

Nonlinear stochastic integration with a non-smooth family of integrators

Christoph Kühn*

Abstract

We consider nonlinear stochastic integrals of Itô-type w.r.t. a family of semimartingales which depend on a spatial parameter. These integrals were introduced by Carmona and Nualart [2], Kunita [8], and Le Jan [9]. The extension of the elementary nonlinear integral is based on the condition that the semimartingale kernel has nice continuity properties in the spatial parameter. We investigate the case that continuity is not available and suggest different directions of generalization. This brings us beyond the case that any integral can be approximated by integrals with integrands taking only finitely many values.

Keywords: Nonlinear stochastic integral, Itô integral, parameter depending semimartingales.

Mathematics Subject Classification (2000): 60H05, 60G44, 60G48.

1 Introduction

In this article we consider nonlinear stochastic integrals of Itô-type for a family of integrators which may depend on a spatial parameter in a *discontinuous* way. So far, to the best of our knowledge, only nonlinear integrals w.r.t. *smooth* families of integrators have

*Frankfurt MathFinance Institute, Johann Wolfgang Goethe-Universität, D-60054 Frankfurt a.M., Germany, e-mail: ckuehn@math.uni-frankfurt.de

Acknowledgements. I would like to thank Tahir Choulli and Jan Kallsen for helpful discussions and comments. In addition, comments by an anonymous referee are gratefully acknowledged.

been investigated in the literature; see the pioneering work of Kunita [8] and Le Jan [9] where the integrators are *continuous* semimartingales and Carmona and Nualart [2] for the general case. Further contributions to the theory of nonlinear stochastic integrals (also called stochastic line integrals) can e.g. be found in Sznitman [11] and Chitashvili and Mania [3, 4].

We allow general semimartingales as integrators and predictable processes as integrands. To our mind smoothness in the spatial parameter is an appropriate assumption, but it does not seem to be essential. Therefore there remains the interesting question how and at which price it can be avoided.

In the last years nonlinear stochastic integration theory has attracted attention by mathematical finance researchers as it is a mathematical tool to model trading gains of a “large” investor whose trades move the market price of the stock, see e.g. [1]. In the standard model the price per share of some stock does not depend on investor’s holdings and its dynamic is exogenous given by some semimartingale X . Then, the trading gains are modeled by the *linear* integral $\int_0^\cdot H_t dX_t$ where H_t is the number of shares the investor holds at time t . Nonlinear integrals arise by a feedback effect. The holdings H_t have some permanent impact on the stock price. However, it is not the scope of the present paper to discuss this application to finance.

Throughout the paper we fix a terminal time $T \in \mathbb{R}_+$ and a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, P)$ satisfying the usual conditions. We use the notation of Jacod and Shiryaev [6]. By $\mathbf{fP}$ we denote the set of real-valued, predictable processes. Let $(X(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$ be a family of semimartingales. We do not assume continuity in ϑ , i.e. for $\vartheta_1 \approx \vartheta_2$ the semimartingales $X(\vartheta_1, \cdot)$ and $X(\vartheta_2, \cdot)$ can behave quite differently. Instead, we firstly only assume that X is *jointly* measurable, namely the mapping $(\vartheta, \omega, t) \mapsto X_\omega(\vartheta, t)$ is $\mathcal{B}(\mathbb{R}) \otimes \mathcal{O}$ -measurable, where $\mathcal{O}$ is the optional σ -algebra on $\Omega \times [0, T]$. Recall that $\mathcal{O}$ is generated by all adapted processes with càdlàg paths (considered as mappings on $\Omega \times [0, T]$).

Definition 1.1. Let $\mathcal{E}$ denote the set of simple predictable real-valued processes, i.e.

$$\begin{aligned} \mathcal{E} := \left\{ \theta : \Omega \times [0, T] \rightarrow \mathbb{R} \mid \theta = H^{-1}1_{[0]} + \sum_{k=1}^n H^{k-1}1_{[T_{k-1}, T_k]} \text{ for some } n \in \mathbb{N}, \right. \\ \left. 0 = T_0 \leq T_1 \leq \dots \leq T_n = T \text{ stopping times, } H^{-1} \in L^0(\Omega, \mathcal{F}_0, P), \right. \\ \left. \text{and } H^{k-1} \in L^0(\Omega, \mathcal{F}_{T_{k-1}}, P), \ k = 1, \dots, n \right\}. \end{aligned}$$

For $\theta = H^{-1}1_{[0]} + \sum_{k=1}^n H^{k-1}1_{[T_{k-1}, T_k]}$ the process

$$t \mapsto \int_0^t X(\theta_s, ds) := \sum_{k=1}^n (X(H^{k-1}, t \wedge T_k) - X(H^{k-1}, t \wedge T_{k-1})) \quad (1.1)$$

is called the elementary integral of θ with respect to $(X(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$.

(1.1) is well-defined, i.e. it does not depend on the representation of θ . By the joint measurability of X the càdlàg process (1.1) is adapted.

The aim is to extend the integral (1.1) to all predictable locally bounded integrands in a reasonable way. The main step is to show that an arbitrary (locally bounded) predictable process θ can be approximated in a suitable way by simple integrands $(\theta^n)_{n \in \mathbb{N}}$ in order to extend the integral. The main problem is that by contrast to the smooth or the standard linear case it is not sufficient to make the distance $|\theta_t^n - \theta_t|$ small (in measure).

In this article we suggest different directions of generalization of the nonlinear integrals in [2, 8]. It is divided into two parts. In Section 2 we show that under the condition that the elementary integral (1.1) is onesided continuous, say right-continuous, in the spatial parameter there exists a unique right-continuous extension to all locally bounded predictable integrands (Theorem 2.7). In Section 3 we deal with the more delicate case that no onesided continuity is available. Then, there is in general no approximation with strategies taking only finitely many values. Within a finite factor model Theorem 3.7 gives conditions under which the integral can be approximated by integrals with simple integrands.

2 Right-continuous nonlinear integrators

In this section we assume that the elementary integral (1.1) is *onesided* continuous in ϑ , say continuous from the right. This assumption is in the spirit of Carmona and Nualart [2]

and has the advantage that it can be stated in terms of the integrators themselves rather than their densities.

Definition 2.1. Let $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ and $a \in \mathbb{R}$. Throughout the article we say that $a_n \searrow a$ iff for any $\varepsilon > 0$, $a_n \in [a, a + \varepsilon]$ holds for $n \in \mathbb{N}$ large enough.

Note that for random variables $P(Y_n \searrow Y) = 1$ is essentially weaker than $P(Y_n \rightarrow Y \text{ and } Y_n \geq Y, \forall n \in \mathbb{N}) = 1$.

Assume that the semimartingales $(X(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$ possess decompositions

$$X(\vartheta, \cdot) = X(\vartheta, 0) + M(\vartheta, \cdot) + B(\vartheta, \cdot), \quad \vartheta \in \mathbb{R}, \quad (2.1)$$

where $(M(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$ are locally square integrable martingales and $(B(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$ are càdlàg processes of finite variation and predictable. $\langle M(\vartheta_1, \cdot), M(\vartheta_2, \cdot) \rangle$ denotes the predictable quadratic covariation of $M(\vartheta_1, \cdot)$ and $M(\vartheta_2, \cdot)$.

Definition 2.2. For every $\vartheta \in \mathbb{R}$ define the σ -finite measure μ^ϑ by

$$\mu^\vartheta(A) := E \left(\int_0^T 1_A d\langle M(\vartheta, \cdot), M(\vartheta, \cdot) \rangle \right) + E \left(\int_0^T 1_A d\text{Var}(B(\vartheta, \cdot)) \right), \quad \forall A \in \mathcal{P},$$

where $\mathcal{P}$ is the predictable σ -algebra on $\Omega \times [0, T]$. Let $(\vartheta^n)_{n \in \mathbb{N}}$ be a dense sequence in $\mathbb{R}$ and let μ be a finite measure on $\mathcal{P}$ with $d\mu^{\vartheta^n} \ll d\mu$ for all $n \in \mathbb{N}$ (of course such a measure always exists).

For càdlàg processes X and Y let

$$d(X, Y) := E \left(\sup_{t \in [0, T]} |X_t - Y_t| \wedge 1 \right).$$

d metrizes the convergence “uniformly in probability”, cf. e.g. Protter [10].

Definition 2.3. A family $X = (X(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$ of semimartingales is called a right-continuous nonlinear integrator, if for any $K \in (\mathbf{bP})_{\text{loc}}$ and any sequence $(\theta^n)_{n \in \mathbb{N}}$ of simple predictable processes taking only finitely many values with $|\theta^n| \leq K$ and

$$\theta^n \searrow \theta, \quad \mu - \text{a.s.}, \quad n \rightarrow \infty \quad (2.2)$$

for some $\theta \in (\mathbf{bP})_{\text{loc}}$, the sequence of elementary integrals $(\int_0^\cdot X(\theta^n, ds))_{n \in \mathbb{N}}$ is a Cauchy-sequence w.r.t. the metric d , cf. Remarks 2.8 and 2.9 (by $(\mathbf{bP})_{\text{loc}}$ we denote the set of real-valued locally bounded predictable processes).

