Orlicz [20, Theorem 4.11] gave another proof of Theorem 2.8 for the case when the functions Φ, Ψ are log-convex and one of f, g is continuous, so that the RS integral exists in this case. They also proved that condition (2.7) cannot be weakened in general (cf. Theorem 4.21 of Leśniewicz and Orlicz [20]). Further results concerning optimality of conditions of type (2.7) are due to L. C. Young [41]. Necessary and sufficient conditions for the existence of the Riemann–Stieltjes integral in terms of the Φ-variation are due to D'yačkov [9].

3. The case $p=1$. In this section we consider the composition $\phi\circ f$ when f is a function of bounded (total) variation, that is, when f has bounded p-variation with $p=1$. In this case a chain rule formula more general than (1.3) holds. The function ϕ in this formula need not have a derivative at each point. For example, ϕ may be a convex function.

We start with the case when ϕ has continuous derivative ϕ' everywhere.

LEMMA 3.1. *If f is of bounded variation on $[a,b]$, and if ϕ is continuously differentiable, then $\phi'\circ f$ is Young–Stieltjes integrable with respect to f, and* (1.3) *holds.*

Proof. The integral $(\mathrm{YS})\int_a^b(\phi'\circ f)\,df$ exists because $\phi'\circ f$ is a regulated function on $[a,b]$. To prove (1.3) let $\varepsilon>0$. Then there are a bounded interval $[c,d]$ containing the range of f and a $\delta>0$ such that $|\phi'(u)-\phi'(v)|<\varepsilon$ whenever $|u-v|<\delta$ and $u,v\in[c,d]$. Also, since f is regulated there is a partition $\lambda=\{z_j : j=0,\dots,m\}\in P([a,b])$ such that $\mathrm{Osc}(f;(z_{j-1},z_j))<\delta$

for $j = 1, \dots, m$. Let $\tau = \{((x_{i-1}, x_i), y_i) : i = 1, \dots, n\}$ be a Young tagged partition of $[a, b]$ which is a refinement of λ. By a telescoping sum, we have

$$\begin{aligned}(3.1)\quad & \Delta(\phi\circ f)[a,b] \\ &= \sum_{i=1}^{n}\{\Delta^{+}(\phi\circ f)(x_{i-1}) + \Delta(\phi\circ f)(x_{i-1},x_i) + \Delta^{-}(\phi\circ f)(x_i)\} \\ &= S_{\mathrm{YS}}(\phi'\circ f, f;\tau) + \sum_{i=1}^{n}(\Delta^{+}(\phi\circ f) - (\phi'\circ f)\Delta^{+}f)(x_{i-1}) \\ &\quad + \sum_{i=1}^{n}(\Delta^{-}(\phi\circ f) - (\phi'\circ f)\Delta^{-}f)(x_i) + R(\tau),\end{aligned}$$

where

$$R(\tau) := \sum_{i=1}^{n}\{\Delta(\phi\circ f)(x_{i-1},x_i) - (\phi'\circ f)(y_i)\Delta f(x_{i-1},x_i)\}.$$

By the mean value theorem, we have

$$R(\tau) = \sum_{i=1}^{n}[\phi'(\theta_i) - \phi'(f(y_i))]\Delta f(x_{i-1},x_i),$$

where for each i, θ_i belongs to the interval with endpoints $f(x_{i-1}+)$ and $f(x_i-)$. Since $y_i \in (x_{i-1}, x_i) \subset (z_{j-1}, z_j)$ and $|\theta_i - f(y_i)| < \delta$ for $i = 1, \dots, n$, it follows that $|R(\tau)| \le \varepsilon v_1(f; [a, b])$. Since ε is arbitrary we have $\lim_{\tau,\mathfrak{R}} R(\tau) = 0$. To show that the sums in (1.3) converge unconditionally, let $\nu \subset (a, b)$ be a finite set. By the mean value theorem again,

$$\sum_{\nu}|\Delta^{\pm}(\phi\circ f) - (\phi'\circ f)\Delta^{\pm}f| \le 2\|\phi'\|_{\infty}v_1(f;[a,b]) < \infty.$$

Thus taking the limit in (3.1) under the refinements of partitions yields (1.3). ∎

We extend the preceding chain rule formula to a composition $\phi\circ f$ with ϕ having regulated one-sided derivatives. To this end we recall some facts about one-sided derivatives. The incrementary ratios I_ϕ of a real-valued function ϕ on $[c, d]$ are defined by $I_\phi(y, x) := [\phi(y) - \phi(x)]/(y - x)$ for $x, y \in [c, d]$ such that $x \neq y$. The *upper left derivative* $D^-\phi$ on $(c, d]$ and the *upper right derivative* $D^+\phi$ on $[c, d)$ are defined respectively by

$$(3.2)\qquad D^{-}\phi(x) := \limsup_{\xi\uparrow x} I_\phi(\xi, x) \quad\text{and}\quad D^{+}\phi(y) := \limsup_{\xi\downarrow y} I_\phi(\xi, y)$$

for each $c \le y < x \le d$. The *lower left derivative* $D_-\phi$ on $(c, d]$ and the *lower right derivative* $D_+\phi$ on $[c, d)$ are defined as in (3.2) except that limsup is replaced by liminf. We extend the definitions of the four derivatives over the whole interval $[c,d]$ by setting $D^-\phi(c) := D^+\phi(c)$, $D^+\phi(d) := D^-\phi(d)$ and similarly for the lower derivatives. Let $\mathcal{DR}^- = \mathcal{DR}^-([c,d])$ be the set of all real-valued functions ϕ on $[c,d]$ having the upper left derivative $D^-\phi$ bounded and such that the left limit $D^-\phi(x-)$ exists for each $x \in (c,d]$. Similarly, let $\mathcal{DR}^+ = \mathcal{DR}^+([c,d])$ be the set of all real-valued functions ϕ on $[c,d]$ having the upper right derivative $D^+\phi$ bounded and such that the right limit $D^+\phi(y+)$ exists for each $y \in [c,d)$.

LEMMA 3.2. *For a continuous function ϕ on $[c,d]$ the following hold*:

(1) *If $\phi \in \mathcal{DR}^-$ then $D^-\phi$, $D_-\phi$ are left-continuous, $D^+\phi$, $D_+\phi$ have left limits, and $D^-\phi = D_-\phi = (D^+\phi)_- = (D_+\phi)_-$ on $(c,d]$.*

(2) *If $\phi \in \mathcal{DR}^+$ then $D^+\phi$, $D_+\phi$ are right-continuous, $D^-\phi$, $D_-\phi$ have right limits, and $D^+\phi = D_+\phi = (D^-\phi)_+ = (D_-\phi)_+$ on $[c,d)$.*

(3) *If $\phi \in \mathcal{DR}^- \cap \mathcal{DR}^+$ then ϕ is a Lipschitz function, and the sets*

$$\begin{aligned}\{y \in (c,d) : D^-\phi(y) \neq D^+\phi(y)\} &= \{y \in (c,d) : \Delta^+(D^-\phi) \neq 0\} \\ &= \{y \in (c,d) : \Delta^-(D^+\phi) \neq 0\}\end{aligned} \tag{3.3}$$

are at most countable.

Proof. Let $\phi \in \mathcal{DR}^-$. By statement 280 of Hobson [13, p. 382], for each $c \le x_0 < x_1 \le d$ we have

$$\begin{aligned}\sup_{x_0 \le x \le x_1} D\phi(x) &= \sup_{x_0 \le x < y \le x_1} I_\phi(y,x), \\ \inf_{x_0 \le x \le x_1} D\phi(x) &= \inf_{x_0 \le x < y \le x_1} I_\phi(y,x),\end{aligned} \tag{3.4}$$

where $D\phi$ can be any of the four derivatives $D^-\phi$, $D_-\phi$, $D^+\phi$, $D_+\phi$. Let $x \in (c,d]$, $\lambda := D^-\phi(x-)$, and $\varepsilon > 0$. Then there exists $\delta \in (0, x-c)$ such that $D^-\phi(u) \in [\lambda-\varepsilon, \lambda+\varepsilon]$ for all $u \in [x-\delta, x)$. By (3.4), the other three derivatives when restricted to $[x-\delta, x)$ also have their values in $[\lambda-\varepsilon, \lambda+\varepsilon]$. By statement 281 of Hobson [13, p. 383], the values $D^-\phi(x)$ and $D_-\phi(x)$ are in $[\lambda-\varepsilon, \lambda+\varepsilon]$. Since ε is arbitrary it then follows that $D^-\phi$ and $D_-\phi$ are left-continuous at x with value λ at x. Similarly, $D^+\phi$ and $D_+\phi$ have left limits at x equal to λ. Therefore (1) holds. We omit the proof of (2) which is based on symmetric arguments.

Let $\phi \in \mathcal{DR}^- \cap \mathcal{DR}^+$. Then ϕ is a Lipschitz function by (3.4). The three sets (3.3) are the same by (1) and (2), and they are at most countable by an analogous fact for regulated functions. ■

Define $\mathcal{DR} = \mathcal{DR}([c,d])$ to be the set of all functions $\phi \in \mathcal{DR}^- \cap \mathcal{DR}^+$ which are continuous on $[c,d]$. By the preceding lemma, the lower left and upper left derivatives of each element of $\mathcal{DR}$ coincide, and are equal to the left derivative also denoted by $D^-\phi$. The same is true about the right derivatives $D^+\phi$ for $\phi \in \mathcal{DR}$. Clearly, convex functions with domains containing

$[c,d]$ belong to $\mathcal{DR}([c,d])$. If ϕ is convex then $D^-\phi$, $D^+\phi$ are non-decreasing, and $D^-\phi(x) \le D^+\phi(x)$ for $x \in [c,d]$.

THEOREM 3.3. *Let f be a function of bounded variation on $[a,b]$, and let $\phi \in \mathcal{DR}([c,d])$ with (c,d) containing the range of f. For ψ being either $D^-\phi$ or $D^+\phi$, if $\psi \circ f$ is Young–Stieltjes integrable with respect to f then*

$$(3.5)\qquad (\phi\circ f)(b) - (\phi\circ f)(a) = (\mathrm{YS})\int_a^b (\psi\circ f)\,df + \sum_{(a,b]}[\Delta^-(\phi\circ f) - (\psi\circ f)\Delta^- f] + \sum_{[a,b)}[\Delta^+(\phi\circ f) - (\psi\circ f)\Delta^+ f],$$

where the two sums are unconditional.

Proof. By Lemma 3.2(3), ϕ is a Lipschitz function. Therefore the two sums in (3.5) converge unconditionally. To prove (3.5) we choose a sequence $\{\phi_n : n \ge 1\}$ of smooth functions in such a way that

$$(3.6)\qquad (\phi_n\circ f)(b) - (\phi_n\circ f)(a) = (\mathrm{YS})\int_a^b (\phi_n'\circ f)\,df + \sum_{(a,b]}[\Delta^-(\phi_n\circ f) - (\phi_n'\circ f)\Delta^- f] + \sum_{[a,b)}[\Delta^+(\phi_n\circ f) - (\phi_n'\circ f)\Delta^+ f]$$

for all $n \ge 1$, and each term of (3.6) converges as $n\to\infty$ to the corresponding term of (3.5). Let $\psi = D^-\phi$, and let ϱ be a probability density on the real line $\mathbb{R}$ continuously differentiable and with support in $[0,1]$. We extend ϕ and ψ from $(c,d]$ to the whole real line by periodicity. For each integer $n \ge 1$ and $r \in \mathbb{R}$, let $\varrho_n(r) := n\varrho(nr)$ and

$$(3.7)\qquad \phi_n(r) := \int_{\mathbb{R}} \varrho_n(r-s)\phi(s)\,ds = \int_{\mathbb{R}} \varrho(s)\phi(r-s/n)\,ds.$$

Then for each $r \in (c,d)$, $\lim_{n\to\infty}\phi_n(r) = \phi(r)$, and since $D^-\phi$ is left-continuous,

$$\lim_{n\to\infty}\phi_n'(r) = \lim_{n\to\infty}\int_{[0,1]} \varrho(s)D^-\phi(r-s/n)\,ds = D^-\phi(r).$$

Each ϕ_n' is continuous on $[c,d]$. Hence by Lemma 3.1, $(\mathrm{YS})\int_a^b(\phi_n'\circ f)\,df$ exists and (3.6) holds for each $n\ge 1$. To show

$$(3.8)\qquad \lim_{n\to\infty}(\mathrm{YS})\int_a^b(\phi_n'\circ f)\,df = (\mathrm{YS})\int_a^b (D^-\phi\circ f)\,df$$

we use the Osgood convergence theorem for the YS integrals proved by Hildebrandt [14, Theorem 19.3.14]. Since $\|\phi_n'\|_\infty \le \|D^-\phi\|_\infty$, the $\phi_n'\circ f$ are uniformly bounded. Moreover, they converge to $\psi\circ f$ pointwise. Since

$D^-\phi \circ f$ is Young–Stieltjes integrable by assumption, the Osgood convergence theorem applies, and hence (3.8) holds. To show

$$\lim_{n\to\infty} \sum_{(a,b]} [\Delta^-(\phi_n \circ f) - (\phi_n' \circ f)\Delta^- f] = \sum_{(a,b]} [\Delta^-(\phi \circ f) - (D^-\phi \circ f)\Delta^- f], \tag{3.9}$$

again we use the fact that the $\|\phi_n'\|_\infty$ are uniformly bounded. Since f is of bounded variation, each sum on the left side of (3.9) converges absolutely and uniformly in n. Since $\Delta^-(\phi_n \circ f)$ and $\phi_n' \circ f$ converge as $n \to \infty$ to $\Delta^-(\phi \circ f)$ and $\psi \circ f$, respectively, (3.9) holds (cf. the proof of Lemma 19.3.15 in Hildebrandt [14]). Similarly, the second sum in (3.6) converges as $n \to \infty$ to the second sum in (3.5). Thus the relation (3.5) with $\psi = D^-\phi$ holds. If instead of (3.7) one uses the smoothing

$$\phi_n(r) := \int_{\mathbb{R}} \varrho_n(s-r)\phi(s)\,ds = \int_{\mathbb{R}} \varrho(s)\phi(r+s/n)\,ds$$

then the above approximation argument yields (3.5) with $\psi = D^+\phi$ because the latter is right-continuous. ∎

In the preceding theorem we assumed that the YS integral exists; this was needed to apply the Osgood convergence theorem. However, under the same assumptions on f and ϕ, the HK integral exists and satisfies the chain rule formula. The proof of the following statement is the same as the proof of Theorem 3.3 except that one needs to use the dominated convergence theorem for the HK integral (see e.g. Section 7.8 in McLeod [25]) instead of the Osgood convergence theorem for the YS integral.

