Weak Convergence of Stochastic Integrals and Differential Equations

Thomas G. Kurtz[1]* and Philip E. Protter[2] **

[1] Departments of Mathematics and Statistics University of Wisconsin - Madison, Madison, WI 53706-1388, USA
[2] Departments of Mathematics and Statistics, Purdue University, West Lafayette, IN 47907-1395, USA

1. Semimartingales

Let W denote a standard Wiener process with $W_0 = 0$. For a variety of reasons, it is desirable to have a notion of an integral $\int_0^1 H_s dW_s$, where H is a stochastic process; or more generally an indefinite integral $\int_0^t H_s dW_s$, $0 \le t < \infty$. If H is a process with continuous paths, an obvious way to define a stochastic integral is by a limit of sums: let $\pi^n[0,t]$ be a sequence of partitions of $[0,t]$, with mesh$(\pi^n) = \sup_i(t_{i+1} - t_i)$, where $0 = t_0 < t_1 < \ldots < t_n = t$ are the successive points of the partition. Then one could define

$$\int_0^t H_s dW_s = \lim_{n\to\infty} \sum_{t_i \in \pi^n[0,t]} H_{t_i}(W_{t_{i+1}} - W_{t_i}) \tag{1.1}$$

when $\lim_{n\to\infty}$ mesh$(\pi^n) = 0$. If one wants the natural condition that (1.1) holds for all continuous processes H, then it is an elementary consequence of the Banach-Steinhaus theorem that W must have a.s. paths of finite variation on compacts. Of course this is precisely not the case for the Wiener process. The key insight of K. Itô in the 1940's was to ask for condition (1.1) to hold only for *adapted* continuous stochastic processes. We will both explain this idea and extend it to a large class of stochastic processes: exactly those for which both the integral exists as a limit of sums, and for which we also have a dominated convergence theorem.

We suppose given a filtered probability space $(\Omega, \mathcal{F}, P, \mathbf{F})$, where $\mathcal{F}$ is a P-complete σ-algebra and where $\mathbf{F} = (\mathcal{F}_t)_{0\le t<\infty}$ is a filtration of σ-algebras: i.e., $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \le t$. We also assume that $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}_0$ and that $\mathbf{F}$ is right continuous: that is, $\mathcal{F}_t = \mathcal{F}_{t+} = \cap_{u>t}\mathcal{F}_u$. (Note that if W is a standard Wiener process with its natural filtration $\mathbf{F}^0 = (\mathcal{F}_t^0)_{0\le t<\infty}$, where $\mathcal{F}_t = \sigma(W_s; s \le t)$, then if one adds the P-null sets of $\mathcal{F}_t^0$ to $\mathcal{F}_t^0$, all t, the resulting filtration $\mathbf{F}$ satisfies the preceding hypotheses, which are known

* Research supported in part by NSF grants DMS 92-04866 and DMS 95-04323
** Research supported in part by NSF grant INT 94-01109 and NSA grant MDR 904-94-H-2049

as the *usual hypotheses*. The same holds for Lévy processes and for most strong Markov processes.)

Let X be an *adapted* process with càdlàg paths: that is, X_t is $\mathcal{F}_t$-measurable, each $t > 0$, and a.s. has paths which are right continuous with left limits.[1]

Definition 1.1. *A process H is* simple predictable *if H has a representation*

$$H_t = H_0 1_{\{0\}}(t) + \sum_{i=1}^{n} H_i 1_{(T_i, T_{i+1}]}(t) \tag{1.2}$$

where $0 = T_1 \leq \ldots \leq T_{n+1} < \infty$ is a finite sequence of stopping times, $H_i \in \mathcal{F}_{T_i}$, $|H_i| < \infty$ a.s., $0 \leq i \leq n$. The collection of simple predictable processes is denoted $\mathbf{S}$.

Let L^0 *denote all a.s. finite random variables. We topologize* L^0 *with convergence in probability, and we topologize* $\mathbf{S}$ *with uniform convergence (in (t,ω)) and denote it* $\mathbf{S}_u$. *For a given X we define an operator I_X mapping* $\mathbf{S}$ *to* L^0 *by (with H as in (1.2)):*

$$I_X(H) = H_0 X_0 + \sum_{i=1}^{n} H_i (X_{T_{i+1}} - X_{T_i}). \tag{1.3}$$

Definition 1.2. *A process X is a* semimartingale *if $I_X : \mathbf{S}_u \to \mathrm{L}^0$ is continuous on compact time sets.*

Definition (1.2) is not customary. We give the customary definition here, and to distinguish it from ours we call it a "classical" semimartingale.

Definition 1.3. *A process X is a* classical semimartingale *if it is adapted, càdlàg, and has a decomposition $X = M + A$, where M is a local martingale, and A (is adapted, càdlàg, and) has paths of finite variation on compacts.*

One of the deepest results in the theory of semimartingales is the following, proved around 1978, primarily by C. Dellacherie and K. Bichteler.

Theorem 1.4 (Bichteler-Dellacherie). *An adapted, càdlàg process X is a semimartingale if and only if it is a classical semimartingale.*

We remark that the deeper implication is the "only if".

Also note that the Bichteler-Dellacherie theorem gives us many *examples of semimartingales:*

(i) Any local martingale, such as the Wiener process, is a semimartingale.
(ii) Any finite variation process, such as the Poisson process, is a semimartingale.

[1] "càdlàg" is the French acronym for right continuous with left limits

(iii) The Doob-Meyer decomposition theorem states that any submartingale Y can be written $Y = M + A$, where M is a local martingale and A is an adapted, càdlàg process with nondecreasing paths. Thus, any submartingale (and hence any supermartingale) is a semimartingale.

(iv) If Z is a Lévy process (i.e., a càdlàg process with stationary and independent increments), then if $E\{|Z_t|\} < \infty$, each t, one has $E\{|Z_t|\} = \alpha t$ (assuming $Z_0 = 0$) and thus $Z_t = (Z_t - \alpha t) + \alpha t$ is a decomposition for Z,a nd Z is a semimartingale. More generally it can be shown that any Lévy process is a semimartingale.

(v) Most "reasonable" real valued strong Markov processes are semimartingales.

(vi) An illustrative example of a Lévy process that is a martingale is as follows: let N^i be a sequence of i.i.d. Poisson processes with arrival intensities $\alpha_i (\alpha_i > 0)$. Let $|\beta_i| \leq c$ and assume $\sum_{i=1}^{\infty} \beta_i^2 \alpha_i < \infty$. Then

$$M_t = \sum \beta_i (N_t^i - \alpha_i t)$$

is a Lévy process. Note that if, for example, $\alpha_i = 1$ (all i) and $\beta_i = \frac{1}{i}$, then if $\Delta M_s = M_s - M_{s-}$ (the jump at time s), we have $\sum_{0<s\leq t} |\Delta M_s| = \sum_{0<s\leq t} \Delta M_s = \sum_{i=1}^{\infty} \frac{1}{i} N_t^i = \infty$ a.s. This is an example of a martingale that cannot be used, path by path, as a classical differential because of behavior arising purely from the jumps; that is, M has paths of infinite variation on compacts and one cannot define a Lebesgue-Stieltjes pathwise integral for M.

Finally let us note some simple but important properties of semimartingales.

Theorem 1.5. *The set of semimartingales is a vector space.*

Theorem 1.6. *If Q is another probability absolutely continuous with respect to P, then every P-semimartingale is a Q-semimartingale.*

Theorem 1.7 (Stricker). *If X is a semimartingale for a filtration $\mathbf{F}$, and if $\mathbf{G}$ is a subfiltration such that X is adapted to $\mathbf{G}$, then X is a $\mathbf{G}$-semimartingale as well.*

Proof. Theorem 1.5 is immediate from the definition. For Theorem 1.6 it is enough to remark that if $Q \ll P$, then convergence in P-probability implies convergence in Q probability. For Theorem 1.7, let $\mathbf{S}(\mathbf{F})$ denote $\mathbf{S}$ for the filtration $\mathbf{F}$. Since $\mathbf{S}(\mathbf{G}) \subset \mathbf{S}(\mathbf{F})$, if I_X is continuous for $I_X : \mathbf{S}_u(\mathbf{F}) \to \mathbf{L}^0$, then it is *a fortiori* continuous for $\mathbf{S}_u(\mathbf{G})$. □

Stricker's theorem shows one can easily shrink the filtration since one is only shrinking the domain of a continuous operator. Expanding the filtration, on the other hand, is more delicate, since one is then asking a continuous operator to remain continuous for a larger domain. An elementary result in this direction is the following:

Theorem 1.8 (P. A. Meyer). *Let $\mathcal{A}$ be a countable collection of disjoint sets in $\mathcal{F}$. Let* $\mathbf{H}$ *be the filtration given by $\mathcal{H}_t = \sigma(\mathcal{F}_t, \mathcal{A})$. Then every* $\mathbf{F}$ *semimartingale is an* $\mathbf{H}$ *semimartingale.*

Proof. Without loss of generality assume $\mathcal{A}$ is a partition of Ω, and $P(A_n) > 0$, each $A_n \in \mathcal{A}$. Define $Q_n \ll P$ by $Q_n(\Lambda) = P(\Lambda | A_n)$. Then X is a Q_n-semimartingale by Theorem 1.6. Let $\mathbf{I}^n$ be the filtration generated by $\mathbf{F}$ and all Q_n null sets. Let X be a $(\mathbf{I}^n, Q_n)$-semimartingale, each n. Moreover $\mathbf{F} \subset \mathbf{H} \subset \mathbf{I}^n$. By Stricker's theorem, X is an $\mathbf{H}$ semimartingale under Q_n. Note that $dP = \sum_{n=1}^{\infty} P(A_n) dQ_n$. Suppose $H^n \in \mathbf{S}(\mathbf{H})$ converges to $H \in \mathbf{S}(\mathbf{H})$ uniformly. Then $I_X(H^n)$ converges to $I_X(H)$ in Q_n-probability for each n, and it follows that it converges in P-probability as well. Thus X is an $(\mathbf{H}, P)$-semimartingale. □

2. Stochastic Integration

We wish to define a stochastic integral of the form $\int_0^t H_{s-} dX_s$, where H is càdlàg, adapted, and H_{s-} represents its left continuous version; and X is a semimartingale. We recall $\mathbf{S}$ is the space of simple predictable processes and $\mathbf{L}^0$ is the space of finite valued random variables.

We also define:

$\mathbf{D}$ = the space of adapted processes with càdlàg paths
$\mathbf{L}$ = the space of adapted processes with càdlàg paths (left continuous with right limits)

Note that if $H \in \mathbf{D}$, then H_- (its left continuous version) is in $\mathbf{L}$; and if $H \in \mathbf{L}$, then H_+ is in $\mathbf{D}$. We next define a new topology, ucp, which will replace uniform convergence.

Definition 2.1. *A sequence of processes Y^n converges to a process Y uni*formly on compacts in probability *(denoted ucp) if for each $t > 0$, $\sup_{s \leq t} |Y_s^n - Y_s| = (Y^n - Y)_t^*$ tends to 0 in probability as n tends to ∞.*

We note that this topology is metrizable.

Theorem 2.2. $\mathbf{S}$ *is dense in* $\mathbf{L}$ *under ucp .*

Proof. By stopping, $b\mathbf{L}$ is dense in $\mathbf{L}$, where $b\mathbf{L}$ denotes the bounded processes in $\mathbf{L}$. For $Y \in b\mathbf{L}$, let $Z = Y_+$, and for $\varepsilon > 0$, define $T_0^\varepsilon = 0$ and

$$T_{n+1}^\varepsilon = \inf\{t : t > T_n^\varepsilon \text{ and } |Z_t - Z_{T_n^\varepsilon}| > \varepsilon\}.$$

Then T_n^ε are stopping times and they are increasing since Z is càdlàg. Pose $Z_1^\varepsilon = Y_0 1_{\{0\}} + \sum_{i=1}^n Z_{T_i^\varepsilon} 1_{(T_i^\varepsilon \wedge n, T_{i+1}^\varepsilon \wedge n]}$. This can be made arbitrarily close to $Y \in b\mathbf{L}$ by taking ε small enough and n large enough. □

The operator I_X defined in (1.3) was, effectively, an operator giving a definite integral for processes $H \in \mathbf{S}$ and semimartingales X. We now wish to define an operator which will be an indefinite integral operator. Thus its range should be processes rather than random variables. Therefore for a given process X and a process $H \in \mathbf{S}$ as given in (1.2) we define the operator $J_X : \mathbf{S} \to \mathbf{D}$ by:

$$J_X(H) = H_0 X_0 + \sum_{i=1}^{n} H_i (X^{T_{i+1}} - X^{T_i}), \tag{2.1}$$

where the notation X^T, for a stopping time T, denotes the process $X_t^T = X_{t\wedge T} (t \geq 0)$.

Definition 2.3. *For an adapted, càdlàg process X and $H \in \mathbf{S}$, the process $J_X(H)$ is called the* stochastic integral *of H with respect to X.*

We will also use the notations $\int_0^t H_s dX_s$ and $H \cdot X$ or $H \cdot X_t$ to denote the stochastic integral. That is

$$\begin{aligned} J_X(H) &= \int H dX &&= H \cdot X \\ J_X(H)_t &= \int_0^t H_s dX_s &&= H \cdot X_t. \end{aligned}$$

Theorem 2.4. *Let X be a semimartingale. Then $J_X : \mathbf{S}_{\text{ucp}} \to \mathbf{D}_{\text{ucp}}$ is continuous.*

Proof. Suppose $H^k \in \mathbf{S}$ tends to H uniformly. By linearity, we can suppose without loss H^k tends to 0. Let $T^k = \inf\{t : |(H^k \cdot X)_t| \geq \delta\}$. Then $H^k 1_{[0,T^k]} \in \mathbf{S}$ tends to 0 uniformly as k tends to ∞. Thus for every t

$$\begin{aligned} P\{(H^k \cdot X)_t^* > \delta\} &\leq P\{|H^k \cdot X_{T^k \wedge t}| \geq \delta\} \\ &= P\{|(H^k 1_{[0,T^k]} \cdot X)_t| \geq \delta\} \\ &= P\{|I_X(H^k 1_{[0,T^k \wedge t]})| \geq \delta\} \end{aligned}$$

which tends to 0 by definition because X is a semimartingale. Therefore $J_X : \mathbf{S}_u \to \mathbf{D}_{\text{ucp}}$ is continuous. We next show $J_X : \mathbf{S}_{\text{ucp}} \to \mathbf{D}_{\text{ucp}}$ is continuous. Let $\delta > 0$, $\varepsilon > 0$, $t > 0$. We now know there exists η such that $\|H\|_u \leq \eta$ implies $P(J_X(H)_t^* > \delta) < \varepsilon/2$. Let $R^k = \inf\{s : |H_s^k| > \eta\}$, and set $\widetilde{H}^k = H^k 1_{[0,R_k]} 1_{\{R_k > 0\}}$. Then $\widetilde{H}^k \in \mathbf{S}$ and $\|\widetilde{H}^k\|_u \leq \eta$ by left continuity. When $R^k \geq t$ we have $(\widetilde{H}^k \cdot X)_t^* = (H^k \cdot X)_t^*$, whence

$$\begin{aligned} P\left((H^k \cdot X)_t^* > \delta\right) &\leq P\left((\widetilde{H}^k \cdot X)_t^* \delta\right) + P(R^k < t) \\ &\leq \varepsilon/2 + P\left((H^k)_t^* > \eta\right) \\ &< \varepsilon, \end{aligned}$$

if k is large enough, since $\lim_{k\to\infty} P((H^k)_t^* > \eta) = 0$. $\square$

Definition 2.5. *Let X be a semimartingale. The continuous linear mapping $J_X : \mathbf{L}_{\mathrm{ucp}} \to \mathbf{D}_{\mathrm{ucp}}$ obtained as the extension of $J_X : \mathbf{S} \to \mathbf{D}$ is called the* stochastic integral.

Suppose H is a process in $\mathbf{D}$. We can write the stochastic integral $H_{s-} \cdot X = (\int_0^t H_{s-} dX_s)_{t \geq 0}$ as defined above, as a limit of sums. Let σ denote a finite sequence of stopping times:

$$0 = T_0 \leq T_1 \leq \ldots \leq T_k < \infty \text{ a.s.} \tag{2.2}$$

Such a sequence is called a *random partition*. A sequence of random partitions σ_n

$$\sigma_n : T_0^n \leq T_1^n \leq \ldots \leq T_{k_n}^n$$

is said to *tend to the identity* if

(i) $\lim_n \sup_i T_i^n = \infty$ a.s.
(ii) $\|\sigma_n\| = \sup_i |T_{i+1}^n - T_i^n|$ converges to 0 a.s.