Proposition 2.4. *Assume that X is a right-continuous nonlinear integrator in the sense of Definition 2.3, but with pointwise convergence in (2.2) without an exceptional μ -null set (note that this is a weaker right-continuity condition than Definition 2.3). Then, for all $\vartheta \in \mathbb{R}$ and $A \in \mathcal{P}$ the following implication holds*

$$\mu^{\vartheta^n}(A) = 0, \quad \forall n \in \mathbb{N} \quad \implies \quad \mu^\vartheta(A) = 0.$$

Remark 2.5. *By Proposition 2.4, the existence of an appropriate finite measure μ that let X be a right-continuous nonlinear integrator in the sense of Definition 2.3 does not depend on the choice of the dense sequence $(\vartheta^n)_{n \in \mathbb{N}}$ in Definition 2.2.*

Proof of Proposition 2.4. Let $\vartheta \in \mathbb{R}$. There exists a subsequence $(\tilde{\vartheta}^n)_{n \in \mathbb{N}}$ of $(\vartheta^n)_{n \in \mathbb{N}}$ with $\tilde{\vartheta}^n > \vartheta$ and $\tilde{\vartheta}^n \rightarrow \vartheta$. Assume that there is an $N \in \mathcal{P}$ with

$$\mu^{\tilde{\vartheta}^n}(N) = 0 \quad \forall n \in \mathbb{N} \quad \text{and} \quad \mu^\vartheta(N) \in (0, \infty). \quad (2.3)$$

We have to lead (2.3) to a contradiction (note that we can restrict ourselves to the case that $\mu^\vartheta(N) \neq \infty$ as μ^ϑ is σ -finite).

By $H \bullet Z$ we denote the (linear) stochastic integral of H w.r.t. Z . As $B(\vartheta, \cdot)$ is predictable, (2.3) implies that

$$Y := 1_N \bullet X(\vartheta, \cdot) = 1_N \bullet M(\vartheta, \cdot) + 1_N \bullet B(\vartheta, \cdot)$$

does not vanish, i.e. $\varepsilon := d(Y, 0) > 0$. As $\mu^{\tilde{\vartheta}^n}(N) = 0$, $n \in \mathbb{N}$, one can find simple predictable sets Γ_n such that $\mu^{\tilde{\vartheta}^n}(\Gamma_n)$ and $\mu^\vartheta(\Gamma_n \Delta N)$, $n \in \mathbb{N}$, get arbitrarily small, where Δ denotes the symmetric difference (see e.g. the proof of Proposition 4.7(ii) in [5] that applies Dynkin's theorem and note that for all $\tilde{\vartheta} \in \mathbb{R}$ the processes $\langle M(\tilde{\vartheta}, \cdot), M(\tilde{\vartheta}, \cdot) \rangle$ and $B(\tilde{\vartheta}, \cdot)$ are predictable and thus locally bounded, cf. Lemma A.1 in [7]). By Burkholder-Davis' inequality this implies that we can choose $(\Gamma_n)_{n \in \mathbb{N}}$ s.t. $d(1_{\Gamma_n} \bullet X(\tilde{\vartheta}^n, \cdot), 0) \leq \varepsilon/3$ and $d(1_{\Gamma_n} \bullet X(\vartheta, \cdot), 1_N \bullet X(\vartheta, \cdot)) \leq \varepsilon/3$. We obtain

$$\begin{aligned} & d\left(X(\tilde{\vartheta}^n, \cdot), 1_{(\Omega \times [0, T]) \setminus \Gamma_n} \bullet X(\tilde{\vartheta}^n, \cdot) + 1_{\Gamma_n} \bullet X(\vartheta, \cdot)\right) \\ & \geq d\left(X(\tilde{\vartheta}^n, \cdot), X(\tilde{\vartheta}^n, \cdot) + 1_{\Gamma_n} \bullet X(\vartheta, \cdot)\right) - \frac{\varepsilon}{3} \\ & \geq d(1_N \bullet X(\vartheta, \cdot), 0) - \frac{2\varepsilon}{3} = \frac{\varepsilon}{3}, \quad \forall n \in \mathbb{N}. \end{aligned} \quad (2.4)$$

Define the simple predictable processes

$$\theta^m := \begin{cases} \tilde{\vartheta}^{m/2} & : \text{if } m \text{ is even} \\ \tilde{\vartheta}^{(m-1)/2} 1_{(\Omega \times [0, T]) \setminus \Gamma_{(m-1)/2}} + \vartheta 1_{\Gamma_{(m-1)/2}} & : \text{if } m \text{ is odd} \end{cases}$$

By (2.4) we have that for all *even* m $d \left(\int_0^\cdot X(\theta^m, ds), \int_0^\cdot X(\theta^{m+1}, ds) \right) \geq \varepsilon/3$ and thus the sequence $(\int_0^\cdot X(\theta^m, ds))_{m \in \mathbb{N}}$ of elementary nonlinear integrals is not a d -Cauchy sequence. As $\theta^m \searrow \vartheta$, pointwise, this contradicts the right-continuity of the elementary integral. $\square$

Definition 2.6. *Let $\mathcal{X}$ be the set of real-valued stochastic processes. We say that a mapping $I : (\mathbf{bP})_{\text{loc}} \rightarrow \mathcal{X}$ is additive in time iff for any $H^1, H^2 \in (\mathbf{bP})_{\text{loc}}$, $A \in \mathcal{P}$ we have that*

$$I(H^1 1_A + H^2 1_{A^c}) + I(H^2 1_A + H^1 1_{A^c}) = I(H^1) + I(H^2), \quad \text{up to evanescence,} \quad (2.5)$$

where $A^c := (\Omega \times [0, T]) \setminus A$.

If I is linear, then (2.5) is obviously satisfied. But, taking H^1, H^2 and A simple, the elementary *nonlinear* integral (1.1) also satisfies (2.5).

Theorem 2.7. *Let X be a right-continuous nonlinear integrator in the sense of Definition 2.3. The elementary integral from (1.1) possesses an up to evanescence unique extension*

$$I : (\mathbf{bP})_{\text{loc}} \rightarrow \{Y : \Omega \times [0, T] \rightarrow \mathbb{R} \mid Y \text{ is adapted and càdlàg}\}$$

satisfying the following properties

(i) *I is additive in time*

(ii) *(right-continuity)*

For all $(\theta^n)_{n \in \mathbb{N}} \subset (\mathbf{bP})_{\text{loc}}$ with $|\theta^n| \leq K$ for some nonnegative process $K \in (\mathbf{bP})_{\text{loc}}$, $\theta \in (\mathbf{bP})_{\text{loc}}$, we have that if $\theta^n \searrow \theta$ pointwise on $\Omega \times [0, T]$, then $I(\theta^n) \rightarrow I(\theta)$ uniformly in probability for $n \rightarrow \infty$.

Remark 2.8. *As for linear Itô-integrals (cf. e.g. Theorem I.4.31 in [6]) the uniqueness of the extension of the elementary integral already holds under the weaker (right-)continuity (ii)*

in Theorem 2.7, i.e. requiring convergence of the integrals only for integrands with $\theta^n \searrow \theta$, pointwise on $\Omega \times [0, T]$ (without an exceptional null set). However, for the existence of the extension we need that the elementary integral is right-continuous in a slightly stronger sense – including sequences $(\theta^n)_{n \in \mathbb{N}}$ which may diverge on an exceptional μ -nullset, cf. Definition 2.3 (by Itô’s isometry this property of the elementary integral is satisfied in the linear case). As not every predictable process can be approximated pointwise by simple processes, the exceptional null set in (2.2) becomes necessary for the construction of the integral.

Remark 2.9. In order to extend the elementary nonlinear integral to the smaller set of left-continuous, adapted integrands one can leave out the μ -nullset in Definition 2.3 as well as condition (i) in Theorem 2.7. This is because left-continuous, adapted processes can be approximated uniformly in time by simple predictable integrands. Furthermore, note that right-continuous functions are not uniformly right-continuous on compacta (by contrast to continuous functions). This is the reason why Definition 2.3 differs from the condition in Definition II.1.2 of [2] (which is solely stated in terms of simple predictable processes).

The nonlinear integral of Kunita [8], Chapter 3.2, is obviously right-continuous in the sense of Definition 2.3.