THEOREM 3.4. *Let f be a function of bounded variation on $[a,b]$, and let $\phi \in \mathcal{DR}([c,d])$ with (c,d) containing the range of f. For ψ being either $D^-\phi$ or $D^+\phi$, the integral* (HK) $\int_a^b (\psi \circ f)\,df$ *exists and* (3.5) *holds with the YS integral replaced by the HK integral.*

By Theorem 2.6, the analogous theorem holds for the WPS integral.

4. The case $1 < p < 2$. In this section we consider the composition $\phi \circ f$ when f is a function of bounded p-variation for some $1 < p < 2$. The main result is Theorem 4.2 from which the chain rule formula of Theorem 1.1 follows. Also, Theorem 4.2 yields an extension of the integration by parts formula for Young–Stieltjes integrals (Corollary 4.3). At the end of this section we show that the p-variation condition in Theorem 1.1 is sharp (Proposition 4.4).

We start with a useful characterization of the class $\mathcal{W}_p^*$, $1 < p < \infty$, defined in Section 1 by (1.2). Its formulation requires the p-variation over an open interval defined by

$$v_p(f;(u,v)) := \sup\{s_p(f;\kappa) : \kappa = \{x_i\}_{i=0}^n,\ u < x_0 < \ldots < x_n < v\}$$
$$= \lim_{x\downarrow u,\, y\uparrow v} v_p(f;[x,y]), \qquad a \le u < v \le b.$$

Also let $\sigma_p(f)$ be the right side of (1.1), and let $v_p^*(f)$ be the right side of (1.2).

LEMMA 4.1. *Let f be a regulated function on $[a,b]$, and let $1 < p < \infty$. Then the following statements are equivalent*:

(1) $f \in \mathcal{W}_p^*([a,b])$;

(2) *for every $\varepsilon > 0$ there is a partition $\{z_j : j = 0,\ldots,m\}$ of $[a,b]$ such that $\sum_{j=1}^m v_p(f;(z_{j-1},z_j)) < \varepsilon$*;

(3) $\sigma_p(f) < \infty$, *and for every $\varepsilon > 0$ there is a partition λ of $[a,b]$ such that $\sum_{i=1}^n |\Delta f(x_{i-1},x_i)|^p < \varepsilon$ for each refinement $\{x_i : i = 0,\ldots,n\}$ of λ.*

Proof. (1)⇒(2). Let $f \in \mathcal{W}_p^*$ and $\varepsilon > 0$. By (1.1) and (1.2), there exists $\lambda = \{z_j : j = 0,\ldots,m\} \in P([a,b])$ such that

$$\sup_{\kappa\in P(\lambda)} s_p(f;\kappa) = \sum_{j=1}^m v_p(f;[z_{j-1},z_j]) < v_p^*(f) + \varepsilon/2$$

and

$$\sum_{j=1}^m [|\Delta^+ f(z_{j-1})|^p + |\Delta^- f(z_j)|^p] > \sigma_p(f) - \varepsilon/2 = v_p^*(f) - \varepsilon/2.$$

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

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

for each $j = 1,\ldots,m$. If we take $u_{j-1} \downarrow z_{j-1}$ and $v_j \uparrow z_j$, by Lemma 2.19 of Dudley and Norvaiša [6, Part II], it follows that

$$\sum_{j=1}^m v_p(f;(z_{j-1},z_j)) \le \sum_{j=1}^m \{v_p(f;[z_{j-1},z_j]) - |\Delta^+ f(z_{j-1})|^p - |\Delta^- f(z_j)|^p\}$$
$$< v_p^*(f) + \varepsilon/2 - v_p^*(f) + \varepsilon/2 = \varepsilon.$$

This proves (2).

(2)⇒(1). Assume (2) holds. Then it is easy to see that $f \in \mathcal{W}_p$. Assume however that $f \notin \mathcal{W}_p^*$. Therefore, since $\sigma_p(f) \le v_p^*(f)$ always holds, $v_p^*(f) - \sigma_p(f) \ge C$ for some positive constant C. Let $\kappa = \{x_i : i = 0,\ldots,n\} \in P([a,b])$. Then

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

It then follows that

$$\sum_{i=1}^{n} v_p(f;(x_{i-1},x_i)) \geq v_p^*(f) - \sum_{i=1}^{n}\{\Delta^+ f(x_{i-1})|^p + |\Delta^- f(x_i)|^p\}$$
$$\geq v_p^*(f) - \sigma_p(f) \geq C > 0.$$

Since κ is arbitrary, (2) cannot hold. This contradiction implies that $f \in \mathcal{W}_p^*$.

(2)⇒(3). Assume (2) holds. Then it holds for any refinement of $\lambda := \{z_j : j = 0,\dots,m\}$. It is easy to see that $f \in \mathcal{W}_p$, and hence $\sigma_p(f) < \infty$. Let $\varepsilon > 0$ and let $\{x_i : i = 0,\dots,n\}$ be a refinement of λ. Then

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

Therefore (3) holds.

(3)⇒(2). Let $\varepsilon > 0$ and let $\lambda \in P([a,b])$ satisfy (3). There exists a finite point set $\nu \subset [a,b]$ such that $\sum_{x\in\mu}[|\Delta^- f(x)|^p + |\Delta^+ f(x)|^p] < \varepsilon$ for each finite set $\mu \subset (a,b)$ with $\mu \cap \nu = \emptyset$. Let $\{z_j : j = 0,\dots,m\} := \lambda \cup \nu \in P([a,b])$, and for each $j = 1,\dots,m$, let $\kappa(j) := \{x_i^j : i = 0,\dots,n(j)\}$ be a set of points in (a,b) such that $z_{j-1} < x_0^j < \dots < x_{n(j)}^j < z_j$, that is, $\kappa(j)$ is a partition of (z_{j-1},z_j). Then applying convexity of $t \mapsto t^p$, $t \geq 0$, twice, we have

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

Since all partitions $\kappa(j)$ of (z_{j-1},z_j) are arbitrary, it follows that

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

This proves (2). ∎

Let d be a positive integer and let ϕ be a real-valued function on $\mathbb{R}^d$. We write $\phi \in \mathit{\Lambda}_{1,0}(\mathbb{R}^d)$ if ϕ satisfies the condition:

(1) ϕ is differentiable with continuous partial derivatives $\phi_l'(u) := \frac{\partial\phi}{\partial u_l}(u)$ for $l = 1,\dots,d$.

For a bounded set $U \subset \mathbb{R}^d$, let $\|\phi_l'\|_\infty := \sup\{|\phi_l'(u)| : u \in U\}$ and $K_0 := 2\max_{1\le l\le d} \|\phi_l'\|_\infty$. Also, for $\alpha \in (0,1]$, we write $\phi \in \Lambda_{1,\alpha}(\mathbb{R}^d)$ if, in addition to condition (1), ϕ satisfies the condition:

(2) for any bounded set $U \subset \mathbb{R}^d$ there is a finite constant K_α such that

$$\max_{1\le l\le d} |\phi_l'(u) - \phi_l'(v)| \le K_\alpha \sum_{k=1}^{d} |u_k - v_k|^\alpha \tag{4.1}$$

for all $u = (u_1,\dots,u_d), v = (v_1,\dots,v_d) \in U$.

Now we are ready to prove the main result.

THEOREM 4.2. *For $\alpha \in [0,1]$, let $\mathbf{f} = (f_1,\dots,f_d) : [a,b] \to \mathbb{R}^d$ be a vector-valued function with coordinate functions $f_l \in \mathcal{W}^*_{1+\alpha}([a,b])$ for $l = 1,\dots,d$, let $\phi \in \Lambda_{1,\alpha}(\mathbb{R}^d)$ and let $h \in \mathcal{R}([a,b])$. Then*

$$(4.2)\qquad (\mathrm{YS})\int_a^b h\,d(\phi\circ\mathbf{f}) = \sum_{l=1}^{d} (\mathrm{YS})\int_a^b h(\phi_l'\circ\mathbf{f})\,df_l$$

$$+\sum_{(a,b]} h\Big\{\Delta^-(\phi\circ\mathbf{f}) - \sum_{l=1}^{d}(\phi_l'\circ\mathbf{f})\Delta^- f_l\Big\} + \sum_{[a,b)} h\Big\{\Delta^+(\phi\circ\mathbf{f}) - \sum_{l=1}^{d}(\phi_l'\circ\mathbf{f})\Delta^+ f_l\Big\},$$

meaning that all the $d+1$ integrals exist provided any d integrals exist, *and the two sums are unconditional.*

Proof. We start by showing that the two sums in (4.2) converge unconditionally. For finite sets $\mu \subset [a,b)$ and $\nu \subset (a,b]$, let

$$V^+(\mu) := \sum_{x\in\mu} h(x)\Big(\Delta^+(\phi\circ\mathbf{f}) - \sum_{l=1}^{d}(\phi_l'\circ\mathbf{f})\Delta^+ f_l\Big)(x), \tag{4.3}$$

$$V^-(\nu) := \sum_{x\in\nu} h(x)\Big(\Delta^-(\phi\circ\mathbf{f}) - \sum_{l=1}^{d}(\phi_l'\circ\mathbf{f})\Delta^- f_l\Big)(x). \tag{4.4}$$

Let $x \in [a,b)$ be such that $\Delta^+ f_l(x) \ne 0$ for some $l \in \{1,\dots,d\}$. By the mean value theorem, there is a vector $\theta = (\theta_1,\dots,\theta_d)$ with $\theta_l \in [f_l(x)\wedge f_l(x+), f_l(x)\vee f_l(x+)]$ $(= \{f_l(x)\}$ if $\Delta^+ f_l(x) = 0)$, $l = 1,\dots,d$, such that

$$\phi(\mathbf{f}(x+)) - \phi(\mathbf{f}(x)) = \sum_{l=1}^{d} \phi_l'(\theta)\Delta^+ f_l(x). \tag{4.5}$$

By the W. H. Young inequality, for any $u, v \ge 0$ and $\alpha > 0$, we have

$$u^\alpha v \le (\alpha/(1+\alpha))u^{1+\alpha} + (1/(1+\alpha))v^{1+\alpha}. \tag{4.6}$$

For $\alpha \in (0,1]$, by (4.1), (4.5) and (4.6), we obtain

$$\delta^+(x) := \Big|[\Delta^+(\phi\circ\mathbf{f}) - \sum_{l=1}^d (\phi_l'\circ\mathbf{f})\Delta^+ f_l](x)\Big|$$

$$\leq K_\alpha \sum_{l,k=1}^d |\Delta^+ f_k(x)|^\alpha |\Delta^+ f_l(x)| \leq dK_\alpha \sum_{l=1}^d |\Delta^+ f_l(x)|^{1+\alpha}.$$

For $\alpha = 0$ we have the bound $\delta^+(x) \leq K_0 \sum_{l=1}^d |\Delta^+ f_l(x)|$. Hence for the sum of the absolute values of each term in (4.3), by the first part of Lemma 4.1(3), it follows that

$$\sum_{x\in\mu} |h(x)|\delta^+(x) \leq dK_\alpha \|h\|_\infty \sum_{l=1}^d \sigma_{1+\alpha}(f_l) < \infty$$

for $\alpha \in [0,1]$ and any finite set $\mu \subset [a,b)$. A similar bound holds for the sum of the absolute values of each term in (4.4). Thus the two sums in (4.2) converge absolutely, and let

$$V^- := \lim_{\mu,\mathfrak{U}_1} V^-(\mu), \qquad V^+ := \lim_{\mu,\mathfrak{U}_2} V^+(\mu) \tag{4.7}$$

be their unconditional sums, where $\mathfrak{U}_1$, $\mathfrak{U}_2$ are the directions formed by inclusions of finite sets in $(a,b]$ and $[a,b)$, respectively.

