47

PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 12, Fasc. 2 (1991), pp. 165–200

SEMIMARTINGALE INTEGRALS VIA DECOUPLING INEQUALITIES AND TANGENT PROCESSES*

BY

STANISŁAW KWAPIEŃ (WARSZAWA)
AND WOJBOR A. WOYCZYŃSKI (CLEVELAND, OHIO)

Dedicated to
Professor KAZIMIERZ URBANIK
on his 60-th birthday

Abstract. A previsible process F is integrable with respect to a semimartingale X if and only if F belongs to a randomized Musielak–Orlicz space $L_{\varphi(\omega)}$, where φ is explicitly expressed in terms of the Grigelionis characteristics of X. Decoupling inequalities and tangent processes are the main tool used in the proof.

1. Introduction. Semimartingales form a natural class of processes with respect to which the stochastic integration is feasible. More precisely, by results of C. Dellacherie and K. Bichteler, classical by now, they are a maximal space of processes X for which the stochastic integral operator $F \to \int F dX$ is a continuous operator from the space of bounded predictable processes with the supremum norm into $L^0(\Omega, \mathscr{F}, P)$. Each semimartingale has random characteristics B, μ and C (Section 4) which were originally inroduced by B. Grigelionis. If X has independent increments (and only in this case), B, μ and C become deterministic, and in the case of a stochastically continuous process X with independent increments they coincide with the usual Lévy characteristics which appear in the Lévy–Hinchin formula for the characteristic function of X.

The principal goal of the present paper is to describe analytically the space of X-integrable predictable processes in terms of the Grigelionis characteristics B, μ and C of a semimartingale X. The space turns out to be a randomized Musielak–Orlicz space $\mathscr{L}_{\varphi}$, and an explicit formula for φ as a functional of B, μ and C is obtained (Section 6).

The above goal is achieved by:

(i) constructing a decoupled tangent process $\tilde{X}$ to X (à la Jacod [11] and following an old idea of Itô [8]), which, in a sense, behaves as if it had independent increments (Section 4);

* Supported in part by an NSF Grant DMS87-13103 while the first-named author was visiting at Case Western Reserve University.

(ii) showing, via decoupling inequalities, that F is X-integrable if and only if F is $\tilde{X}$-integrable, and proving that $\tilde{X}$-integrability of F is equivalent to pathwise $\tilde{X}$-integrability of F (Section 6);

(iii) obtaining a complete description of deterministic functions integrable with respect to a process with independent increments with given Lévy characteristics (Section 5).

Our exposition here is deliberately as elementary as we could make it and uses only basic martingale properties and stopping time techniques. Essentially, no prior knowledge of the semimartingale theory or of the general theory of stochastic integration is necessary here.

A primitive version of the ideas developed in this paper was used previously by the authors (cf. [16] and [17]) to study single and double stochastic integrals with respect to stochastically continuous, symmetric processes with independent increments.

The underlying ideas of this paper can be traced to many papers which exploited the "conditioning" techniques. Let us just mention here the papers by Burkholder [3], Hill [5], Itô [8], Jacod [11], Jakubowski [13], Jakubowski and Słonimski [14], Kallenberg [15], Szulga [23], which have closest connections to our subject matter.

2. Decoupling inequalities and tangent sequences. The present section is a collection of basic inequalities and definitions which are necessary for the development of material of Sections 3–6, and which are also of independent interest. The techniques are essentially those of Burkholder [2].

Let $(\Omega, \mathscr{F}, P)$ be a probability space and let $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots$ be an ascending sequence of sub-σ-fields (filtration) of $\mathscr{F}$. For a sequence $\xi = (\xi_i)$ of random variables we put

$$\xi^* := \sup_i |\xi_i|, \quad \sum{}^* \xi_i := \sup_k \Big|\sum_{i=1}^{k} \xi_i\Big|, \quad \sum{}_*^* \xi_i := \sup_{k,l} \Big|\sum_{i=k}^{l} \xi_i\Big|.$$

Definition 2.1. We shall say that two $(\mathscr{F}_i)$-adapted sequences of random variables (ξ_i) and (η_i) are *tangent* if for each $i = 1, 2, \ldots$

$$\xi_i \sim_{\mathscr{F}_{i-1}} \eta_i,$$

i.e. for each $c \in \boldsymbol{R}$ we have $P(\xi_i < c \mid \mathscr{F}_{i-1}) = P(\eta_i < c \mid \mathscr{F}_{i-1})$ a.s.

Equivalently, (ξ_i) and (η_i) are tangent if, and only if, for each $(\mathscr{F}_i)$-predictable, bounded sequence (v_i) (i.e. (v_i) is $(\mathscr{F}_{i-1})$-adapted) and for each sequence of Borel measurable, bounded functions (φ_i),

$$\mathrm{E} v_i \varphi_i(\xi_i) = \mathrm{E} v_i \varphi_i(\eta_i), \quad i = 1, 2, \ldots,$$

so that, in particular, for any stopping time τ and any tangent sequences (ξ_i) and (η_i),

$$\mathrm{E} \sum_{i=1}^{\tau} v_i \varphi_i(\xi_i) = \mathrm{E} \sum_{i=1}^{\tau} v_i \varphi_i(\eta_i).$$

If (ξ_i) and (η_i) are tangent, then for any $(\mathscr{F}_i)$-predictable sequence (v_i), the sequences $(v_i\xi_i)$ and $(v_i\eta_i)$ are tangent as well.

A simple example of tangent sequences (which also motivated our definition) is given by $(\xi_i) = (v_i\beta_i)$ and $(\eta_i) = (v_i\beta_i')$, where (β_i), (β_i') are independent copies of a sequence of independent random variables and the sequence (v_i) is predictable with respect to the filtration

$$\mathscr{F}_i = \sigma(\beta_1, \ldots, \beta_i, \beta_1', \ldots, \beta_i'), \quad i = 1, 2, \ldots$$

In all the inequalities of this paper, sequences of random variables will always have only finitely many non-zero terms, so we need not concern ourselves with the question of convergence of their series. In this context the sign $\sum$ will always mean that the summation extends over all $i = 1, 2, \ldots$ Also, by definition,

$$[\xi_i]^c = \begin{cases} c & \text{if } \xi_i > c, \\ \xi_i & \text{if } |\xi_i| \leqslant c, \\ -c & \text{if } \xi_i < -c. \end{cases}$$

Routinely, the standard truncation $[\ ]^1$ will be written as $[\ ]$ and the brackets will not be used here for any other purpose. The first result of this section is a basic inequality which compares tail probabilities of maximal functions of sums of tangent sequences.

THEOREM 2.1. *Let (d_i) and (e_i) be two tangent sequences of random variables. Then, if $f_n = d_1 + \ldots + d_n$, $g_n = e_1 + \ldots + e_n$, $n = 1, 2, \ldots$, then, for each $a, b, c > 0$ with $c \geqslant 2b$,*

$$(2.1)\qquad P(f^* > a) \leqslant \frac{5}{4}\,\frac{2b+c}{a} + 3P(g^* > b) + 2P\big(\textstyle\sum^* \mathrm{E}([e_i]^c \,|\, \mathscr{F}_{i-1}) > b\big).$$

Proof. Let a, b, c be arbitrary positive real numbers with $c \geqslant 2b$, and let us define stopping times

$$\sigma := \inf\{k\colon |e_k| > c\}, \qquad \tau := \inf\{k\colon \Big|\sum_{i=1}^{k} ([d_i]^c - \mathrm{E}([d_i]^c \,|\, \mathscr{F}_{i-1}))\Big| > \tfrac{4}{5}a\},$$

$$\varrho := \inf\{k\colon \Big|\sum_{i=1}^{k} [e_i]^c\Big| > b\}, \qquad \lambda := \inf\{k\colon \Big|\sum_{i=1}^{k+1} \mathrm{E}([e_i]^c \,|\, \mathscr{F}_{i-1})\Big| > b\}.$$

Clearly,

$$P(f^* > a) \leqslant P(d^* > c) + P\big(\textstyle\sum^* ([d_i]^c - \mathrm{E}([d_i]^c \,|\, \mathscr{F}_{i-1})) > \tfrac{4}{5}a\big)$$
$$+ P\big(\textstyle\sum^* \mathrm{E}([d_i]^c \,|\, \mathscr{F}_{i-1}) > \tfrac{1}{5}a\big),$$

and, under the above notation, the three terms on the right-hand side can be estimated from above as follows:

$$P(d^* > c) \leqslant \mathrm{E} \sum_{i=1}^{\sigma} I(|d_i| > c) + P(\sigma < \infty) = \mathrm{E} \sum_{i=1}^{\sigma} I(|e_i| > c) + P(\sigma < \infty)$$
$$= 2P(e^* > c) \leqslant 2P(g^* > c/2) \leqslant 2P(g^* > b).$$

Next

$$P\big(\sum\nolimits^* ([d_i]^c - \mathrm{E}([d_i]^c \mid \mathscr{F}_{i-1})) > \tfrac{4}{5} a\big)$$
$$\leqslant P(\varrho < \infty) + P(\lambda < \infty) + \frac{5}{4} a^{-1} \Big(\mathrm{E}\big(\sum_{i=1}^{\tau \wedge \varrho \wedge \lambda} ([d_i]^c - \mathrm{E}([d_i]^c \mid \mathscr{F}_{i-1}))\big)^2\Big)^{1/2}$$
$$= P\big(\sum\nolimits^* [e_i]^c > b\big) + P\big(\sum\nolimits^* \mathrm{E}([e_i]^c \mid \mathscr{F}_{i-1}) > b\big)$$
$$+ \frac{5}{4} a^{-1} \Big(\mathrm{E}\big(\sum_{i=1}^{\tau \wedge \varrho \wedge \lambda} ([e_i]^c - \mathrm{E}([e_i]^c \mid \mathscr{F}_{i-1}))\big)^2\Big)^{1/2}$$
$$\leqslant P(g^* > b) + P\big(\sum\nolimits^* \mathrm{E}([e_i]^c \mid \mathscr{F}_{i-1}) > b\big) + \frac{5}{4} \frac{2b+c}{a},$$

because, by the definitions of ϱ and λ,

$$\Big|\sum_{i=1}^{\tau \wedge \varrho \wedge \lambda} ([e_i]^c - \mathrm{E}([e_i]^c \mid \mathscr{F}_{i-1}))\Big| \leqslant 2b + c.$$

Inequality (2.1) is trivial in the case

$$\frac{5}{4} \frac{2b+c}{a} \geqslant 1,$$

so suppose that

$$\frac{5}{4} \frac{2b+c}{a} < 1.$$

Then $a/5 > b$, and

$$P\big(\sum\nolimits^* \mathrm{E}([d_i]^c \mid \mathscr{F}_{i-1}) > a/5\big) \leqslant P\big(\sum\nolimits^* \mathrm{E}([e_i]^c \mid \mathscr{F}_{i-1}) > b\big),$$

which, together with two previous estimates, gives inequality (2.1). Q.E.D.

Remark 2.1. In [18] we obtained a modified version of the inequality in Theorem 2.1 which is more useful for some applications. We proved there that if a, b, c and (d_i), (e_i) are as in Theorem 2.1, then

$$P(f^* > a) \leqslant 6\frac{b}{a} + 3\frac{a+c}{b} P(g^* > b) + 2P\big(\sum\nolimits^* \mathrm{E}([e_i]^c \mid \mathscr{F}_{i-1}) > b\big).$$

It is also proved there that if, additionally, either (d_i) (and hence also (e_i)) is a conditionally symmetric sequence, i.e., $d_i \sim_{\mathscr{F}_{i-1}} -d_i$, $i = 1, 2, \ldots,$ or (d_i) is a non-negative sequence, then we simply have

$$P(f^* > a) \leqslant 3\big(b/a + P(g^* > b)\big).$$

An interested reader is also referred to that paper for more results on tangent sequences.

The last term in the basic inequality (2.1) of Theorem 2.1 is, in general, quite complicated to evaluate. However, in the special case of conditionally independent sequences, to be defined below, it is controlled by the term $P(g^* > b)$, which considerably simplifies applying of Theorem 2.1.

DEFINITION 2.2. An $(\mathscr{F}_i)$-adapted sequence (e_i) is said to satisfy *condition* (CI) if there exists a σ-field $\mathscr{G} \subset \mathscr{F}$ such that $\mathscr{L}(e_i \mid \mathscr{F}_{i-1}) = \mathscr{L}(e_i \mid \mathscr{G})$ a.s. for $i = 1, 2, \ldots$, and such that (e_i) is a sequence of $\mathscr{G}$-conditionally independent random variables.

If (e_i) satisfies condition (CI), then the σ-field $\mathscr{G}$ can always be selected to be equal to $\sigma\big(\mathscr{L}(e_i \mid \mathscr{F}_{i-1}), i = 1, 2, \ldots\big)$.

LEMMA 2.1. *If an $(\mathscr{F}_i)$-adapted sequence (e_i) satisfies condition* (CI), *then for each $b > 0$, $c \geqslant b/4$*

$$P\big(\textstyle\sum^* \mathrm{E}([e_i]^c \mid \mathscr{F}_{i-1}) > b\big) \leqslant \frac{8(c+b)}{b} P\left(g^* > \frac{b}{8}\right),$$

where $g_n = e_1 + \ldots + e_n$.

Proof. If (ξ_i) is a sequence of independent random variables, then, following [6], for any $s, t, a > 0$ we have

$$P\big(\textstyle\sum^* \xi_i > s+t+a\big) \leqslant P(\xi^* > a) + P\big(\sum^* \xi_i > s\big) P\big(\sum^*_* \xi_i > t\big),$$

and

$$\int_0^\infty P\big(\textstyle\sum^* \xi_i > 2s+t\big)ds \leqslant \int_0^\infty P(\xi^* > s)ds + P\big(\sum^*_* \xi_i > t\big) \int_0^\infty P\big(\sum^* \xi_i > s\big)ds,$$

so that

$$\mathrm{E}\textstyle\sum^* \xi_i \leqslant \dfrac{t/2 + \mathrm{E}\xi^*}{1/2 - P(\sum^*_* \xi_i > t)} \tag{2.2}$$

for each t for which the right-hand side is positive. Therefore, for each $c > t > 0$,

$$\mathrm{E}\textstyle\sum^* [\xi_i]^c \leqslant \dfrac{t/2 + \mathrm{E}\sup_i |[\xi_i]^c|}{1/2 - P(\sum^*_* [\xi_i]^c > t)}$$

$$\leqslant \frac{t/2 + cP(\sup_i |[\xi_i]^c| > t) + t}{1/2 - P(\sum^*_* [\xi_i]^c > t)} \leqslant \frac{3t/2 + cP(\sum^* \xi_i > t/2)}{1/2 - P(\sum^* \xi_i > t/2)},$$

because, for $c > t$, $\{\sum_*^* [\xi_i]^c > t\} \subset \{\sum^* \xi_i > t/2\}$. So, if $\sum^* \mathrm{E}[\xi_i]^c > 4t$, then

$$P(\sum^* \xi_i > t/2) \geqslant t/(2(c+4t)).$$

Inserting $t = b/4$ and applying the above implication to the $\mathscr{G}$-conditionally independent sequence (e_i), we see that a.s. if $\sum^* \mathrm{E}([e_i]^c | \mathscr{G}) \geqslant b$, then

$$P(\sum^* e_i > b/8 | \mathscr{G}) \geqslant b(8(c+b))^{-1}.$$

Hence, integrating the last inequality over the event $\{\sum^* \mathrm{E}([e_i]^c | \mathscr{G}) \geqslant b\}$ we get

$$P(\sum^* \mathrm{E}([e_i]^c | \mathscr{G}) > b) \leqslant \frac{8(c+b)}{b} P(\sum^* e_i > b/8),$$

which, in view of condition (CI), concludes the proof of Lemma 2.1.