Remark 2.10. For suitable $\theta \in \mathbf{bP}$ and $(M(\vartheta, \cdot))_{\vartheta \in \mathbb{R}} \subset \mathcal{H}^2$ Sznitman [11] defines the nonlinear integral $\int_0^\cdot M(\theta_t, dt)$ as the up to evanescence unique square integrable martingale s.t. for all $\widetilde{M} \in \mathcal{H}^2$ the covariation process $\langle \int_0^\cdot M(\theta_t, dt), \widetilde{M} \rangle$ possesses the form “one would expect” by the covariation processes $\langle M(\vartheta, \cdot), \widetilde{M} \rangle$, $\vartheta \in \mathbb{R}$ (see Proposition 4 in [11]). Chitashvili and Mania [4] show that this approach is still valid under weaker assumptions, especially if $M(\vartheta, \cdot)$ is only continuous in ϑ instead of Hölder continuous as it is required in [11]. An inspection of the proof of Theorem 1 in [4] reveals that continuity can be replaced by one-sided continuity. However, in this approach the integral is not characterized as the unique extension of the elementary integral (and of course it is not shown that the integrands can be approximated from the right by simple integrands). Thus Theorem 2.7 may be also seen as a complement to [4].

Proof of Theorem 2.7. Step 1 (uniqueness): Assume that there are two extensions of (1.1),

I and $\tilde{I}$, satisfying conditions (i) and (ii). We want to show by “algebraic induction” that they coincide up to evanescence. For this let us firstly prove the assertion for all predictable θ taking at most two values, i.e. fix $a, b \in \mathbb{R}$ with $a \leq b$ and show that

$$I(a1_A + b1_{A^c}) = \tilde{I}(a1_A + b1_{A^c}) \quad \text{up to evanescence} \quad (2.6)$$

holds for all $A \in \mathcal{P}$. For $A \in \mathcal{E}$ the assertion holds. As $\mathcal{E}$ is a generator of $\mathcal{P}$ which is stable under intersection, by Dynkin’s theorem it is enough to show that

(I) For any $B \subset C$ s.t. $B, C \in \mathcal{P}$ and (2.6) is satisfied for $A \in \{B, C\}$ (2.6) holds for $A = C \setminus B$ as well.

(II) If (2.6) holds for all A_n , $n \in \mathbb{N}$, and $A_n \uparrow A$, $n \uparrow \infty$, (2.6) holds for A as well.

Ad (I): Let $B \subset C$ both satisfy (2.6). Applying (2.5) to $H^1 := a1_B + b1_{B^c}$, $H^2 := a1_{C \setminus B} + b1_{(C \setminus B)^c}$ and $A := B$ yields

$$\begin{aligned} I(H^1) + I(H^2) &= I(H^1 1_A + H^2 1_{A^c}) + I(H^2 1_A + H^1 1_{A^c}) \\ &= I(a1_B + a1_{C \setminus B} + b1_{C^c}) + I(b) \\ &= I(a1_C + b1_{C^c}) + I(b). \end{aligned}$$

As the same holds for $\tilde{I}$ and $I(b) = \tilde{I}(b)$ we arrive at $I(H^2) = \tilde{I}(H^2)$, i.e. $C \setminus B$ satisfies (2.6).

Ad (II): Let $(A_n)_{n \in \mathbb{N}}$ satisfy (2.6) and $A_n \uparrow A$. By $a \leq b$, $a1_{A_n} + b1_{(A_n)^c} \downarrow a1_A + b1_{A^c}$ which implies by condition (ii) that A satisfies (2.6).

Having established $I(\theta) = \tilde{I}(\theta)$ up to evanescence for all predictable integrands θ taking at most two values, we want to show by induction on the number of values that the assertion holds as well for all integrands taking finitely many values. Let $n \geq 2$ and $H = \sum_{i=1}^{n+1} a_i 1_{A_i}$ with A_i disjoint and $\bigcup_{i=1}^{n+1} A_i = \Omega \times [0, T]$. Applying (2.5) to $H^1 := H$, $H^2 := a_1$, and $A := A_{n+1}$ yields

$$\begin{aligned} I(H) + I(a_1) &= I(H^1) + I(H^2) \\ &= I(H^1 1_A + H^2 1_{A^c}) + I(H^2 1_A + H^1 1_{A^c}) \\ &= I(a_{n+1} 1_{A_{n+1}} + a_1 1_{A_{n+1}^c}) + I(a_1 1_{\{H=a_1\} \cup A_{n+1}} + H 1_{\{H \neq a_1\} \cap A_{n+1}^c}). \end{aligned} \quad (2.7)$$

As the same holds for $\tilde{I}$ and both integrands in the last line of (2.7) take at most n values (note that $n \geq 2$) the induction step $n \rightsquigarrow n+1$ is proven.

Finally, let $\theta \in (\mathbf{bP})_{\text{loc}}$. Define $\theta^n := \sum_{k=-n2^n}^{n2^n} \frac{k}{2^n} 1_{\{\theta \in ((k-1)/2^n, k/2^n]\}}$. We have that $|\theta^n| \leq |\theta| + 1$. As $I(\theta^n) = \tilde{I}(\theta^n)$ up to evanescence and $\theta^n \searrow \theta$, pointwise on $\Omega \times [0, T]$, and by right-continuity of I and $\tilde{I}$, it follows that $I(\theta) = \tilde{I}(\theta)$ up to evanescence.

Step 2 (construction):

Let $\theta \in (\mathbf{bP})_{\text{loc}}$ and define $\tilde{\theta}^n := \sum_{k=-n2^n}^{n2^n} \frac{k}{2^n} 1_{\{\theta \in ((k-1)/2^n, k/2^n]\}}$. Observe that

$$\tilde{\theta}^n \searrow \theta, \quad \text{pointwise on } \Omega \times [0, T], \quad n \rightarrow \infty. \quad (2.8)$$

The predictable sets

$$B_{k,n} := \begin{cases} \{\theta \in ((k-1)/2^n, k/2^n]\}, & \text{for } k = -n2^n, \dots, -1, 1, \dots, n2^n \\ \{\theta \in ((-1/2^n, 0] \cup (-\infty, -(n2^n+1)/2^n] \cup (n, \infty))\}, & \text{for } k = 0 \end{cases}$$

can be approximated in μ -measure by elements of $\mathcal{E}$, i.e. there are $\Gamma_{k,n} \in \mathcal{E}$ s.t.

$$\mu(\Gamma_{k,n} \Delta B_{k,n}) \leq \frac{1}{n2^{n+1} + 1} \frac{1}{2^n}, \quad n \in \mathbb{N}, \quad k = -n2^n, \dots, n2^n,$$

see again the proof of Proposition 4.7(ii) in [5]. Define

$$\theta_t^n(\omega) := \begin{cases} \frac{k}{2^n}, & \text{where } k \text{ is the smallest number with } (\omega, t) \in \Gamma_{k,n} \\ 0, & \text{if no such } k \text{ exists.} \end{cases}$$

As intersections of stochastic intervals are stochastic intervals, θ^n is a simple predictable process. Denote $\Gamma_n := \bigcup_k ((\Gamma_{k,n} \cap B_{k,n}) \setminus \bigcup_{l \neq k} \Gamma_{l,n})$. As $(B_{k,n})_{k=-n2^n, \dots, n2^n}$ is a partition of $\Omega \times [0, T]$ we have $\mu((\Omega \times [0, T]) \setminus \Gamma_n) \leq (n2^{n+1} + 1) \frac{1}{n2^{n+1} + 1} \frac{1}{2^n} = \frac{1}{2^n}$. But, on the set Γ_n we have that $\theta^n = \tilde{\theta}^n$ and thus

$$\mu(\theta^n \neq \tilde{\theta}^n) \leq \frac{1}{2^n}, \quad \forall n \in \mathbb{N}. \quad (2.9)$$

By the theorem of Borel-Cantelli, (2.9) implies that

$$\mu(\theta^n \neq \tilde{\theta}^n \text{ for infinitely many } n \in \mathbb{N}) = 0. \quad (2.10)$$

Putting (2.8) and (2.10) together we arrive at $\theta^n \searrow \theta$, μ -a.s., $n \rightarrow \infty$ (cf. Definition 2.1 for the exact meaning of this statement). By Definition 2.3, $(\int_0^\cdot X(\theta_s^n, ds))_{n \in \mathbb{N}}$ is a

Cauchy-sequence w.r.t. the convergence “uniformly in probability”. By completeness we can define

$$\int_0^\cdot X(\theta_s, ds) := \lim_{n \rightarrow \infty} \int_0^\cdot X(\theta_s^n, ds). \quad (2.11)$$

$\int_0^\cdot X(\theta_s, ds)$ is well-defined up to evanescence, i.e. it does not depend on the special approximating sequence $(\theta^n)_{n \in \mathbb{N}}$ (as two sequences $(\theta^{n,1})_{n \in \mathbb{N}}, (\theta^{n,2})_{n \in \mathbb{N}} \subset \mathcal{E}$ with $\theta^{n,1} \searrow \theta$ and $\theta^{n,2} \searrow \theta$, μ -a.s. can be merged to a joint sequence whose elementary integrals build a Cauchy-sequence). In particular (2.11) is indeed an extension of the elementary integral.