Given a Young tagged partition $\tau = \{((x_{i-1},x_i),y_i) : i = 1,\dots,n\}$, let $S(\tau) := S_{\mathrm{YS}}(h,\phi\circ\mathbf{f};\tau)$ be equal to

$$\sum_{i=1}^n \{[h\Delta^+\phi\circ\mathbf{f}](x_{i-1}) + h(y_i)(\phi\circ\mathbf{f})(x_{i-1},x_i) + [h\Delta^-\phi\circ\mathbf{f}](x_i)\}$$

and, for each $l = 1,\dots,d$, let $S_l(\tau) := S_{\mathrm{YS}}(h(\phi_l'\circ\mathbf{f}), f_l;\tau)$ be equal to

$$\sum_{i=1}^n \{[h(\phi_l'\circ\mathbf{f})\Delta^+ f_l](x_{i-1}) + [h(\phi_l'\circ\mathbf{f})](y_i) f_l(x_{i-1},x_i) + [h(\phi_l'\circ\mathbf{f})\Delta^- f_l](x_i)\}.$$

Also let

$$U(\tau) := \Big|S(\tau) - \sum_{l=1}^d S_l(\tau) - V^+ - V^-\Big|, \tag{4.8}$$

where V^+, V^- are defined by (4.7). To prove the theorem it is enough to show that given $\varepsilon > 0$ there is a partition $\lambda \in P([a,b])$ such that $U(\tau) < \varepsilon$ for each refinement τ of λ. For each $\kappa = \{x_i : i = 0,\dots,n\} \in P([a,b])$ and a Young tagged partition $\tau = \tau(\kappa) = \{((x_{i-1},x_i),y_i) : i = 1,\dots,n\}$ associated to κ, we have the identity

$$S(\tau) = \sum_{l=1}^d S_l(\tau) + V^+(\kappa) + V^-(\kappa) + R(\tau), \tag{4.9}$$

where $V^+(\kappa) := V^+(\{x_0, \dots, x_{n-1}\})$, $V^-(\kappa) := V^-(\{x_1, \dots, x_n\})$ are defined by (4.3), (4.4), respectively, and

$$R(\tau) = \sum_{i=1}^{n} h(y_i)\Big\{(\phi \circ \mathbf{f})(x_{i-1}, x_i) - \sum_{l=1}^{d}(\phi_l' \circ \mathbf{f})(y_i) f_l(x_{i-1}, x_i)\Big\}.$$

Let $\varepsilon > 0$. By (4.7), there exist a finite set $\mu^+ \subset [a, b)$ and a finite set $\mu^- \subset (a, b]$ such that

$$|V^+ - V^+(\kappa)| < \varepsilon/3 \quad \text{and} \quad |V^- - V^-(\kappa)| < \varepsilon/3 \tag{4.10}$$

for each $\kappa \in P([a, b])$ containing $\mu^+ \cup \mu^-$. Let $U \subset \mathbb{R}^d$ be a bounded set containing the range of $\mathbf{f}$. Since each ϕ_l', $l = 1, \dots, d$, is continuous, there is a $\delta > 0$ such that

$$\max_{1\le l\le d} |\phi_l'(u) - \phi_l'(v)| < \varepsilon\Big/\Big(3\|h\|_\infty \sum_{l=1}^{d} v_1(f_l; [a, b])\Big) \tag{4.11}$$

whenever $\max_{1\le k\le d} |u_k - v_k| < \delta$ for $u = (u_1, \dots, u_d), v = (v_1, \dots, v_d) \in U$. By Lemma 4.1 and since f_l, $l = 1, \dots, d$, are regulated functions on $[a, b]$, one can find a partition $\chi := \{z_j : j = 0, \dots, m\}$ of $[a, b]$ such that, respectively,

$$\sum_{l=1}^{d}\sum_{i=1}^{n} v_{1+\alpha}(f_l; (x_{i-1}, x_i)) < \varepsilon/(3dK_\alpha\|h\|_\infty) \quad \text{for } \alpha \in (0, 1] \tag{4.12}$$

for each refinement $\{x_i : i = 1, \dots, n\}$ of χ, and

$$\max_{1\le l\le d}\max_{1\le j\le m} \operatorname{Osc}(f_l; (z_{j-1}, z_j)) < \delta \quad \text{for } \alpha = 0. \tag{4.13}$$

By the mean value theorem, there are vectors $\theta_i = (\theta_{1,i}, \dots, \theta_{d,i})$, $i = 1, \dots, n$, with coordinates $\theta_{l,i} \in [f_l(x_{i-1}+) \wedge f_l(x_i-), f_l(x_{i-1}+) \vee f_l(x_i-)]$, $l = 1, \dots, d$, such that

$$R(\tau) = \sum_{i=1}^{n} h(y_i)\sum_{l=1}^{d}[\phi_l'(\theta_i) - \phi_l'(\mathbf{f}(y_i))][f_l(x_i-) - f_l(x_{i-1}+)].$$

For $\alpha \in (0, 1]$, by (4.1), (4.6) and (4.12), we have

$$\begin{aligned} |R(\tau)| &\le \|h\|_\infty \sum_{i=1}^{n}\sum_{l=1}^{d} |\phi_l'(\theta_i) - \phi_l'(\mathbf{f}(y_i))| \cdot |f_l(x_i-) - f_l(x_{i-1}+)| \\ &\le dK_\alpha\|h\|_\infty \sum_{l=1}^{d}\sum_{i=1}^{n} v_{1+\alpha}(f_l; (x_{i-1}, x_i)) < \varepsilon/3 \end{aligned} \tag{4.14}$$

whenever τ is a refinement of χ. For $\alpha = 0$, by (4.11) and (4.13), we get

$$|R(\tau)| \le \|h\|_\infty \sum_{l=1}^{d} v_1(f_l; [a, b]) \max_{1\le i\le n} |\phi_l'(\theta_i) - \phi_l'(\mathbf{f}(x_i-))| < \varepsilon/3 \tag{4.15}$$

whenever τ is a refinement of χ. Let $\lambda := \mu^+ \cup \mu^- \cup \chi \in P([a,b])$, and let $\tau = \tau(\kappa)$ be a Young tagged partition which is a refinement of λ. Then by (4.9), (4.10), (4.14) and (4.15), the desired bound of (4.8) follows:

$$U(\tau) \le |V^+(\kappa) - V^+| + |V^-(\kappa) - V^-| + |R(\tau)| < \varepsilon. \ \blacksquare$$

An integration by parts formula was proved for functions with bounded variation by Kurzweil [18] using the HK integral, and by Hildebrandt [14, Theorem II.19.3.13] using the YS integral (which was already used in [40] by L. C. Young). Next we show that Theorem 4.2 yields the same formula for a larger class of functions.

Corollary 4.3. *Let $f, g \in \mathcal{W}_2^*([a,b])$ and let $\sharp$ be either* YS, *or* HK, *or* CY. *If $(\sharp)\int_a^b f\,dg$ exists then so does $(\sharp)\int_a^b g\,df$, and*

$$(\sharp)\int_a^b g\,df + (\sharp)\int_a^b f\,dg = \Delta(fg)[a,b] + \sum_{(a,b]} \Delta^- f \Delta^- g - \sum_{[a,b)} \Delta^+ f \Delta^+ g,$$

where the two sums are unconditional.

Proof. Consider the function $\phi : \mathbb{R}^2 \to \mathbb{R}$ given by $\phi(u) := u_1 u_2$ for $u = (u_1, u_2)$. Then $\phi \in \Lambda_{1,1}(\mathbb{R}^2)$ with $\phi_1'(u) = u_2$ and $\phi_2'(u) = u_1$. Thus $\phi_l'((f,g)) \in \mathcal{W}_2^*$ for $l = 1,2$. The claim for the YS integral follows from Theorem 4.2 by taking $h \equiv 1$, because $\Delta^-(fg) - g\Delta^- f - f\Delta^- g = -\Delta^- f \Delta^- g$ and $\Delta^+(fg) - g\Delta^+ f - f\Delta^+ g = \Delta^+ f \Delta^+ g$. The claim for the CY and HK integrals then follows from Theorems 2.5 and 2.7, respectively, with $p = 2$. $\blacksquare$

Notice that the two integrals in Corollary 4.3 may not exist under the stated conditions. Indeed, let $\Phi(u) := u^2|\log u|$ for $u > 0$ and $\Phi(0) := 0$. If $f, g \in \mathcal{W}_\Phi$ then $f, g \in \mathcal{W}_2^*$ but the series (2.7) is divergent. Condition (2.7) is shown by Leśniewicz and Orlicz [20] to be the best possible in general for the existence of the Riemann–Stieltjes integral, and hence for the Young–Stieltjes integral. However according to Theorem 4.2, the Young–Stieltjes integrals with a *special* form of integrand can exist even if the condition (2.7) fails to hold.

The next statement shows that the p-variation condition in Theorem 4.2 is best possible.

Proposition 4.4. *Let f be a regulated function on $[a,b]$, and let $\sharp$ be either* YS *or* CY. *The integral $(\sharp)\int_a^b f\,df$ exists and*

$$(\sharp)\int_a^b f\,df = \frac{1}{2}\Big\{f(b)^2 - f(a)^2 + \sum_{(a,b]} (\Delta^- f)^2 - \sum_{[a,b)} (\Delta^+ f)^2\Big\},$$

with the two sums converging unconditionally, if and only if $f \in \mathcal{W}_2^*([a,b])$.

Proof. The "if" part follows from Theorem 4.2 with $h \equiv 1$, $d = 1$ and $\phi(u) = u^2$ for $u \in \mathbb{R}$, and from Theorem 2.4. By the same Theorem 2.4, it is enough to prove the "only if" part for $\sharp = \mathrm{CY}$. Since unconditional convergence yields absolute convergence, we have $\sigma_2(f) < \infty$. Therefore it is enough to prove the second part of statement (3) of Lemma 4.1 with $p = 2$. Let $\varepsilon > 0$. There exists a set $\mu = \{u_j : j = 1, \dots, m\} \subset (a, b]$ such that

$$\sum_{j=1}^{m} [\Delta^- f_+^{(b)}(u_j)]^2 \geq \sum_{(a,b]} [\Delta^- f_+^{(b)}]^2 - \varepsilon/2,$$

where $\Delta^- f_+^{(b)}(x) = f(x+) - f(x-)$ for $x \in (a, b)$ and $\Delta^- f_+^{(b)}(b) = f(b) - f(b-)$. Thus for each finite set $\nu \subset (a, b]$ we have

$$(4.16) \qquad -\sum_{\nu} [\Delta^- f_+^{(b)}]^2 \geq -\sum_{(a,b]} [\Delta^- f_+^{(b)}]^2 \geq -\sum_{j=1}^{m} [\Delta^- f_+^{(b)}(u_j)]^2 - \varepsilon/2.$$

Choose a set $\{v_j : j = 1, \dots, m\} \subset (a, b]$ such that $u_{j-1} < v_j < u_j$, $j = 1, \dots, m$, with $u_0 := a$, and

$$\sum_{j=1}^{m} \inf\{[f_+^{(b)}(u_j) - f(x)]^2 : v_j \leq x < u_j\} \geq \sum_{j=1}^{m} [\Delta^- f_+^{(b)}(u_j)]^2 - \varepsilon/2.$$

By the definition of the CY integral, (MPS) $\int_a^b f_-^{(a)} \, df_+^{(b)}$ exists. Therefore one can choose a partition λ of $[a, b]$ such that $\lambda \supset \{v_j, u_j : j = 1, \dots, m\}$ and

$$(4.17) \qquad \Big| \sum_{i=1}^{n} [f_-^{(a)}(y_i') - f_-^{(a)}(y_i'')][f_+^{(b)}(x_i) - f_+^{(b)}(x_{i-1})] \Big| < \varepsilon$$

for each refinement $\kappa = \{x_i : i = 0, \dots, n\}$ of λ and $y_i', y_i'' \in [x_{i-1}, x_i]$, $i = 1, \dots, n$. Take any such κ, and for each $j = 1, \dots, m$, let ξ_j be the point of κ nearest to u_j from the left. Then $\xi_j \in [v_j, u_j)$ for $j = 1, \dots, m$, and

$$(4.18) \qquad \sum_{j=1}^{m} [\Delta f_+^{(b)}[\xi_j, u_j]]^2 \geq \sum_{j=1}^{m} [\Delta^- f_+^{(b)}(u_j)]^2 - \varepsilon/2.$$