The above lemma and Theorem 2.1 immediately give

THEOREM 2.2. *If (d_i) and (e_i) are tangent sequences and (e_i) satisfies condition* (CI), *then for each $a, b > 0$*

$$P(f^* > a) \leqslant 40\frac{b}{a} + 51P(g^* > b),$$

where $f_n = d_1 + \ldots + d_n$ and $g_n = e_1 + \ldots + e_n$.

The next lemma shows how the last term in the basic inequality (2.1) of Theorem 2.1 can also be controlled by another expression which will prove to be useful later on.

LEMMA 2.2. *If (e_i) is an $(\mathscr{F}_i)$-adapted sequence, then, for each $a, b, c > 0$ and $c \geqslant 2a$,*

$$P(\sum |\mathrm{E}([e_i]^c | \mathscr{F}_{i-1})| > b) \leqslant \frac{a}{b} + 2\frac{b+c}{b} \sup_{(v_i)\in\mathscr{P}_1} P(|\sum v_i e_i| > a),$$

where $\mathscr{P}_1$ is the class of all $(\mathscr{F}_i)$-predictable sequences (v_i) such that $|v_i| \leqslant 1$, $i = 1, 2, \ldots$

Proof. Let $u_i = \operatorname{sgn} \mathrm{E}([e_i]^c | \mathscr{F}_{i-1})$, $i = 1, 2, \ldots$ Then $(u_i) \in \mathscr{P}_1$. Let

$$\lambda := \inf_k \{k\colon |\sum_{i=1}^k u_i [e_i]^c| > a\}.$$

Then we have

$$P(\sum |\mathrm{E}([e_i]^c | \mathscr{F}_{i-1})| > b) = P(\sum u_i \mathrm{E}([e_i]^c | \mathscr{F}_{i-1}) > b)$$

$$\leqslant P(\lambda < \infty) + \frac{\mathrm{E}(\sum_{i=1}^{\lambda} u_i \mathrm{E}([e_i]^c | \mathscr{F}_{i-1}))}{b} = P(\lambda < \infty) + \frac{\mathrm{E}(\sum_{i=1}^{\lambda} u_i [e_i]^c)}{b}$$

$$\leqslant P(\textstyle\sum^* u_i[e_i]^c > a) + \frac{a}{b} P(\sum^* u_i[e_i]^c \leqslant a) + \frac{a+c}{b} P(\sum^* u_i[e_i]^c > a)$$

$$= \frac{a}{b} + \frac{b+c}{b} P(\textstyle\sum^* u_i[e_i]^c > a) \leqslant \frac{a}{b} + \frac{b+c}{b} [P(\sum^* u_i e_i > a) + P(e_i^* > c)]$$

$$\leqslant \frac{a}{b} + \frac{b+c}{b} \left[P(\textstyle\sum^* u_i e_i > a) + P\left(\sum^* e_i > \frac{c}{2}\right)\right]$$

$$\leqslant \frac{a}{b} + \frac{b+c}{b} 2 \sup_{(v_i) \in \mathscr{P}_1} P(|\textstyle\sum v_i e_i| > a),$$

where the last inequality follows from the following simple Lemma 2.3. Q.E.D.

LEMMA 2.3. *For each $(\mathscr{F}_i)$-adapted sequence of random variables (h_i)*

$$P(\textstyle\sum^* h_i > a) \leqslant \sup_{(v_i) \in \mathscr{P}_1} P(|\sum v_i h_i| > a).$$

The proof follows immediately by taking $v_i = I(\tau \geqslant i)$, where $\tau = \inf\{k\colon |\sum_{i=1}^k h_i| > a\}$.

Lemmas 2.2 and 2.3 and Theorem 2.1 immediately give

COROLLARY 2.1. *For each $\varepsilon > 0$ there exists a $\delta > 0$ such that if (d_i) and (e_i) are tangent sequences and if the inequality $P(|\sum v_i d_i| > \delta) < \delta$ holds true for every $(\mathscr{F}_i)$-predictable sequence (v_i) with $|v_i| \leqslant 1$, $i = 1, 2, \ldots$, then $P(\sum^* e_i > \varepsilon) < \varepsilon$.*

In general, the last term in the inequality (2.1) of Theorem 2.1 cannot be omitted. This can be seen from the following

EXAMPLE 2.1. Let (ξ_i) be a sequence of independent random variables and let (ξ_i') be its independent copy. Define $d_{2i+1} = \xi_i'$, $d_{2i+2} = -\xi_i$, $e_{2i+1} = \xi_i$, $e_{2i+2} = -\xi_i$, $i = 0, 1, 2, \ldots$ Both sequences (e_i) and (d_i) are $(\mathscr{F}_i)$-adapted and tangent for $\mathscr{F}_i := \sigma(d_j; e_j; j = 1, 2, \ldots, i)$. Moreover, (d_i) has property (CI) with $\mathscr{G} = \sigma((\xi_i))$. In view of the construction, the partial sums of $\sum e_i$ are either 0 or ξ_i and $\sum d_i = \sum(\xi_i' - \xi_i)$. Hence, for each $a, b > 0$,

$$P(f^* > a) \geqslant P(\textstyle\sum^*(\xi_i' - \xi_i) > a) \quad \text{and} \quad P(g^* > b) = P(\xi_i^* > b),$$

so that it is possible for $P(g^* > b)$ to be small without $P(f^* > a)$ being small.

3. Decoupled tangent sequences. Tangent sequences with property (CI) share many properties with sequences of independent random variables and the main idea of the present paper is to construct, for a given sequence (d_i) (or, in subsequent sections, for a process), a tangent sequence (or process) with property (CI) (or its analogue for a process), and then, via inequalities of Section 2, deduce results about (d_n) from results about independent random variables.

DEFINITION 3.1. Let (d_i) be an $(\mathscr{F}_i)$-adapted sequence on a filtered probability space $(\Omega, \mathscr{F}, P; (\mathscr{F}_i))$. For any filtered space $(\Omega', \mathscr{F}', (\mathscr{F}_i'))$ and any probability transition function $P'\colon \Omega \times \mathscr{F} \to \boldsymbol{R}^+$, a sequence $(\tilde{d}_i)$ defined on $\tilde{\Omega} = \Omega \times \Omega'$ and adapted to the filtration $(\tilde{\mathscr{F}}_i) = (\mathscr{F}_i \otimes \mathscr{F}_i')$ is said to be a *decoupled tangent* sequence to (d_i) if

(a) for each $\omega \in \Omega$, $(\tilde{d}_i(\omega, \cdot))$ is a sequence of independent random variables on $(\Omega', \mathscr{F}', P'(\omega, \cdot))$,

(b) the sequences $(\tilde{d}_i)$ and $(\bar{d}_i)$, where $\bar{d}_i(\omega, \omega') := d_i(\omega)$, $(\omega, \omega') \in \Omega \times \Omega'$, $i = 1, 2, \ldots$, are tangent on the filtered probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}; (\tilde{\mathscr{F}}_i))$, where $\tilde{P}$ is defined by the formula

$$\tilde{P}(A \times B) := P \otimes P'(A \times B) = \int_A P'(\omega, B) P(dw), \qquad A \in \mathscr{F},\ B \in \mathscr{F}'.$$

(In the sequel, the trivial extension $\bar{d}_i$ of d_i will be simply denoted by d_i without any risk of misunderstanding.)

Clearly, a decoupled tangent sequence satisfies condition (CI) with respect to the σ-field $\mathscr{G} = \mathscr{F}$ (or, more precisely, with respect to $\mathscr{G} = \mathscr{F} \otimes \{\Omega, \emptyset\}$).

For a given sequence (d_i) there is a canonical way to construct a decoupled tangent sequence: Let $\Omega' = \boldsymbol{R}^N$, $\mathscr{F}_i'$ be the σ-field generated by the first i coordinates in $\boldsymbol{R}^N$ and, finally, let

$$P'(\omega, B) = \Big(\bigotimes_{i=1}^{\infty} \mathscr{L}(d_i \mid \mathscr{F}_{i-1})(\omega)\Big)(B).$$

The sequence $\tilde{d}_i(\omega, (x_j)) = x_i$, $i = 1, 2, \ldots$, is a decoupled tangent sequence to d_i, $i = 1, 2, \ldots$

EXAMPLE 3.1. Let

$$(\Omega, \mathscr{F}, P) = \bigotimes_{i=1}^{\infty} (\Omega_i, \mathscr{H}_i, P_i)$$

be an infinite product probability space with $\omega = (\omega_1, \omega_2, \ldots)$ and let $\mathscr{F}_i$ be the σ-field which depends only on the first i coordinates $\omega_1, \ldots, \omega_i$. If $d_i = d_i(\omega_1, \ldots, \omega_i)$, $i = 1, 2, \ldots$, then the sequence

$$\tilde{d}_i(\omega, \omega') := d_i(\omega_1, \omega_2, \ldots, \omega_{i-1}, \omega_i'), \qquad i = 1, 2, \ldots,$$

defined on $(\Omega \times \Omega, \mathscr{F} \otimes \mathscr{F}, P \otimes P; (\mathscr{F}_i \otimes \mathscr{F}_i))$ is a decoupled tangent sequence to (d_i). Note that in this case $P \otimes P$ is just a product measure.

In particular, if $\xi_1, \xi_2, \ldots$ is a sequence of independent random variables and $v_1, v_2, \ldots$ is a predictable sequence with respect to $\mathscr{F}_i = \sigma(\xi_1, \ldots, \xi_i)$ (i.e., v_i is $\mathscr{F}_{i-1}$-measurable, $i = 1, 2, \ldots$) and $(d_i) = (v_i \xi_i)$, then $(\tilde{d}_i) := (v_1 \xi_1')$, where (ξ_1') is an independent copy of (ξ_i), is a decoupled tangent sequence to (d_i) (cf. [16]).

The following corollary, which parallels Corollary 2.1 (but formally does not follow from it), will play a pivotal role in Section 6 in applications of the notion of a decoupled tangent sequence to theory of stochastic integrals.

Corollary 3.1. *For any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $(\tilde{d}_i)$ is a decoupled tangent sequence to (d_i) and if the inequality $P(|\sum v_i d_i| > \delta) < \delta$ holds true for every $(\mathscr{F}_i)$-predictable sequence (v_i) with $|v_i| \leqslant 1$, $i = 1, 2, \ldots$ (i.e. $(v_i) \in \mathscr{P}_1$), then*

$$\tilde{P}(\sum^* v_i \tilde{d}_i > \varepsilon) < \varepsilon$$

for every such sequence (v_i).

Proof. By Lemma 2.2 we have

$$P(\sum^* \mathrm{E}([v_i d_i]^c \,|\, \mathscr{F}_{i-1}) > b) \leqslant \frac{a}{b} + 2\frac{a+c}{b} \sup_{(w_i) \in \mathscr{P}_1} P(|\sum w_i d_i| > a),$$

and, in view of the definition of $\tilde{P}$,

$$\tilde{P}(\sum^* \mathrm{E}([v_i \tilde{d}_i]^c \,|\, \tilde{\mathscr{F}}_{i-1}) > b) = P(\sum^* \mathrm{E}([v_i d_i]^c \,|\, \mathscr{F}_{i-1}) > b).$$

Similarly,

$$P(\sum^* v_i d_i > a) \leqslant \sup_{(w_i) \in \mathscr{P}_1} P(|\sum w_i d_i| > a)$$

by Lemma 2.3, and $\tilde{P}(\sum^* v_i \tilde{d}_i > a) = P(\sum^* v_i d_i > a)$. Since the sequences $(v_i \tilde{d}_i)$ and $(v_i \tilde{d}_i)$ are tangent on $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}; (\tilde{\mathscr{F}}_i))$, an application of Theorem 2.1 concludes the proof. Q.E.D.

4. Decoupled tangent processes. Let $T = [0, t_\infty]$, let $(\Omega, \mathscr{F}, P; (\mathscr{F}(t)))$ be a probability space with filtration satisfying the standard assumptions (i.e., right continuity and completeness), and let $X(t)$, $t \in T$, be a process with sample paths in the Skorohod space $D(T)$ and adapted to $(\mathscr{F}(t))$. In the present section we describe a concept of a decoupled tangent process to $X(t)$ (which was introduced by Jacod [11] as a tangent process), study conditions for its existence and its properties.

Let

$$\pi^n = \{(t_k^n)\colon\ 0 = t_0^n < \ldots < t_{k_n}^n = t_\infty\} \tag{4.1}$$

be a normal sequence of partitions of T (i.e., $\lim_{n\to\infty} \max_{1 \leqslant k \leqslant k_n} |t_k^n - t_{k-1}^n| = 0$). For each n, consider the sequence

$$d_k^n = X(t_k^n) - X(t_{k-1}^n), \qquad k = 1, \ldots, k_n, \tag{4.2}$$

which is $(\mathscr{F}(t_k^n))$-adapted. Let $(\tilde{d}_k^n)$ be a decoupled tangent sequence to (d_k^n) defined on the filtered probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}; (\tilde{\mathscr{F}}_k))$ (depending on n) as described in Definition 3.1.

Next, let us define processes on $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$ by the formulas

$$\tilde{X}^n(t) := \sum_{k: t_k^n \leqslant t} \tilde{d}_k^n, \qquad n = 1, 2, \ldots,$$

and random variables on Ω, with values in the space of probability distributions on $D(T)$, by the formula

$$M_X^n(\omega) := \mathscr{L}(\tilde{X}^n(\omega, \cdot, t), t \in T), \quad n = 1, 2, \ldots$$

Notice that by the definition of a decoupled tangent sequence, for each $\omega \in \Omega$, the stochastic process $\tilde{X}^n(\omega, \cdot, t)$, $t \in T$, on Ω', has independent increments with respect to the probability $P'(\omega, \cdot)$.

DEFINITION 4.1. We shall say that $X(t)$, $t \in T$, *admits a decoupled* (π^n)*-tangent process* (the partition (π^n) usually will not be explicitly mentioned, although see Remark 4.5) if the sequence of random variables M_X^n, $n = 1, 2, \ldots$, on $(\Omega, \mathscr{F}, P)$, with values being measures on $D(T)$, converges in probability P. As usual, the space of measures on $D(T)$ is equipped with the topology of weak convergence.