Step 3 (right-continuity and additivity in time): Let us firstly show that the mapping $\theta \mapsto \int_0^\cdot X(\theta, ds)$ is right-continuous in the sense of (ii). Assume that $(\theta^n)_{n \in \mathbb{N}} \subset (\mathbf{bP})_{\text{loc}}$ with $|\theta^n| \leq K$ for some nonnegative process $K \in (\mathbf{bP})_{\text{loc}}$ and

$$\theta^n \searrow \theta, \quad \text{pointwise on } \Omega \times [0, T], \quad n \rightarrow \infty. \quad (2.12)$$

By Step 2, there exist sequences $(\psi^{n,m})_{m \in \mathbb{N}}$ of simple predictable processes with $|\psi^{n,m}| \leq K + 1$ and $\psi^{n,m} \searrow \theta^n$, μ -a.s., for $m \rightarrow \infty$.

Let $\varepsilon > 0$. By Definition 2.3 and construction (2.11) there exists a $\delta > 0$ such that for all *simple* $\tilde{\theta}$ with $|\tilde{\theta}| \leq K + 1$ the following implication holds

$$\int_{\Omega \times [0, T]} \frac{|\tilde{\theta} - \theta|}{K + 1} \vee 1_{\{\tilde{\theta} < \theta\}} d\mu \leq \delta \implies d\left(\int_0^\cdot X(\tilde{\theta}_s, ds), \int_0^\cdot X(\theta_s, ds)\right) \leq \varepsilon/2. \quad (2.13)$$

Namely, assume that the implication (2.13) does *not* hold, i.e. there exists a sequence $(\tilde{\theta}^n)_{n \in \mathbb{N}} \subset \mathcal{E}$ with $|\tilde{\theta}^n| \leq K + 1$ and

$$\int_{\Omega \times [0, T]} \frac{|\tilde{\theta}^n - \theta|}{K + 1} \vee 1_{\{\tilde{\theta}^n < \theta\}} d\mu \leq \frac{1}{2^n} \quad \text{and} \quad d\left(\int_0^\cdot X(\tilde{\theta}_s^n, ds), \int_0^\cdot X(\theta_s, ds)\right) > \varepsilon/2 \quad (2.14)$$

for all $n \in \mathbb{N}$. By a Borel-Cantelli argument this implies

$$\mu\left(\tilde{\theta}^n < \theta \quad \text{for infinitely many } n \in \mathbb{N}\right) = 0, \quad \text{and}$$

$$\mu\left(\frac{|\tilde{\theta}^n - \theta|}{K + 1} > \tilde{\varepsilon} \quad \text{for infinitely many } n \in \mathbb{N}\right) = 0, \quad \forall \tilde{\varepsilon} > 0.$$

This means that $\tilde{\theta}^n \searrow \theta$, μ -a.s. By Step 2 this is a contradiction to the second condition in (2.14).

By (2.12) there exists a $n^\varepsilon \in \mathbb{N}$ such that $\int_{\Omega \times [0, T]} \frac{|\theta^n - \theta|}{K+1} \vee 1_{\{\theta^n < \theta\}} d\mu \leq \delta/2$ for all $n \geq n^\varepsilon$. In addition, for every n there exists a $m(n) \in \mathbb{N}$ s.t. $\int_{\Omega \times [0, T]} \frac{|\psi^{n,m} - \theta^n|}{K+1} \vee 1_{\{\psi^{n,m} < \theta^n\}} d\mu \leq \delta/2$ for all $n \in \mathbb{N}$ and $m \geq m(n)$. By a triangle inequality this implies

$$\int_{\Omega \times [0, T]} \frac{|\psi^{n,m} - \theta|}{K+1} \vee 1_{\{\psi^{n,m} < \theta\}} d\mu \leq \delta, \quad \forall n \geq n^\varepsilon, m \geq m(n)$$

and thus by (2.13)

$$d\left(\int_0^\cdot X(\psi_s^{n,m}, ds), \int_0^\cdot X(\theta_s, ds)\right) \leq \varepsilon/2, \quad \forall n \geq n^\varepsilon, m \geq m(n). \quad (2.15)$$

In addition, for every $n \in \mathbb{N}$ there exists a $\tilde{m}(n) \in \mathbb{N}$ s.t.

$$d\left(\int_0^\cdot X(\psi_s^{n,m}, ds), \int_0^\cdot X(\theta_s^n, ds)\right) \leq \varepsilon/2, \quad \forall n \in \mathbb{N}, m \geq \tilde{m}(n). \quad (2.16)$$

Putting (2.15) and (2.16) together, taking $m = m(n) \vee \tilde{m}(n)$ for each n , implies that $d(\int_0^\cdot X(\theta_s^n, ds), \int_0^\cdot X(\theta_s, ds)) \leq \varepsilon$ for $n \geq n^\varepsilon$ which implies (ii).

Property (2.5) is obviously satisfied for the elementary integral, i.e. with H^1, H^2 , and A simple. For arbitrary H^1, H^2 , and A there exist by Step 2 simple sequences $(\theta^{1,n})_{n \in \mathbb{N}}$, $(\theta^{2,n})_{n \in \mathbb{N}}$, and $(A_n)_{n \in \mathbb{N}}$ with $\theta^{1,n} \searrow H^1$, $\theta^{2,n} \searrow H^2$, μ -a.s., and $\mu((A \setminus A_n) \cup (A_n \setminus A)) \rightarrow 0$, $n \rightarrow \infty$. Thus also $\theta^{1,n} 1_A + \theta^{2,n} 1_{A^c} \searrow H^1 1_A + H^2 1_{A^c}$, μ -a.s., $n \rightarrow \infty$, etc. (cf. Definition 2.1) and by (ii) we are done. $\square$

3 Beyond onesided continuity

In this section we go beyond onesided continuity in the spatial parameter ϑ . This comes at the price that we restrict ourselves to a model where all integrators $X(\vartheta, \cdot)$, $\vartheta \in \mathbb{R}$, are driven by the same (finitely many) semimartingales $Y^1, \dots, Y^d$, $d \in \mathbb{N}$, i.e.

$$X(\vartheta, \cdot) = X(\vartheta, 0) + \sum_{i=1}^d \int_0^\cdot H^i(\vartheta, t) dY_t^i, \quad P\text{-a.s.}, \vartheta \in \mathbb{R}, \quad (3.17)$$

where $(\vartheta, \omega, t) \mapsto H_\omega^i(\vartheta, t)$ are $\mathcal{B}(\mathbb{R}) \otimes \mathcal{P}$ -measurable and $H^i(\vartheta, \cdot) \in L(Y^i)$, $\forall \vartheta \in \mathbb{R}$. Assume that Y^i possess decompositions $Y^i = Y_0^i + M^i + B^i$ where $M^i \in \mathcal{H}_{\text{loc}}^2$ and $B^i \in \mathcal{V} \cap \mathbf{fP}$. Denote

$$A^i := \langle M^i, M^i \rangle + \text{Var}(B^i). \quad (3.18)$$

Let $\theta \in \mathbf{fP}$. By the bi-measurability of H^i , the process $H^i(\theta, \cdot) : \Omega \times [0, T] \rightarrow \mathbb{R}$, $(\omega, t) \mapsto H^i_\omega(\theta_t(\omega), t)$ is predictable. If $H^i(\theta, \cdot) \in L(Y^i)$, $i = 1, \dots, d$, it suggests itself to define the nonlinear integral w.r.t. X by

$$\int_0^\cdot X(\theta_t, dt) := \sum_{i=1}^d \int_0^\cdot H^i(\theta_t, t) dY_t^i. \quad (3.19)$$

So, the nonlinear integral can be defined using the theory of *linear* stochastic integration. For a simple process θ taking only finitely many values (3.19) coincides with the elementary integral (1.1). Similar constructions arise frequently in the literature. But, if H is not (onesided) continuous in ϑ , there remains the interesting question whether the nonlinear integral (3.19) with a general predictable strategy θ can be approximated by corresponding nonlinear integrals with simple predictable strategies θ^n . As we want to understand (3.19) as the limit of basic objects – namely the corresponding nonlinear integrals with piecewise constant strategies which have a clear interpretation in applications – this question is of essential importance.

Notice that from the linear theory we only know that $t \mapsto H(\theta_t, t)$ can be approximated in measure by a sequence $(f^n)_{n \in \mathbb{N}}$ of simple predictable processes. But, there is in general no sequence $(\theta^n)_{n \in \mathbb{N}}$ with $f_t^n = H(\theta_t^n, t)$. Consider the following examples for (not onesided continuous) integrators.

Example 3.1. Take $d = 1$, $H^1(\vartheta, t) = 1_{(\vartheta=t)}$, and $Y^1 = B$, where B is a standard Brownian motion. For the strategy $\theta_t = t$ we obtain that $\int_0^\cdot X(\theta_t, dt) = B$, whereas any elementary integral vanishes.