For a process H and a random partition σ as in (2.2) we define

$$H^\sigma = H_0 1_{\{0\}} + \sum_{i=1}^k H_{T_i} 1_{(T_i, T_{i+1}]}. \tag{2.3}$$

Thus if H is in $\mathbf{L}$ or $\mathbf{D}$, we have

$$\int_0^t H_s^\sigma dX_s = H_0 X_0 + \sum_{i=1}^k H_{T_i}(X^{T_{i+1}} - X^{T_i}). \tag{2.4}$$

Theorem 2.6. *Let X be a semimartingale and let $H \in \mathbf{D}$. Let $(\sigma_n)_{n \geq 1}$ be a sequence of random partitions tending to the identity. Then*

$$H_- \cdot X = \lim_{n \to \infty} \sum_i H_{T_i^n}(X^{T_{i+1}^n} - X^{T_i^n})$$

with convergence in ucp.

Proof. Let $H^k \in \mathbf{S}$ converge to H in ucp. Then

$$(H_- - H^{\sigma_n}) \cdot X = (H_- - H^k) \cdot X + (H^k - (H_+^k)^{\sigma_n}) \cdot X + ((H_+^k)^{\sigma_n} - H^{\sigma_n}) \cdot X. \tag{2.5}$$

The first term on the right side of (2.5) equals $J_X(H_- - H^k)$, which goes to 0 because J_X is continuous on $\mathbf{L}_{\mathrm{ucp}}$. The same applies to the third term for fixed k as n tends to ∞. Indeed, $(H_+^k)^{\sigma_n} - H^{\sigma_n}$ tends to 0 as $k \to \infty$ uniformly in n. As for the middle term on the right side of (2.5), for fixed k it tends to 0 as n tends to ∞. Thus one need only choose k so large that the first and third terms are small, and then choose n so large that the middle term is small. □

Theorem 2.6 gives an appealing intuitive description of the stochastic integral as a limit of Riemann-type sums. Of course one can only do this because of the path regularity of the integrands.

Let us next note some simple and quite nice properties of the stochastic integral. H will be assumed to be in $\mathbf{D}$, and X a semimartingale in Theorems 2.7 through 2.11.

Theorem 2.7. *If X has paths of finite variation a.s., then $H_- \cdot X$ agrees with the Lebesgue-Stieltjes integral, denoted $\int_{LS} H_{s-} dX_s$.*

Proof. The result is evident for $H \in \mathbf{S}$. For $H \in \mathbf{D}$, let $H^n \in \mathbf{S}$ converge to H_- in ucp. Then there exists a subsequence n_k such that H^{n_k} converges uniformly on compacts a.s. to $H_- \cdot X$. Since the convergence is uniform, $\int_{LS} H_s^{n_k} dX_s$ converges as well to $\int_{LS} H_{s-} dX_s$, whence the result. □

Recall that for a process $Y \in \mathbf{D}$, $\Delta Y_t = Y_t - Y_{t-}$, and ΔY denotes the process $(\Delta Y_t)_{0 \leq t < \infty}$. An important feature of the stochastic integral is that the jumps behave "correctly" — that is, in the same manner as they do for the Lebesgue-Stieltjes integral. This is part of the reason we use $\mathbf{L}$, rather than, for example, $\mathbf{D}$, as our space of integrands. (See Pratelli [14] or Ahn-Protter [1] for more on this subject.)

Theorem 2.8. *The jump process $\Delta(H_- \cdot X)_s$ is indistinguishable[2] from the process $H_{s-} \Delta X_s$.*

Theorem 2.9. *Let $Q \ll P$. Then $H_{-Q} \cdot X$ is Q-indistinguishable from $H_{-P} \cdot X$.*

Theorem 2.10. *Let P and Q be any two probabilities and X a semimartingale for each. Then there exists $H_- \cdot X$ which is a version of both $H_{-P} \cdot X$ and $H_{-Q} \cdot X$.*

Theorem 2.11. *Let $\mathbf{G}$ be another filtration and suppose $H \in \mathbf{D}(\mathbf{G}) \cap \mathbf{D}(\mathbf{F})$, and that X is semimartingale for both $\mathbf{F}$ and $\mathbf{G}$. Then $H_{-\mathbf{G}} \cdot X = H_{-\mathbf{F}} \cdot X$.*

Proof. For Theorem 2.8 and 2.9, the result is clear for $H \in \mathbf{S}$ and follows for H_- with $H \in \mathbf{D}$ by taking limits in ucp (convergence in P-probability implies convergence in Q-probability). For Theorem 2.10, let $R = \frac{1}{2}(P+Q)$, and apply Theorem 2.9. For Theorem 2.11, we can use the construction in the proof of Theorem 2.2 to approximate $H \in \mathbf{D}$ constructively from H; thus the approximations $H^n \in \mathbf{S}$ are in $\mathbf{S}(\mathbf{F}) \cap \mathbf{S}(\mathbf{G})$; the result is clearly true for H in $\mathbf{S}$ and thus it follows by again taking limits. □

Theorem 2.9 can be used to show that many global results also hold locally.

We give an example.

[2] Two processes Y and Z are *indistinguishable* if $P\{\omega : t \to Y_t(\omega) \neq t \to Z_t(\omega)\} = 0$.

Theorem 2.12. *Let X, Y be two semimartingales and H, J be two processes in* $\mathbf{D}$. *Let*

$$A = \{\omega : H.(\omega) = J.(\omega) \text{ and } X.(\omega) = Y.(\omega)\}$$

where $H.(\omega)$ denotes the path of $H : t \to H_t(\omega)$. Let

$$B = \{\omega : X.(\omega) \text{ is finite variation on compacts}\}.$$

Then $H_- \cdot X = J_- \cdot X$ on A a.s., and $H_- \cdot X = \int_{LS} H_{s-} dX$ on B a.s.

Proof. Without loss of generality assume $P(A) > 0$. Define a new Q by $Q(\Lambda) = P(\Lambda|A)$. Then $H_- = J_-$ and $X = Y$ under Q. Note that X and Y are also semimartingales under Q. Thus $H_{-Q} \cdot X = H_{-P} \cdot X$, and one need only apply Theorem 2.9. The second assertion is a combination of the above idea with Theorems 2.7 and 2.9. □

The next result is quite important.

Theorem 2.13. *Let $H \in \mathbf{D}$ and X be a semimartingale. Then $Y = H_- \cdot X$ is again a semimartingale. Moreover if $G \in \mathbf{D}$ as well, then*

$$G_- \cdot Y = G_- \cdot (H_- \cdot X) = (GH)_- \cdot X.$$

Proof. If $G, H \in \mathbf{S}$, then clearly $Y = H_- \cdot X$ is a semimartingale, and $J_Y(G) = J_X(GH)$. The associativity property extends to H_-, G_- with $G, H \in \mathbf{D}$ by continuity. Therefore it remains only to show $Y = H_- \cdot X$ is a semimartingale. By taking subsequences if necessary, assume $H^n \in \mathbf{S}$ converges to H_- in ucp and also $H^n \cdot X$ converges a.s. to $H_- \cdot X$. For $G \in \mathbf{S}$, $J_Y(G)$ is defined for any process Y and hence makes sense *a priori*. Thus

$$\begin{aligned} J_Y(G) &= \lim_{n\to\infty} G \cdot Y^n = \lim_{n\to\infty} G \cdot (H^n \cdot X) \\ &= \lim_{n\to\infty} (GH^n) \cdot X = J_X(GH_-), \end{aligned}$$

since X is a semimartingale. Next let G^n converge to G in $\mathbf{S}_u$. We wish to show $I_Y(G^n)$ converges to $I_Y(G)$. But

$$\lim_{n\to\infty} J_Y(G^n) = \lim_{n\to\infty} J_X(G^n H_-) = J_X(GH_-)$$

since $G^n H_-$ converges to GH_- in ucp. Then since $J_X(GH_-) = J_Y(G)$ we have the result. □

3. Quadratic Variation

A process which plays a key role in the theory of stochastic integration is the quadratic variation process. We define it using stochastic integration:

Definition 3.1. *Let X be a semimartingale. The* quadratic variation process, *$[X,X]$, is defined to be*

$$[X,X]_t = X_t^2 - 2\int_0^t X_{s-}dX_s. \tag{3.1}$$

If X and Y are two semimartingales, the quadratic covariation process *is defined to be*

$$[X,Y]_t = X_tY_t - \int_0^t X_{s-}dY_s - \int_0^t Y_{s-}dX_s. \tag{3.2}$$

Note that if X is of finite variation, then (3.1) is simply integration by parts, and if X is also continuous then $[X,X]_t = X_0^2$, and in particular it is constant. Note further that the bracket $[\cdot,\cdot]$ satisfies a polarization identity:

$$[X,Y] = \frac{1}{2}\{[X+Y,X+Y] - [X,X] - [Y,Y]\}.$$

We make the convention that $X_{0-} = 0$ a.s. always.