If $X(t)$, $t \in T$, admits a decoupled tangent process, then any process $\tilde{X}(t)$, $t \in T$, defined on $(\Omega \times \Omega', \mathscr{F} \otimes \mathscr{F}', P \otimes P'; (\mathscr{F}(t) \otimes \mathscr{F}'(t)))$, where $(\Omega', \mathscr{F}', (\mathscr{F}'(t)))$ is a filtered space and $P'\colon \Omega \times \mathscr{F} \to \mathbf{R}^+$ is a probability transition function, is called a *decoupled tangent process* to X as long as the following two conditions are satisfied:

(i) $\tilde{X}$ has sample paths in $D(T)$ and is adapted to $\mathscr{F}(t) \otimes \mathscr{F}'(t)$, $t \in T$;

(ii) for P-a.a. $\omega \in \Omega$

$$\mathscr{L}(\tilde{X}(\omega, \cdot, t), t \in T) = M_X^\infty(\omega), \quad \text{where } M_X^\infty = P\text{-}\lim_{n \to \infty} M_X^n.$$

Once $X(t)$ admits a decoupled tangent process, a canonical way to construct it is as follows: choose $\Omega' = D(T)$, $\mathscr{F}(t) = \sigma(\omega'(s), s \leqslant t)$, $\mathscr{F}' = \mathscr{F}'(t_\infty)$, $P(\omega, \cdot) = M_X^\infty(\omega, \cdot)$, and set $\tilde{X}(\omega, \omega', t) = \omega'(t)$.

Remark 4.1. It is clear that, for P-a.a. $\omega \in \Omega$, $\tilde{X}(\omega, \cdot, t)$, $t \in T$, is a process defined on $(\Omega', \mathscr{F}', P'(\omega, \cdot))$, which has independent increments and sample paths in $D(T)$.

If $X(t)$, $t \in T$, is itself a process with independent increments, then it admits a decoupled tangent process $\tilde{X}$ which can be taken to be an independent copy of X. More precisely, $\tilde{X}$ can be defined by the formula

$$\tilde{X}(\omega, \omega') = X(\omega'), \quad (\omega, \omega') \in \Omega \times \Omega.$$

Strongly predictable processes X (i.e., processes such that there exists and $\varepsilon > 0$ such that, for all $t \in T$, $X(t)$ is $\mathscr{F}(t-\varepsilon)$-measurable) also admit a decoupled tangent process. In this case we can take $\tilde{X}(\omega, \omega', t) = X(\omega, t)$, $t \in T$, $(\omega, \omega') \in \Omega \times \Omega$.

A process X with sample paths in $D(T)$ is said to be *left quasi-continuous* if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that for any stopping times τ, σ, $\tau \leqslant \sigma$, such that $P(\sigma - \tau > \delta) < \delta$ we have

$$P\left(\sup_{\tau \leqslant s, t \leqslant \sigma} |X(s) - X(t)| > \varepsilon\right) < \varepsilon.$$

This definition, which is handy for our purposes, is equivalent to the usual one (cf., e.g., [9]).

The class of left quasi-continuous processes in $D(T)$ admitting a decoupled tangent process was characterized by Jacod [11]. In particular, he proved that any left quasi-continuous semimartingale admits a decoupled tangent process as well. This result is what we need in Section 6, and we discuss it below in some detail sketching also its strengthening with a proof based on inequalities of Section 2.

We begin by introducing what we call *Grigelionis characteristics* of the process X.

The *first characteristic* is a predictable process $B(t)$, $t \in T$, in $D(T)$ defined by the formula

$$B := P\text{-}\lim_{n\to\infty} B_n, \tag{4.3}$$

where the convergence is in the space $D(T)$, and

$$B_n(t) = \sum_{k:t_k^n \leqslant t} \mathrm{E}\big([d_k^n] \mid \mathscr{F}(t_{k-1}^n)\big), \qquad t \in T. \tag{4.4}$$

As in (4.1), $t_1^n, \ldots, t_{k_n}^n \in \pi^n$ and (d_k^n) is defined by (4.2).

The *second* and *third characteristics* are a predictable random measure μ supported on $T \times (\boldsymbol{R}\setminus\{0\})$ and a predictable process $C(t)$, $t \in T$, which are simultaneously defined by the condition that for each $t \in T$ and each $f \in \mathscr{R} := \{f\colon f$ is bounded and continuous on $\boldsymbol{R}$ and $\lim_{x\to 0} f(x)/x^2 = \frac{1}{2}f''(0)$ exists and is finite$\}$

$$\tfrac{1}{2}f''(0)C(t) + \int_{\boldsymbol{R}}\int_0^t f(x)\mu(ds, dx) = K(f)(t), \tag{4.5}$$

where $K(f)(t) = P\text{-}\lim_{n\to\infty} K_n(f)(t)$, the limit is in $D(T)$, and

$$K_n(f)(t) := \sum_{k:t_k^n \leqslant t} \mathrm{E}\Big(f\big(d_k^n - \mathrm{E}([d_k^n] \mid \mathscr{F}(t_{k-1}^n))\big) \mid \mathscr{F}(t_{k-1}^n)\Big). \tag{4.6}$$

Of course, the above characteristics need not exist in general, but the following proposition shows that once B exists, μ and C do exist as well. The proposition can be proved using the machinery of the theory of semimartingales but we propose here an elementary proof which can be found in the Appendix.

PROPOSITION 4.1. *If $X(t)$, $t \in T$, is a left quasi-continuous process and (π^n) is a nested normal sequence of partitions of T, and the characteristic B exists, then the characteristics μ and C are also well defined, the convergence in* (4.3), (4.5) *and* (4.6) *is uniform, rather than only in $D(T)$, and the processes B, C and $K(f)$, $f \in \mathscr{R}$, have sample paths in the space $C(T)$ of continuous functions.*

Remark 4.2. If $X(t)$, $t \in T$, is a process with independent increments, then B, μ and C are deterministic and are well known as the so-called *Lévy characteristics* of X. It is easy to see that X is stochastically continuous if and only if the functions B and $K(f)$, $f \in \mathscr{R}$, are continuous. In this case B, μ, and C are, of course, the quantities appearing in the Lévy–Hinchin formula for X (cf. Section 5).

The next theorem, which explains a connection between the existence of Grigelionis characteristics and the existence of decoupled tangent processes for processes in $D(T)$, extends a result of Jacod [11].

THEOREM 4.1. *A left quasi-continuous process X in $D(T)$ admits a decoupled tangent process if and only if the Grigelionis characteristic B exists for it.*

Proof. We shall restrict ourselves to the proof of the "if" part which is the only one needed in what follows. Besides, a proof of the "only if" part can be found in [11]. So, assume that B exists for X. To show that X admits a decoupled tangent process (Definition 4.1) we need to show the weak convergence of M_X^n.

For each fixed $\omega \in \Omega$ and for the process with independent increments $\tilde{X}^n(\omega, \cdot, \cdot)$, the first characteristic $B(\cdot)$ and the functional $K(f)(\cdot)$, $f \in \mathscr{R}$, coincide with $B_n(\omega, \cdot)$ and $K_n(f)(\omega, \cdot)$. If they converge uniformly to continuous functions $B(\omega, \cdot)$ and $K(f)(\omega, \cdot)$ (for f in a countable dense set in $\mathscr{R}$, then, by a criterion of weak convergence for processes with independent increments (cf. [10]), the processes $\tilde{X}^n(\omega, \cdot, \cdot)$ converge weakly to a process with independent increments with first characteristic $B(\omega, \cdot)$ and the functional $K(f)(\omega, \cdot)$, which uniquely determine the distribution of this process. This, and the standard subsequence argument, immediately conclude the proof. Q.E.D.

Remark 4.3. The above proof follows the original Jacod's [11] approach. However, the basic inequality of Theorem 2.1 creates also a possibility for the following, a little more direct, proof. It follows from Theorem 2.1 that for each two stopping times σ and τ on $(\Omega, \mathscr{F}, P; (\mathscr{F}(t)))$ with values in π^n we have, for each $\varepsilon, \delta > 0$,

$$P \otimes P_n'\big(\sup_{\sigma < t \leqslant \tau} |\tilde{X}^n(t) - \tilde{X}^n(\sigma)| > \varepsilon\big)$$

$$\leqslant \frac{5\delta}{\varepsilon} + 3P\big(\sup_{\sigma < t \leqslant \tau} |X(t) - X(\sigma)| > \delta\big) + 2P\Big(\sum_{k: \sigma < t_k^n \leqslant \tau}^{*} \mathrm{E}\big([d_k^n]^{2\delta} \,|\, \mathscr{F}(t_{k-1}^n)\big) > \delta\Big).$$

Now, the Aldous [1] method can be used to prove the tightness of processes $\tilde{X}^n$ and the convergence of characteristics permits identification of the limit.

In the next few paragraphs we discuss the situation when X is a semimartingale, i.e. a sum of an $(\mathscr{F}_t)$-local martingale and an $(\mathscr{F}_t)$-adapted process with sample paths with finite variation on finite intervals.

Since for any left quasi-continuous semimartingale the characteristic B exists (cf. Appendix), we obtain thus the following result of J. Jacod.

COROLLARY 4.1. *Any left quasi-continuous semimartingale admits a decoupled tangent process for each normal sequence of partitions (π^n).*

Remark 4.4. If $X(t)$, $t \in T$, is a left quasi-continuous semimartingale on $(\Omega, \mathscr{F}, \mathscr{F}(t), P)$, then the Grigelionis characteristics B, μ, and C are uniquely characterized by the following property: For each $u \in \boldsymbol{R}$ the process

$$(4.7)\qquad A(u, t) = iuB(t) - \frac{u^2 C(t)}{2} + \int_0^t \int_{\boldsymbol{R}} (e^{iux} - 1 - iu[x])\mu(ds, dx)$$

is a unique predictable process of bounded variation such that $A(u, 0) = 0$ and such that $\exp(iuX(t) - A(u, t))$, $t \in T$, is a local martingale (cf. [11]). Note that if X is not left quasi-continuous, then the characteristic B need not exist as defined (cf. [4]) and another definition of B is necessary.

Remark 4.5. In the case where a semimartingale X is left quasi--continuous the Grigelionis characteristics μ and C may be defined by the equation $(f \in \mathscr{R})$

$$\tfrac{1}{2} f''(0) C(t) + \int_{\boldsymbol{R}} \int_0^t f(x)\mu(ds, dx) = \lim_{n\to\infty} \sum_{k: t_k^n \leqslant t} \mathrm{E}\big(f(d_k^n) \mid F(t_{k-1}^n)\big)$$

and, moreover, the above limit is in $C(T)$ (cf. Appendix 7.O). This is the definition of μ and C that we will use in Sections 5 and 6.

Remark 4.6. In general, the existence of a tangent process, as well as the existence of the Grigelionis characteristic B, depends on the initial choice of the sequence of partitions (π^n). In fact, it is possible to construct a process X with continuous sample paths and a nested sequence of partitions (π^n) such that on two of its subsequences, say (π_1^n) and (π_2^n), X admits decoupled (π_1^n)- and (π_2^n)-tangent processes with different Grigelionis characteristics. (This solves a problem posed by Jacod [11].)

Indeed, let r_k^n, $n = 1, 2, \ldots$, $k = 1, 2, \ldots, 4^{n-1}$, be independent Rademacher random variables on $(\Omega, \mathscr{F}, P)$. Define

$$X_t^n = \frac{1}{n^2} \sum_{k=1}^{4^{n-1}} r_k^n \{(1 - |4^n t - 4k + 2|) \vee 0\} \qquad \text{for } t \in [0, 1] = T,$$

and then put

$$X_t = \sum_{n=1}^{\infty} X_t^n,$$

where the series is uniformly convergent in view of $\|X_t^n\|_\infty \leqslant 1/n^2$, so that (X_t) is in $C(T)$ because each (X_t^n) is in $C(T)$, and

$$\mathscr{F}_t = \bigcap_{s>t} \mathscr{G}_s = \sigma\big((X_u^n)_{u \leqslant s},\ n = 1, 2, \ldots\big).$$

If $\pi_1^n = (k/4^n)$, $k = 0, 1, 2, \ldots, 4^n$, is a sequence of partitions of $[0, 1]$, then

$$\mathrm{E}(\Delta X_{k/4^n} \mid \mathscr{F}_{(k-1)/4^n}) = \Delta X_{k/4^n}$$

so that

$$\lim_{n\to\infty}\sum_{k/4^n\leqslant t} \mathrm{E}(\Delta X_{k/4^n}\,|\,\mathscr{F}_{(k-1)/4^n}) = \lim_{n\to\infty} X_{[t4^n]/4^n} = X_t.$$

Hence B exists and $B = X$. Also

$$\sum_{k/4^n\leqslant t} \mathrm{E}\big((\Delta X_{k/4^n}-\mathrm{E}(\Delta X_{k/4^n}\,|\,\mathscr{F}_{(k-1)/4^n}))^2\,|\,\mathscr{F}_{k/4^{n-1}}\big) = 0,$$

so that $\mu \equiv 0$ and $C \equiv 0$.

On the other hand, if $\pi_2^n = (2k/4^n)$, $k = 0, 1, \ldots, 4^{n/2}$,

$$\mathrm{E}(\Delta X^m_{2k/4^n}\,|\,\mathscr{F}_{2(k-1)/4^n})$$
$$= \begin{cases} \Delta X^m_{2k/4^n} & \text{if } m \neq n \text{ or if } m = n \text{ and } k \text{ is odd},\\ 0 = \Delta X^n_{2k/4^n}-n^{-2}r^n_k & \text{if } m = n \text{ and } k \text{ is even},\end{cases}$$

$$\mathrm{E}(\Delta X_{2k/4^n}\,|\,\mathscr{F}_{2(k-1)/4^n}) = \begin{cases} \Delta X_{2k/4^n} & \text{for odd } k,\\ \Delta X_{2k/4^n}-n^{-2}r^n_k & \text{for even } k,\end{cases}$$

and

$$B^n(t) = \sum_{2k/4^n\leqslant t} \mathrm{E}(\Delta X_{2k/4^n}\,|\,\mathscr{F}_{2(k-1)/4^n}) = X_t - n^{-2}\sum_{2k/4^n\leqslant t} r^n_k.$$

Clearly, $B^n(t)$ does not converge in any sense as $n\to\infty$.

The above construction can be easily fine-tuned to yield the desired counterexample.