Example 3.1 shows that in general (3.19) cannot be approximated by corresponding nonlinear integrals with simple integrands.

Example 3.2. Let

$$X(\vartheta, \cdot) = 1_{\{Z=\vartheta\} \times (t_1, t_2]} B, \quad \vartheta \in \mathbb{R},$$

where B is a one-dimensional standard Brownian motion, $t_1 < t_2$, and Z is a continuously distributed $\mathcal{F}_{t_1}$ -measurable random variable.

In this example (3.19) can be approximated by corresponding nonlinear integrals with $(\theta^n)_{n \in \mathbb{N}} \subset \mathcal{E}$ (as for any $\theta \in \mathbf{fP}$ the predictable set $\{(\omega, t) \in \Omega \times (t_1, t_2] \mid \theta_t(\omega) = Z(\omega)\}$ can be approximated in $(P \otimes A^1)$ -measure by unions of stochastic intervals). However, e.g. the elementary integral (1.1) for the simple strategy $\theta_t(\omega) := Z(\omega)1_{(t_1, t_2]}(t)$ cannot be approximated by elementary integrals with simple strategies taking only finitely many values. Namely, as Z is continuously distributed, for any strategy ψ taking only finitely many values we have that $\int_{t_1}^{t_2} X(\psi_t, dt) = 0$, P -a.s., whereas $\int_{t_1}^{t_2} X(\theta_t, dt) = B_{t_2} - B_{t_1}$. This effect cannot occur if integrators are right-continuous in the sense of Definition 2.3.

Remark 3.3. By contrast to the approach in Section 2, (3.19) is not determined by $(X(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$ alone, but it also depends on the versions of the integrands $(H^i(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$. Take $d = 1$. The problem is that for fixed $\vartheta \in \mathbb{R}$ the integrand $H^1(\vartheta, \cdot)$ is uniquely determined from the integral $X(\vartheta, \cdot)$ in (3.17) only up to a $(P \otimes A^1)$ -null set and there is no joint null set for all $\vartheta \in \mathbb{R}$. Put differently, the choice of the versions is needed to clarify the precise meaning of the nonlinear integral.

For convenience and to avoid effects which seem to be of minor relevance for our problem, we make the following assumption.

Assumption 3.4. Assume that A^i , $i = 1, \dots, d$, defined in (3.18) are deterministic and possess $[a_{\min}, a_{\max}]$ -valued densities w.r.t. the Lebesgue measure for some $a_{\min}, a_{\max} \in \mathbb{R}_+ \setminus \{0\}$.

This and the following assumptions on A^i and H^i should be understood in the sense that H^i and Y^i in (3.17) can be chosen s.t. these properties hold – as of course the same nonlinear integral (3.19) can result from different semimartingales $Y^1, \dots, Y^d$ (note however that not all assertions in the following are invariant under this choice).

Assumption 3.5. We assume that for every $K \in \mathbb{R}_+$, $i = 1, \dots, n$, there exists a $G^i \in L^2(A^i)$ (i.e. $G^i \in \mathbf{fP}$ and $P\left(\int_0^T (G^i)^2 dA^i < \infty\right) = 1$) s.t. $|H^i(\vartheta, \cdot)| \leq G^i$ for all $\vartheta \in [-K, K]$.

Assumption 3.6. Assume that for every $K \in \mathbb{R}_+$

$$\int_{\delta}^T \sup_{\vartheta \in [-K, K]} \left(\frac{\int_{t-\delta}^t H^i(\vartheta, s) dA_s^i}{A_t^i - A_{t-\delta}^i} - H^i(\vartheta, t) \right)^2 dA_t^i \quad (3.20)$$

converges to 0 in probability for $\delta \rightarrow 0$

(note that by the bi-measurability of H^i the process

$$t \mapsto \sup_{\vartheta \in [-K, K]} \left(\frac{\int_{t-\delta}^t H^i(\vartheta, s) dA_s^i}{A_t^i - A_{t-\delta}^i} - H^i(\vartheta, t) \right)^2$$

is predictable).

Assumption 3.6 rules out Example 3.1, but it is satisfied by Example 3.2. It is discussed in detail in Subsection 3.1.

Theorem 3.7. *Under Assumptions 3.4, 3.5, and 3.6 any $\theta \in \mathbf{bP}$ can be approximated by a uniformly bounded sequence $(\theta^n)_{n \in \mathbb{N}}$ of simple predictable processes in the sense that for $i = 1, \dots, d$*

$$\int_0^T (H^i(\theta_t^n, t) - H^i(\theta_t, t))^2 dA_t^i \rightarrow 0, \quad \text{in probability, } n \rightarrow \infty. \quad (3.21)$$

Remark 3.8. *By the theory of linear stochastic integration Theorem 3.7 implies that for any locally bounded θ the integral $\int_0^\cdot X(\theta_t, dt)$ can be approximated uniformly in probability by integrals $\int_0^\cdot X(\theta_t^n, dt)$ with simple predictable processes θ^n .*

Remark 3.9. *The assumptions of Theorem 3.7 remain satisfied if one replaces H by $\tilde{H} = (H^1, \dots, H^d, H^{d+1})$ with $H^{d+1}(\vartheta, \cdot) = \vartheta$. Then, it is guaranteed that also the distance $|\theta^n - \theta|$ becomes small.*

Proof of Theorem 3.7. For ease of notation we assume that $d = 1$ in this proof and we write $H := H^1$ and $A := A^1$. The general case follows analogously. Let $\theta \in \mathbf{bP}$ with $|\theta| \leq \tilde{K}$ for some $\tilde{K} \in \mathbb{R}_+$.

Let $\delta > 0$. Define a family of stochastic processes $(f(\vartheta, t))_{\vartheta \in \mathbb{R}, \delta \leq t \leq T}$ by

$$f(\vartheta, t) := \frac{1}{A_t - A_{t-\delta}} \int_{t-\delta}^t H(\vartheta, s) dA_s. \quad (3.22)$$

f is jointly measurable and for every $\vartheta \in \mathbb{R}$ the process $f(\vartheta, \cdot)$ is continuous. Consequently, $t \mapsto f(\theta_t, t) = \frac{1}{A_t - A_{t-\delta}} \int_{t-\delta}^t H(\theta_t, s) dA_s$ is a predictable process on $[\delta, T]$.

Firstly, let us give a short idea of the proof. It will turn out that by Assumption 3.6 it is sufficient to prove (3.21) for the families of smoothed integrands $(f(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$ instead of

$(H(\vartheta, \cdot))_{\vartheta \in \mathbb{R}}$ itself (for all fixed $\delta > 0$). Then, the continuity of $f(\vartheta, \cdot)$ in time (for ϑ fixed) will play the key role as it let $t \mapsto f(\theta_t^n, t)$ be piecewise continuous for $\theta^n \in \mathcal{E}$ (however, $t \mapsto f(\theta_t, t)$ is in general still irregular).

Denote by $\mathcal{S}$ the set of $[\delta, T]$ -valued stopping times. We fix $K, m \in \mathbb{N}$ and define for $l = -2^m K, -2^m K + 1, \dots, 2^m K$ the predictable sets

$$B^l := \left\{ (\omega, t) \in \Omega \times [\delta, T] \mid f_\omega(\theta_t(\omega), t) \in \left(\frac{l-1}{2^m}, \frac{l}{2^m} \right] \right\} \quad \text{and} \quad (3.23)$$

$$\begin{aligned} \tilde{B}^l = \text{ess} \sup_{T_1, T_2 \in \mathcal{S} \text{ with } T_1 \leq T_2, Y \in L^0(\Omega, \mathcal{F}_{T_1}, P)} & \left\{ (\omega, t) \in \Omega \times [\delta, T] \mid \right. \\ & \left. f_\omega(\theta_t(\omega), t) \in \left(\frac{l-1}{2^m}, \frac{l}{2^m} \right], T_1(\omega) < t \leq T_2(\omega), \text{ and } f_\omega(Y(\omega), t) \in \left[\frac{l-2}{2^m}, \frac{l+1}{2^m} \right] \right\}, \end{aligned}$$

where the essential supremum is taken w.r.t. $(P \otimes A)$ and the predictable σ -algebra. For sets M_i, M the assertion $M = \text{ess sup}_{i \in I} M_i$ means that $1_M = \text{ess sup}_{i \in I} 1_{M_i}$. This definition makes sense as there exists a $\{0, 1\}$ -valued version of the essential supremum taken over a set of $\{0, 1\}$ -valued functions. Of course $\tilde{B}^l$ is well-defined only up to a $(P \otimes A)$ -null set.