Theorem 3.2. *Let X be a semimartingale. Then $[X,X]$ is in* **D** *and has non-decreasing paths. Moreover $[X,X]_0 = X_0^2$ and*

(i) $\Delta[X,X] = (\Delta X)^2$;

(ii) If σ_n is a sequence of random partitions tending to the identity as defined in (2.2), then

$$\lim_{n\to\infty}\left\{X_0^2 + \sum_i (X^{T_{i+1}^n} - X^{T_i^n})^2\right\} = [X,X]$$

with convergence in ucp;

(iii) for a stopping time T, $[X^T,X] = [X,X^T] = [X^T,X^T] = [X,X]^T$.

Proof. $[X,X]$ is in **D** since the right side of (3.1) is in **D**. It is nondecreasing a consequence of (ii) above. That (i) holds follows from (3.1) and Theorem 2.8. Property (ii) is an elementary consequence of Theorem 2.6. Finally (iii) follows easily from (ii). □

Theorem 3.2 gives a method of extending the notion of quadratic variation to a wider class of processes than semimartingales; namely, those for which a limit of sums exists in ucp, as given in Theorem 3.2 (ii). This would include, for example, the Dirichlet processes.

It is worthwhile to calculate the quadratic variation of some basic processes. Theorem 3.2 (ii) allows one to deduce that $[W,W]_t = t$ a.s., where W is standard Wiener process. If A is of finite variation, again Theorem 3.2 (ii) allows one to conclude that $[A,A]_t = \sum_{0<s\le t}(\Delta A_s)^2$. In particular if N is the Poisson process then $[N,N]_t = N_t$. If A is continuous and of finite variation, $[A,A]_t = A_0^2$, and thus if $A_0 = 0$ then $[A,A] \equiv 0$.

The quadratic variation process has a particularly nice property with respect to stochastic integrals:

Theorem 3.3. *Let X and Y be two semimartingales, and let $H, K \in \mathbf{D}$. Then*

$$[H_- \cdot X, K_- \cdot Y]_t = \int_0^t H_{s-}K_{s-}d[X,Y]_s. \tag{3.3}$$

In particular,

$$[H_- \cdot X, H_- \cdot X]_t = \int_0^t (H_{s-})^2 d[X,X]_s.$$

Proof. Without loss we assume $X_0 = Y_0 = 0$. By symmetry it suffices to prove

$$[H_- \cdot X, Y]_t = \int_0^t H_{s-}d[X,Y]_s. \tag{3.4}$$

First assume that $H = 1_{[0,T]}$. Then (3.4) follows from Theorem 3.2 (iii). Next let $H = V1_{(S,T]}$ where X and T are stopping times and $V \in \mathcal{F}_S$. Then $H \cdot X = V(X^T - X^S)$, and by Theorem 3.2

$$\begin{aligned}[H \cdot X, Y] &= V\{[X^T, Y] - [X^S, Y]\} \\ &= V\{[X,Y]^T - [X,Y]^S\} = \int H_s d[X,Y]_s.\end{aligned}$$

The result now holds for $H \in \mathbf{S}$ by linearity. For $H \in \mathbf{D}$, let $H^n \in \mathbf{S}$ approximate H_- in ucp. Let $Z^n = H^n \cdot X$. Then $[Z^n, Y] = \int H^n_s d[X,Y]_s$, and since $H^n \in \mathbf{S}$ we have:

$$\begin{aligned}[Z^n, Y] &= YZ^n - \int Y_- dZ^n - \int Z^n_- dY \\ &= YZ^n - \int Y_- H^n dX - \int Z^n_- dY\end{aligned}$$

which converges to

$$YZ - \int Y_- H_- dX - \int Z_- dY = YZ - \int Y_- dZ - \int Z_- dY = [Z,Y].$$

Thus,

$$[Z,Y] = \lim_{n\to\infty}[Z^n, Y] = \lim_{n\to\infty}\int H^n_s d[X,Y]_s = \int H_{s-}d[X,Y]_s.$$

But $Z = \lim_{n\to\infty} Z^n = \lim_{n\to\infty} H^n \cdot X = H_- \cdot X$, and we have the result. □

An important special case is that of martingales. If M is a martingale and $E\{\sup_{s\le t}|M_s|^2\} < \infty$, then $E\{M_t^2\} = E\{[M,M]_t\}$. Therefore Doob's maximal quadratic inequality for martingales can be expressed as follows (see [15]):

Theorem 3.4. *Let M be a local martingale. Then*

$$E\{\sup_{s\le t}(M_s)^2\} \le 4E\{[M,M]_t\}.$$

In particular if $E\{[M,M]_t\} < \infty$, then M is a square integrable martingale on $[0,t]$.

Let W be a standard Wiener process, and let $H \in \mathbf{D}$. Then as we previously remarked $[W,W]_t = t$. Hence

$$[H_- \cdot W, H_- \cdot W]_t = \int_0^t H_{s-}^2 ds = \int_0^t H_s^2 ds.$$

Therefore,

$$E\left\{\left(\int_0^t H_{s-} dW_s\right)^2\right\} = \int_0^t E\{H_s^2\} ds,$$

by Fubini's theorem. It is this isometry that K. Itô used when he originally defined the stochastic integral for the Wiener process.

4. Change of Variables

The change of variables formula in the general case (that is, the case for semimartingales with jumps) often looks strange, but actually it is close to the formula for Lebesgue-Stieltjes integration. The problem is that the latter formula is not well known. Of course it is a corollary of the general formula, but we nevertheless state it first.

Theorem 4.1. *Let V be a process with càdlàg paths of finite variation on compacts, and let f be C^1. Then*

$$f(V_t) - f(V_0) = \int_{0+}^t f'(V_{s-}) dV_s + \sum_{0<s\leq t} \{f(V_s) - f(V_{s-}) - f'(V_{s-})\Delta V_s\}$$

and in particular $f(V)$ is again a process with paths of finite variation on compacts.

Note that it is not *a priori* obvious that the infinite sum above converges. Before we state the general theorem recall that for a semimartingale X the process $[X,X]$ is in $\mathbf{D}$ and is non-decreasing. Therefore ω by ω the paths $t \to [X,X]_t(\omega)$ have a Lebesgue decomposition into a continuous part and a pure jump part. Indeed, in light of Theorem 3.2 we can write

$$[X,X]_t = [X,X]_t^c + \sum_{0<s\leq t} (\Delta X_s)^2,$$

and we call $[X,X]^c$ the *continuous part of the quadratic variation.* Note that we also have for semimartingales X, Y:

$$[X,Y]_t = [X,Y]_t^c + \sum_{0<s\leq t} \Delta X_s \Delta Y_s,$$

and in particular we deduce $\sum_{0<s\leq t} (\Delta X_s)^2 < \infty$ a.s. for any semimartingale X. (It is of course not true in general that $\sum_{0<s\leq t} |\Delta X_s|$ is finite a.s.; see example (vi) where such a term is ∞ a.s., each $t > 0$.)

Theorem 4.2 (Change of Variables). *Let $X = (X^1, \dots, X^d)$ be a d-dimensional semimartingale, and let $f : \mathbf{R}^d \to \mathbf{R}$ be C^2. Then $f(X)$ is a semimartingale and moreover*

$$\begin{aligned} f(X_t) - f(X_0) \quad &= \quad \sum_{i=1}^{d} \int_{0+}^{t} \frac{\partial f}{\partial x_i}(X_{s-})dX_s^i \\ &\quad + \frac{1}{2} \sum_{1 \le i,j \le d} \int_{-+}^{t} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{s-})d[X^i, X^j]_s^c \\ &\quad + \sum_{0<s\le t} \left\{ f(X_s) - f(X_{s-}) - \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(X_{s-})\Delta X_s^i \right\}. \end{aligned} \tag{4.1}$$

Proof. We give the proof for $d = 1$; the case for $d > 1$ is analogous but messier. Thus we want to establish:

$$\begin{aligned} f(X_t) - f(X_0) \quad &= \int_{0+}^{t} f'(X_{s-})dX_s + \frac{1}{2}\int_{0}^{t} f''(X_{s-})d[X,X]_s^c \\ &\quad + \sum_{0<s\le t} \{f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\}. \end{aligned} \tag{4.2}$$

We further assume $X_0 = 0$, to eliminate the plus symbols. First suppose f is a polynomial on $\mathbf{R}$. Obviously (4.2) holds for f a constant function. We will use induction and thus it suffices to prove the following: let g be such that $g(X)$ is a semimartingale and that (4.2) holds; then if $f(x) = xg(x)$, also $f(X)$ is a semimartingale and (4.2) holds for f.