5. Stochastic integrals: deterministic integrands, integrators with independent increments. In the present section, which is independent of the remainder of the paper, we collect basic facts on Wiener-type integrals of the form $\int f\,dX$, where f is a deterministic function and X is a process with independent increments. As we shall see in Section 6 understanding of such integrals is critical for explaining the structure of general predictable processes which are integrable with respect to semimartingales. A number of authors studied these integrals in a similar spirit under various restrictions on X (cf. Urbanik and Woyczyński [25] for symmetric, stationary and stochastically continuous X, Urbanik [24] for X symmetric, stochastically continuous but not necessarily stationary, and Rosiński [21] for stationary, stochastically continuous, but not necessarily symmetric X). Since what we need in Section 6 are results in the general case, with no assumptions on stationarity or symmetry of X, for the benefit of the reader, we provide these results below (even without assumption of stochastic continuity) with concise proofs.

Let $X(t)$, $t\in T$, be a process with independent increments with sample paths in the Skorohod space $D(T)$, and let, as before, $\pi^n = (t^n_k)$ be a normal nested sequence of partitions of T, i.e., $\max_k|t^n_k - t^n_{k-1}|\to 0$ as $n\to\infty$ and $\pi^n\subset\pi^{n+1}$, $n = 1, 2, \ldots$ We will assume that $\bigcup_n \pi^n$ contains all the points of stochastic discontinuity of the process X. The Lévy characteristics B, μ and C (compare with Section 4, Remark 4.5) of X are defined as follows:

$$B(t) = \lim B_n = \lim_{n\to\infty} \sum_{k:t_k^n \leq t} \mathrm{E}[d_k^n], \qquad t \in T, \tag{5.1}$$

where $d_k^n = X(t_k^n) - X(t_{k-1}^n)$. $B(t)$ is a well-defined function in $D(T)$ and the convergence is uniform (see Appendix 7.P). From now on, in this section, we will assume that $B(t)$ is a function of bounded variation on T ($B \in \mathrm{BV}(T)$). It was noticed by Prekopa [20] that the condition $B \in \mathrm{BV}(T)$ is a necessary and sufficient condition for the stochastic integral with respect to X to exist (under any reasonable definition of the integral). This fact also follows easily from the methods developed below. If $B \in \mathrm{BV}(T)$, then (see Appendix 7.P)

$$\operatorname{Var} B(t) = \lim_{n\to\infty} \sum_{k:t_k^n \leq t} |\mathrm{E}[d_k^n]|, \qquad t \in T. \tag{5.2}$$

In the case of $B \in \mathrm{BV}(T)$, the condition

$$\tfrac{1}{2} f''(0) C(t) + \int_{\boldsymbol{R}} \int_0^t f(x) \mu(ds, dx) = \lim_{n\to\infty} \sum_{k:t_k^n \leq t} \mathrm{E} f(d_k^n) \tag{5.3}$$

satisfied for each $f \in \mathscr{R} := \{f\colon f$ is bounded and continuous on $\boldsymbol{R}$ and $\lim_{x\to 0} f(x)/x^2 = \frac{1}{2} f''(0)$ exists and is finite$\}$ uniquely determines a positive measure μ supported on $T \times (\boldsymbol{R}\setminus\{0\})$ and a function $C\colon T \to \boldsymbol{R}^+$. The measure μ fulfills the condition

$$\int_{\boldsymbol{R}} \int_T (1 \wedge x^2) \mu(ds, dx) < \infty, \tag{5.4}$$

and the function C is continuous and non-decreasing.

Recall that if X is stochastically continuous, then the Lévy characteristics of X introduced above determine its characteristic functions and

$$\mathrm{E} e^{iuX(t)} = \exp\left\{ iB(t)u - \frac{C(t)}{2} u^2 + \int_{\boldsymbol{R}} \int_0^t (e^{ixu} - 1 - iu[x]) \mu(ds, dx) \right\}.$$

Let ν be a measure on T defined by the formula

$$\nu(ds) = |dB(s)| + dC(s) + \int_{\boldsymbol{R}} (1 \wedge x^2) \mu(ds, dx). \tag{5.5}$$

If f is a step function, i.e. a linear combination of functions $I_{(s,t]}$, we define $\int_T f\, dX$ in an obvious way, and in this case define

$$\varrho_X^{\det}(f) := \sup\left\{ \left\| \int_T v f\, dX \right\|_0 \colon v \text{ is a step function}, |v| \leq 1 \right\}. \tag{5.6}$$

Here and thereafter $\|\xi\|_0 \overset{\mathrm{df}}{=} \mathrm{E}|[\xi]| = \mathrm{E}|\xi| \wedge 1$.

DEFINITION 5.1. A deterministic function $f\colon T \to \boldsymbol{R}$ is said to be *X-integrable* (in short, $f \in L^{\det}(dX)$) if there exists a sequence (f_n) of step functions such that

(i) $f_n \to f$ ν-a.e.;
(ii) $\varrho_X^{\det}(f_n - f_m) \to 0$ as $n, m \to \infty$.
If $f \in L^{\det}(dX)$, then, by definition, $\int_T f\,dX = P\text{-}\lim_n \int_T f_n\,dX$.

Remark 5.1. It follows from Theorem 5.1 that $\int f\,dX$ defined in such a way is independent of the choice of (f_n).

Our main goal in this section is to characterize functions in $L^{\det}(dX)$ in terms of the three Lévy characteristics of X, and in order to accomplish this we need to introduce the following notation. Let, by the Randon–Nikodým Theorem,

$$dB(s) = b(s)\nu(ds), \qquad dC(s) = c(s)\nu(ds), \tag{5.7}$$

and let

$$\mu(ds, dx) = \hat{\nu}(s, dx)\nu(ds) \tag{5.8}$$

represent the desintegration of μ. Furthermore, let

$$\begin{aligned} k(s, x) &= \int_{\boldsymbol{R}} \{1 \wedge (xu)^2\}\,\hat{\nu}(s, du) + c(s)x^2, \\ l(s, x) &= \int_{\boldsymbol{R}} ([xu] - x[u])\,\hat{\nu}(s, du) + b(s)x, \\ \mathrm{l}(s, x) &= \sup_{|y| \leqslant |x|} l(s, y) \end{aligned} \tag{5.9}$$

and, finally, let

$$\varphi(s, x) = k(s, x) + \mathrm{l}(s, x). \tag{5.10}$$

The function φ is measurable in s, continuous and symmetric in x, and increasing in x for $x \geqslant 0$. Moreover, for each s, φ fulfills the condition (Δ_2):

$$\varphi(s, 2x) \leqslant 5\varphi(s, x), \qquad x \in \boldsymbol{R}^+. \tag{5.11}$$

Indeed, we have

$$l(s, 2x) = 2l(s, x) + \int_{\boldsymbol{R}} ([2xu] - 2[xu])\,\hat{\nu}(s, dx) \leqslant 2l(s, x) + k(s, x),$$

because $|[2xu] - 2[xu]| \leqslant 1 \wedge (xu)^2$ for all x, u. Hence $\mathrm{l}(s, 2x) \leqslant 2\mathrm{l}(s, x) + k(s, x)$, and, since $k(s, 2x) \leqslant 4k(s, x)$, we get (5.11). Under these circumstances, the so-called Musielak–Orlicz space

$$L^{\varphi}(d\nu) = \Big\{f\colon\ \Phi(f) := \int_T \varphi(s, f(s))\,\nu(ds) < \infty\Big\} \tag{5.12}$$

is a linear space. The modular Φ induces on $L^{\varphi}(d\nu)$ a topology of a complete linear metric space in which step functions are dense (cf., e.g., [19]).

The following Theorem 5.1 gives a complete characterization of functions which are X-integrable, and we precede it with recall of a version of the three series theorem and the Lévy–Octaviani inequality.

PROPOSITION 5.1. *Let ξ_i, $i = 1, \ldots, n$, be a sequence of independent random variables. Then:*

(i) *If* $\|\sum^* \xi_i\|_0$ *is small, then* $\sum \mathrm{E}[\xi_i]$, $\sum\{\mathrm{E}([\xi_i])^2-(\mathrm{E}[\xi_i])^2\}$ *and* $\sum P(|\xi_i| > 1)$ *are small.*

(ii) *If* $\sum \mathrm{E}[\xi_i]$ *and* $\sum \mathrm{E}([\xi_i])^2$ *are small, then* $\|\sum \xi_i\|_0$ *is small.*

(iii) $P(\sum^* \xi_i > \varepsilon) \leqslant 3\sup_{1\leqslant j\leqslant n} P(|\sum_{i=1}^{j} \xi_i| > \varepsilon/3)$ *for each* $\varepsilon > 0$.

Here and thereafter the generic implication "if A is small, then B is small" means that for each $\varepsilon > 0$ there exists a $\delta > 0$, depending only on ε, such that $A < \delta$ implies $B < \varepsilon$.

THEOREM 5.1. $f \in L^{\det}(dX)$ *if and only if* $f \in L^{\varphi}(d\nu)$. *Moreover,* $\varrho_X^{\det}(f)$ *is small if and only if* $\Phi(f)$ *is small.*

It is clear that to prove Theorem 5.1 it suffices to demonstrate that for a step function f both $\varrho_X^{\det}(f)$ and $\Phi(f)$ are simultaneously small, and to accomplish this we need the following

LEMMA 5.1. *Let* $\Lambda(f) := \int_T \mathrm{l}(s, f(s))\nu(ds)$ *and* $\tilde{\Lambda}(f) := \int_T l(s, f(s))\nu(ds)$. *Then, for any step function* f,

$$\Lambda(f) = \sup\{\tilde{\Lambda}(vf)\colon\ v \text{ is a step function}, |v| \leqslant 1\}. \tag{5.13}$$

Proof. Let us define

$$v_X(s, x) = v(s, x) := \min\{r\colon\ |r| \leqslant 1, \mathrm{l}(s, x) = l(s, rx)\}. \tag{5.14}$$

Then $v(s, x)$ is measurable in (s, x) because, for each α, $-1 \leqslant \alpha \leqslant 1$,

$$\{v(s, x) \leqslant \alpha\} = \{\mathrm{l}_\alpha(s, x) = \mathrm{l}(s, x)\},$$

where $\mathrm{l}_\alpha(s, x) = \max_{-1\leqslant r\leqslant\alpha} l(s, rx)$ is measurable in s and continuous in x. So $v(f)(s) = v(s, f(s))$ is also measurable in s and $|v(s, f(s))| \leqslant 1$ for all $s \in T$. Now, by the definition of v, we have $\Lambda(f) = \tilde{\Lambda}(v(f)f)$, and an approximation of $v(f)$ by step functions gives (5.13). Q.E.D.

Proof of Theorem 5.1. Let f be a step function. Without loss of generality, we may assume that all jumps of f are contained in π^n for some n. Then, for each $m > n$ we have

$$\sum_i \mathrm{E}[f(t_i^m)d_i^m] = \sum_k \sum_{t_{k-1}^n < t_i^m \leqslant t_k^n} \mathrm{E}[f(t_i^m)d_i^m]. \tag{5.15}$$

The function $[ux]-u[x]$ belongs to $\mathscr{R}$ (see (5.3)) for each u. So, for each fixed k, taking $u = f(t_k^m)$ (wich happens to be equal to $f(t_k^n)$ as long as $t_{k-1}^n < t_i^m \leqslant t_k^n$), by (5.1), (5.2), and by the definition of l (see (5.9)), we obtain

$$\lim_{m\to\infty} \sum_{t_{k-1}^n < t_i^m \leqslant t_k^n} \mathrm{E}[f(t_i^m)d_i^m]$$
$$= \lim_{m\to\infty} \sum_{t_{k-1}^n < t_i^m \leqslant t_k^n} \big(\mathrm{E}([f(t_k^n)d_i^m]-f(t_k^n)[d_i^m]) + \mathrm{E}f(t_k^n)[d_i^m]\big)$$

$$= \int_{R} \int_{t_{k-1}^n < s \leq t_k^n} ([f(t_k^n)x] - f(t_k^n)[x])\mu(ds, dx)$$
$$+ \int_{t_{k-1}^n < s \leq t_k^n} f(t_k^n)dB(s) = \int_{R} \int_{(t_{k-1}^n, t_k^n]} l(s, f(s))\nu(ds).$$

Hence we conclude that

$$\lim_{m\to\infty} \sum_i \mathrm{E}[f(t_i^m)d_i^m] = \tilde{\Lambda}(f). \tag{5.16}$$

Similarly, by (5.2), for each $k = 1, 2, \ldots, k_n$ we have

$$\lim_{m\to\infty} \sum_{t_{k-1}^n < t_i^m \leq t_k^n} \mathrm{E}([f(t_i^m)d_i^m])^2 = \int_{(t_{k-1}^n, t_k^n]} ([f(t_k^n)x])^2 \mu(ds, dx)$$
$$= \int_{(t_{k-1}^n, t_k^n]} k(s, f(s))\nu(ds),$$

so that

$$\lim_{m\to\infty} \sum_i \mathrm{E}([f(t_i^m)d_i^m])^2 = K(f) := \int_T k(s, f(s))\nu(ds). \tag{5.17}$$

Let us also observe that, by the Cauchy-Schwartz inequality, for each m

$$\sum_i \mathrm{E}([f(t_i^m)d_i^m])^2 \leq \sum_i \{(\mathrm{E}([f(t_i^m)d_i^m])^2 - (\mathrm{E}[f(t_i^m)d_i^m]))^2\} + (\sum_i \mathrm{E}[f(t_i^m)r_i^m d_i^m])^2, \tag{5.18}$$

where $r_i^m = \operatorname{sgn} \mathrm{E}[f(t_i^m)d_i^m]$.

Now, the rest of the proof follows quickly from Proposition 5.1. Indeed, if $\varrho_X^{\det}(f)$ is small, then for each step function v, $|v| \leq 1$, with jumps in π^r, $r \geq n$, it follows that for $m > r$

$$\|\int_T vf\,dX\|_0 = \|\sum_i v(t_i^m) f(t_i^m)d_i^m\|_0$$

is small (uniformly in v). Hence, by Proposition 5.1 (iii), $\|\sum^* v(t_i^m) f(t_i^m)d_i^m\|_0$ is small uniformly in v. Therefore, by Proposition 5.1 and (5.16) we see that $\tilde{\Lambda}(vf)$ is small, and so is $\Lambda(f)$ by Lemma 5.1. Inserting $v(t_i^m) = r_i^m$, as in (5.18), and interpolating it to a step function $v(t)$, we infer from (5.18) and from Proposition 5.1 (i) that $\sum_i \mathrm{E}([f(t_i^m)d_i^m])^2$ is small. Hence, by (5.17), $K(f)$ is small, and we conclude that $\Phi(f) = \Lambda(f) + K(f)$ is small. This completes the proof of the first implication.

The converse implication is even simpler to demonstrate. Indeed, if $\Phi(f)$ is small, then both $\Lambda(f)$ and $K(f)$ are small, and so is $\tilde{\Lambda}(f)$. Therefore, by (5.16), (5.17) and Proposition 5.1 (ii), we see that $\|\int f\,dX\|_0$ is small. The same argument applies to $v \cdot f$, which implies that $\varrho_X^{\det}(f)$ is small. Q.E.D.