Step 1: Let us show that

$$(P \otimes A)(B^l \setminus \tilde{B}^l) = 0, \quad l = -2^m K, -2^m K + 1, \dots, 2^m K. \quad (3.24)$$

Having established (3.24) the main part of the proof is done. By (3.24), the predictable sets where $f(\theta_t, t)$ falls into the interval $((l-1)2^{-m}, l2^{-m}]$, $l = -2^m K, -2^m K + 1, \dots, 2^m K$, can be approximated in measure by sets where *both* $f(\theta_t, t)$ falls into $((l-1)2^{-m}, l2^{-m}]$ and $f(\tilde{\theta}_t, t)$ falls into $[(l-2)2^{-m}, (l+1)2^{-m}]$ for some simple predictable $\tilde{\theta}$ not depending on l (cf. Step 2 for the construction of $\tilde{\theta}$).

Assume that (3.24) does not hold for some l , i.e. $(P \otimes A)(C) > 0$ with $C := B^l \setminus \tilde{B}^l$. We want to lead this to a contradiction to the maximality of $\tilde{B}^l$ in (3.23). Define $\mu(D) := (P \otimes A)((\Omega \times D) \cap C)$, $\forall D \in \mathcal{B}([\delta, T])$ and $C_t := \{\omega \in \Omega \mid (\omega, t) \in C\}$, $\forall t \in [0, T]$. μ is absolutely continuous w.r.t. A with density $P(C_t)$. By Lebesgue's differentiation theorem,

μ is differentiable w.r.t. A at A -almost every t and the derivative coincides with $P(C_t)$ for A -almost every t . As $\mu([\delta, T]) > 0$ there exists a $t_0 \in [\delta, T]$ s.t. $P(C_{t_0}) > 0$ and for any $\lambda \in (0, P(C_{t_0}))$

$$\mu((t_0, t_0 + \varepsilon)) \geq \lambda (A_{t_0 + \varepsilon} - A_{t_0}) \geq \lambda a_{\min} \varepsilon, \quad \text{for } \varepsilon > 0 \text{ small enough} \quad (3.25)$$

(remember that $a_{\min} > 0$). A first candidate to disprove maximality in (3.23) is the triple $T_1 = t_0, T_2 = \tau_0 \wedge (t_0 + \varepsilon)$ with some appropriate $\varepsilon > 0$, where

$$\tau_0 := \inf \left\{ t \geq t_0 \mid f(\theta_{t_0}, t) \notin \left[\frac{l-2}{2^m}, \frac{l+1}{2^m} \right] \right\}, \quad (3.26)$$

and $Y = \theta_{t_0}$. We make the following case differentiation.

Case 1: Assume that

$$\limsup_{t \downarrow t_0} \frac{(P \otimes A)((C_{t_0} \times (t_0, t)) \cap C)}{A_t - A_{t_0}} > 0. \quad (3.27)$$

By continuity of $t \mapsto f(\theta_{t_0}, t)$ we have $C_{t_0} \subset \{\tau_0 > t_0\}$ and thus

$$\lim_{\varepsilon \downarrow 0} P(C_{t_0} \cap \{\tau_0 \leq t_0 + \varepsilon\}) = 0. \quad (3.28)$$

(3.28) implies that under (3.27)

$$\limsup_{\varepsilon \downarrow 0} \frac{(P \otimes A)((C_{t_0} \times (t_0, t_0 + \varepsilon)) \cap C \cap]t_0, \tau_0])}{(P \otimes A)((C_{t_0} \times (t_0, t_0 + \varepsilon)) \cap C)} = 1 \quad (3.29)$$

and $(P \otimes A)(]t_0, \tau_0 \wedge (t_0 + \varepsilon)] \cap C) > 0$ for some $\varepsilon > 0$. Here, we are already done. Namely, $]t_0, \tau_0 \wedge (t_0 + \varepsilon)] \cap \{f(\theta, \cdot) \in (\frac{l-1}{2^m}, \frac{l}{2^m}]\} \subset \tilde{B}^l$, $(P \otimes A)$ -a.s., and

$$\begin{aligned} 0 &< (P \otimes A)(]t_0, \tau_0 \wedge (t_0 + \varepsilon)] \cap C) \\ &\leq (P \otimes A) \left(\left(]t_0, \tau_0 \wedge (t_0 + \varepsilon)] \cap \left\{ f(\theta, \cdot) \in \left(\frac{l-1}{2^m}, \frac{l}{2^m} \right] \right\} \right) \setminus \tilde{B}^l \right), \end{aligned} \quad (3.30)$$

which is a contradiction to the maximality of $\tilde{B}^l$ in (3.23). Namely, the essential supremum would be enlarged by the triple $(t_0, \tau_0 \wedge (t_0 + \varepsilon), \theta_{t_0})$. Thus $(P \otimes A)(B^l \setminus \tilde{B}^l) = 0$ in Case 1.

Case 2: Assume that (3.27) does not hold. In view of (3.25) we have that

$$\liminf_{t \downarrow t_0} \frac{(P \otimes A)((\Omega \setminus C_{t_0} \times (t_0, t)) \cap C)}{A_t - A_{t_0}} \geq P(C_{t_0}), \quad (3.31)$$

i.e. nearly the entire mass contributing to (3.25) has to come from $\Omega \setminus C_{t_0}$. In this case it becomes necessary to vary the triple $(t_0, \tau_0 \wedge (t_0 + \varepsilon), \theta_{t_0})$ a bit. In this case the measure

$\tilde{\mu} \leq \mu$ defined by $\tilde{\mu}(D) := (P \otimes A)((\Omega \setminus C_{t_0} \times D) \cap C)$ is also differentiable (w.r.t. A) from the right at t_0 with the same derivative $P(C_{t_0})$. We obtain

$$\begin{aligned} & \limsup_{t_1 \downarrow t_0} \liminf_{t \downarrow t_1} \frac{(P \otimes A)((\Omega \setminus C_{t_0} \times (t_1, t)) \cap C)}{A_t - A_{t_1}} \\ & \geq \liminf_{t \downarrow t_0} \frac{(P \otimes A)((\Omega \setminus C_{t_0} \times (t_1, t)) \cap C)}{A_t - A_{t_0}} = P(C_{t_0}). \end{aligned} \quad (3.32)$$

For the inequality we use the fact that for a function g with $g(x) = \int_0^x r(t) dt$, $r(t) \in [0, 1]$, we have

$$\limsup_{x_0 \downarrow 0} \liminf_{x \downarrow x_0} \frac{g(x) - g(x_0)}{x - x_0} \geq \liminf_{x \downarrow 0} \frac{g(x)}{x}$$

(Taking x_0 not from the null set where $r(x_0)$ is not the right-sided derivative of g at x_0 , we have that $(g(x) - g(x_0))/(x - x_0)$ is close to $r(x_0)$. Given $\varepsilon' > 0$ one can find $x_0 > 0$ arbitrary small with $r(x_0) \geq \liminf_{x \downarrow 0} \frac{g(x)}{x} - \varepsilon'$).

Given an $\tilde{\varepsilon} > 0$, by (3.32), we can find $t_1 > t_0$ arbitrary close to t_0 with

$$\tilde{\mu}((t_1, t_1 + \varepsilon)) \geq (P(C_{t_0}) - \tilde{\varepsilon})(A_{t_1 + \varepsilon} - A_{t_1}), \quad \text{for } \varepsilon \text{ small enough.} \quad (3.33)$$

In addition t_1 can be chosen such that $\tilde{\mu}$ is differentiable w.r.t. A at t_1 with derivative $P(\tilde{C}_{t_1})$, where $\tilde{C}_t := \{\omega \in \Omega \mid (\omega, t) \in ((\Omega \setminus C_{t_0}) \times [0, T]) \cap C\}$. We have

$$P(\tilde{C}_{t_1}) \geq P(C_{t_0}) - \tilde{\varepsilon} \quad \text{and} \quad P(C_{t_0} \cap \tilde{C}_{t_1}) = 0. \quad (3.34)$$

(3.33) implies that either

$$\limsup_{t \downarrow t_1} \frac{(P \otimes A)((\tilde{C}_{t_1} \times (t_1, t)) \cap C)}{A_t - A_{t_1}} > 0 \quad (3.35)$$

or

$$\liminf_{t \downarrow t_1} \frac{(P \otimes A)((\Omega \setminus (C_{t_0} \cup \tilde{C}_{t_1})) \times (t_1, t)) \cap C)}{A_t - A_{t_1}} \geq P(C_{t_0}) - \tilde{\varepsilon}. \quad (3.36)$$

In case of (3.35) we are done for the same reasons as in Case 1. Otherwise nearly the entire mass contributing to $\tilde{\mu}((t_1, t_1 + \varepsilon))$ has to come from $\Omega \setminus (C_{t_0} \cup \tilde{C}_{t_1})$. Proceeding with $\tilde{\tilde{\mu}} := (P \otimes A)((\Omega \setminus (C_{t_0} \cup \tilde{C}_{t_1}) \times D) \cap C)$ an iteration in this manner has to terminate after finitely many steps as (3.34) implies $P(C_{t_0} \cup \tilde{C}_{t_1}) \geq 2P(C_{t_0}) - \tilde{\varepsilon}$ and such an estimation can

only hold for finitely many t_i 's (namely, if $\tilde{\varepsilon} > 0$ is chosen small enough, $\left\lceil \frac{1}{P(C_{t_0})} \right\rceil + 1$ steps are needed) and we obtain for at least one of these t_i 's that $(P \otimes A)(\llbracket t_i, \tau_i \wedge (t_i + \varepsilon) \rrbracket \cap C) > 0$, for some $\varepsilon > 0$, where analogously to (3.26)

$$\tau_i = \inf \left\{ t \geq t_i \mid f(\theta_{t_i}, t) \notin \left[\frac{l-2}{2^m}, \frac{l+1}{2^m} \right] \right\}.$$

Again, as in (3.30), $\llbracket t_i, \tau_i \wedge (t_i + \varepsilon) \rrbracket \cap \{f(\theta, \cdot) \in (\frac{l-1}{2^m}, \frac{l}{2^m}]\} \subset \tilde{B}^l$, $(P \otimes A)$ -a.s. leads to a contradiction. Thus we have (3.24).