Note that the product of two semimartingales is a semimartingale by integration by parts (formula (3.2) and Theorem 3.2), thus $Xg(X)$ is a semimartingale. Again using integration by parts (formula (3.2)) we have

$$f(X_t) = f(X_0) + \int_0^t X_{s-}dg(X_s) + \int_0^t g(X_{s-})dX_s + [X, g(X)]_t.$$

By hypothesis $g'(X)$ satisfies (4.2), hence by Theorems 2.8 and 2.13 we obtain:

$$\begin{aligned} f(X_t) \quad &= f(X_0) + \int_0^t X_{s-}g'(X_{s-})dX_s + \frac{1}{2}\int_0^t X_{s-}g''(X_{s-})d[X,X]_s^c \\ &\quad + \sum_{0<s\le t} X_{s-}\{g(X_s) - g(X_{s-}) - g'(X_{s-})\Delta X_s\} \\ &\quad + \int_0^t g(X_{s-})dX_s + [X, g(X)]_t. \end{aligned} \tag{4.3}$$

Next, using Theorem 3.3 we see that

$$\begin{aligned}
[X, g(X)]_t &= \int_0^t g(X_{s-})d[X,X]_s \\
&\quad + \sum_{0<s\le t} (\Delta X_s)\{g(X_s) - g(X_{s-}) - g'(X_{s-})\Delta X_s\} \\
&= \int_0^t g(X_{s-})d[X,X]_s^c + \sum_{0<s\le t} \Delta X_s\{g(X_s) - g(X_{s-})\}.
\end{aligned} \tag{4.4}$$

Combining (4.3) and (4.4) we see that f satisfies (4.2).

Now we consider the case where f is not a polynomial. Let

$$T_n = \inf\{t > 0 : |X_t| > n\}.$$

Then for each fixed n we can find a sequence $(g_{nm})_{m\ge 1}$ of polynomials that converge, together with their first and second derivatives, respectively to f and its first two derivatives, uniformly on $\{x : |x| \le n\}$. There exists a constant K_n such that for $|x|, |y| \le n$,

$$|f(x) - f(y) - f'(y)(x-y)| \le K_n|x-y|^2 \tag{4.5}$$

and

$$|g_{nm}(x) - g_{nm}(y) - g'_{nm}(y)(x-y)| \le K_n|x-y|^2. \tag{4.6}$$

Recall we remarked just before stating this theorem that $\sum_{s\le t}(\Delta X_s)^2 < \infty$ a.s. for any semimartingale X; therefore using (4.5) and (4.6) and taking limits as m increases to ∞ we deduce that for $t < T_n$:

$$\sum_{0<s\le t} |f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s| < \infty \text{ a.s.}$$

and moreover

$$\begin{aligned}
\lim_{m\to\infty} &\sum_{0<s\le t} \{g_{nm}(X_s) - g_{nm}(X_{s-}) - g'_{nm}(X_{s-})\Delta X_s\} \\
&= \sum_{0<s\le t} \{f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\}.
\end{aligned}$$

Furthermore $\lim_{m\to\infty} g_{nm}(X_t) = f(X_t)$, for $t < T_n$. Note further that $\int_0^t g'_{nm}(X_{s-})dX_s$ tends to $\int_0^t f'(X_{s-})dX_s$ since J_X is continuous in ucp on L, and also $\int_0^t g''_{nm}(X_{s-})d[X,X]_s^c$ converges in ucp to $\int_0^t f''(X_{s-})d[X,X]_s^c$.

Since T_n increases to ∞ a.s., the process $\sum_{0<s\le t}\{f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\}$ is of finite variation (and thus absolutely convergent as a series a.s. for $t > 0$) on compacts. Since the other terms on the right side of (4.2) are all well defined semimartingales, we conclude that $f(X)$ is a semimartingale and that (4.2) indeed holds. □

We remark that the preceding proof, while quick, simple, and elegant, is not particularly intuitive. A more intuitive proof, using Taylor expansions, can be found for example in [15, pp. 71ff].

If $X = (W^1, \dots, W^d)$ is a d-dimensional Wiener process, then W^i is independent of W^j for $i \neq j$, and one can check that $[W^i, W^j] \equiv 0$ for $i \neq j$ (we assume $W_0 = 0$). In this case for $f \in C^2$ we can write the change of variables formula (known here as *Itô's formula*) in the form

$$f(W_t) - f(W_0) = \int_0^t \nabla f(W_s) \cdot dW_s + \frac{1}{2} \int_0^t \Delta f(W_s) ds.$$

In particular if $\Delta f = 0$, then $f(W)$ is a local martingale.

Also note that if $X = (X^1, \dots, X^d)$ is such that some of the components of X are finite variation processes, then f need only to be C^1 for the corresponding coordinates.

5. Stochastic Differential Equations

We consider here fairly general stochastic differential equations which are sufficient for most applications. Let $\mathbf{D}^d$ denote d-dimensional vectors of processes in $\mathbf{D}$.

Definition 5.1. *An operator $F : \mathbf{D}^d \to \mathbf{D}$ is said to be* functional Lipschitz *if for any processes X, Y in $\mathbf{D}^d$ we have*

(i) for any stopping time T, $X^{T-} = Y^{T-}$ implies $F(X)^{T-} = F(Y)^{T-}$.[3]
(ii) $|F(X)_t - F(Y)_t| \leq M \sup_{s \leq t} |X_s - Y_s|$.

The following theorem is a special case of a more general result to be proved in Section 7 of Part II of these notes (see also [15]).

Theorem 5.2. *Let $Y = (Y^1, \dots, Y^d)$ be a vector of semimartingales and let J^i, $1 \leq i \leq k$, be processes in $\mathbf{D}$. Let F_j^i, $1 \leq i \leq k$, $1 \leq j \leq d$ be functional Lipschitz operators. Then the system of equations*

$$X_t^i = J_t^i + \sum_{i=1}^{d} \int_0^t F_j^i(X)_{s-} dZ_s^j \tag{5.1}$$

has a solution in $\mathbf{D}$, and it is unique. Moreover if the processes J^i are semimartingales, then X^i are semimartingales as well.

The reader may wonder if the condition $F(X)_{s-}$, instead of $F(X)_s$ is merely a technicality to ensure that the integrand of the stochastic integral is in $\mathbf{L}$. It is not, but rather is essential if one considers driving terms with jumps, and it corresponds to one's physical intuition: a jump at time t "kicks" the process according to where it was just before t. Indeed, if one takes the

[3] The notation X^{T-} denotes $X_t 1_{\{t<T\}} + X_{T-} 1_{\{t \geq T\}}$.

non-random example where $Z_t = 1_{\{t \geq 2\}}$ and $X_t = 1 + \int_0^t X_s dZ_s$, then one has $X_t = 1$ for $0 \leq t < 2$, and $X_2 = 1 + X_2$, which gives $0 = 1$.

A particularly important special case of (5.1) is the exponential equation:

$$X_t = X_0 + \int_0^t \alpha X_{s-} dY_s. \tag{5.2}$$

One can use the change of variables formula to give an explicit solution of (5.2), called the ***stochastic exponential*** and denoted $\mathcal{E}(Y)$:

$$\mathcal{E}(Y)_t = \exp\left(\alpha Y_t - \frac{\alpha^2}{2}[Y,Y]_t^c\right) \prod_{0<s\leq t} (1 + \alpha \Delta Y_s) \exp(-\alpha \Delta Y_s). \tag{5.3}$$

The stochastic exponential has behavior similar to a true exponential, but of course slightly different due to the semimartingale calculus; for example we have the following pretty result:

Theorem 5.3 (Yor). *Let X and Y be semimartingales with $X_0 = Y_0 = 0$. Then $\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X + Y + [X,Y])$.*

Proof. Let $U_t = \mathcal{E}(X)_t$ and $V_t = \mathcal{E}(Y)_t$. By integration by parts $U_t V_t - 1 = \int_{0+}^t U_{s-} dV_s + \int_{0+}^t V_{s-} dU_s + [U,V]_t$. Using that U and V are exponentials and letting $W = UV$ this becomes

$$W_t = 1 + \int_0^t W_{s-} d(X + Y + [X,Y])_s,$$

whence the result. □

Using a variation of constants technique we can generalize the stochastic exponential results.