The above discussion also justifies introduction of the stochastic integral as a process $(\int_0^t f\,dX, t \in T)$ with sample paths in $D(T)$ as follows. If f is a step function, then the definition of the process is obvious. If $f \in L_X^{\det}$, and f_n is a sequence of simple functions as in Definition 5.1, then the defining condition $\varrho_X(f_n - f_m) \to 0$ and Proposition 5.1 (iii) imply that

$$\sup_{t \in T} \Big| \int_0^t (f_n - f_m) dX \Big| \to 0,$$

which gives the uniform convergence of processes $\left(\int_0^t f_n dX,\ t \in T\right)$, which in the limit gives the desired process $\left(\int_0^t f dX,\ t \in T\right)$. Therefore, for $f \in L_X^{\det}$ we can write

$$\int_T^* f dX = \sup_{t \in T} \Big| \int_0^t f dX \Big|.$$

Remark 5.2. If $\varrho_X(f)$ is small, then also $\|\int_T^* f dX\|_0$ is small. This follows immediately from the above definitions and Proposition 5.1 (iii).

In the next section we will need the following

PROPOSITION 5.2. *Let f be a simple function. Then $\varrho_X(f)$ is small if and only if both $\|\int_T^* f dX\|_0$ and $\|\int^* v_X(f) f dX\|_0$ are small, where v_X is defined by* (5.14).

Proof. To show the "if" part let us begin with an observation that if g is bounded and $\|\int^* g dX\|_0$ is small, then $\tilde{\Lambda}(g)$ is small. For a simple g this is an immediate consequence of (5.16) and Proposition 5.1 (i). For a bounded g this follows by standard approximation techniques using the Dominated Convergence Theorem. Applying the above to $g = v_X(f) \cdot f$ we see that $\Lambda(f)$ is small.

To prove that $K(f)$ is small let us observe that, for the process $f \circ X = \left(\int_0^t f dX,\ t \in T\right)$, the first characteristic is equal to

$$B_{f \circ X}(t) = \int_0^t l(s, f(s)) dv(s).$$

So

$$\operatorname{Var} B_{f \circ X} \leqslant \int_T L(s, f(s)) dv(s) \leqslant \Lambda(f).$$

By (5.2), it follows that, for large n, $\sum_i |E[f(t_i^n) d_i^n]|$ is small because $\Lambda(f)$ was proved to be small so that we obtain the smallness of $K(f)$ by (5.18).

The proof of the "only if" part is obvious. Q.E.D.

6. Stochastic integral: predictable integrands, semimartingale integrators. Let $X(t)$, $t \in T$, be a left quasi-continuous semimartingale on $(\Omega, \mathcal{F}, P; \mathcal{F}(t))$ with three predictable Grigelionis characteristics $B(t)$, $C(t)$, $t \in T$, and $\mu(ds, dx)$ (cf. (4.4)–(4.6)). Let us define a measure $v \circ P$ on $\Omega \times T$ by the formula

$$v \otimes P(d\omega, ds) = v(\omega, ds) P(d\omega),$$

where

$$v(\omega, ds) = |dB(s)| + dC(s) + \int_R (1 \wedge x^2) \mu(ds, dx)$$

is a random version of measure ν introduced for processes with independent increments at the beginning of Section 5.

A process $F(t)$, $t \in T$, is said to be a *predictable step process* if it is a finite linear combination of processes of the form $\xi I_{(s_1,s_2]}(t)$, $t \in T$, where ξ is $\mathscr{F}(s_1)$-measurable. For a predictable step process F the integral $\int FdX$ is defined in an obvious manner. A process $F(t)$, $t \in T$, is said to be *predictable* if it is measurable as a function on $\Omega \times T$ with respect to the σ-field generated on $\Omega \times T$ by all predictable step processes.

DEFINITION 6.1. A predictable process F is said to be *X-integrable* (in short, $F \in L(dX)$) if there exists a sequence of predictable step processes (F_n) such that

(i) $F_n \to F$ $\nu \otimes P$-a.e. as $n \to \infty$,

(ii) $\varrho_X(F_n - F_m) \to 0$ as $n, m \to \infty$, where

$$\varrho_X(G) := \sup \| \int_T VGdX \|_0,$$

and where the supremum is taken over all predictable step processes V such that $|V| \leqslant 1$. If $F \in L(dX)$, then

$$\int_T FdX := P\text{-}\lim_{n \to \infty} \int_T F_n dX.$$

Remark 6.1. With some effort, one can show that the above definition coincides with the standard definition of a stochastic integral (cf., e.g., [22]) and that the above definition of the integral is correct, i.e., that $\lim_n \int_T F_n dX$ does not depend on the choice of F_n. The correctness also follows from Theorem 6.1.

Remark 6.2. It is easy to see that, for each predictable step process F, if $\varrho_X(F)$ is small, then not only $\| \int_T FdX \|_0$ is small, but also $\| \int_T^* FdX \|_0$ is small (cf. Lemma 2.3), where

$$\int_T^* FdX := \sup_{t \in T} \left| \int_T I_{(0,t]}(s) F(s) dX(s) \right|.$$

Therefore, if F is a predictable process and F_n is a sequence of predictable step processes defining $\int FdX$, then the sequence of processes

$$Y_n(t) := \int_T I_{(0,t]}(s) F_n(s) dX(s), \qquad t \in T,$$

converges uniformly in $t \in T$, in probability, to the process

$$Y(t) := \int_T I_{(0,t]}(s) F(s) dX(s), \qquad t \in T.$$

So, as a result, the stochastic integral can also be viewed as a stochastic process with sample paths in $D(T)$, and the first statement of this remark extends to all $F \in L(dX)$.

To formulate our main result which provides a characterization of predictable X-integrable processes in terms of Grigelionis characteristics B,

C and μ of a semimartingale X, we need to adjust (randomize) the quantities introduced in Section 5. Since $B(t)$, $C(t)$ and $\mu(ds, dx)$ are all predictable, there exist predictable processes $b(s)$ and $c(s)$, and a predictable random measure $\hat{v}(\omega, s, dx)$ such that

$$dB(s) = b(s)v(ds), \qquad dC(s) = c(s)\,v(ds)$$

and

$$\mu(\omega, ds, dx) = \hat{v}(\omega, s, dx)v(\omega, ds).$$

Therefore, in the semimartingale case, the Musielak–Orlicz function φ defined by formulas (5.10) is random ($\varphi = \varphi(\omega, s, x)$) because all of the functions k, l, l are now random. This permits us to introduce the space $\mathscr{L}_\varphi(dv)$ which, by definition, is the space of all predictable processes F for which

$$\Phi(F)(\omega) := \int \varphi(\omega, s, F(\omega, s))v(\omega, ds) < \infty \quad P\text{-a.s.}$$

The modular $\|\Phi(F)\|_0$ equips $\mathscr{L}_\varphi(dv)$ with a topology of a complete linear metric space and the predictable step processes are dense in L^φ (cf., e.g., Hudson's [7] explicit work on a particular case; the general case can be verified in a similar fashion).

Another, equivalent, way to introduce the random functions φ, k, l, l is via the decoupled tangent process $\tilde{X}(t)$. Recall (Section 4) that a decoupled tangent (to $X(t)$) process $\tilde{X}(t)$ (obtained with the use of a partition π^n which, for convenience, is assumed to be nested) is defined on the probability space $(\Omega \times \Omega', \mathscr{F} \otimes \mathscr{F}', P \otimes P')$ with filtration $\mathscr{F}(t) \otimes \mathscr{F}'(t)$ and enjoys the following property: for each fixed $\omega \in \Omega$, the process $\tilde{X}(\omega) = (\tilde{X}(t, \omega, \cdot), t \in T)$ defined on $(\Omega', \mathscr{F}', P'(\omega))$ is a process with independent increments. For each $\omega \in \Omega$, the functions $\varphi(\omega)$, $k(\omega)$, $l(\omega)$, $\mathrm{l}(\omega)$ introduced above are exactly the functions corresponding to the process $\tilde{X}(\omega)$ in Section 5 (cf. (5.9) and (5.10)).

Let us also note that, for a predictable process F, $F \in \mathscr{L}_\varphi(dv)$ if and only if, for P-almost all $\omega \in \Omega$, $F(\omega) \in L_{\varphi(\omega)}(v(\omega, ds))$.

THEOREM 6.1. *A predictable process F is X-integrable ($F \in L(dX)$) if and only if $F \in \mathscr{L}_\varphi(dv)$.*

Moreover, for an $F \in L(dX)$, $\varrho_X(F)$ is small if and only if $\|\Phi(F)\|_0$ is small.

Alternatively, F is X-integrable if and only if, for P-a.a. $\omega \in \Omega$, the deterministic function $F(\omega, \cdot)$ is integrable (in the sense of Definition 5.1) with respect to the process (with independent increments) $\tilde{X}(\omega)$.

Proof. It suffices to check that, for a predictable step function F, $\varrho_X(F)$ is small if and only if $\Phi(F)(\omega)$ is small with large probability, the last statement, by Theorem 5.1, being equivalent to saying that $\varrho^{\det}_{\tilde{X}(\omega)}(F(\omega, \cdot))$ is small with large probability $P(d\omega)$.

So, let F be a predictable step process such that (without loss of generality) all jumps of F are contained in all partitions π^n for n large enough.

Now, assume that $\varrho_X(F)$ is small or, more precisely, that

$$P(|\int_T VFdX| > \delta) < \delta$$

for each predictable step process V with $|V| \leqslant 1$. If the jumps of V are contained in π^n, then we can write that

$$\int_T VFdX = \sum_i V(t_i^n)F(t_i^n)d_i^n,$$

where $d_i^n = X(t_i^n) - X(t_{i-1}^n)$. Hence Corollary 2.3, applied to the σ-fields $\mathscr{F}(t_i^n)$ and the sequence $(F(t_i^n)d_i^n)$, implies that, with large probability $P \otimes P'_n$,

$$\sum^* V(t_i^n)F(t_i^n)\tilde{d}_i^n = \int_T^* VFd\tilde{X}^n$$

is small ($\tilde{d}_i^n$ and $\tilde{X}^n$ are as defined in Section 4), i.e. for a given δ' if δ is sufficiently small, we have

$$P \otimes P'_n(|\int_T^* VFd\tilde{X}^n| > \delta') < \delta'.$$

Hence, by the Chebyshev inequality,

$$P\big(P'_n(|\int_T^* VFd\tilde{X}^n| > \delta') \leqslant \sqrt{\delta'}\big) > 1 - \sqrt{\delta'}.$$

So, for a fixed V with jumps contained in π^n for some n, we obtain

$$P\big(P'(|\int_T^* VFd\tilde{X}| > \delta') \leqslant \sqrt{\delta'}\big) \geqslant 1 - \sqrt{\delta'}$$

because the distributions $\mathscr{L}(\tilde{X}^n(\omega)) \to \mathscr{L}(\tilde{X}(\omega))$ in probability P (cf. Section 4). The integral inside the above inequality is understood in the sense of Section 5 for each ω. The above inequality extends to all $\mathscr{F}$-predictable processes V bounded by 1 (by the Monotone Class Theorem).

Finally, take $V(\omega, s, x) = v_{\tilde{X}(\omega)}(\omega, s, x)$, where v is as in (5.14) but defined using Grigelionis characteristics. The $V(\omega, s, F(s))$ is predictable and bounded by 1 (by the argument similar to the one following (5.5) which was used to prove the measurability of v). So by Proposition 5.2 we see that, for a given ε, if δ' is small enough, then

$$P\big(\varrho_{\tilde{X}(\omega)}^{\text{det}}(F(\omega)) > \varepsilon\big) < \varepsilon.$$

This concludes the proof of the first implication.

Conversely, let F be a predictable step process with jumps contained in some π^m and such that

$$P\big(\varrho_{\tilde{X}(\omega)}^{\text{det}}(F(\omega)) > \varepsilon\big) < \varepsilon.$$

By Remark 5.2, for a given $\delta' > 0$, if ε is small enough, we have

$$P\big(P'(\int_T^* VF(\omega)d\tilde{X}(\omega) > \delta') < \delta'\big) > 1 - \varepsilon$$

for all $\mathscr{F}$-predictable step processes V bounded by 1. Now, if V is a fixed $\mathscr{F}$-predictable step process with jumps contained in π^m, then for n large enough

$$P\big(P'_n(\int_T^* VFd\tilde{X}^n(\omega) > \delta') < \delta'\big) > 1-\varepsilon,$$

because $\mathscr{L}(\tilde{X}^n(\omega)) \to \mathscr{L}(\tilde{X}(\omega))$ in P, and hence we get

$$P\otimes P'_n(\int_T^* VFd\tilde{X}^n > \delta') < \varepsilon+(1-\varepsilon)\delta'.$$

Since

$$\int_T^* VFd\tilde{X}^n = \sum^* V(t_i^n)F(t_i^n)\tilde{d}_i^n,$$

by Theorem 2.2, we see that, for a given $\delta > 0$, if δ' and ε are small enough, then

$$P(\sum^* V(t_i^n)F(t_i^n)d_i^n > \delta) < \delta,$$

so that, finally,

$$P(|\int VFdX| > \delta) < \delta$$

for any $\mathscr{F}$-predictable step process V, which, in view of the arbitrariness of V, gives $\varrho_X(F) < \delta$. Q.E.D.

As a consequence of Theorem 6.1 we also obtain the following

COROLLARY 6.1. *A predictable process $F(t)$, $t\in T$, is X-integrable if and only if, for P-almost every $\omega\in\Omega$, the deterministic function $F(\omega, t)$, $t\in T$ is X-integrable.*

Remark 6.3. After a preprint of the original version of this paper has been distributed Jacod and Sadi [12] extended the concept of tangent processes to processes which include the class of all semimartingales. This extension was possible due to the introduction of random predictable partition (π^n). Namely, for a given semimartingale X one can find a nested sequence of random predictable partitions (i.e., t_k^n are predictable stopping times) such that $\max_k|t_k^n-t_{k-1}^n| \to 0$ in P as $n\to\infty$, and such that $\bigcup_n \pi^n$ contains all predictable jump times of X. For such a sequence π^n and each sequence of stopping times λ_n, we have

$$\sup_{\lambda_n \leqslant t < \sigma_n(\lambda_n)} |X(\lambda_n)-X(t)| \to 0$$

in P as $n\to\infty$, where $\sigma_n(t) = \min\{t_k^n\in\pi^n\colon\ t_k^n \geqslant t\}$, and this property can be used, instead of left quasicontinuity, and after some modifications, in the proof of the existence of characteristic B and other characteristics as uniform limits (see Section 4 and Appendix). As a result the main theorem of Section 6 can be extended to the case of general semimartingale integrators.