Step 2: In this step we construct a $\tilde{\theta} \in \mathcal{E}$ such that $f(\tilde{\theta}_t, t)$ approximates $f(\theta_t, t)$ (with a given accuracy).

By Assumption 3.5 there exists a $G \in L^2(A)$ with $|H(\vartheta, \cdot)| \leq G$ for all $\vartheta \in [-(\tilde{K} + 1), \tilde{K} + 1]$. As any predictable, nondecreasing process is locally bounded, there is a localizing sequence $(T_n)_{n \in \mathbb{N}}$ s.t. $E \left(\int_0^T 1_{\llbracket 0, T_n \rrbracket} G^2 dA \right) < \infty$. Aiming at convergence in probability, we can assume w.l.o.g. that $E \left(\int_0^T G^2 dA \right) < \infty$. Define $G_t^{(\delta)} := \frac{1_{(t \geq \delta)}}{A_t - A_{t-\delta}} \int_{t-\delta}^t G_s dA_s$. By Assumption 3.4 we have that

$$E \left(\int_0^T (G_t^{(\delta)})^2 dA_t \right) \leq E \left(\int_\delta^T \frac{\int_{t-\delta}^t G_s^2 dA_s}{A_t - A_{t-\delta}} dA_t \right) \leq \frac{a_{\max}}{a_{\min}} E \left(\int_0^T G_t^2 dA_t \right) < \infty.$$

This allows us to define a finite measure $\nu : \mathcal{P} \rightarrow \mathbb{R}_+$, absolutely continuous to $(P \otimes A)$, by

$$\nu(B) := E \left(\int_0^T (G_t^{(\delta)})^2 1_B dA \right), \quad \forall B \in \mathcal{P}. \quad (3.37)$$

Let $\varepsilon > 0$. As $(P \otimes A)$ is a finite measure we have

$$(P \otimes A)(\{|f(\theta, \cdot)| > K\}) \leq \varepsilon \quad (3.38)$$

for $K \in \mathbb{N}$ big enough. For such a K we will apply Step 1 to construct an approximating simple strategy.

Let $m \in \mathbb{N}$. Denote $\mathcal{I} := \{-2^m K, -2^m K + 1, \dots, 2^m K\}$. For each $l \in \mathcal{I}$ the essential supremum in (3.23) can be approximated by finitely many triples (T_1, T_2, Y) . As $\{|f(\theta, \cdot)| \leq K\} \cap (\Omega \times [\delta, T]) \subset \bigcup_{l \in \mathcal{I}} B^l$, there are finitely many simple processes

$Y_1 1_{\llbracket T_1^1, T_2^1 \rrbracket}, Y_2 1_{\llbracket T_1^2, T_2^2 \rrbracket}, \dots, Y_k 1_{\llbracket T_1^k, T_2^k \rrbracket}$, where $T_1^i \leq T_2^i$ are stopping times and Y_i are $\mathcal{F}_{T_1^i}$ -measurable real-valued random variables, with

$$\begin{aligned} (P \otimes A) \Big(\Big(\left\{ |f(\theta, \cdot)| \leq K \right\} \cap (\Omega \times [\delta, T]) \Big) \setminus \left\{ (\omega, t) \in \Omega \times [\delta, T] \mid f_\omega(\theta_t(\omega), t) \in \left(\frac{l-1}{2^m}, \frac{l}{2^m} \right], \right. \\ \left. T_1^i(\omega) < t \leq T_2^i(\omega), \text{ and } f(Y_i(\omega), t) \in \left[\frac{l-2}{2^m}, \frac{l+1}{2^m} \right], \text{ for some } l \in \mathcal{I} \text{ and some } i = 1, \dots, k \right\} \Big) \\ \leq \varepsilon. \end{aligned} \quad (3.39)$$

For $i = 1, \dots, k$ define

$$\begin{aligned} D_i := \Big\{ (\omega, t) \in \Omega \times [\delta, T] \mid i = \min \left\{ j = 1, \dots, k \mid T_1^j(\omega) < t \leq T_2^j(\omega), \right. \\ \left. f_\omega(Y_j(\omega), t) \in \left[\frac{l-2}{2^m}, \frac{l+1}{2^m} \right] \text{ and } f_\omega(\theta_t(\omega), t) \in \left(\frac{l-1}{2^m}, \frac{l}{2^m} \right] \right. \right. \\ \left. \left. \text{for some } l = -2^m K, \dots, 2^m K \right\} \right\}, \end{aligned}$$

with the convention that $\min \emptyset := \infty \notin \{1, \dots, k\}$. $(D_i)_{i=1, \dots, k}$ are disjoint predictable sets with

$$D_i \subset \llbracket T_1^i, T_2^i \rrbracket. \quad (3.40)$$

There is a finite union Γ_i of disjoint stochastic intervals s.t.

$$(P \otimes A) (\Gamma_i \Delta D_i) \leq \frac{\varepsilon}{k}, \quad i = 1, \dots, k, \quad (3.41)$$

where Δ denotes the symmetric difference (see again the proof of Proposition 4.7(ii) in [5]). By (3.40), Γ_i can be chosen (and actually are chosen) as subsets of $\llbracket T_1^i, T_2^i \rrbracket$. Denote $\Gamma := \bigcup_i \left((D_i \cap \Gamma_i) \setminus \left(\bigcup_{j \neq i} \Gamma_j \right) \right)$. As $(D_i)_{i=1, \dots, k}$ are disjoint we have by (3.38), (3.39), and (3.41) that

$$(P \otimes A) ((\Omega \times [\delta, T]) \setminus \Gamma) \leq 3\varepsilon. \quad (3.42)$$

Define the simple predictable $[-\tilde{K} - 1, \tilde{K}]$ -valued process

$$\tilde{\theta}_t(\omega) := \begin{cases} Y_i(\omega), & \text{where } i \text{ is the smallest number with } (\omega, t) \in \Gamma_i \\ 0, & \text{if no such } i \text{ exists.} \end{cases} \quad (3.43)$$

On the set Γ we have $|f_\omega(\tilde{\theta}_t(\omega), t) - f_\omega(\theta_t(\omega), t)| \leq 2^{-(m-1)}$ and arrive at

$$\begin{aligned} E \left(\int_\delta^T \left(\frac{\int_{t-\delta}^t H(\tilde{\theta}_t, s) dA_s}{A_t - A_{t-\delta}} - \frac{\int_{t-\delta}^t H(\theta_t, s) dA_s}{A_t - A_{t-\delta}} \right)^2 dA_t \right) \\ \leq A_T 2^{-2(m-1)} + 4E \left(\int_0^T 1_{(\Omega \times [\delta, T]) \setminus \Gamma} (G^{(\delta)})^2 dA_t \right). \end{aligned} \quad (3.44)$$

By (3.42) and as ν is a finite measure with $\nu \ll (P \otimes A)$, the rhs of (3.44) can be made arbitrarily small by the choice of $\varepsilon > 0$ and $m \in \mathbb{N}$.