Letting $y_i' \uparrow x_i$ and $y_i'' \downarrow x_{i-1}$ for each $i = 1, \dots, n$ in (4.17), we get

$$\begin{aligned} \varepsilon &\geq \Big| \sum_{i=1}^{n} [\Delta f(x_{i-1}, x_i)][\Delta f_+^{(b)}[x_{i-1}, x_i]] \Big| \\ &= \frac{1}{2} \Big| \sum_{i=1}^{n} [\Delta f(x_{i-1}, x_i)]^2 + \sum_{i=1}^{n} [\Delta f_+^{(b)}[x_{i-1}, x_i]]^2 - \sum_{i=1}^{n} [\Delta^- f_+^{(b)}(x_i)]^2 \Big| \end{aligned}$$

$$\geq \frac{1}{2}\Big(\sum_{i=1}^{n}[\Delta f(x_{i-1},x_i)]^2+\sum_{j=1}^{m}[\Delta f_+^{(b)}[\xi_j,u_j]]^2-\sum_{j=1}^{m}[\Delta^- f_+^{(b)}(u_j)]^2-\varepsilon/2\Big)$$
$$\geq \frac{1}{2}\Big(\sum_{i=1}^{n}[\Delta f(x_{i-1},x_i)]^2-\varepsilon\Big),$$

where the last two inequalities follow from (4.16) and (4.18). Hence

$$0\leq \sum_{i=1}^{n}[\Delta f(x_{i-1},x_i)]^2\leq 3\varepsilon$$

for any refinement $\kappa=\{x_i : i=1,\dots,n\}$ of λ. ■

5. The case $2\leq p<3$. In this section we consider the composition $\phi\circ f$ when f is a function of bounded p-variation for some $2\leq p<3$. To extend the chain rule formula (1.3) to this case we cannot use the YS integral. Indeed, by the preceding Proposition 4.4, if (YS) $\int_a^b f\,df$ exists then $f\in\mathcal{W}_2^*$. If in addition f is continuous then this conclusion follows easily once we recall that then (RS) $\int_a^b f\,df$ must exist by Proposition 2.1. For a partition $\kappa=\{x_i : i=0,\dots,n\}\in P([a,b])$, let $\tau^l=\tau^l(\kappa)=\{([x_{i-1},x_i],x_{i-1}) : i=1,\dots,n\}$ and $\tau^r=\tau^r(\kappa)=\{([x_{i-1},x_i],x_i) : i=1,\dots,n\}$ be two tagged partitions with tags on the left and on the right, respectively, of each tagged interval. Then the difference between the two Riemann–Stieltjes sums based on τ^r and τ^l is

$$S_{\mathrm{RS}}(f,f;\tau^r)-S_{\mathrm{RS}}(f,f;\tau^l)=s_2(f;\kappa).$$

Thus if (RS) $\int_a^b f\,df$ exists then $\lim_{|\kappa|\to 0}s_2(f;\kappa)=0$. Almost every sample function of a standard Brownian motion $B=\{B(t) : t\geq 0\}$ provides an example of a continuous function $f\notin\mathcal{W}_2^*$. By Théorème 9 of Lévy [21, p. 516], we have almost surely

$$\lim_{\varepsilon\downarrow 0}\sup\{s_2(B;\kappa) : \kappa\in P([0,1]),\ |\kappa|<\varepsilon\}=+\infty.$$

Thus almost surely the integral (RS) $\int_0^1 B\,dB$ does not exist.

One possibility to extend the Riemann–Stieltjes integral is to consider a limit of Riemann–Stieltjes sums along a *fixed* sequence of tagged partitions. Let $\{\kappa_m : m\geq 1\}$ be a sequence of partitions of $[a,b]$. For the tagged partitions $\tau^r(\kappa_m)$ and $\tau^l(\kappa_m)$ with tags on the left and on the right, respectively, of each tagged interval, and for each $m\geq 1$, we have

$$S_{\mathrm{RS}}(f,f;\tau^r(\kappa_m))=\frac{1}{2}\{f(b)^2-f(a)^2+s_2(f;\kappa_m)\},$$
$$S_{\mathrm{RS}}(f,f;\tau^l(\kappa_m))=\frac{1}{2}\{f(b)^2-f(a)^2-s_2(f;\kappa_m)\}.$$

Therefore the left sides have limits provided the limits on the right sides exist, that is, provided $\lim_{m\to\infty} s_2(f;\kappa_m)$ exists and is finite. The latter property of f is a pathwise variant of the quadratic variation, usually defined for stochastic processes as a limit in probability. For almost all sample functions of a Brownian motion B and any $t>0$, the limit $\lim_{m\to\infty} s_2(B;\kappa_m)$, $\kappa_m \in P([0,t])$, exists and is equal to t either if the sequence $\{\kappa_m : m\geq 1\}$ is nested and $\bigcup_m \kappa_m$ is dense in $[0,t]$ (by Théorème 5 of Lévy [21, Section 4]), or if the mesh $|\kappa_m| \to 0$ sufficiently fast (by Theorem 4.5 of Dudley [4]). Föllmer [10] used this extended Riemann–Stieltjes integral to prove a variant of Itô's formula for the composition $\phi\circ f$ whenever f has a quadratic variation in a sense he defined, and ϕ is a C^2 function.

In this paper we extend the Riemann–Stieltjes integral in a different direction. Once again rearranging the Riemann–Stieltjes sums for a function f with respect to itself, we have

$$\begin{aligned}S_{\mathrm{RS}}(f,f;\tau^l(\kappa)) + \frac{1}{2}s_2(f;\kappa) &= S_{\mathrm{RS}}(f,f;\tau^r(\kappa)) - \frac{1}{2}s_2(f;\kappa)\\ &= \sum_{i=1}^n \frac{f(x_{i-1})+f(x_i)}{2}[f(x_i)-f(x_{i-1})] = \frac{1}{2}\{f(b)^2 - f(a)^2\}.\end{aligned}$$

This is the sum defining the mean Stieltjes integral when the integrand and integrator coincide. For any functions g and f on $[a,b]$, the (mesh) *mean Stieltjes* integral, or *MS integral*, (MS) $\int_a^b g\,df$, is defined as the limit, if it exists, of the directed function $(S_{\mathrm{MS}}(g,f;\cdot),\mathfrak{M})$, where for a partition $\kappa = \{x_i : i=0,\ldots,n\}$,

$$S_{\mathrm{MS}}(g,f;\kappa) := \sum_{i=1}^n \frac{g(x_{i-1})+g(x_i)}{2}\,[f(x_i)-f(x_{i-1})]. \tag{5.1}$$

The (mesh) MS integral and its (refinement) variant when the direction $\mathfrak{M}$ is replaced by $\mathfrak{R}$ were introduced by Smith [35], and were later used by many authors. P. Lévy extended the MS integral in a series of papers by considering random partitions and showed that the extension may exist when the L. C. Young condition (2.7) does not hold (see, for example, [22]). A further extension of the MS integral is possible when the functions f and g are replaced by stochastic processes $X=\{X(t): 0\leq t\leq 1\}$ and $Y=\{Y(t): 0\leq t\leq 1\}$, respectively. Then it is said that the symmetric stochastic integral of Stratonovich $\int_0^1 Y\circ dX$ is defined and equals the limit in probability of the sums $S_{\mathrm{MS}}(Y,X;\kappa)$ if it exists when the mesh $|\kappa|\to 0$ (see Itô [15]). The above identity suggests that a limit of sums (5.1) may exist when g is close to f in some sense. As follows from the results of Lyons [24], this is indeed the case when $g=\psi\circ f$, f is a continuous function with bounded p-variation for some $p\in[2,3)$ and ψ has a derivative satisfying a

Lipschitz condition of order $\alpha > p-2$. On the other hand, Lyons [23] showed that the sums $S_{\mathrm{MS}}(Y,X;\kappa)$ do not converge to a finite limit if X and Y are certain jointly Gaussian but dependent Brownian motions.

We modify the definition of the MS integral as follows. For a tagged partition $\tau = \{([x_{i-1},x_i],y_i) : i=1,\ldots,n\}$ of $[a,b]$, let

$$C_{\mathrm{RS}}(g,f;\tau) := \sum_{i=1}^{n}\{[\Delta g\Delta f][y_i,x_i] - [\Delta g\Delta f][x_{i-1},y_i]\},$$

and let

$$\begin{aligned}(5.2)\qquad \varXi_{\mathrm{RS}}(g,f;\tau) &:= S_{\mathrm{RS}}(g,f;\tau) + \frac{1}{2}C_{\mathrm{RS}}(g,f;\tau)\\ &= \frac{1}{2}\sum_{i=1}^{n}\{[g(x_{i-1})+g(y_i)][f(y_i)-f(x_{i-1})] + [g(y_i)+g(x_i)][f(x_i)-f(y_i)]\}.\end{aligned}$$

Notice that the sum $\varXi_{\mathrm{RS}}(g,f;\tau)$ agrees with (5.1) either when all tags are $y_i = x_{i-1}$, or when all tags are $y_i = x_i$.

Definition 5.1. Let g and f be real-valued functions on $[a,b]$. Define the *symmetric Riemann–Stieltjes* integral, or the *SRS integral* (SRS) $\int_a^b g\,df$ to exist and equal the limit of the directed function $(\varXi_{\mathrm{RS}}(g,f;\cdot),\mathfrak{M})$. Define the *Riemann–Stieltjes quadratic covariation* $C_{\mathrm{RS}}(g,f)$ *over* $[a,b]$ to exist and equal the limit of the directed function $(C_{\mathrm{RS}}(g,f;\cdot),\mathfrak{M})$.

Usually a quadratic covariation is defined for stochastic processes X and Y to be the limit in probability, if it exists, of the sums $C_{\mathrm{RS}}(X,Y;\tau^l(\kappa_m))$ as $m\to\infty$ (see e.g. [11]).

Proposition 5.2. *Let g and f be functions on $[a,b]$. The integral* (RS) $\int_a^b g\,df$ *exists if and only if both the integral* (SRS) $\int_a^b g\,df$ *exists and the quadratic covariation $C_{\mathrm{RS}}(g,f)$ is* 0. *The two integrals have the same value whenever both statements hold.*

Proof. Suppose that (RS) $\int_a^b g\,df$ exists. Given a tagged partition $\tau = \{([x_{i-1},x_i],y_i)\colon i=1,\ldots,n\}$, let

$$\begin{aligned}\tau_1 &:= \{([x_{i-1},y_i],y_i),([y_i,x_i],x_i) : i=1,\ldots,n\},\\ \tau_2 &:= \{([x_{i-1},y_i],x_{i-1}),([y_i,x_i],x_i) : i=1,\ldots,n\},\\ \tau_3 &:= \{([x_{i-1},y_i],x_{i-1}),([y_i,x_i],y_i) : i=1,\ldots,n\}.\end{aligned}$$

Then $C_{\mathrm{RS}}(g,f;\tau) = [S_{\mathrm{RS}}(\tau_2)-S_{\mathrm{RS}}(\tau_3)]-[S_{\mathrm{RS}}(\tau_1)-S_{\mathrm{RS}}(\tau_2)]$. By assumption the directed function $(C_{\mathrm{RS}}(g,f;\cdot),\mathfrak{M})$ has a zero limit. The second part of the implication now follows from (5.2). The converse implication is clear. ∎

Several extensions of the Riemann–Stieltjes integral coincide if the integrand and integrator have bounded p- and q-variation, respectively, with conjugate p and q (see e.g. Theorems 2.5 and 2.7). We show next that the same is true for the SRS integral. Let $\mathcal{CW}_p^*$ be the subclass of all continuous functions in $\mathcal{W}_p^*$.

PROPOSITION 5.3. *Let $1/p + 1/q = 1$, and let one of the two functions g, f be in $\mathcal{CW}_p^*([a,b])$ and the other in $\mathcal{W}_q([a,b])$. If either* (RS) $\int_a^b g\,df$ *or* (SRS) $\int_a^b g\,df$ *exists then so does the other, and the two integrals are equal.*

Proof. Let $f \in \mathcal{CW}_p^*$ and let $g \in \mathcal{W}_q$. By Hölder's inequality, for any tagged partition $\tau = \{([x_{i-1}, x_i], y_i) : i = 1, \dots, n\} \in TP([a,b])$, we have

$$\begin{aligned}
&|\varXi_{\mathrm{RS}}(g,f;\tau) - S_{\mathrm{RS}}(g,f;\tau)| \\
&\qquad \le \frac{1}{2}\sum_{i=1}^{n}\{|[\Delta g\Delta f][y_i, x_i]| + |[\Delta g\Delta f][x_{i-1}, y_i]|\} \\
&\qquad \le v_q(g;[a,b])^{1/q}\Big(\sum_{i=1}^{n}\{|f(y_i) - f(x_{i-1})|^p + |f(x_i) - f(y_i)|^p\}\Big)^{1/p}.
\end{aligned}$$

Since $f \in \mathcal{CW}_p^*$, by Lemma B.1 and Theorem 4.7 of Dudley and Norvaiša [6, Part I] and [6, Part II], respectively, we have $\lim_{|\kappa|\to 0} s_p(f;\kappa) = \lim_{\kappa,\mathfrak{P}} s_p(f;\kappa) = 0$. Thus the directed functions $(S_{\mathrm{RS}}(g,f;\cdot), \mathfrak{M})$ and $(\varXi_{\mathrm{RS}}(g,f;\cdot), \mathfrak{M})$ both converge or not simultaneously, and if they converge then both have the same limit. The same conclusion follows by the same arguments when f and g are interchanged. ∎

Next we modify further the definition of the symmetric stochastic integral so that the new variant extends the YS integral. Let g and f be regulated functions on $[a,b]$. For a Young tagged partition $\tau = \{((x_{i-1}, x_i), y_i) : i = 1, \dots, n\}$ of $[a,b]$, let

$$\begin{aligned}
C_{\mathrm{YS}}(g,f;\tau) := \sum_{i=1}^{n}\{&[\Delta^+ g\Delta^+ f](x_{i-1}) - [\Delta g\Delta f](x_{i-1}, y_i] \\
&+ [\Delta g\Delta f][y_i, x_i) - [\Delta^- g\Delta^- f](x_i)\},
\end{aligned}$$

and let

$$\begin{aligned}
\varXi_{\mathrm{YS}}(g,f;\tau) &:= S_{\mathrm{YS}}(g,f;\tau) + \frac{1}{2}C_{\mathrm{YS}}(g,f;\tau) \\
&= \frac{1}{2}\sum_{i=1}^{n}\{[(g_+ + g)\Delta^+ f](x_{i-1}) + [g_+(x_{i-1}) + g(y_i)]\Delta f(x_{i-1}, y_i] \\
&\qquad + [g(y_i) + g_-(x_i)]\Delta f[y_i, x_i) + [(g + g_-)\Delta^- f](x_i)\},
\end{aligned} \tag{5.3}$$

where the Young–Stieltjes sum S_{YS} is defined by (2.2).