7. Appendix. This appendix contains proofs of facts we referred to in Sections 4 and 5. We felt that these facts were too technical to include in the main body of the paper and would distort what we judged to be the proper balance of exposition of the principal resuls. Also, the nature of these facts is such that, despite the tediousness of their proofs, they are not surprising and could be anticipated by an expert. Finally, once again we would like to call the reader's attention to the fact that Jacod and Sadi [12], in their recent paper, obtained results from which many results given below can be deduced. However, our proofs are more direct and elementary, and make this paper self-contained.

The standing assumption in what follows is that the process X is left quasi-continuous.

Subsections 7.A through 7.M contain a complete proof of Proposition 4.1, subsection 7.N completes the proof of Corollary 4.1, subsection 7.O proves Remark 4.5, and, finally, subsection 7.P contains proofs omitted in Section 5.

7.A. *If* $\varphi\colon \boldsymbol{R}\to\boldsymbol{R}$ *is a bounded continuous function such that* $\varphi(0)=0$, *then*

$$P\text{-}\lim_{n\to\infty}\max_{1\leqslant k\leqslant k_n}\left|\mathrm{E}\big(\varphi(d_k^n)\,|\,\mathscr{F}(t_{k-1}^n)\big)\right| = 0.$$

Indeed, let

$$\lambda_n = \min\{k\colon 1\leqslant k\leqslant k_n,\ \left|\mathrm{E}\big(\varphi(d_k^n)\,|\,\mathscr{F}(t_{k-1}^n)\big)\right| > a\},$$

and $\lambda_n = k_n$ if the set is empty. Then

$$\begin{aligned}
&P\big(\max_k\left|\mathrm{E}\big(\varphi(d_k^n)\,|\,\mathscr{F}(t_{k-1}^n)\big)\right| > a\big)\\
&\quad\leqslant a^{-1}\mathrm{E}\sum_k I(\lambda_n=k)\left|\mathrm{E}\big(\varphi(d_k^n)\,|\,\mathscr{F}(t_{k-1}^n)\big)\right| \leqslant a^{-1}\mathrm{E}\sum_k I(\lambda_n=k)|\varphi(d_k^n)|\\
&\quad= a^{-1}\mathrm{E}\left|\varphi\big(X(\tau_n)-X(\tau_n')\big)\right| \leqslant a^{-1}\big(cP(|X(\tau_n)-X(\tau_n')|>\delta)+\varepsilon\big),
\end{aligned}$$

where $\tau_n = t_{\lambda_n-1}^n$, $\tau_n' = t_{\lambda_n}^n$, $c=\sup_{x\in\boldsymbol{R}}|\varphi(x)|$, and ε, δ are such that $|x|\leqslant\delta$ implies $|\varphi(x)|\leqslant\varepsilon$. Now, since τ_n and τ_n' are stopping times with $|\tau_n-\tau_n'|$ less than the mesh of partition π^n, the left quasi-continuity gives statement 7.A.

7.B. By 7.A we have

$$P\text{-}\lim_{n\to\infty}\sup_{t\in T}|B_n(t)-B_n(t-)| = 0.$$

Thus, *if the sequence B_n converges in $D(T)$, then it converges uniformly in P and the limit B is continuous.*

7.C. *If the characteristic B exists for X, then for each $f\in\mathscr{R}$ the sequence of processes $K_n(f)$ is uniformly convergent in P to a continuous process $K(f)$.*

The proof of this fact will take several paragraphs, and the first step in it consists in the observation that

7.D. *The assertion in* 7.C *holds true if and only if $K_n(f)$ is uniformly convergent in P to a continuous process for $f(x)=h(x)=:[x^2]$ and for each $f\in\mathscr{R}_0:=\{f_0\in\mathscr{R}\colon f_0(x)=0$ for $|x|<\delta$ for some $\delta>0\}$.*

This reduction follows immediately from the fact that for each $\varepsilon>0$ and $f\in\mathscr{R}$ there exists an $f_0\in\mathscr{R}_0$ such that $(\tfrac{1}{2}f''(0)-\varepsilon)h+f_0\leqslant f\leqslant(\tfrac{1}{2}f''(0)+\varepsilon)h+f_0$.

7.E. *Let*

$$K_n^0(f)(t) = \sum_{t_k^n \leqslant t} \mathrm{E}(f(d_k^n) \mid \mathscr{F}(t_{k-1}^n)).$$

If the characteristic B exists for X, then

$$P\text{-}\lim_{n\to\infty} \sup_{t\in T} |K_n^0(f)(t) - K_n(f)(t)| = 0 \quad \textit{for each } f \in \mathscr{R}_0.$$

Indeed, by Lemma 2.2,

$$P(\sup_{t\in T} |K_n^0(f)(t) - K_n(f)(t)| > b)$$

$$\leqslant \frac{a}{b} + 2\frac{b+2c}{b} P\big(\sum |f(d_k^n - \mathrm{E}([d_k^n] \mid \mathscr{F}(t_{k-1}^n))) - f(d_k^n)| > a\big),$$

where $c = \sup_{x\in R} |f(x)|$. Also, by 7.A,

$$P\text{-}\lim_{n\to\infty} \sup_k |\mathrm{E}([d_k^n] \mid \mathscr{F}(t_{k-1}^n))| = 0.$$

Thus, since

$$\lim_{n\to\infty} \sum_k |f(d_k^n - \varepsilon_k^n) - f(d_k^n)| = 0$$

for each (ε_k^n) such that

$$\lim_{n\to\infty} \sup_k |\varepsilon_k^n| = 0,$$

and for each ω such that $X(\omega, \cdot) \in D(T)$, we immediately get 7.E.

7.F. *Suppose that* $f \in \mathscr{R}_0$ *and that X has the property*

$$\sum_k |f(d_k^n)| \leqslant M \tag{7.1}$$

for some $M > 0$ *and all* $n \in N$. *Then* $K_n^0(f)$ *converges uniformly in P.*

Without loss of generality, we can also assume that $f \geqslant 0$. For $m > n$, let us introduce processes

$$K_{n,m}^0(f)(t) := \sum_{t_k^n \leqslant t} \mathrm{E}\Big(\sum_{t_{k-1}^n < t_i^m \leqslant t_k^n} f(d_i^m) \mid \mathscr{F}(t_{k-1}^n)\Big). \tag{7.2}$$

Since $K_m^0(f)$ is a non-decreasing process in t, and $K_n^0(f)$ has only jumps in π^n, we have

$$P(\sup_{t\in T} |K_m^0(f)(t) - K_n^0(f)(t)| > b)$$

$$\leqslant P(\sup_{t\in\pi^n} |K_m^0(f)(t) - K_n^0(f)(t)| > b/2) + P(\sup_k \mathrm{E}(f(d_k^n) \mid \mathscr{F}(t_{k-1}^n)) > b/2).$$

By 7.A, the second term converges to 0 as $n \to \infty$. The first term is estimated from above by

$$(7.3)\qquad P\big(\sup_{t\in\pi^n}|K_m^0(f)(t)-K_{n,m}^0(f)(t)| > b/4\big) + P\big(\sup_{t\in\pi^n}|K_n^0(f)(t)-K_{n,m}^0(f)(t)| > b/4\big).$$

By Lemma 2.2, the second term in (7.3) is estimated from above by

$$(7.4)\qquad \frac{4a}{b}+\frac{2b+8M}{b}P\Big(\sum_k\Big|\sum_{t_{k-1}^n<t_i^m\leqslant t_k^n} f(d_i^m)-f(d_k^n)\Big| > a\Big).$$

Hence, since for each ω for which the sample path of X is in $D(T)$

$$\sum_k\Big|\sum_{t_{k-1}^n<t_i^m\leqslant t_k^n} f(d_i^m)-f(d_k^n)\Big| \to 0,$$

the second term in (7.3) converges to 0.

Now, since $\{K_m^0(f)(t_k^n)-K_{m,n}^0(f)(t_k^n)\}$ is an $\mathscr{F}(t_k^n)$-martingale, we get, by the maximal inequality, that the first term in (7.3) is estimated from above by

$$\frac{16}{b^2}\mathrm{E}|K_m^0(f)(t_\infty)-K_{n,m}^0(f)(t_\infty)|^2$$

$$=\frac{16}{b^2}\sum_{t_k^n\in\pi^n}\Big\{\mathrm{E}\Big(\sum_{t_{k-1}^n<t_i^m\leqslant t_k^n}\mathrm{E}(f(d_i^m)\,|\,\mathscr{F}(t_{i-1}^m))\Big)^2-\Big(\mathrm{E}\Big(\sum_{t_{k-1}^n<t_i^m\leqslant t_k^n} f(d_i^m)\,|\,\mathscr{F}(t_{k-1}^n)\Big)\Big)^2\Big\}$$

$$\leqslant\frac{16}{b^2}\sum_{t_k^n\in\pi^n}\mathrm{E}\Big(\sum_{t_{k-1}^n<t_i^m\leqslant t_k^n}\mathrm{E}(f(d_i^m)\,|\,\mathscr{F}(t_{i-1}^m))\Big)^2$$

$$\leqslant\frac{16}{b^2}\mathrm{E}\sum_{t_k^m\in\pi^m}2\mathrm{E}(f(d_i^m)\,|\,\mathscr{F}(t_{i-1}^m))\mathrm{E}\Big(\sum_{t_i^m\leqslant t_j^m\leqslant\sigma_n(t_i^m)} f(d_j^m)\,|\,\mathscr{F}(t_{i-1}^m)\Big)$$

$$=\frac{32}{b^2}\mathrm{E}\sum_{t_i^m\in\pi^m} f(d_i^m)\mathrm{E}\Big(\sum_{t_i^m\leqslant t_j^m\leqslant\sigma_n(t_i^m)} f(d_j^m)\,|\,\mathscr{F}(t_{i-1}^m)\Big)$$

$$\leqslant\frac{32}{b^2}M^2P\Big(\sup_{t_i^m\in\pi^m}\mathrm{E}\Big(\sum_{t_i^m\leqslant t_j^m\leqslant\sigma_n(t_i^m)} f(d_j^m)\,|\,\mathscr{F}(t_{i-1}^m)\Big)>\varepsilon\Big)+\frac{32}{b^2}M\varepsilon,$$

where $\sigma_n(t)=\min\{t_k^n\in\pi^n\colon t\leqslant t_k^n\}$. To estimate the above probability we proceed as in 7.A. Let

$$\lambda=\inf\Big\{t_{i-1}^m\colon \mathrm{E}\Big(\sum_{t_i^m\leqslant t_j^m\leqslant\sigma_n(t_i^m)} f(d_j^m)\,|\,\mathscr{F}(t_{i-1}^m)\Big)\geqslant\varepsilon\Big\}.$$

Then

$$\text{(7.5)}\quad P\Big(\sup_{t_i^m\in\pi^m} \mathrm{E}\Big(\sum_{t_i^m\leqslant t_j^m\leqslant\sigma_n(t_i^m)} f(d_j^m)\,|\,\mathscr{F}(t_{i-1}^m)\Big)\geqslant\varepsilon\Big)$$

$$\leqslant \varepsilon^{-1}\mathrm{E}\sum_{t_i^m\in\pi^m} I(\lambda=t_i^m)\mathrm{E}\Big(\sum_{t_i^m\leqslant t_j^m\leqslant\sigma_n(t_i^m)} f(d_i^m)\,|\,\mathscr{F}(t_{i-1}^m)\Big)$$

$$=\varepsilon^{-1}\mathrm{E}\sum_{\lambda\leqslant t_j^m\leqslant\sigma_n(\lambda)} f(d_j^m)$$

$$\leqslant M\varepsilon^{-1}P\Big(\sup_{\lambda\leqslant s,t\leqslant\sigma_n(\lambda)} |X(t)-X(s)|>\delta\Big),$$

where δ is such that $f(x)=0$ for $|x|<\delta$. Therefore, in view of the left quasi-continuity of X, the probability in (7.5) converges to 0 as $n\to\infty$.

Thus, the first term in (7.3) also converges to 0 as $n, m\to\infty$, and this proves 7.F.

7.G. *For each $f\in\mathscr{R}_0$, $K_n^0(f)$ converges uniformly in P to a process with continuous sample paths.*

In the general case, where X need not satisfy property (7.1), fix $M>0$ and let $\hat{X}(t)=X(t\wedge\tau)$, where

$$\tau=\inf\Big\{t\in T\colon \sum_{i=1}^{n}|f(X(t_i)-X(t_{i-1}))|>M$$

$$\text{for some } n\in N \text{ and } 0\leqslant t_1<t_2<\ldots<t_n\leqslant t\Big\}.$$

Now, the process $\hat{X}$ has property (7.1) and is also left quasi-continuous. Let

$$\hat{K}_n^0(f)(t):=\sum_{t_k^n\leqslant t}\mathrm{E}\big(f(\hat{d}_k^n)\,|\,\mathscr{F}(t_{k-1}^n)\big),$$

where $\hat{d}_k^n=\hat{X}(t_k^n)-\hat{X}(t_{k-1}^n)$. Then, in view of 7.F, $\hat{K}_n^0(f)$ is uniformly convergent in P. On the other hand, by Lemma 2.2, for any $b>0$ and $c=\sup_{x\in R}|f(x)|$,

$$P\big(\sup_{t\in T}|\hat{K}_n^0(f)(t)-K_n^0(f)(t)|>b\big)\leqslant 2\frac{b+2c}{b}P\big(\sum|f(\hat{d}_k^n)-f(d_k^n)|>0\big)$$

$$\leqslant 2\frac{b+2c}{b}P\big(\hat{X}_n(t)\neq X(t) \text{ for some } t\in T\big)<2\frac{b+2c}{b}\varepsilon$$

for M sufficiently large, so that

$$\lim_{n,m\to\infty} P\big(\sup_{t\in T}|K_n^0(f)(t)-K_m^0(f)(t)|>3b\big)\leqslant 6\frac{b+2c}{b}\varepsilon,$$

which gives, for general X, the desired uniform convergence of $K_n^0(f)$ for $f\in\mathscr{R}_0$. Now, the continuity of $\lim_n K_n^0(f)=K^0(f)$ is assured by 7.A used as in the proof of the continuity of B. This completes the proof of 7.G.

Hence by 7.E the assertion in 7.C holds true for each $f \in \mathscr{R}_0$. It is worth of noticing that to prove this we did not use the assumption on the existence of the characteristic B.

The case of $h = [x^2]$ is technically more complex, however the main ideas are similar to those in the case of $f \in \mathscr{R}_0$.

7.H. *Let us define* $w_k^n = [d_k^n] - \mathrm{E}([d_k^n] | \mathscr{F}(t_{k-1}^n))$ *and let*

$$\bar{K}_n(h)(t) = \sum_{t_k^n \leq t} \mathrm{E}((w_k^n)^2 | \mathscr{F}(t_{k-1}^n)).$$

Then $\bar{K}_n(h) - K_n(h)$ *converges uniformly to zero in* P.