Theorem 5.4. *Let H and Z be semimartingales and assume $P\{\Delta Z_t \neq -1, t \geq 0\} = 1$. Let X be the unique solution of*

$$X_t = H_t + \int_0^t X_{s-} dZ_s.$$

Then $X_t = \mathcal{E}_H(Z)_t$ has the form:

$$\mathcal{E}_H(Z)_t = \mathcal{E}(Z)_t \left\{ H_0 + \int_{0+}^t \mathcal{E}(Z)_{s-}^{-1} d\left\{ H_s - [H,Z]_s - \sum_{0<u\leq s} \left(\frac{\Delta H_u (\Delta Z_u)^2}{1 + \Delta Z_u} \right) \right\} \right\}.$$

Proof. Let us assume the solution is of the form $C_t \mathcal{E}(Z)_t$. Let $U_t = \mathcal{E}(Z)_t$, and we wish to determine C. Note that

$$\Delta X_t = \Delta H_t + X_{t-} \Delta Z_t. \tag{5.4}$$

Integration by parts yields:

$$\begin{aligned} dX_t &= C_t - dU_t + U_{t-}dC_t + d[C,U]_t \\ &= C_{t-}U_{t-}dZ_t + U_{t-}d\{C_t + [C,Z]_t\} \end{aligned} \tag{5.5}$$

from which we deduce

$$\Delta X_t = X_{t-}\Delta Z_t + U_{t-}\Delta C_t + U_{t-}\Delta C_t \Delta Z_t. \tag{5.6}$$

Combining (5.4) and (5.5) yields

$$\Delta C_t = \frac{\Delta H_t}{U_{t-}(1+\Delta Z_t)} = \frac{\Delta H_t}{\mathcal{E}(Z)_{t-}(1+\Delta Z_t)}. \tag{5.7}$$

¿From (5.6) we have

$$dH_t = U_{t-}d\{C_t + [C,Z]_t\}$$

which implies

$$\begin{aligned} \frac{1}{U_-}\cdot H &= C + [C,Z], \\ \left[\frac{1}{U_-}\cdot H, Z\right] &= [C,Z]_t + [\,[C,Z],Z]_t\,, \end{aligned} \tag{5.8}$$

and since $[\,[C,Z],Z] = \sum \Delta C(\Delta Z)^2$, and since we know ΔC by (5.7), we obtain

$$[C,Z] = \frac{1}{U_-}\cdot[H,Z] - \sum \frac{\Delta H(\Delta Z)^2}{U_-(1+\Delta Z)}. \tag{5.9}$$

Using (5.8) and (5.9) we get:

$$C_t = \int_0^t \frac{1}{U_{s-}}dH_s - d[C,Z]_t = \int_0^t \frac{1}{U_{s-}}d\left\{H_s - [H,Z]_s - \sum_{0<u\leq s} \frac{\Delta H_u(\Delta Z_u)^2}{(1+\Delta Z_u)}\right\},$$

and the result follows. □

6. The Skorohod Topology and Weak Convergence

In this section we recall the essentials of the Skorohod topology and weak convergence. Since this material is by now classic, we omit the proofs except for Theorem 6.5 which is recent and the reader can consult any of several expository treatments both in books and research articles.

Recall that **D** has been used to denote adapted, càdlàg stochastic processes.

We now let $D = D_{\mathbf{R}^d} = D_{\mathbf{R}^d}[0,\infty)$ denote the space of càdlàg *functions* from $[0,\infty)$ to $\mathbf{R}^d$. A process in **D** has almost all of its sample paths in D. We wish to endow D with a topology for which it is a complete separable metric space. A natural candidate would be *uc* (uniform convergence on compacts), and the corresponding metric would be:

$$d_{uc}(x,y) = \sum_{n=1}^{\infty} \frac{1}{2^n}\left(\min\left(1, \sup_{s\le n}|x(s)-y(s)|\right)\right).$$

Such a topology, however, is not separable: for example, the family of functions $x_s(t) = 1_{[s,\infty)}(t)$, $0 \le s < 1$, is uncountable, while $d_{uc}(x_s, x_u) = 1$ for $s \neq u$.

Let $\mathbf{R}_+$ denote $[0,\infty)$.

Definition 6.1. *A* time change function λ *is an increasing, bijective function from* $\mathbf{R}_+$ *to* $\mathbf{R}_+$. *We let* Γ *denote the class of these functions.*

Definition 6.2. *A sequence of functions* $x_n \in D$ *converges in the* Skorohod topology *to* $x \in D$ *if there exists* $\lambda_n \in \Gamma$ *such that* $\lambda_n(t)$ *converges to* $\lambda(t) = t$ *uniformly and* $x_n(\lambda_n(t))$ *converges to* $x(t)$ *uniformly on compacts.*

In the *uc* topology, if x_n converges to x, then for large enough n the jumps of s_n must occur at the same time as those of x, and of course the sizes must also converge.

With the Skorohod topology the sizes of the jumps must still converge, but the jumps need not occur at the same time. The Skorohod topology also allows the *times of occurrence* of the jumps to converge. Note that if the limit process x is continuous, then $x_n \to x$ in the Skorohod topology if and only if it converges in *uc*.

Theorem 6.3. *The Skorohod topology is metrizable, and the resulting metric space is separable and complete.*

To prove Theorem 6.3 one can construct a compatible metric. A metric analogous to the *uc* metric is:

$$d^1(x,y) = \inf_{\lambda\in\Gamma} d^1_\lambda(x,y)$$

where

$$d^1_\lambda(x,y) = \sup_{t\ge 0}|\lambda(t)-t| + \sum_{n=1}^{\infty}\frac{1}{2^n}\left(1 \wedge \sup_{t\ge 0}|x(n\wedge\lambda(t)) - y(n\wedge t)|\right). \quad (6.1)$$

The metric d^1 is a compatible metric, but it is not complete.

We give two other compatible metrics which are in fact complete. We define Γ' to be the set of Lipschitz continuous functions $\lambda \in \Gamma$ such that

$$\gamma(\lambda) = \operatorname*{ess\,sup}_{t\ge 0} |\log \lambda'(t)| = \sup_{0\le s<t}\left|\log\frac{\lambda(t)-\lambda(s)}{t-s}\right| < \infty.$$

Next define $d^2(x,y,\lambda,u) = \sup_{t\ge 0}|x(\lambda(t)\wedge u) - y(t\wedge u)|$. Finally we can define

$$d^2(x,y) = \inf_{\lambda\in\Gamma'}\left\{\gamma(\lambda) \vee \int_0^\infty e^{-u} d(x,y,\lambda,u)\,du\right\}.$$

For the third metric, we define

$$k_n(t) = \begin{cases} 1 & \text{if } t \leq n, \\ n+1-t & \text{if } n < t < n+1, \\ 0 & \text{if } t \geq n+1 \end{cases},$$

$$d^3(x,y,n) = \inf_{\lambda \in \Gamma'} (\gamma(\lambda) + \|(k_n x) \circ \lambda - k_n y\|_{L^\infty}),$$

$$d^3(x,y) = \sum_{n \geq 0} 2^{-n}(1 \wedge d^3(x,y,n)).$$

The first distance d^1 is close to that of the original distance proposed by Skorohod. The second distance d^2 is taken from Ethier and Kurtz (1986) and is a modernized variation of Prokhorov's distance. The third distance d^3 is actually that of Prokhorov. Note that if x_n converges to x in the Skorohod topology and if the limit function x is continuous, then convergence in the Skorohod topology is equivalent to convergence in the *uc* topology.

While the Skorohod topology seems to be quite nice from the above description, it has a few traits which can create problems.

Example 6.4. Let $z_n(s) = x(s)1_{\{r_n \leq s\}} + y(s)1_{\{t_n \leq s\}}$, with $r_n < t_n$, and x and y continuous, not zero. Then z_n converges to a limit z in the Skorohod topology if and only if:

(i) $\lim_{n\to\infty} r_n = \infty$; whence $z = 0$;
(ii) $\lim_{n\to\infty} r_n = r < \infty$; $\lim_{n\to\infty} t_n = \infty$; then $z(s) = x(s)1_{\{t \leq s\}}$;
(iii) $\lim_{n\to\infty} r_n = r < \infty$; $\lim_{n\to\infty} t_n = t < \infty$; and $r < t$, then $z(s) = x(s)1_{\{t \leq s\}} + y(s)1_{\{t \leq s\}}$.

From Example 6.4 two important properties are clear:

> It can happen that $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$, but $\lim_{n\to\infty}(x_n + y_n) \neq x + y$. (Note that if y is continuous, then the above does hold.) Thus D with the Skorohod topology *is not a topological vector space.*
>
> $D(\mathbf{R}^d) \neq \prod_{i=1}^d D(\mathbf{R})$ in the sense of Cartesian products as topological spaces. Indeed, the topology of $D(\mathbf{R}^d)$ *is finer* than the product topology $D(\mathbf{R})^d$.