DEFINITION 5.4. Let $g, f \in \mathcal{R}([a,b])$. Define the *symmetric Young–Stieltjes* integral (SYS) $\int_a^b g\,df$ to be the limit (if it exists) of the directed function $(\Xi_{\mathrm{YS}}(g,f;\cdot),\mathfrak{R})$. Define the *Young–Stieltjes quadratic covariation* $C_{\mathrm{YS}}(g,f)$ *over* $[a,b]$ to be the limit (if it exists) of the directed function $(C_{\mathrm{YS}}(g,f;\cdot),\mathfrak{R})$.

For any Young tagged partition $\tau=\{((x_{i-1},x_i),y_i):i=1,\dots,n\}$ and $x_{i-1}<u_{i-1}\le y_i\le v_i<x_i$, $i=1,\dots,n$, if we let $u_{i-1}\downarrow x_{i-1}$ and $v_i\uparrow x_i$ for $i=1,\dots,n$, then $\Xi_{\mathrm{YS}}(g,f;\tau)$ can be approximated arbitrarily closely by sums $\Xi_{\mathrm{RS}}(g,f;\tau')$ with $\kappa'=\{([x_{i-1},y_i],u_{i-1}),([y_i,x_i],v_i)\colon i=1,\dots,n\}$. Thus the SYS integral based on Ξ_{YS} extends the SRS integral based on Ξ_{RS}, just as the usual Young–Stieltjes integral extends the Riemann–Stieltjes integral.

It is remarkable that unlike the YS integral, the SYS integral satisfies a simple integration by parts formula without any condition on jumps (cf. Corollary 4.3 above). To see this for a given τ, it is enough to add the two sums $\Xi_{\mathrm{YS}}(g,f;\tau)$ and $\Xi_{\mathrm{YS}}(f,g;\tau)$, which gives

$$\Xi_{\mathrm{YS}}(g,f;\tau)+\Xi_{\mathrm{YS}}(f,g;\tau)=(fg)(b)-(fg)(a).$$

From this the next statement immediately follows.

COROLLARY 5.5. *Let* $f,g\in\mathcal{R}([a,b])$. *If* (SYS) $\int_a^b f\,dg$ *exists then so does* (SYS) $\int_a^b g\,df$, *and*

$$(\mathrm{SYS})\int_a^b g\,df+(\mathrm{SYS})\int_a^b f\,dg=(fg)(b)-(fg)(a).$$

Suppose that the values of f and g at points of discontinuity satisfy

$$\begin{cases} f=(f_{+}+f_{-})/2 \quad \text{on } (a,b),\\ g(a+)=g(a),\ g(b-)=g(b), \quad g=(g_{+}+g_{-})/2 \quad \text{on} (a,b).\end{cases} \tag{5.4}$$

In this case the terms of S_{YS} and Ξ_{YS} corresponding to a given singleton of a Young tagged partition agree because

$$\frac{g+g_{-}}{2}\,\Delta^{-}f+\frac{g+g_{+}}{2}\,\Delta^{+}f=g[f_{+}-f_{-}]=g\Delta^{-}f+g\Delta^{+}f$$

on (a,b). Also, because in this case $\Delta^{+}g\Delta^{+}f=\Delta^{-}g\Delta^{-}f$ on (a,b) and $\Delta^{+}g(a)=\Delta^{-}g(b)=0$, the two integration by parts formulas for the YS and SYS integrals agree provided (5.4) holds. Next we prove that the SYS integral extends the YS integral under the above stated conditions on jumps of the integrand and integrator.

THEOREM 5.6. *Let* $g,f\in\mathcal{R}([a,b])$ *be such that* (5.4) *holds. The integral* (YS) $\int_a^b g\,df$ *exists if and only if both* (SYS) $\int_a^b g\,df$ *exists and the quadratic covariation* $C_{\mathrm{YS}}(g,f)$ *is* 0.

Proof. It is enough to show that $C_{\mathrm{YS}}(g,f)=0$ provided $(\mathrm{YS})\int_a^b g\,df$ exists. Indeed, the conclusion then follows from (5.3). Thus suppose that $(\mathrm{YS})\int_a^b g\,df$ exists. Let $\tau=\{((x_{i-1},x_i),y_i):i=1,\dots,n\}$ be a Young tagged partition of $[a,b]$. By the assumption (5.4), we have

$$\begin{cases}\Delta^-f=\Delta^+f=(f_+-f_-)/2 \quad \text{on } (a,b),\\ \Delta^+g(a)=\Delta^-g(b)=0, \quad \Delta^-g=\Delta^+g=(g_+-g_-)/2 \quad \text{on } (a,b).\end{cases}$$

Therefore we have the representation

$$C_{\mathrm{YS}}(g,f;\tau)=\sum_{k=1}^{3}D_k(\tau)-\sum_{k=4}^{6}D_k(\tau),$$

where $D_k(\tau)=\sum_{i=1}^n d_{k,i}$, $k=1,\dots,6$, and for each $i=1,\dots,n$,

$$\begin{aligned}
d_{1,i}&:=[\Delta g\Delta f](y_i,x_i), & d_{4,i}&:=[\Delta g\Delta f](x_{i-1},y_i),\\
d_{2,i}&:=\Delta^+g(y_i)\Delta f(y_i,x_i), & d_{5,i}&:=\Delta^-g(y_i)\Delta f(x_{i-1},y_i),\\
d_{3,i}&:=\Delta g(y_i,x_i)\Delta^+f(y_i), & d_{6,i}&:=\Delta g(x_{i-1},y_i)\Delta^-f(y_i).
\end{aligned}$$

We show that $C_{\mathrm{YS}}(g,f;\tau)$ can be approximated arbitrarily closely by a sum of six differences of YS sums based on refinements of τ. First we have

$$\begin{aligned}
D_1(\tau)&=\sum_{i=1}^n g(x_i-)\Delta f(y_i,x_i)-\sum_{i=1}^n g(y_i+)\Delta f(y_i,x_i)\\
&=\lim_{\forall s_i\uparrow x_i}S_{\mathrm{YS}}(\tau_1(s))-\lim_{\forall t_i\downarrow y_i}S_{\mathrm{YS}}(\tau_2(t)),
\end{aligned}$$

where

$$\begin{aligned}
\tau_1(s)&=\{((x_{i-1},y_i),z_i),((y_i,x_i),s_i):i=1,\dots,n\},\\
\tau_2(t)&=\{((x_{i-1},y_i),z_i),((y_i,x_i),t_i):i=1,\dots,n\}.
\end{aligned}$$

Notice that each term in the two YS sums corresponding to the tagged intervals $((x_{i-1},y_i),z_i)$ cancel each other. Similarly we have

$$\begin{aligned}
D_4(\tau)&=\sum_{i=1}^n g(y_i-)\Delta f(x_{i-1},y_i)-\sum_{i=1}^n g(x_{i-1}+)\Delta f(x_{i-1},y_i)\\
&=\lim_{\forall s_i\uparrow y_i}S_{\mathrm{YS}}(\tau_3(s))-\lim_{\forall t_i\downarrow x_{i-1}}S_{\mathrm{YS}}(\tau_4(t)),
\end{aligned}$$

where

$$\begin{aligned}
\tau_3(s)&=\{((x_{i-1},y_i),s_i),((y_i,x_i),w_i):i=1,\dots,n\},\\
\tau_4(t)&=\{((x_{i-1},y_i),t_i),((y_i,x_i),w_i):i=1,\dots,n\}.
\end{aligned}$$

Next let $D_7(\tau):=4^{-1}\sum_{i=1}^n[\Delta^{\pm}g\Delta^{\pm}f](y_i)$. Then we have:

$$\begin{aligned}
D_2(\tau) &= \frac{1}{2}\sum_{i=1}^{n} \Delta^{\pm} g(y_i)\Delta f(y_i, x_i)\\
&= \frac{1}{2}\sum_{i=1}^{n} \{g(y_i-)\Delta f(x_{i-1}, y_i) + [g\Delta^{\pm} f](y_i) + g(y_i+)\Delta f(y_i, x_i)\\
&\qquad - g(y_i-)\Delta f(x_{i-1}, x_i)\} - \frac{1}{2}\sum_{i=1}^{n} \Delta^{-} g(y_i)\Delta^{\pm} f(y_i)\\
&= \frac{1}{2}\{\lim_{\forall t_i\uparrow y_i,\, s_i\downarrow y_i} S_{\mathrm{YS}}(\tau_5(t,s)) - \lim_{\forall t_i\uparrow y_i} S_{\mathrm{YS}}(\tau_6(t))\} - D_7(\tau),
\end{aligned}$$

where

$$\begin{aligned}
\tau_5(t,s) &= \{((x_{i-1}, y_i), t_i), ((y_i, x_i), s_i) : i = 1, \ldots, n\},\\
\tau_6(t) &= \{((x_{i-1}, x_i), t_i) : i = 1, \ldots, n\};
\end{aligned}$$

$$\begin{aligned}
D_5(\tau) &= \frac{1}{2}\sum_{i=1}^{n} \Delta^{\pm} g(y_i)\Delta f(x_{i-1}, y_i)\\
&= \frac{1}{2}\sum_{i=1}^{n} \{g(y_i+)\Delta f(x_{i-1}, x_i) - g(y_i-)\Delta f(x_{i-1}, y_i) - [g\Delta^{\pm} f](y_i)\\
&\qquad - g(y_i+)\Delta f(y_i, x_i)\} - \frac{1}{2}\sum_{i=1}^{n} \Delta^{+} g(y_i)\Delta^{\pm} f(y_i)\\
&= \frac{1}{2}\{\lim_{\forall s_i\downarrow y_i} S_{\mathrm{YS}}(\tau_7(s)) - \lim_{\forall t_i\uparrow y_i,\, s_i\downarrow y_i} S_{\mathrm{YS}}(\tau_5(t,s))\} - D_7(\tau),
\end{aligned}$$

where $\tau_7(s) := \{((x_{i-1}, x_i), s_i) : i = 1, \ldots, n\}$;

$$\begin{aligned}
D_3(\tau) &= \frac{1}{2}\sum_{i=1}^{n} \Delta g(y_i, x_i)\Delta^{\pm} f(y_i) = -\frac{1}{2}\sum_{i=1}^{n} \Delta^{+} g(y_i)\Delta^{\pm} f(y_i)\\
&\quad + \frac{1}{2}\sum_{i=1}^{n} \{g(x_i-)[f(x_i-) - f(y_i-)] + g(z_i)\Delta f(x_{i-1}, y_i)\\
&\quad - g(x_i-)\Delta f(y_i, x_i) - [g\Delta^{\pm} f](y_i) - g(z_i)\Delta f(x_{i-1}, y_i)\}\\
&= \frac{1}{2}\{\lim_{\forall u_i\uparrow y_i,\, t_i\uparrow x_i} S_{\mathrm{YS}}(\tau_8(u,t)) - \lim_{\forall t_i\uparrow x_i} S_{\mathrm{YS}}(\tau_9(t))\} - D_7(\tau),
\end{aligned}$$

where

$$\begin{aligned}
\tau_8(u,t) &= \{((x_{i-1}, u_i), z_i), ((u_i, x_i), t_i) : i = 1, \ldots, n\},\\
\tau_9(t) &= \{((x_{i-1}, y_i), z_i), ((y_i, x_i), t_i) : i = 1, \ldots, n\};
\end{aligned}$$

$$D_6(\tau) = \frac{1}{2}\sum_{i=1}^{n} \Delta g(x_{i-1}, y_i)\Delta^{\pm} f(y_i) = -\frac{1}{2}\sum_{i=1}^{n} \Delta^{-} g(y_i)\Delta^{\pm} f(y_i)$$
$$+ \frac{1}{2}\sum_{i=1}^{n}\{g(x_{i-1}+)\Delta f(x_{i-1}, y_i) + [g\Delta^{\pm} f](y_i) + g(z_i)\Delta f(y_i, x_i)$$
$$- g(x_{i-1}+)[f(y_i+) - f(x_{i-1}+)] - g(z_i)\Delta f(y_i, x_i)\}$$
$$= \frac{1}{2}\{\lim_{\forall s_i \downarrow x_{i-1}} S_{\mathrm{YS}}(\tau_{10}(s)) - \lim_{\forall s_i \downarrow x_{i-1},\, v_i \downarrow y_i} S_{\mathrm{YS}}(\tau_{11}(s, v))\} - D_7(\tau),$$

where

$$\tau_{10}(s) = \{((x_{i-1}, y_i), s_i), ((y_i, x_i), z_i) : i = 1, \dots, n\},$$
$$\tau_{11}(s, v) = \{((x_{i-1}, v_i), s_i), ((v_i, x_i), z_i) : i = 1, \dots, n\}.$$

Notice that the $D_7(\tau)$ enter into the representation of $C_{\mathrm{YS}}(g, f; \tau)$ with alternating signs. Since the Young partition τ is arbitrary, and $\tau_1, \dots, \tau_{11}$ are refinements of τ, we have shown that $C_{\mathrm{YS}}(g, f) = 0$, proving Theorem 5.6. ■

Next we prove a statement analogous to Proposition 5.3.