The proof of this statement is almost the same as in the case of 7.E. We have only to observe that since for all $x, u \in \boldsymbol{R}$

$$|([x]-[u])^2 - ([x-u])^2| \leq 3\big|[x]-[u]-[x-[u]]\big|,$$

we have

$$\sup_{t \in T} |\bar{K}_n(h)(t) - K_n(h)(t)| \leq 3 \sum_k \mathrm{E}\big(|[d_k^n] - \varepsilon_k^n - [d_k^n - \varepsilon_k^n]| \,\big|\, \mathscr{F}(t_{k-1}^n)\big),$$

where $\varepsilon_k^n = \mathrm{E}([d_k^n] | \mathscr{F}(t_{k-1}^n))$, and that $\sum_k |[d_k^n] - \varepsilon_k^n - [d_k^n - \varepsilon_k^n]|$ converges to zero in P since $\max_k |\varepsilon_k^n|$ converges to zero in P.

7.I. *Let* $\hat{X}(t) = X(t \wedge \tau)$ *for* $t \in T$, *where* τ *is a stopping time. Then for each* $b > 0$

$$P\big(\sup_{t \in T} |\hat{B}_n(t) - B_n(t)| > b\big) \leq 2\frac{b+2}{b} P(\tau < t_\infty),$$

$$P\big(\sup_{t \in T} |\hat{\bar{K}}_n(h)(t) - \bar{K}_n(h)(t)| > b\big) \leq 2\frac{b+16}{b} P(\tau < t_\infty).$$

Moreover, for each $a, b > 0$ *there exists* n_0 *which does not depend on* τ *such that for all* $n > n_0$

$$P\big(\sup_{t \in T} |\hat{B}_n(t) - B_n(t)| > b\big) < a.$$

Since for each $x, y, u, v \in \boldsymbol{R}$

$$|([x]-[u])^2 - ([y]-[v])^2| \leq 4|[x]-[y]+[v]-[u]|,$$

we have

$$|\hat{\bar{K}}_n(h)(t) - \bar{K}_n(h)(t)| \leq 8 \sum_k \mathrm{E}\big(|[\hat{d}_k^n] - [d_k^n]| \,\big|\, \mathscr{F}(t_{k-1}^n)\big).$$

Hence Lemma 2.2 with $c = 16$, $a = 0$ proves the second inequality of 7.I. The first inequality is even easier and it is proved in a similar way. To prove the last statement of 7.I let us observe that if τ is a stopping time, then on the set

$\{t_{l-1}^n < \tau \leqslant t_l^n\}$ we have

$$\hat{B}_n(t) - B_n(t \wedge \tau) = \sum_{t_k^n \leqslant t_l^n \wedge t} \mathrm{E}\big([\hat{d}_k^n] - [d_k^n] | \mathscr{F}(t_{k-1}^n)\big) + I(\tau < t_l^n \wedge t)\mathrm{E}\big([d_l^n] | \mathscr{F}(t_{k-1}^n)\big)$$

because

$$\mathrm{E}\big([\hat{d}_k^n] | \mathscr{F}(t_{k-1}^n)\big) = 0 \quad \text{on } \{\tau \leqslant t_l^n\} \text{ for } k > l,$$

$$\mathrm{E}\big(I(\tau \leqslant t_{k-1}^n)([\hat{d}_k^n] - [d_k^n]) | \mathscr{F}(t_{k-1}^n)\big) = 0 \quad \text{on } \{t_{l-1}^n < \tau\} \text{ for } k \leqslant l$$

and $I(t_k^n < \tau)([\hat{d}_k^n] - [d_k^n]) = 0$. Hence we get

$$\sup_{t \in T} |\hat{B}_n(t) - B_n(t \wedge \tau)| \leqslant \max_k \big|\mathrm{E}\big([d_k^n] | \mathscr{F}(t_{k-1}^n)\big)\big| + \sum_k \mathrm{E}\big(I(t_{k-1}^n < \tau \leqslant t_k^n) 2 \big|[X(\sigma_n(\tau)) - X(\tau)]\big| \,|\, \mathscr{F}(t_{k-1}^n)\big)$$

because

$$I(t_{k-1}^n < \tau \leqslant t_k^n) |[\hat{d}_k^n] - [d_k^n]| \leqslant 2 \big|[X(\sigma_n(\tau)) - X(\tau)]\big|.$$

Hence, by Lemma 2.2, we have

$$P\big(\sup_{t \in T} |\hat{B}_n(t) - B_n(t)| > b\big) \leqslant P\big(\max_k \big|\mathrm{E}\big([d_k^n] | \mathscr{F}(t_{k-1}^n)\big)\big| > b/2\big) + \frac{4a}{b} + 2\frac{b+4}{b} P\big(|X(\sigma_n(\tau)) - X(\tau)| > a\big) \quad \text{for each } a, b > 0,\ a \leqslant 1.$$

This, the left quasi-continuity of X and 7.A prove the last statement of 7.I.

7.J. *If the characteristic B of X exists, then $\bar{K}_n(h)$ is uniformly convergent in P.*

This and 7.H imply 7.C since

$$\sup_{t \in T} |\bar{K}_n(h)(t) - \bar{K}_n(h)(t-)| \leqslant \max_k \mathrm{E}\big(([d_k^n])^2 | \mathscr{F}(t_{k-1}^n)\big),$$

and by 7.A the uniform limit of $\bar{K}_n(h)$ is a continuous process.

For fixed $m > n$ and a stopping time τ let $\hat{X}(t) = X(t \wedge \tau)$ and let

$$\hat{\bar{K}}_{n,m}(h)(t) = \sum_{t_k^n \leqslant t} \mathrm{E}\Big(\Big(\sum_{t_{k-1}^n < t_i^m \leqslant t_k^n} \hat{w}_i^m\Big)^2 | \mathscr{F}(t_{k-1}^n)\Big),$$

where $\hat{w}_i^m$ are defined as w_i^m in 7.H with $\hat{X}$ instead of X. By arguments as in 7.E we have

$$\sup_{t \in T} |\bar{K}_n(h)(t) - \bar{K}_m(h)(t)| \leqslant \sup_{t \in \pi^n} |\bar{K}_n(h)(t) - \bar{K}_m(h)(t)| + \max_k \mathrm{E}\big(([d_k^n])^2 | \mathscr{F}(t_{k-1}^n)\big).$$

This yields the following inequality:

$$(7.6)\quad P\big(\sup_{t\in T}|\bar{K}_n(h)(t)-\bar{K}_m(h)(t)|>5b\big)$$
$$\leqslant P\big(\sup_{t\in T}|\bar{K}_n(h)(t)-\hat{\bar{K}}_n(h)(t)|>b\big)+P\big(\sup_{t\in T}|\bar{K}_m(h)(t)-\hat{\bar{K}}_m(h)(t)|>b\big)$$
$$+P\big(\sup_{t\in\pi^n}|\hat{\bar{K}}_m(h)(t)-\hat{\bar{K}}_{n,m}(h)(t)|>b\big)+P\big(\sup_{t\in\pi^n}|\hat{\bar{K}}_n(h)(t)-\hat{\bar{K}}_{n,m}(h)(t)>b\big)$$
$$+P\big(\max_k \mathrm{E}(([d_k^n])^2|\mathscr{F}(t_{k-1}^n))>b\big).$$

By 7.A the last probability is convergent to zero as $n\to\infty$. Therefore, to prove 7.J it suffices to show that, given a, $b>0$, if n, m are large enough, we can find a stopping time τ such that the first four probabilities on the right-hand side of (7.6) are less than a. By 7.I the first two probabilities are estimated by

$$2\frac{b+16}{b}P(\tau<t_\infty).$$

Now we will make the stopping time τ more specific. For a positive integer m and $\bar{M}>0$ let

$$(7.7)\quad \begin{aligned} &\tau_1=\min\{t\in\pi^m:\ t\geqslant t_\infty\ \text{ or }\ |\sum_{t_i^m\leqslant t}[d_i^m]|>\bar{M}\ \text{ or }\ |B_m(t)|>\bar{M}\},\\ &\tau_2=\min\{t\in\pi^m:\ t\geqslant t_\infty\ \text{ or }\ \sum_{t_i^m\leqslant t}(w_i^m)^2>\bar{M}^3\},\\ &\tau:=\tau_{m,\bar{M}}=\tau_1\wedge\tau_2.\end{aligned}$$

Since

$$\mathrm{E}\big(\sum_{t_i^m\leqslant\tau_1}(w_i^m)^2\big)=\mathrm{E}\big(\sum_{t_i^m\leqslant\tau_1}w_i^m\big)^2=\mathrm{E}\big(\sum_{t_i^m\leqslant\tau_1}[d_i^m]-B_m(\tau_1)\big)^2\leqslant(M+2)^2,$$

we obtain

$$P(\tau_2<t_\infty)\leqslant P(\tau_1<t_\infty)+(\bar{M}+2)^2/\bar{M}^3.$$

Moreover, we have

$$P(\tau_1<t_\infty)\leqslant P\big(\sup_{t\in T}|\sum_{t_i^m\leqslant t}[d_i^m]|>\bar{M}\big)+P\big(\sup_{t\in T}|B_m(t)|>\bar{M}\big).$$

Since B_m is uniformly convergent in P, the above inequalities imply that if $\bar{M}$ is large enough, then $P(\tau<t_\infty)<a$ for all m. Thus 7.J will be proved if we manage to show that for each fixed $\bar{M}$ if τ is defined by (7.7), then the third and the fourth probabilities on the right-hand side of (7.6) are convergent to zero when n, $m\to\infty$.

7.K. *If the characteristic B exists and the stopping times $\tau_{m,\bar{M}}$ are defined by* (7.7), *then*

$$\lim_{n,m\to\infty} P\big(\sup_{t\in\pi^n} |\hat{\bar{K}}_m(h)(t) - \hat{\bar{K}}_{n,m}(h)(t)| > b\big) = 0 \quad \textit{for each } \bar{M} > 0.$$

This can be demonstrated by mimicking the proof of the estimate of the first term of (7.3) in 7.F. With $\hat{\bar{K}}_n(h)$ replacing $K_n^0(f)$, $\hat{\bar{K}}_{n,m}(h)$ replacing $K_{n,m}^0(f)$, and $(\hat{w}_i^m)^2$ replacing $f(d_i^m)$ we have

$$\sum_i (\hat{w}_i^m)^2 \leqslant \bar{M}^3 + 4 =: M$$

and the proof carries over to the current situation with the following change in the replacement of (7.5):

$$\begin{aligned} &\varepsilon^{-1} \mathrm{E} \sum_i I(\lambda = t_i^m) \sum_{t_i^m \leqslant t_j^m \leqslant \sigma_n(t_i^m)} (\hat{w}_j^m)^2 \\ &\quad = \varepsilon^{-1} \mathrm{E}\big(\sum_{\lambda \leqslant t_j^m \leqslant \sigma_n(\lambda)} \hat{w}_j^m \big)^2 \leqslant \varepsilon^{-1} \big((4M+2)^2 P\big(\big| \sum_{\lambda \leqslant t_j^m \leqslant \sigma_n(\lambda)} \hat{w}_j^m \big| > \delta \big) + \delta^2 \big). \end{aligned}$$

Since $\hat{B}_m(t) = B_m(t \wedge \tau)$, the last probability is estimated by

$$\begin{aligned} &P\big(\max_{\substack{s,t\in\pi^m \\ |s-t| \leqslant \operatorname{mesh}\pi^n}} \big| \sum_{s \leqslant t_i^m \leqslant t} [d_i^m] - [\sum_{s \leqslant t_i^m \leqslant t} d_i^m] \big| > \delta/3 \big) \\ &\qquad + P\big(|X(\lambda_2) - X(\lambda_1)| > \delta/3\big) + P\big(|B_m(\lambda_2) - B_m(\lambda_1)| > \delta/3\big), \end{aligned}$$

where $\lambda_2 = \sigma_n(\lambda) \wedge \tau$ and $\lambda_1 = \max\{t_i^m \colon t_i^m < \lambda \text{ or } t_i^m \leqslant \tau\}$ (λ_1 is a stopping time since $\{\lambda = t_i^m\}$ is in $\mathscr{F}(t_{i-1}^m)$). Since $|\lambda_2 - \lambda_1| \leqslant \operatorname{mesh}\pi^n$ and since B_m is uniformly convergent to a continuous process in P, the last three probabilities are convergent to zero when $n, m \to \infty$.

7.L. *If the characteristic B exists and the stopping times $\tau_{m,\bar{M}}$ are defined by* (7.7), *then*

$$\lim_{n,m\to\infty} P\big(\sup_{t\in\pi^n} |\hat{\bar{K}}_n(h)(t) - \hat{\bar{K}}_{n,m}(h)(t)| > b\big) = 0 \quad \textit{for each } \bar{M}.$$

If $d_k = f_k - f_{k-1}$, $e_k = g_k - g_{k-1}$, $k = 1, 2, \ldots, l$, are martingale differences with respect to $(\mathscr{F}_k)$, then

$$\mathrm{E} \sum_{k=1}^{l} |\mathrm{E}(d_k e_k | \mathscr{F}_{k-1})| \leqslant (\mathrm{E} f_l^2)^{1/2} (\mathrm{E} g_l^2)^{1/2}.$$

Hence we easily infer that if λ is a stopping time with respect to $(\mathscr{F}_k)$ such that $|f_\lambda|$, $|g_\lambda| \leqslant M'$, then for each $b, c > 0$

$$P\big(\sum_{k=1}^{l} |\mathrm{E}(d_k e_k | \mathscr{F}_{k-1})| > b\big) \leqslant P(\lambda < l) + \frac{M'}{b}\big(M'(P(|f_l^*| > c))^{1/2} + c\big). \tag{7.8}$$

We apply this to the following situation: for $k = 1, 2, \ldots, l = k_n$

$$d_k = \hat{w}_k^n - \sum_{t_{k-1}^n < t_i^m \leqslant t_k^n} \hat{w}_i^m, \quad e_k = \hat{w}_k^n + \sum_{t_{k-1}^n < t_i^m \leqslant t_k^n} \hat{w}_i^m, \quad \mathscr{F}_k = \mathscr{F}(t_k^n),$$

$$\lambda = \min\{j\colon j \geqslant l \text{ or } |\sum_{k=1}^{j} [\hat{d}_k^n]| > \bar{\bar{M}} \text{ or } |\hat{B}_n(t_j^n)| > \bar{\bar{M}}\}.$$

By the definition of λ and $\tau_{m,\bar{M}}$ we have $|f_\lambda|, |g_\lambda| \leqslant \bar{M} + \bar{\bar{M}} + 4 =: M'$. Since for each $t \in \pi^n$

$$\hat{\hat{K}}_n(h)(t) - \hat{\hat{K}}_{n,m}(h)(t) = \sum_{t_k^n \leqslant t} \mathrm{E}(d_k e_k | \mathscr{F}_{k-1}), \tag{7.9}$$

7.L will follow by (7.8) if we show that for given $a > 0$ there exists $\bar{\bar{M}}$ such that $P(\lambda < l) < a$ for all n, m and that $\lim_{n,m\to\infty} P(f_l^* > c) = 0$ for each $c > 0$. However, this follows easily by 7.I, the uniform convergence of B_m in P and the estimates

$$\begin{aligned} P(\lambda < l) &\leqslant P\big(\max_j |\sum_{k=1}^{j} [\hat{d}_k^n]| > \bar{\bar{M}}\big) + P\big(\sup_{t\in T} |\hat{B}_n(t)| > \bar{\bar{M}}\big) \\ &\leqslant P\big(\sup_{t\in T} |\sum_{t_k^n \leqslant t} [d_k^n]| > \bar{\bar{M}} - 1\big) + P\big(\sup_{t\in T} |B_n(t \wedge \tau) - \hat{B}_n(t)| > \bar{\bar{M}}/2\big) \\ &\quad + P\big(\sup_{t\in T} |B_n(t)| > \bar{\bar{M}}/2\big) \end{aligned}$$

and

$$\begin{aligned} P(f_l^* > c) \leqslant\; & P\big(\max_{t\in\pi^m} |\sum_{t_i^m \leqslant t} [d_i^m] - \sum_{t_k^n \leqslant \sigma^n(t)} [d_k^n] - [\sum_{\sigma^n(t) < t_i^m \leqslant t} d_i^m]| > c/3\big) \\ & + P\big(\sup_{t\in T} |\hat{B}_n(t) - B_n(t \wedge \tau)| > c/3\big) + P\big(\sup_{t\in T} |B_n(t) - B_m(t)| > c/3\big), \end{aligned}$$

where $\sigma^n(t) = \max\{t_k^n \in \pi^n\colon t_k^n \leqslant t\}$. This concludes the proof of 7.L and, thus, by 7.K, we have also completed the proof of 7.C.