We will say that a subset A of D is *relatively compact* if it has compact closure. Note that if A is relatively compact then every sequence has a convergent subsequence (in the Skorohod topology, of course).

We now wish to pass from convergence in the space of functions to convergence of stochastic processes. There is a minor problem to make this procedure measurable. We have the following result. (See [7] for more results of this type.)

Theorem 6.5. *Let X_n and X be E-valued stochastic processes, where E is a Polish space, and suppose $\lim_{n\to\infty} X_n = X$ a.s. in the Skorohod topology. Then there exists a sequence $(\Lambda_n)_{n\geq 1}$ of measurable processes with paths in Γ such that $\lim_{n\to\infty} d_{\Lambda_n}(X_n, X) = 0$ a.s.*

Proof. Recall $d^1_\lambda(x,y)$ is defined in (6.1).

$$U_n = \{(\omega,\lambda) \in \Omega \times \Gamma : d^1_\lambda(X_n(\omega), X(\omega)) \leq d^1(X_n(\omega), X(\omega)) + 2^{-n}\}.$$

Denote by $\mathcal{D}$ and $\mathcal{G}$ the Borel fields of D and Γ respectively, where Γ is endowed with the uniform topology. Then $(x,\lambda) \to x \circ \lambda$ is Borel from $D \times \Gamma$ into D, and $(x,y) \to d^1(x(t), y(t); t \leq n)$ is Borel from $D \times D$ into $\mathbf{R}$. Since X_n and X are measurable from $(\Omega, \mathcal{F})$ into $(D, \mathcal{D})$, by composition we have $U_n \in \mathcal{F} \otimes \mathcal{G}$.

The projection $\pi_n(U_n) = \{\omega : \exists \lambda \in \Gamma \text{ with } (\omega,\lambda) \in U_n\}$ is equal to all of Ω. By the measurable section theorem (cf, e.g., [2, p. 18], there exists a random variable Λ_n with values in $(\Gamma, \mathcal{G})$ such that $P\{\omega : (\omega, \Lambda_n(\omega)) \in U_n\} = 1$. Since $d^1(x,y) = \inf_{\lambda\in\Gamma} d^1_\lambda(x,y)$, we have $d^1_{\Lambda_n}(X_n, X) \leq d^1(X_n, X) + 2^{-n}$ a.s., whence the result. □

We remark that one can improve upon this result to obtain a sure result (instead of "almost sure"); the proof is complicated and uses a measurable selection theorem (see, e.g., [7]).

Let us now turn to *weak convergence*. For a given Polish space E (for our purposes one can think of E as a complete, separable metric space), let $\mathcal{P}(E)$ denote the space of all probability measures on $(E, \mathcal{E})$. We endow $\mathcal{P}(E)$ with the *weak topology:* this is the smallest topology making all the mappings

$$f \to \int f \, d\mu$$

continuous for all *bounded continuous functions* f defined on E. We have that $\mathcal{P}(E)$ is also a Polish space for this topology. We have the following elementary properties:

Theorem 6.6. *Let E, E' be Polish spaces. Suppose μ_n, $\mu \in \mathcal{P}(E)$ and μ_n converges to μ weakly. Then*

(i) if F is a closed subset of E then $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$;

(ii) if f is a bounded function on E that is $\mu-$ a.s. continuous then $\lim_{n\to\infty} \mu_n(f) = \mu(f)$;

(iii) if $h : E \to E'$ then $\mu \to \mu \circ h^{-1}$ is continuous from $\mathcal{P}(E)$ to $\mathcal{P}(E')$ at each point μ such that h is $\mu-$ a.s. continuous.

Definition 6.7. *A subset A of $\mathcal{P}(E)$ is called* "tight" *if for every $\varepsilon > 0$ there exists a compact subset K of E such that $\mu(K^c) \leq \varepsilon$ for all $\mu \in A$.*

Perhaps the most important result in weak convergence is the following:

Theorem 6.8 (Prokhorov). *A subset A of $\mathcal{P}(E)$ is relatively compact for the weak topology if and only if it is tight.*

We now wish to consider the *weak convergence of stochastic processes.* Let X_n, X be $\mathbf{R}^d$-valued stochastic processes with paths in D. That is, X_n and X have càdlàg paths. (One could replace $\mathbf{R}^d$ with a Polish space E if desired.) Two obvious ways X_n could converge to X are:

$$X^n(\omega) \to X(\omega) \text{ in } D \text{ for the Skorohod topology for all } \omega; \quad (6.2)$$

$$X^n \to X \text{ in } D \text{ for the Skorohod topology almost surely.} \quad (6.3)$$

We wish to consider a third way, namely,

$$E\{f(X^n)\} \to E\{f(X)\} \text{ for all bounded Skorohod continuous functions } f. \quad (6.4)$$

Note that if we let μ_n, μ be the distributions respectively of X^n, X, then (6.4) is the same as

$$\int f\, d\mu_n \to \int f\, d\mu \text{ for all bounded Skorohod continuous } f. \quad (6.5)$$

The third type ((6.4) above) will be called *the convergence in distribution of* X_n to X and it will be implicitly understood that *we are always using the Skorohod topology.* We denote $X^n \Rightarrow X$ to mean X^n converges in distribution to X.

Observe that for convergence types (6.2) and (6.3), X^n and X must all be defined on the same probability space, whereas for convergence in distribution (6.4) each X^n and X can be defined on a different space. Such a nuance is important for limit theorems, since the limit X may of necessity "live" on a strictly bigger space than the converging sequence X^n. On the other hand, combining Theorem 6.5 with the classical Skorohod representation theorem, one can prove the following:

Theorem 6.9. *Let $X^n \Rightarrow X$. Then there exists a probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P})$ such that there exists processes $(\widehat{X}^n)_{n\geq 1}$, $\widehat{X}$ defined on $\widehat{\Omega}$ with $\mathcal{L}(\widehat{X}^n) = \mathcal{L}(X^n)$; $\mathcal{L}(\widehat{X}) = \mathcal{L}(X)$; and furthermore there exists a sequence of measurable processes Λ_n with paths in Γ such that $\lim_{n\to\infty} d_{\Lambda_n}(X^n, \widehat{X}) = 0$, $\widehat{P}$ a.s.*[4]

(Recall that the metric $d^1_\lambda(x,y)$ is given in (6.1)).

The definition (6.4) of convergence in distribution is stated in terms of functions which are "continuous for the Skorohod topology". We can relate (6.4) to the more familiar continuous functions (for the uniform topology) using the ideas of Theorem 6.5 or Theorem 6.9:

Theorem 6.10. *$X^n \Rightarrow X$ if and only if there exists a sequence of measurable processes Λ^n with paths in Γ such that $\lim_{n\to\infty} E\{f(X^n_{\Lambda^n})\} = E\{f(X)\}$ for all bounded, continuous f (f continuous in the uniform topology).*

---

[4] $\mathcal{L}(X)$ denotes the law of X; that is, the distribution of X

Since we will be concerned with the convergence in distribution of càdlàg processes (and not probability measures), it is useful to reformulate Prokhorov's theorem in terms of them:

Definition 6.11. *A sequence $(X^n)_{n\geq 1}$ of stochastic processes with paths in D is said to be* relatively compact in distribution *if the sequence $\mathcal{L}(X^n)$ of its distribution measures is relatively compact.*

Note that Definition 6.11 says essentially that (X^n) is tight if there exists a compact subset K of D such that $P^n(X^n \notin K) \leq \varepsilon$ for all n.

Theorem 6.12 (Prokhorov). *Let $(X^n)_{n\geq 1}$ be a sequence of stochastic processes with paths in D. The sequence $(\mathcal{L}(X^n))_{n\geq 1}$ is relatively compact in $\mathcal{P}(D_E)$ if and only if the collection of distribution measures of $(X^n)_{n\geq 1}$ is tight.*

7. Weak Convergence of Stochastic Integrals

Let (H^n, X^n) be a sequence of processes in $\mathbf{D}$. If we assume X^n are semimartingales, each n, a natural question to pose — which is useful in many applications — is when do the stochastic integrals $\int H^n_{s-}dX^n_s$ converge, and to what do they converge? If $(H^n, X^n) \Rightarrow (H, X)$, it would be desirable to have sufficient conditions such that X is a semimartingale too and $\int H^n_{s-}dX^n_s \Rightarrow \int H_{s-}dX_s$. We will see that we have a surprisingly nice answer to this question.