PROPOSITION 5.7. *Let $1/p + 1/q = 1$, and let one of the two functions g, f be in $\mathcal{W}_p^*([a, b])$ and the other in $\mathcal{W}_q([a, b])$. Suppose that g and f have values satisfying* (5.4) *at points of discontinuity. If either of the two integrals* (YS) $\int_a^b g\,df$ *and* (SYS) $\int_a^b g\,df$ *exists then so does the other*, *and the two integrals are equal.*

Proof. Let $f \in \mathcal{W}_p^*$ and let $g \in \mathcal{W}_q$. By the assumption (5.4), $\Delta^+ g\Delta^+ f = \Delta^- g\Delta^- f$ on (a, b) and $\Delta^+ g(a) = \Delta^- g(b) = 0$. For a Young tagged partition $\tau = \{((x_{i-1}, x_i), y_i) : i = 1, \dots, n\} \in TP([a, b])$, by Hölder's inequality, we have the bound

$$|\Xi_{\mathrm{YS}}(g, f; \tau) - S_{\mathrm{YS}}(g, f; \tau)| = \frac{1}{2}\Big|\sum_{i=1}^{n}\{[\Delta g\Delta f][y_i, x_i) - [\Delta g\Delta f](x_{i-1}, y_i]\}\Big|$$
$$\le v_q(g; [a, b])^{1/q}\Big(\sum_{i=1}^{n} v_p(f; (x_{i-1}, x_i))\Big)^{1/p}.$$

By Lemma 4.1, given $\varepsilon > 0$ the right side of the last bound is ultimately less than ε along the direction $\mathfrak{P}$. Thus the directed functions $(S_{\mathrm{YS}}(g, f; \cdot), \mathfrak{R})$ and $(\Xi_{\mathrm{YS}}(g, f; \cdot), \mathfrak{R})$ both converge or not simultaneously, and if they do then both have the same limit. The same conclusion follows by the same arguments when f and g are interchanged. ■

Now we are prepared for the main result of this section. We write $\phi \in \Lambda_{2,\alpha}$ if ϕ has a second derivative satisfying a Lipschitz condition of order α on every bounded interval.

THEOREM 5.8. *For $\alpha \in (0,1]$, let $\phi \in \Lambda_{2,\alpha}$, and let $g := f + h$, where $f \in \mathcal{W}^*_{2+\alpha}([a,b])$ and $h \in \mathcal{W}_p([a,b])$ for some $p < (2+\alpha)/(1+\alpha)$. The composition $\phi' \circ g$ is SYS and YS integrable with respect to f and h, respectively, and the two integrals satisfy*

$$\begin{aligned}(5.5)\qquad & (\phi \circ g)(b) - (\phi \circ g)(a) \\ &= (\mathrm{SYS})\int_a^b (\phi' \circ g)\,df + (\mathrm{YS})\int_a^b (\phi' \circ g)\,dh \\ &\quad + \sum_{[a,b)} \Big\{ \Delta^+\phi \circ g - \phi' \circ g \Delta^+ g - \frac{\Delta^+\phi' \circ g}{2}\Delta^+ f \Big\} \\ &\quad + \sum_{(a,b]} \Big\{ \Delta^-\phi \circ g - \phi' \circ g \Delta^- g + \frac{\Delta^-\phi' \circ g}{2}\Delta^- f \Big\},\end{aligned}$$

where the two sums are unconditional. If in addition f and h are both continuous, then the composition $\phi' \circ g$ is SRS and RS integrable with respect to f and h, respectively, and the two integrals satisfy

$$(5.6)\qquad (\phi \circ g)(b) - (\phi \circ g)(a) = (\mathrm{SRS})\int_a^b (\phi' \circ g)\,df + (\mathrm{RS})\int_a^b (\phi' \circ g)\,dh.$$

REMARK. This theorem shows that the SRS and SYS integrals can exist and satisfy a suitable chain rule formula when the RS and YS integrals are not defined. To see this, take $h \equiv 0$, $f \in \mathcal{W}^*_{2+\alpha} \setminus \mathcal{W}^*_2$, and recall Proposition 4.4.

Proof. Let $\kappa = \{x_i : i = 0, \ldots, n\} \in P([a,b])$, and let $\tau = \tau(\kappa) = \{((x_{i-1}, x_i), y_i) : i = 1, \ldots, n\}$ be a Young tagged partition of $[a,b]$. By a telescoping sum, we have

$$\begin{aligned}(5.7)\qquad & \Delta(\phi \circ g)[a,b] \\ &= \sum_{i=1}^n \{\Delta^+(\phi \circ g)(x_{i-1}) + \Delta(\phi \circ g)(x_{i-1}, x_i) + \Delta^-(\phi \circ g)(x_i)\} \\ &= \Xi_{\mathrm{YS}}(\phi' \circ g, f; \tau) + S_{\mathrm{YS}}(\phi' \circ g, h; \tau) + S_+(\kappa) + S_-(\kappa) + R(\tau),\end{aligned}$$

where

$$S_+(\kappa) := \sum_{i=1}^n \Big(\Delta^+(\phi \circ g) - (\phi' \circ g)\Delta^+ g - \frac{1}{2}\Delta^+(\phi' \circ g)\Delta^+ f \Big)(x_{i-1}),$$

$$S_-(\kappa) := \sum_{i=1}^n \Big(\Delta^-(\phi \circ g) - (\phi' \circ g)\Delta^- g + \frac{1}{2}\Delta^-(\phi' \circ g)\Delta^- f \Big)(x_i)$$

and $R(\tau) := \sum_{i=1}^n r_i$ with

$$r_i := \Delta(\phi\circ g)(x_{i-1},x_i) - (\phi'\circ g)(y_i)\Delta g(x_{i-1},x_i) + \frac{\Delta(\phi'\circ g)(x_{i-1},y_i]}{2}\Delta f(x_{i-1},y_i] - \frac{\Delta(\phi'\circ g)[y_i,x_i)}{2}\Delta f[y_i,x_i).$$

To bound r_i we use Taylor's theorem with Lagrange's form of the remainder:

$$\phi(v) = \phi(u) + \phi'(u)[v-u] + \frac{1}{2}\phi''(\theta)[v-u]^2, \tag{5.8}$$

where θ is in the interval with endpoints u, v, and then apply the mean value theorem. It then follows that for each $i = 1,\dots,n$,

$$\begin{aligned}
2r_i &= \phi''(\theta_{i,1})(\Delta g[y_i,x_i))^2 - \Delta(\phi'\circ g)[y_i,x_i)\Delta f[y_i,x_i)\\
&\quad - \phi''(\theta_{i,2})(\Delta g(x_{i-1},y_i])^2 + \Delta(\phi'\circ g)(x_{i-1},y_i]\Delta f(x_{i-1},y_i]\\
&= [\phi''(\theta_{i,1}) - \phi''(\vartheta_{i,1})](\Delta g[y_i,x_i))^2 + \phi''(\vartheta_{i,1})\Delta g[y_i,x_i)\Delta h[y_i,x_i)\\
&\quad - [\phi''(\theta_{i,2}) - \phi''(\vartheta_{i,2})](\Delta g(x_{i-1},y_i])^2\\
&\quad - \phi''(\vartheta_{i,2})\Delta g(x_{i-1},y_i]\Delta h(x_{i-1},y_i],
\end{aligned}$$

where $\theta_{i,1},\vartheta_{i,1}$ are in the interval with endpoints $g(y_i), g(x_i-)$, and $\theta_{i,2},\vartheta_{i,2}$ are in the interval with endpoints $g(x_{i-1}+), g(y_i)$. Using Lipschitz continuity of ϕ'', and applying the W. H. Young inequality with the conjugate exponents $\beta := 2+\alpha$ and $\beta' := (2+\alpha)/(1+\alpha)$, it then follows that

$$|R(\tau)| \le \left(\frac{K_\alpha}{2} + \frac{\|\phi''\|_\infty}{2\beta}\right)\sum_{i=1}^n v_\beta(g;(x_{i-1},x_i)) + \frac{\|\phi''\|_\infty}{2\beta'}\sum_{i=1}^n v_{\beta'}(h;(x_{i-1},x_i)).$$

By Lemma 4.1, given $\varepsilon > 0$ the right side of the preceding bound is ultimately less than ε. Therefore the limit of the directed function $(R,\mathfrak{R})$ is zero. To ascertain that the first sum in (5.6) is the limit of the directed function $(S_+,\mathfrak{P})$ it is enough to show its absolute convergence. Using Taylor's theorem (5.8) and the mean value theorem, for $x \in [a,b)$ we get

$$\begin{aligned}
\delta_+(x) &:= \Big(\Delta^+(\phi\circ g) - (\phi'\circ g)\Delta^+ g - \frac{1}{2}\Delta^+(\phi'\circ g)\Delta^+ f\Big)(x)\\
&= \frac{1}{2}[\phi''(\theta) - \phi''(\vartheta)](\Delta^+ g(x))^2 + \frac{1}{2}\phi''(\vartheta)[\Delta^+ g\Delta^+ h](x),
\end{aligned}$$

where θ,ϑ are points in the interval with endpoints $g(x), g(x+)$. Using Lipschitz continuity of ϕ'', and applying the W. H. Young inequality with the same exponents β,β' as above, we get the bound

$$\sum_{x\in\mu}|\delta_+(x)| \le \left(\frac{K_\alpha}{2} + \frac{\|\phi''\|_\infty}{2\beta}\right)\sum_{[a,b)}|\Delta^+ g|^\beta + \frac{\|\phi''\|_\infty}{2\beta'}\sum_{[a,b)}|\Delta^+ h|^{\beta'} < \infty$$

for any finite subset μ of $[a,b)$. Thus the first sum in (5.5) converges absolutely, and hence unconditionally. Likewise, the second sum in (5.5) is the unconditional limit of S_-. Since $\phi'\circ g \in \mathcal{W}_{2+\alpha}$, $h \in \mathcal{W}_p$ and $(2+\alpha)^{-1} + p^{-1}$

> 1, the YS integral in (5.5) exists by Theorem 5.1 of L. C. Young [40] (cf. Proposition 2.8 above). Therefore the limit of the directed function $(S_{\mathrm{YS}}(\phi' \circ g, h; \cdot), \mathfrak{R})$ exists. By (5.7), it then follows that the directed function $(\Xi_{\mathrm{YS}}(\phi' \circ g, f; \cdot), \mathfrak{R})$ has a limit which satisfies the relation (5.5). This proves the first part of the theorem. The second part follows by the same arguments in conjunction with the fact that

$$v_p^*(f; [a, b]) = \lim_{\delta \downarrow 0} \sup\{s_p(f; \kappa) : \kappa \in P([a, b]),\ |\kappa| < \delta\}$$

provided f is continuous (see e.g. Lemma B.1 in Dudley and Norvaiša [6, Part I]). ■

To illustrate the results of the present section consider a fractional Brownian motion $B_H = \{B_H(t) : t \geq 0\}$ with the Hurst exponent $H \in (0, 1)$. It is a mean zero Gaussian stochastic process with the covariance function

$$E\{B_H(t)B_H(s)\} = \frac{1}{2}\{t^{2H} + s^{2H} - |t - s|^{2H}\}, \qquad t, s \geq 0,$$

and $B_H(0) = 0$ almost surely. B_H with $H = 1/2$ is a standard Brownian motion B. If $H \neq 1/2$ then B_H is not a semimartingale, and hence stochastic integration with respect to B_H meets certain difficulties. Almost all sample functions of B_H restricted to a bounded interval satisfy a Lipschitz condition of order α for each $\alpha < H$, thus almost surely $v_p(B_H; [0, 1]) < \infty$ for each $p > 1/H$. But $v_{1/H}(B_H; [0, 1]) = +\infty$ almost surely. If $H > 1/2$ then for almost all sample functions of B_H the quadratic covariation $C_{\mathrm{RS}}(B_H, B_H)$ on $[0, 1]$ exists and is 0. Also in this case, $\int_0^1 B_H\, dB_H$ exists for almost all sample functions both as an RS and SRS integral, and the relation

$$(\mathrm{RS})\int_0^1 B_H\, dB_H = (\mathrm{SRS})\int_0^1 B_H\, dB_H = \frac{1}{2}B_H(1)^2.$$

If $H \leq 1/2$ then the quadratic covariation $C_{\mathrm{RS}}(B_H, B_H)$ is undefined. However by Theorem 5.8, the SRS integral in the preceding relation exists and the second equality holds provided $1/3 < H \leq 1/2$.