Step 3: We are now in the position to complete the proof. For an arbitrary $\theta' \in \mathcal{E}$ with $|\theta'| \leq \tilde{K} + 1$ and $\delta > 0$ we obtain, using that $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$,

$$\begin{aligned}
& \int_0^T (H(\theta'_t, t) - H(\theta_t, t))^2 dA_t \\
& \leq 3 \int_\delta^T \left(\frac{\int_{t-\delta}^t H(\theta'_t, s) dA_s}{A_t - A_{t-\delta}} - H(\theta'_t, t) \right)^2 dA_t \\
& \quad + 3 \int_\delta^T \left(\frac{\int_{t-\delta}^t H(\theta_t, s) dA_s}{A_t - A_{t-\delta}} - H(\theta_t, t) \right)^2 dA_t \\
& \quad + 3 \int_\delta^T \left(\frac{\int_{t-\delta}^t H(\theta'_t, s) dA_s}{A_t - A_{t-\delta}} - \frac{\int_{t-\delta}^t H(\theta_t, s) dA_s}{A_t - A_{t-\delta}} \right)^2 dA_t + 4 \int_0^\delta G_t^2 dA_t \\
& \leq 6 \int_\delta^T \sup_{\vartheta \in [-\tilde{K}-1, \tilde{K}+1]} \left(\frac{\int_{t-\delta}^t H(\vartheta, s) dA_s}{A_t - A_{t-\delta}} - H(\vartheta, t) \right)^2 dA_t \\
& \quad + 3 \int_\delta^T \left(\frac{\int_{t-\delta}^t H(\theta'_t, s) dA_s}{A_t - A_{t-\delta}} - \frac{\int_{t-\delta}^t H(\theta_t, s) dA_s}{A_t - A_{t-\delta}} \right)^2 dA_t + 4 \int_0^\delta G_t^2 dA_t \\
& =: I + II + III.
\end{aligned} \tag{3.45}$$

By Assumption 3.6, I converges to zero in probability when $\delta \rightarrow 0$. III tends to zero pointwise when $\delta \rightarrow 0$ by dominated convergence. Given $\delta > 0$, by Step 2, also II gets arbitrarily small (w.r.t. a metric which metrizes convergence in probability) by taking for θ' the strategy $\tilde{\theta}$ in (3.43) with suitable $\varepsilon > 0$ and $m \in \mathbb{N}$. $\square$

3.1 Discussion of Assumption 3.6

Assumption 3.6 requires that the integrands $(\omega, t) \mapsto H_\omega(\vartheta, t)$, $\vartheta \in \mathbb{R}$, do not vary too much in time “jointly” in $\vartheta \in \mathbb{R}$. Note that if (3.20) were requested only for a fixed $\vartheta \in \mathbb{R}$, it *would* be automatically satisfied by Lebesgue’s differentiation theorem. The following proposition provides a sufficient condition for Assumption 3.6.

Proposition 3.10. *Let Assumptions 3.4 and 3.5 be satisfied. In addition, assume that for any $K \in \mathbb{R}_+$ there are sets $D_i \subset \mathbb{R}$ with $\cup_{i=1}^d D_i = [-K, K]$, $d \in \mathbb{N}$, and a nonincreasing sequence of predictable processes $G^n : \Omega \times [0, T] \rightarrow \mathbb{R}_+$ with $P\left(\int_0^T (G^n)^2 dA < \infty\right) = 1$*

s.t. for all $i = 1, \dots, d$

$$|H_\omega(\vartheta_2, t) - H_\omega(\vartheta_1, t)| \leq G_t^n(\omega), \quad \forall \vartheta_1, \vartheta_2 \in D_i, \quad |\vartheta_2 - \vartheta_1| \leq \frac{1}{n}, \quad \omega \in \Omega, t \in [0, T] \quad (3.46)$$

and $G^n \rightarrow 0$, $(P \otimes A)$ -a.s., $n \rightarrow \infty$. Then Assumption 3.6 is satisfied.

Remark 3.11. If H is continuous in ϑ , it is for fixed (ω, t) uniformly continuous in ϑ on $[-K, K]$. This implies that (3.46) holds with $d = 1$, $D_1 = [-K, K]$, and $G_t^n(\omega) = \sup_{\vartheta_1, \vartheta_2 \in [-K, K] \cap \mathbb{Q}, |\vartheta_2 - \vartheta_1| \leq 1/n} |H_\omega(\vartheta_2, t) - H_\omega(\vartheta_1, t)|$.

(3.46) holds accordingly if $\vartheta \mapsto H_\omega(\vartheta, t)$ possesses only discontinuities of the first kind at finitely many ϑ .

Proof of Proposition 3.10. Let $K \in \mathbb{R}_+$. There exists a sequence $(\vartheta_m)_{m \in \mathbb{N}} \subset [-K, K]$ s.t. for all $n \in \mathbb{N}$ and suitable $m = m(n) \in \mathbb{N}$ the following holds

$$\forall \vartheta \in [-K, K] \exists k \leq m(n) \text{ s.t. } |\vartheta - \vartheta_k| \leq \frac{1}{n} \text{ and } \vartheta, \vartheta_k \in D_i \text{ for some } i = 1, \dots, d.$$

Thus we have

$$\sup_{\vartheta \in [-K, K]} \min_{k=1, \dots, m(n)} |H(\vartheta_k, s) - H(\vartheta, s)| \leq G_s^n, \quad 0 \leq s \leq T.$$

With standard localization arguments we can w.l.o.g. assume that $\int_0^T (G_t^1)^2 dA_t$ is bounded.

We have

$$E \left(\int_\delta^T \left(\frac{\int_{t-\delta}^t G_s^n dA_s}{A_t - A_{t-\delta}} \right)^2 dA_t \right) \leq E \left(\int_\delta^T \frac{\int_{t-\delta}^t (G_s^n)^2 dA_s}{A_t - A_{t-\delta}} dA_t \right) \leq \frac{a_{\max}}{a_{\min}} E \left(\int_0^T (G_t^n)^2 dA_t \right)$$

and

$$\begin{aligned} & E \left(\int_\delta^T \sup_{\vartheta \in [-K, K]} \left(\frac{\int_{t-\delta}^t H(\vartheta, s) dA_s}{A_t - A_{t-\delta}} - H(\vartheta, t) \right)^2 dA_t \right) \\ & \leq 3 \sum_{k=1}^{m(n)} E \left(\int_\delta^T \left(\frac{\int_{t-\delta}^t H(\vartheta_k, s) dA_s}{A_t - A_{t-\delta}} - H(\vartheta_k, t) \right)^2 dA_t \right) \\ & + 3E \left(\int_\delta^T \sup_{\vartheta \in [-K, K]} \min_{k=1, \dots, m(n)} \left(\left(\frac{\int_{t-\delta}^t H(\vartheta_k, s) dA_s}{A_t - A_{t-\delta}} - \frac{\int_{t-\delta}^t H(\vartheta, s) dA_s^i}{A_t - A_{t-\delta}} \right)^2 \right. \right. \\ & \quad \left. \left. + \left(H(\vartheta_k, t) - H(\vartheta, t) \right)^2 \right) dA_t \right) =: I + II. \end{aligned} \quad (3.47)$$

Let $\varepsilon > 0$. Take $n = n_\varepsilon \in \mathbb{N}$ large enough s.t. $\frac{a_{\max}}{a_{\min}} E \left(\int_0^T (G^n)^2 dA_t \right) \leq \frac{\varepsilon}{9}$. This implies that $II \leq \frac{2}{3}\varepsilon$. By Assumption 3.5 and standard localization arguments we can w.l.o.g. assume that $\int_0^T H^2(\vartheta_k, t) dA_t$ are bounded for all $k \in \mathbb{N}$. By Lebesgue's differentiation theorem and dominated convergence we have that there exists a $\delta(\varepsilon) > 0$ s.t. for all $\delta \leq \delta(\varepsilon)$ it holds that $I \leq \frac{\varepsilon}{3}$. Putting this together we obtain

$$E \left(\int_\delta^T \sup_{\vartheta \in [-K, K]} \left(\frac{\int_{t-\delta}^t H(\vartheta, s) dA_s}{A_t - A_{t-\delta}} - H(\vartheta, t) \right)^2 dA_t \right) \leq \varepsilon, \quad \forall \delta \leq \delta(\varepsilon). \quad (3.48)$$

□

References

- [1] P. Bank and D. Baum. Hedging and portfolio optimization in financial markets with a large trader. *Mathematical Finance*, 14(1):1–18, 2004.
- [2] R. A. Carmona and D. Nualart. *Nonlinear Stochastic Integrators, Equations and Flows*, volume 6. Stochastics Monographs, Gordon and Breach Science Publishers, 1990.
- [3] R. Chitashvili and M. Mania. Characterization of a regular family of semimartingales by line integrals. *Georgian Mathematical Journal*, 3(6):525–542, 1996.
- [4] R. Chitashvili and M. Mania. Stochastic line integrals with respect to local martingales and semimartingales. *Proceedings of A. Razmadze, Mathematical Institute*, 120:1–26, 1999.
- [5] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, Cambridge, 1992.
- [6] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, second edition, 2003.
- [7] J. Kallsen. *Semimartingale Modelling in Finance*. PhD thesis, University of Freiburg, 1998.

- [8] H. Kunita. *Stochastic flows and stochastic differential equations*. Cambridge University Press, Cambridge, 1990.
- [9] Y. Le Jan. Flots de diffusions dans $\mathbb{R}^d$. *C. R. Acad. Sc. Paris, Série I*, 294:697–699, 1982.
- [10] P. Protter. *Stochastic Integration and Differential Equations*. Springer, Berlin, second edition, 2004.
- [11] A.S. Sznitman. Martingales dépendant d'un paramètre: une formule d'Itô. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 60:41–70, 1982.