7.M. To complete our proof of Proposition 4.1 it remains to show the existence of the characteristics μ and C. This implication follows from the proved existence of the functional $K(f)$ and requires only a standard observation that the form and obvious properties of K (cf. (4.5)) *uniquely determine* μ *and* C.

7.N. *For any left quasi-continuous semimartingale* X *the characteristic* B *exists.*

Initially, we assume that $X = A + X'$, where a.s.

$$\operatorname{Var} A(t_\infty) \leqslant M \quad \text{and} \quad \sup_{t\in T} |X'(t)| \leqslant M, \tag{7.10}$$

A is left quasi-continuous and X' is a martingale. Since X' is a left quasi-continuous martingale, by 7.G, X' has the characteristic B. So it is sufficient to prove the uniform convergence of B_n in the case of $X = A$. In this case it is a known fact on "laplaciens approches". It can also be directly deduced as follows. Consider, without loss of generality, only non-decreasing A, and then follow the pattern of estimation of the first term in (7.3).

The reduction of the general case to the case considered above proceeds as in 7.G. Take a general semimartingale $X = A+X'$, where A is a left quasi--continuous process of locally bounded variation and X' is a local martingale, and introduce

$$\tau = \inf\{t \in T\colon \operatorname{Var} A(t) \geqslant M \text{ or } |X'(t)| \geqslant M\}.$$

Let $\hat{X}(t) = [X(\tau \wedge t)]^{2M}$. Then the rest of the proof mimicks the remainder of the proof of 7.G.

7.O. To prove Remark 4.5 it suffices to demonstrate that *if X is a left quasi-continuous semimartingale and $f \in \mathscr{R}$, then $K_n(f) - K_n^0(f)$ is uniformly convergent to* 0 *in P* ($K_n^0(f)$ is defined by (7.2)).

As in 7.D it is enough to prove the above statement for $f \in \mathscr{R}_0$ and $f = h$. If $f \in \mathscr{R}_0$, then it follows from 7.N. If $f = h$, then, once again, assume initially that X satisfies (7.10). Then we have

$$\begin{aligned} K_n(h)(t) - K_n^0(h)(t) &= \sum_{t_k^n \leqslant t} \Big(\big(\mathrm{E}(d_k^n \mid \mathscr{F}(t_{k-1}^n))\big)^2 + \big(\mathrm{E}\psi(d_k^n) \mid \mathscr{F}(t_{k-1}^n)\big)^2 \Big) \\ &= \sum_{t_k^n \leqslant t} \Big(\big(\mathrm{E}(a_k^n \mid \mathscr{F}(t_{k-1}^n))\big)^2 + \big(\mathrm{E}\psi(d_k^n) \mid \mathscr{F}(t_{k-1}^n)\big)^2 \Big), \end{aligned}$$

where $a_k^n = A(t_k^n) - A(t_{k-1}^n)$ and $\psi(x) = [x] - [x]^{2M}$, so that

$$\begin{aligned} \sup_{t \in T} |K_n(h)(t) - K_n^0(h)(t)| \leqslant{} & \sup_{t_k^n \in \pi^n} \big|\mathrm{E}(a_k^n \mid \mathscr{F}(t_{k-1}^n))\big| \sum_{t_k^n \in \pi^n} \big|\mathrm{E}(a_k^n \mid \mathscr{F}(t_{k-1}^n))\big| \\ & + \sup_{t_k^n \in \pi^n} \big|\mathrm{E}(\psi(d_k^n) \mid \mathscr{F}(t_{k-1}^n))\big| K_n^0(|\psi|)(t_\infty). \end{aligned}$$

By 7.G the second term converges to 0 in P as $n \to \infty$, and the first term converges to 0 in P because, by 7.A, the sup goes to zero in P and because, in view of Lemma 2.2, we have, for each $b > 0$,

$$P\Big(\sum_{t_k^n \in \pi^n} \big|\mathrm{E}(a_k^n \mid \mathscr{F}(t_{k-1}^n))\big| > b\Big) \leqslant \frac{M}{b} + 2\frac{b+2M}{b} P\big(\operatorname{Var} A(t_\infty) > M\big) = \frac{M}{b}.$$

Now, in the general case, we can take $\hat{X}$ as in 7.N, and the proof can be concluded by estimating both

$$P\big(\sup_{t \in T} |K_n^0(h)(t) - \hat{K}_n^0(h)(t)| > b\big) \quad \text{and} \quad P\big(\sup_{t \in T} |K_n(h)(t) - \hat{K}_n(h)(t)| > b\big)$$

as in 7.I.

7.P. *If $X(t)$, $t\in T$, is a process with independent increments with sample paths in $D(T)$ and such that $\bigcup_n \pi^n$ contains all the points of stochastic discontinuity of the process X, then B_n is uniformly convergent on T to a function B in $D(T)$. If $B\in \mathrm{BV}(T)$, then*

$$\operatorname{Var} B(t) = \lim_{n\to\infty} \sum_{t_k^n \leqslant t} |\mathrm{E}[d_k^n]|, \qquad t\in T.$$

Indeed, define

$$X_n(t) = \sum_{t_k^n \leqslant t} [d_k^n].$$

Then, for each t and for each ω such that $X(\omega,\cdot)\in D(T)$ $\hat{X}(T) := \lim X_n(t)$ is well defined.

Moreover, $\{X_n(t), t\in T\}$ is also convergent in $D(T)$ to $\{\hat{X}(t+), t\in T\}$ for ω as above. Thus, by inequality (2.2), for a large enough and all $n\in N$, we have

$$\mathrm{E}X_n^* \leqslant \frac{a+1}{1/2 - P(X_n^* > a)} \leqslant 4(a+1).$$

Here for a process Y, $Y^* := \sup_{t\in T} |Y(t)|$. Therefore, we get

$$\mathrm{E}\hat{X}^* \leqslant 4(a+1). \tag{7.11}$$

Let $B(t) = \mathrm{E}\hat{X}(t)$ (which is well defined for each t in view of (7.11)) and let $\hat{d}_k^n = \hat{X}(t_k^n) - \hat{X}(t_{k-1}^n)$. Then for each n, any $a>0$, and any selection of signs $\varepsilon_k^n = \pm 1$, we have (by (2.2))

$$\mathrm{E}\sum\nolimits^* \varepsilon_k^n([d_k^n] - \hat{d}_k^n) \leqslant \frac{a + \mathrm{E}\sup_k |[d_k^n] - \hat{d}_k^n|}{1/2 - P(\sum |[d_k^n] - \hat{d}_k^n| > a)}. \tag{7.12}$$

Since, by (7.11),

$$\sup_k |[d_k^n] - \hat{d}_k^n| \leqslant 2\hat{X}^* + 1 \in L^1 \quad \text{and} \quad \lim_{n\to\infty} \sum |[d_k^n] - \hat{d}_k^n| = 0$$

for each ω such that $X(\omega,\cdot)\in D(T)$, the left-hand side of (7.12) converges to 0 as $n\to\infty$.

Hence firstly,

$$\lim_{n\to\infty} \sup_{t\in\pi^n} |B_n(t) - B(t)| = 0$$

and, since B is a continuous function away from $\bigcup_n \pi^n$, we infer that the uniform convergence of B_n to B takes place on all of T, which implies that $B\in D(T)$. The fact that B is continuous away from $\bigcup_n \pi^n$ can be seen as follows: Let $t_0 \notin \bigcup_n \pi^n$. Then for each $\varepsilon > 0$ there exists an interval $I = (t_{k-1}^n, t_k^n)$ for some n, k, containing t_0 and such that

$$P\big(\sup_{t,t'\in I} |X(t) - X(t')| > \varepsilon\big) < \varepsilon,$$

and therefore

$$P\left(\sup_{t,t'\in I}|\hat{X}(t)-\hat{X}(t')| > \varepsilon\right) < \varepsilon.$$

Thus, as before, by (2.2),

$$\sup_{t,t'\in I}|B(t)-B(t')| \leqslant \frac{\varepsilon+\mathrm{E}\sup_{t,t'\in I}|\hat{X}(t)-\hat{X}(t')|}{1/2-P\left(\sup|\hat{X}(t)-\hat{X}(t')| > \varepsilon\right)},$$

and since, by (7.11), $\mathrm{E}\hat{X}^* < \infty$, we get

$$\mathrm{E}\sup_{t,t'\in I}|\hat{X}(t)-\hat{X}(t')| < \varepsilon$$

if I is sufficiently small. Hence

$$\sup_{t,t'\in I}|B(t)-B(t')| \leqslant 2\varepsilon(1/2-\varepsilon)^{-1}$$

if I is sufficiently small.

Secondly, (7.12) also implies that

$$\lim_{n\to\infty}\sum_{t_k^n\in\pi^n}|\mathrm{E}[d_k^n]-\mathrm{E}\hat{d}_k^n| = 0,$$

which proves the second statement of 7.P.

Acknowledgement. We are grateful to W. Hudson, J. Jacod, O. Kallenberg and L. Słominski for drawing our attention to shortcomings of previous versions of this paper.

REFERENCES

[1] D. J. Aldous, *Stopping times and tightness*, Ann. Probab. 6 (1978), pp. 335–340.

[2] D. L. Burkholder, *Distribution function inequalities for martingales*, ibidem 1 (1973), pp. 19–42.

[3] — *A geometric condition that implies the existence of certain singular integrals of Banach-space-valued functions*, Conference on Harmonic Analysis in Honor of Antoni Zygmund, 1983.

[4] C. Dellacherie et C. Doleans-Dade, *Un contre-exemple au problème des laplaciens approches*, Séminaire de Probabilités V, Springer Lecture Notes in Math. 191 (1971), pp. 127–137.

[5] T. P. Hill, *Conditional generalizations of strong laws which conclude that partial sums converge almost surely*, Ann. Probab. 10 (1982), pp. 828–830.

[6] J. Hoffmann-Jørgensen, *Sums of independent Banach space valued random variables*, Studia Math. 52 (1974), pp. 159–186.

[7] W. Hudson, *Stochastic integrals and processes with independent increments*, UNC Center for Stochastic Processes. Technical Report No. 98 (1985).

[8] K. Itô, *Differential equations determining a Markov process*, Journ. Pan-Japan Math. Coll. No. 1077 (1942).

[9] J. Jacod, *Calcul stochastique et problèmes de martingales*, Springer Lecture Notes in Math. 714 (1979).

[10] — *Processus à accroissements indépendants: une condition nécessaire et suffisante de convergence en loi*, Z. Wahrsch. Verw. Gebiete 63 (1983), pp. 109–136.

[11] — *Une généralisation des semimartingales: les processus admettant un processus à accroissement indépendants tangent*, Springer Lecture Notes in Math. 1059 (1984), pp. 91–118.

[12] — et H. Sadi, *Processus admettant un processus à accroissements indépendants tangent; cosi general*, ibidem 1247 (1988), pp. 479–514.

[13] A. Jakubowski, *Principle of conditioning in limit theorems for sums of random variables*, Ann. Probab. 14 (1986), pp. 902–915.

[14] — and L. Słonimski, *Extended convergence to continuous in probability processes with independent increments*, Probab. Theory Related Fields 72 (1986), pp. 55–82.

[15] O. Kallenberg, *On the existence and path properties of stochastic integrals*, Ann. Probab. 3 (1975), pp. 262–280.

[16] S. Kwapień and W. A. Woyczyński, *Decoupling of martingale transforms and stochastic integrals for processes with independent increments*, pp. 139–148 in: *Probability Theory and Harmonic Analysis*, M. Dekker, Inc., 1986.

[17] — *Double stochastic integrals, random quadratic forms and random series in Orlicz spaces*, Ann. Probab. 15 (1987), pp. 1072–1096.

[18] — *Decoupling inequalities, tangent sequences and almost sure convergence*, Proc. Conf. on Almost Everywhere Convergence in Probability and Ergodic Theory, 1988.

[19] J. Musielak, *Orlicz spaces and modular spaces*, Springer Lecture Notes in Math. 1034 (1983).

[20] A. Prekopa, *On stochastic set functions, I, II, III*, Acta Math. Acad. Sci. Hungar. 7 (1956), pp. 215–263; ibidem 8 (1957), pp. 337–374; ibidem 8 (1957), pp. 375–400.

[21] J. Rosiński, *Bilinear random integrals*, Dissertationes Math. (Rozprawy Mat.) 259 (1987), pp. 1–69.

[22] L. Schwartz, *Semi-martingales formelles*, Séminaire de Probabilités XV, Springer Lecture Notes in Math. 850 (1979–1980), pp. 413–489.

[23] J. Szulga, *Three series theorems for martingales*, Bull. Polish Acad. Sci. Math. 25 (1978), pp. 1015–1021.

[24] K. Urbanik, *Random measures and harmonizable sequences*, Studia Math. 31 (1968), pp. 61–88.

[25] — and W. A. Woyczyński, *A random integral and Orlicz spaces*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom Phys. 15 (1967), pp. 161–169.

Warsaw University
Institute of Applied Mathematics and Mechanics
Pałac Kultury i Nauki
00-901 Warszawa

Case Western Reserve University
Cleveland, Ohio 44106
U.S.A.

Received on 28.9.1989