In the case $H = 1/2$, that is, when B_H is a standard Brownian motion B, one can compare (5.6) for $g = f$ equal to a sample function of B, with the Itô formula. To this end we split the SRS integral into two parts by separating an extended RS integral and the RS integral with respect to a quadratic variation. Let $\lambda = \{\kappa_m : m \geq 1\}$ be a nested sequence of partitions $\kappa_m = \{x_i^m : i = 0, \ldots, n(m)\}$ of $[a, b]$ such that $\bigcup_m \kappa_m$ is dense in $[a, b]$, and let each $\tau_m^l = \tau_m^l(\kappa_m) = \{([x_{i-1}^m, x_i^m], x_{i-1}^m) : i = 1, \ldots, n(m)\}$ be a tagged partition of $[a, b]$. Suppose that (SRS) $\int_a^b g\, df$ exists. By (5.2), it then follows that the limits

$$\int_a^b g\, d_\lambda f := \lim_{m\to\infty} S_{\mathrm{RS}}(g,f;\tau_m^l) \quad \text{and} \quad C_\lambda(g,f) := \lim_{m\to\infty} C_{\mathrm{RS}}(g,f;\tau_m^l)$$

both exist or not simultaneously. If they do, then

$$(\mathrm{SRS})\int_a^b g\, df = \int_a^b g\, d_\lambda f + \frac{1}{2} C_\lambda(g,f). \tag{5.9}$$

Suppose that ϕ and f satisfy the assumptions of the second part of Theorem 5.8 with f being continuous and $h \equiv 0$. Then (SRS) $\int_a^b (\phi' \circ f)\, df$ is defined, and so the left side of (5.9) is defined for $g = \phi' \circ f$. To show the existence of $C_\lambda(\phi' \circ f, f)$ we assume that f in addition has the quadratic λ-variation in the sense of Definition 3.1 in [30]. For a given λ, these assumptions on f are satisfied by almost every sample function of a standard Brownian motion. We say that a (continuous) function f has the *quadratic λ-variation* if there exists a continuous non-decreasing function $[f]_\lambda$ on $[a,b]$ such that for all $a \le x < y \le b$,

$$\lim_{m\to\infty} \sum_{i=1}^{n(m)} [f(x_i^m) - f(x_{i-1}^m)]^2 \mathbb{1}_{[x,y]}(x_{i-1}^m) = [f]_\lambda(y) - [f]_\lambda(x).$$

By the mean value theorem, for each $m \ge 1$, we have

$$\begin{aligned} C_{\mathrm{RS}}(\phi' \circ f, f; \tau_m^l) &= \sum_{i=1}^{n(m)} [\phi' \circ f(x_i^m) - \phi' \circ f(x_{i-1}^m)][f(x_i^m) - f(x_{i-1}^m)] \\ &= \sum_{i=1}^{n(m)} \phi'' \circ f(x_{i-1}^m)[f(x_i^m) - f(x_{i-1}^m)]^2 + R_m. \end{aligned}$$

By the assumptions on ϕ and f, one can ascertain that $\lim_{m\to\infty} R_m = 0$. On the other hand, by Lemma 3.8 of [30], since $[f]_\lambda$ is continuous, we have

$$\lim_{m\to\infty} \sum_{i=1}^{n(m)} \phi'' \circ f(x_{i-1}^m)[f(x_i^m) - f(x_{i-1}^m)]^2 = (\mathrm{RS})\int_a^b \phi'' \circ f\, d[f]_\lambda,$$

and so $C_\lambda(\phi' \circ f, f)$ is equal to the right side. By (5.2), (5.6) and (5.9), it then follows that $\int_a^b (\phi' \circ f)\, d_\lambda f$ exists and

$$\begin{aligned} (\phi \circ f)(b) - (\phi \circ f)(a) &= \int_a^b (\phi' \circ f)\, d_\lambda f + \frac{1}{2}(\mathrm{RS})\int_a^b (\phi'' \circ f)\, d[f]_\lambda \\ &= (\mathrm{SRS})\int_a^b (\phi' \circ f)\, df. \end{aligned}$$

Recall that here $\phi \in \Lambda_{2,\alpha}$, and f has both the quadratic λ-variation and bounded p-variation for some $2 < p < 2 + \alpha \leq 3$. The analogous formula for stochastic processes is called the Itô formula, where C^2 smoothness of ϕ can be weakened considerably. For a Brownian motion B, Föllmer, Protter and Shiryayev [11] (see also [32]) proved that $C_{\mathrm{RS}}(\phi' \circ B, B; \tau_m^l)$ converges in probability as $m \to \infty$ provided ϕ' is locally square integrable.

6. A chain rule formula. In this section we discuss the results of the paper in the context of analysis. In particular, we rewrite formula (1.3) using a variant of the notion of differential equivalence of Kolmogorov [16].

Suppose that the functions f and ϕ have finite derivatives f' and ϕ' everywhere on their domains. Then the composition $\phi \circ f$ is also differentiable everywhere on $[a, b]$, and its derivative can be expressed by the formula

$$(\phi \circ f)' = (\phi' \circ f) \cdot f'. \tag{6.1}$$

This is a typical example of a chain rule for differentiation of a composition of functions from the line to the line. We are interested in what sense the integral representation (1.3) of $\phi \circ f$ might be compared with (6.1). It is well known that for mappings between more general normed spaces than the real numbers, the chain rule may fail to hold for certain forms of differentiability (see §2 in Averbukh and Smolyanov [2] for details). Then the chain rule can be given a more general form which includes such cases. Let two forms of differentiation $\mathcal{R}_1$ and $\mathcal{R}_2$ for mappings between normed spaces X, Y, Z be given. Dudley and Norvaiša [6, Part I, Section 8] say that $\mathcal{R}_2$ *preserves* $\mathcal{R}_1$ if whenever $F : X \to Y$ is $\mathcal{R}_1$-differentiable at $x \in X$ with derivative DF_x and $G : Y \to Z$ is $\mathcal{R}_2$-differentiable at $y = F(x)$ with derivative DG_y then $G \circ F$ is $\mathcal{R}_1$-differentiable at x, and the chain rule formula $D(G \circ F)_x = DG_y \circ DF_x$ holds. Several known differentiability facts, which do not have a typical form of the chain rule, can be restated by saying that one form of differentiability preserves another form of differentiability. It is clear that the main result of the present paper can be restated as follows: let f have a property $\mathcal{P}$ (have bounded p-variation) and let ϕ be a smooth function; then $\phi \circ f$ also has the property $\mathcal{P}$ and a variant of a chain rule formula holds (an integral representation of the composition holds). Usually, it is easy to establish the property $\mathcal{P}$ for $\phi \circ f$, while the chain rule formula, different in each case, is the main problem.

6.1. *An integral chain rule formula.* Using classical real analysis arguments, we show that the chain rule formula (6.1) can be given an integral form similar to the formula (1.4). The following chain rule is due to Serrin and Varberg [34, Theorem 2].

THEOREM A. *Let ϕ, f and $\phi \circ f$ have finite derivatives Lebesgue almost everywhere on their domains. If ϕ maps Lebesgue null sets into Lebesgue null sets then the chain rule* (6.1) *holds Lebesgue almost everywhere.*

Serrin and Varberg [34, Theorem 3] used their chain rule to prove the following variant of a change of variables formula:

THEOREM B. *Suppose that f has a finite derivative Lebesgue almost everywhere on $[a,b]$ and that ϕ is absolutely continuous on $[c,d]$. Then $(\phi'\circ f)f'$ is Lebesgue integrable on $[a,b]$, and the equality of the two Lebesgue integrals*

$$\int_{[f(\alpha),f(\beta)]} \phi' = \int_{[\alpha,\beta]} (\phi' \circ f)f' \tag{6.2}$$

holds for all α, β such that $a \le \alpha \le \beta \le b$, if and only if $\phi \circ f$ is absolutely continuous.

If f is *continuous and increasing* on $[a,b]$, and hence has a finite derivative Lebesgue almost everywhere on $[a,b]$, then a form can be given to formula (6.2) which does not depend on the absolute continuity of $\phi \circ f$. Indeed, suppose that $\phi \circ f$ is absolutely continuous. Since ϕ is absolutely continuous, by the fundamental theorem of calculus, the left side of (6.2) is equal to $(\phi\circ f)(\beta) - (\phi\circ f)(\alpha)$. Also, since $(\phi'\circ f)f'$ is Lebesgue integrable, one can apply Theorem 4 of Phillips [31, p. 407] to conclude that $\phi' \circ f$ is Lebesgue–Stieltjes integrable with respect to f and the integral on the right side of (6.2) is equal to the corresponding Lebesgue–Stieltjes integral. Thus, by (6.2), we have

$$(\phi\circ f)(b) - (\phi\circ f)(a) = (\mathrm{LS}) \int_{[a,b]} (\phi' \circ f)\, df. \tag{6.3}$$

To see that formula (6.3) is more general than (6.2), consider a continuous, strictly increasing function f such that $f' = 0$ Lebesgue almost everywhere (cf. Chapter III, §13 of Saks [33]), and let $\phi(x) := x$. Then $\phi \circ f$ is not absolutely continuous because f is not. Thus, (6.2) is not true whereas (6.3) holds for this example.

6.2. *A distributional derivative of a composition.* The following analog of a chain rule is developed in geometric measure theory. We reproduce here only its simplest variant in the case of functions of one variable. Let Ω be a bounded open subset of $\mathbb{R}$. A function $f : \Omega \to \mathbb{R}$ is a *BV function* if its distributional derivative $\mathbf{f}'$ can be represented by a measure on Ω. The latter means that there exists a Radon measure *df* on Ω such that the Lebesgue integral $\int_\Omega h'f = -\int_\Omega h\,df$ for each $h \in C_0^\infty(\Omega)$, i.e. for each infinitely differentiable function h with support in Ω. BV functions are Lebesgue almost everywhere equal to functions of bounded variation. Recall that functions whose distributional derivatives are functions are Lebesgue almost everywhere equal to absolutely continuous functions.

To establish a differentiability formula similar to (6.1) Vol'pert [37, pp. 279, 281] introduced the *functional composition* $\phi' \,\hat{\circ}\, f$ and proved the following statement:

THEOREM C. *Let* $\phi : \mathbb{R} \to \mathbb{R}$ *be a* C^1 *function, and let* f *be a summable* BV *function on* Ω *such that* $\phi' \,\hat{\circ}\, f$ *is integrable with respect to* df *on* Ω. *Then* $\phi \circ f$ *is a* BV *function on* Ω *and* $\phi' \circ f = \phi' \,\hat{\circ}\, f\, df$ *in the sense of measures, i.e., for each* $h \in C_0^\infty(\Omega)$,

$$\int_\Omega h\, d(\phi \circ f) = \int_\Omega h(\phi' \,\hat{\circ}\, f)\, df. \tag{6.4}$$

If the ordinary composition "$\circ$" is used instead of "$\hat{\circ}$" then the right side of (6.4) breaks into two parts. The first part is the integral over $\Omega \setminus D_f$, where f has an approximate limit, and the rest is the sum over jump points D_f. See Ambrosio and Dal Maso [1] for a precise statement and for several extensions of Theorem C.

6.3. *Inequalities for the density of a composition.* For a convex continuous function ϕ on a Banach space, a subdifferential $\partial\phi$ in the sense of convex analysis has been used by Moreau and Valadier [28] to establish inequalities for the density of $\phi \circ f$ when f is a Banach space valued interval function of locally bounded variation. They also obtained a related result using Clarke's generalized gradient when ϕ is a Lipschitz function. We restrict ourselves here to reproducing a special case of their first result when the convex function ϕ is differentiable and f has real values. In this case the subdifferential $\partial\phi(x)$ at a point x reduces to the single element $\nabla\phi(x)$, the gradient of ϕ at this point.

THEOREM D. *Let* Ω *be an open convex subset of* $\mathbb{R}$ *and let* $\phi : \Omega \to \mathbb{R}$ *be a convex differentiable function with gradient* $\nabla\phi$. *Let* $f : I \to \Omega$, $I \subset [a, b]$, *be an interval function of bounded variation with the Stieltjes measure* $df = f'_\mu d\mu$. *Then* $\phi \circ f$ *is of bounded variation and its Stieltjes measure has, relative to* μ, *a density* $(\phi \circ f)'_\mu$ *such that the inequalities*

$$(\nabla\phi \circ f_-) \cdot f'_\mu \le (\phi \circ f)'_\mu \le (\nabla\phi \circ f_+) \cdot f'_\mu \tag{6.5}$$

hold μ*-almost everywhere on* I, *where* f_- *and* f_+ *are the left-continuous and right-continuous modifications of* f. *Instead of* (6.5), *the equality* $(\phi \circ f)'_\mu = (\nabla\phi \circ f) \cdot f'_\mu$ *holds* μ*-almost everywhere on the set* $\{x \in I : df(\{x\}) = 0\}$.

6.4. *Young–Stieltjes differential equivalence.* The integral formula (1.3) can be given a differential form similar to (6.1). We show this by adopting the notion of differential equivalence proposed by Kolmogorov [16, Section II.15].

A different route of development of a differential equivalence has been taken by Leader [19] and Thomson [36]. Consider two pairs (g_1, f_1) and (g_2, f_2) of functions defined on $[a, b]$. Suppose that f_1 and f_2 are regulated. For a partition $\kappa = \{x_i : i = 0, \ldots, n\} \in P([a, b])$, let

$$S_-(\kappa) := \sum_{i=1}^{n} (g_1 \Delta^- f_1 - g_2 \Delta^- f_2)(x_i), \tag{6.6}$$

$$S_+(\kappa) := \sum_{i=1}^{n} (g_1 \Delta^+ f_1 - g_2 \Delta^+ f_2)(x_{i-1}), \tag{6.7}$$

and for a Young tagged partition $\tau = \tau(\kappa) = \{((x_{i-1}, x_i), y_i) : i = 1, \ldots, n\}$, let

$$D(\tau(\kappa)) := \sum_{i=1}^{n} \{g_1(y_i)\Delta f_1(x_{i-1}, x_i) - g_2(y_i)\Delta f_2(x_{i-1}, x_i)\}. \tag{6.8}$$

We define the Young–Stieltjes differential equivalence as a relation between two pairs of functions.

Definition 6.1. Let $f_1, f_2 \in \mathcal{R}([a, b])$ and let $g_1, g_2 : [a, b] \to \mathbb{R}$. We say that the pair (g_1, f_1) of functions *is Young–Stieltjes differentially equivalent on* $[a, b]$ to the pair (g_2, f_2), and write $g_1\, df_1 \overset{\text{(YS)}}{=} g_2\, df_2$ on $[a, b]$, if:

(a) the directed function $(D, \mathfrak{R})$ has the limit $\lim_{\tau, \mathfrak{R}} D(\tau) = 0$;

(b) the directed functions $(S_-, \mathfrak{P})$ and $(S_+, \mathfrak{P})$ have finite limits

$$\lim_{\kappa, \mathfrak{P}} S_-(\kappa) = \sum_{(a,b]} [g_1 \Delta^- f_1 - g_2 \Delta^- f_2]$$

and

$$\lim_{\kappa, \mathfrak{P}} S_+(\kappa) = \sum_{[a,b)} [g_1 \Delta^+ f_1 - g_2 \Delta^+ f_2].$$

The pairs (g_1, f_1) and (g_2, f_2) may be Young–Stieltjes differentially equivalent but neither $(\mathrm{YS})\int_a^b g_1\, df_1$ nor $(\mathrm{YS})\int_a^b g_2\, df_2$ need exist. Next we show that if any one of the two integrals exists then so does the other.

Proposition 6.2. *Let $f_1, f_2 \in \mathcal{R}([a, b])$ and let $g_1, g_2 : [a, b] \to \mathbb{R}$. Suppose that either* $(\mathrm{YS})\int_a^b g_1\, df_1$ *or* $(\mathrm{YS})\int_a^b g_2\, df_2$ *exists. Then the following statements are equivalent*:

(1) $g_1\, df_1 \overset{\text{(YS)}}{=} g_2\, df_2$ *on* $[a, b]$;

(2) *the integrals* $(\mathrm{YS})\int_a^b g_1\, df_1$ *and* $(\mathrm{YS})\int_a^b g_2\, df_2$ *are defined and*

$$(6.9)\qquad (\mathrm{YS})\int_a^b g_1\,df_1 = (\mathrm{YS})\int_a^b g_2\,df_2 + \sum_{(a,b]}[g_1\Delta^- f_1 - g_2\Delta^- f_2] + \sum_{[a,b)}[g_1\Delta^+ f_1 - g_2\Delta^+ f_2],$$

where the two sums converge unconditionally.

Proof. Let $\kappa = \{x_i : i = 0,\dots,n\} \in P([a,b])$, and let $\tau = \tau(\kappa) = \{((x_{i-1},x_i),y_i) : i = 1,\dots,n\}$ be a Young tagged partition. Then

$$(6.10)\qquad S_{\mathrm{YS}}(g_1,f_1;\tau) - S_{\mathrm{YS}}(g_2,f_2;\tau) = D(\tau) + S_-(\kappa) + S_+(\kappa),$$

where D, S_- and S_+ are defined by (6.8), (6.6) and (6.7), respectively. Suppose that (1) holds. By Definition 6.1, the three terms on the right side of (6.10) are directed functions which have limits. Then each Young–Stieltjes sum on the left side of (6.10) has a limit whenever one of them has a limit. Thus (2) holds. To prove the converse it is enough to notice that $\lim_{\tau,\mathfrak{R}} D(\tau) = 0$ by (6.10) and (6.9). ■

If $g_1 \equiv 1$ then g_1 is YS integrable with respect to any regulated function f_1, and $(\mathrm{YS})\int_a^b g_1\,df_1 = f_1(b) - f_1(a)$. Thus we have

COROLLARY 6.3. *Let $f \in \mathcal{R}([a,b])$, and let ϕ, ψ be real-valued functions defined on the range of f. Then the following statements are equivalent*:

(1) $d(\phi\circ f) \overset{(\mathrm{YS})}{=} (\psi\circ f)\,df$ *on* $[a,b]$;

(2) *the integral* $(\mathrm{YS})\int_a^b(\psi\circ f)\,df$ *is defined, and*

$$(\phi\circ f)(b) - (\phi\circ f)(a) = (\mathrm{YS})\int_a^b(\psi\circ f)\,df + \sum_{(a,b]}[\Delta^-(\phi\circ f) - (\psi\circ f)\Delta^- f] + \sum_{[a,b)}[\Delta^+(\phi\circ f) - (\psi\circ f)\Delta^+ f],$$

where the two sums converge unconditionally.

If f and ϕ satisfy the conditions of Theorem 1.1, then the statement (2) of the preceding corollary holds with $\psi = \phi'$. Thus the conclusion of Theorem 1.1 can be restated as the Young–Stieltjes differential equivalence relation

$$d(\phi\circ f) \overset{(\mathrm{YS})}{=} (\phi'\circ f)\,df \quad \text{on } [a,b],$$

which has the desirable differential form similar to (6.1).

References

[1] L. Ambrosio and G. Dal Maso, *A general chain rule for distributional derivatives*, Proc. Amer. Math. Soc. 108 (1990), 691–702.

[2] V. I. Averbukh and O. G. Smolyanov, *The various definitions of the derivative in linear topological spaces*, Math. Surveys 23 (1968), 67–113 = Uspekhi Mat. Nauk 23 (1968), 67–116 (in Russian).

[3] Z. Ciesielski, G. Kerkyacharian et B. Roynette, *Quelques espaces fonctionnels associés à des processus gaussiens*, Studia Math. 107 (1993), 171–204.

[4] R. M. Dudley, *Sample functions of the Gaussian process*, Ann. Probab. 1 (1973), 66–103.

[5] —, *Fréchet differentiability, p-variation and uniform Donsker classes*, ibid. 20 (1992), 1968–1982.

[6] R. M. Dudley and R. Norvaiša, *Differentiability of Six Operators on Nonsmooth Functions and p-Variation*, Lecture Notes in Math. 1703, Springer, Berlin, 1999.

[7] —, —, *An Introduction to p-Variation and Young Integrals*, Lecture Notes, No. 1, MaPhySto, Univ. of Aarhus, 1998.

[8] —, —, *Concrete Functional Calculus*, book in preparation.

[9] A. M. D'yačkov, *Conditions for the existence of Stieltjes integral of bounded generalized variation*, Anal. Math. (Budapest) 14 (1988), 295–313.

[10] H. Föllmer, *Calcul d'Itô sans probabilités*, in: Séminaire de Probabilités XV, 1979/80, J. Azéma and M. Yor (eds.), Lecture Notes in Math. 850, Springer, Berlin, 1981, 143–150.

[11] H. Föllmer, P. Protter and A. N. Shiryayev, *Quadratic covariation and an extension of Itô's formula*, Bernoulli 1(1/2) (1995), 149–169.

[12] R. Henstock, *Theory of Integration*, Butterworths, London, 1963.

[13] E. W. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series*, Vol. 1, Dover, New York, 1957.

[14] T. H. Hildebrandt, *Introduction to the Theory of Integration*, Academic Press, New York, 1963.

[15] K. Itô, *Stochastic differentials*, Appl. Math. Optim. 1 (1975), 374-381. Reprinted in: *Kiyosi Itô Selected Papers*, D. W. Stroock and S. R. S. Varadhan (eds.), Springer, New York, 1987, 558–565.

[16] A. N. Kolmogorov, *Untersuchungen über Integralbegriff*, Math. Ann. 103 (1930), 654–696. English transl. in: Selected Works of A. N. Kolmogorov, Vol. I, Mathematics and Mechanics, V. M. Tikhomirov (ed.), Kluwer, Dordrecht, 1991, 100–143.

[17] J. Kurzweil, *Generalized ordinary differential equations and continuous dependence on a parameter*, Czechoslovak Math. J. 7 (1957), 568–583 (in Russian).

[18] —, *On integration by parts*, ibid. 8 (1958), 356–359.

[19] S. Leader, *The Kurzweil–Henstock Integral and its Differentials*, Dekker, New York, 2001.

[20] R. Leśniewicz and W. Orlicz, *On generalized variations (II)*, Studia Math. 45 (1973), 71–109.

[21] P. Lévy, *Le mouvement brownien plan*, Amer. J. Math. 62 (1940), 487–550.

[22] —, *Intégrales de Stieltjes généralisées*, Ann. Soc. Polon. Math. 25 (1952), 17–26. Reprinted in: Oeuvres de Paul Lévy, Vol. II, Gauthier-Villars, Ecole Polytechnique, Paris, 1974, 473–482.

[23] T. Lyons, *On the non-existence of path integrals*, Proc. Roy. Soc. London Ser. A 432 (1991), 281–290.

[24] T. Lyons, *Differential equations driven by rough signals*, Rev. Mat. Iberoamericana 14 (1998), 215–310.
[25] R. M. McLeod, *The Generalized Riemann Integral*, Carus Math. Monographs 20, Math. Assoc. Amer., 1980.
[26] E. J. McShane, *A theory of limits*, in: Studies in Modern Analysis, Vol. 1, R. C. Buk (ed.), Math. Assoc. Amer., 1962, 7–29.
[27] E. J. McShane and T. A. Botts, *Real Analysis*, Van Nostrand, Princeton, 1959.
[28] J. J. Moreau and M. Valadier, *A chain rule involving vector functions of bounded variation*, J. Funct. Anal. 74 (1987), 333–345.
[29] R. Norvaiša, *p-variation and integration of sample functions of stochastic processes*, in: Probab. Theory and Math. Statist. (Vilnius, 1998), B. Grigelionis *et al.* (eds.), TEV, 1999, 521–540.
[30] —, *Quadratic variation, p-variation and integration with applications to stock price modelling*, preprint, The Fields Institute, Toronto, 2001.
[31] E. R. Phillips, *An Introduction to Analysis and Integration Theory*, with a new historical introduction and notes by the author, Dover, New York, 1984.
[32] F. Russo and P. Vallois, *Itô formula for C^1-functions of semimartingales*, Probab. Theory Related Fields 104 (1996), 27–41.
[33] S. Saks, *Theory of the Integral*, transl. by L. C. Young, Hafner, New York, 1937.
[34] J. Serrin and D. E. Varberg, *A general chain rule for derivatives and the change of variables formula for the Lebesgue integral*, Amer. Math. Monthly 76 (1969), 514–520.
[35] H. L. Smith, *On the existence of the Stieltjes integral*, Trans. Amer. Math. Soc. 27 (1925), 491–515.
[36] B. S. Thomson, *Some remarks on differential equivalence*, Real Anal. Exchange 12 (1986-87), 294–312.
[37] A. I. Vol'pert, *The spaces BV and quasi-linear equations*, Mat. Sb. 73 (115) (1967), 255–302 (in Russian).
[38] A. J. Ward, *The Perron–Stieltjes integral*, Math. Z. 41 (1936), 578–604.
[39] L. C. Young, *An inequality of the Hölder type, connected with Stieltjes integration*, Acta Math. 67 (1936), 251–282.
[40] —, *General inequalities for Stieltjes integrals and the convergence of Fourier series*, Math. Ann. 115 (1938), 581–612.
[41] —, *A converse to some inequalities and approximations in the theory of Stieltjes and stochastic integrals, and for nth derivatives*, Studia Math. 55 (1976), 215–223.
[42] W. H. Young, *On integration with respect to a function of bounded variation*, Proc. London Math. Soc. (2) 13 (1914), 109-150.
[43] M. Zähle, *Integration with respect to fractal functions and stochastic calculus. I*, Probab. Theory Related Fields 111 (1998), 333–374.

Institute of Mathematics and Informatics
Akademijos 4
Vilnius 2600, Lithuania
E-mail: norvaisa@ktl.mii.lt

Received June 24, 1999
Revised version May 28, 2001 (4351)