Since we are dealing with weak convergence we may assume that each (H^n, X^n) is defined on its own space. Let $\Theta^n = (\Omega^n, \mathcal{F}^n, P^n, \mathbf{F}^n)$ where $\mathbf{F}^n = (\mathcal{F}^n_t)_{t\geq 0}$ is a filtration satisfying the usual hypothesis, each $n \geq 1$.

Before we make the next definition, that of goodness, we must clear up an important ambiguity. If (H^n, X^n) is a sequence of processes in $\mathbf{D}^2$ converging to (H, X) in $\mathbf{D}^2$ in the Skorohod topology, then they could be considered to converge either in $D_{\mathbf{R}^2}[0,\infty)$ or in $D_{\mathbf{R}}[0,\infty) \times D_{\mathbf{R}}[0,\infty)$. The former convergence is stronger: for $D_{\mathbf{R}^2}[0,\infty)$ we assume there is *one* sequence Λ_n of changes of time such that $(H^n_{\Lambda_n(t)}, X^n_{\Lambda_n(t)})$ converges uniformly to (H_t, X_t); in $D_{\mathbf{R}}[0,\infty) \times D_{\mathbf{R}}[0,\infty)$ there are *two* changes of time, Λ^1_n and Λ^2_n, such that $(H^n_{\Lambda^1_n(t)}, X^n_{\Lambda^2_n(t)})$ converges uniformly to (H_t, X_t). We will always use the stronger topology $D_{\mathbf{R}^2}[0,\infty)$. That is, *if we write $(H^n, X^n) \Rightarrow (H, X)$ it will be understood that convergence is in the topology $D_{\mathbf{R}^2}[0,\infty)$*, and thus one change of time Λ_n applies to both H^n and X^n. It turns out that this is the natural convergence to use for most applications (eg to stochastic differential equations), since often the jumps of H^n will be intimately related to those of X^n. In addition, the following example shows that the fundamental theorem of weak convergence of stochastic integrals (Theorem 7.10) fails if one takes convergence in $D_{\mathbf{R}}[0,\infty) \times D_{\mathbf{R}}[0,\infty)$.

Example 7.1. Let $X_t^n = X_t = 1_{\{t\geq 1\}}$ for all n, and let $H_t^n = 1_{\{t\geq 1+\frac{1}{n}\}}$. Then $\int_0^t H_{s-}^n dX_s^n = 1$ for $t > 1+\frac{1}{n}$, but the limiting integral $\int_0^t H_{s-}dX_s = 0$ for all t.

Caveat 7.2. To keep our notation simple we make the convention that when we say, for example, that (H^n, X^n), (H, X) are vector processes in $\mathbf{D}$ and $(H^n, X^n) \Rightarrow (H, X)$, we mean that X^n, X are d dimensional vectors of processes with each component a process in $\mathbf{D}$, and H^n is a $k \times d$ matrix of processes with each component in $\mathbf{D}$. The convergence is of course weak convergence in the Skorohod topology $D_{M^{kd}\times\mathbf{R}^d}[0,\infty)$, where M^{kd} denotes $k \times d$ real valued matrices.

Definition 7.3. *Let X^n be a sequence of $\mathbf{R}^d$-valued semimartingales on Θ^n, $n \geq 1$ and assume $X^n \Rightarrow X$. The sequence X^n is* good *if for any sequence $(H^n)_{n\geq 1}$ of $d \times k$ matrix processes in $\mathbf{D}$ defined on Θ^n such that $(H^n, X^n) \Rightarrow (H, X)$, then X is semimartingale and $\int H_{s-}^n dX_s^n \Rightarrow \int H_{s-}dX_s$.*

Observe that in Definition 7.3 we are implicitly assuming that the limit process X is a semimartingale on a space $(\Omega, \mathcal{F}, P, \mathbf{F})$ relative to which H is an *adapted,* càdlàg process. Thus $\mathbf{F}$ may be required to be a bigger filtration than the minimal one generated by X that satisfies the usual hypotheses.

Recall that X was defined to be a semimartingale if for H in $\mathbf{S}$, satisfying (1.2), and I_X defined by (1.3), we have $I_X : \mathbf{S}_u \to \mathbf{L}^0$ were continuous on compact time sets. In other words, if $H^n \in \mathbf{S}$ converged uniformly to $H \in \mathbf{S}$, then $\int H_s^n dX_s$ would converge in probability to $\int H_s dX_s$. The following analogous property for sequences was proposed by Jakubowski, Mémin, and Pagès [9]:

Definition 7.4. *A sequence of semimartingales $(X^n)_{n\geq 1}$, with X^n defined on Θ^n, is said to be* uniformly tight, *denoted* UT, *if for each $t > 0$, the set $\{\int_0^t H_{s-}^n dX_s^n, H^n \in \mathbf{S}^n, |H^n| \leq 1, n \geq 1\}$ is stochastically bounded (uniformly in n).*

In the above definition $\mathbf{S}^n$ denotes the simple predictable processes on Θ^n.

Definition 7.4 gives a theoretically compelling criterion, but it is perhaps not easy to verify in practice. We will give another criterion that is indeed easy to verify in practice and which turns out to be equivalent. A first step is to modify a semimartingale in such a way as to work with processes with *bounded jumps.* A standard procedure in the theory of stochastic integration is simply to subtract away the jumps bigger than a certain size: that is, if X is a given semimartingale, and $\delta > 0$ is given, let

$$X_t = \left\{ X_t - \sum_{0<s\leq t} \Delta X_s 1_{\{|\Delta X_s|>\delta\}} \right\} + \sum_{0<s\leq t} \Delta X_s 1_{\{|\Delta X_s|>\delta\}} \tag{7.1}$$

where of course $\Delta X_s = X_s - X_{s-}$. The sums converge since ω by ω they have only a finite number of terms before each $t > 0$, since X has càdlàg paths. The problem with this approach is that it is not a continuous operation for the Skorohod topology! We will instead propose a similar procedure which — while it is a bit more complicated — is indeed a Skorohod continuous procedure! Instead of removing the large jumps, we shrink them to be no larger than a specified $\delta > 0$. We define $h_\delta : \mathbf{R}_+ \to \mathbf{R}_+$ by $h_\delta(r) = (1-\delta/r)^+$, and $J_\delta : D(\mathbf{R}^d) \to D(\mathbf{R}^d)$ by

$$J_\delta(x)t = \sum_{0<s\leq t} h_\delta(|\Delta x_s|)\Delta x_s. \tag{7.2}$$

For a semimartingale X set $X^\delta = X - J_\delta(X)$, and analogously for a sequence $X^n : X^{n,\delta} = X^n - J_\delta(X^n)$. Then X^δ will have all of its jumps bounded by δ. A semimartingale with bounded jumps has many nice properties. The most important ones for us will be as follows. Let Y *be a semimartingale with jumps bounded by* $\delta > 0$; then we have:

Y is locally bounded; that is, there exist stopping times $(T^k)_{k\geq 1}$ increasing to ∞ a.s. such that $Y^{T^k} = (Y_{t\wedge T^k})_{t\geq 0}$ is bounded a.s.; (7.3)

$[Y,Y]$ is locally bounded; (7.4)

Y has a decomposition $Y = M + A$ where M is a local martingale and $A \in \mathbf{D}$ has paths of finite variation on compacts, and M and A both have bounded jumps (by, e.g., 2δ); (7.5)

The process A in (7.5) can be taken to be "natural" (see [15]), or equivalently, predictably measurable.[5] (7.6)

The process M and A in (7.5) are each locally bounded. Moreover the total variation process of A, denoted $\int |dA_s|$, is also locally bounded. (7.7)

Suppose X^n is a sequence of semimartingales and $\delta > 0$. We can form $X^{n,\delta}$ for each n and we then obtain decompositions such as (7.5) for each n : $X^{n,\delta} = M^{n,\delta} + A^{n,\delta}$. As in (7.4) and (7.7) there will exist stopping times $T^{n,k}$ increasing to ∞ a.s. in k such that $[M^{n,\delta}, M^{n,\delta}]$ and $\int |dA_s|$ are locally bounded; the next definition makes the dependence of each $T^{n,k}$ on k uniform in n; note that this is a little subtle, since each sequence $(T^{n,k})_{k\geq 1}$ is *a priori* defined on a different space Θ^n.

Definition 7.5. *A sequence of semimartingales* $(X^n)_{n\geq 1}$ *is said to have* uniformly controlled variations (UCV) *if there exists* $\delta > 0$, *and for each* $\alpha > 0$, $n \geq 1$, *there exist decompositions* $X^{n,\delta} = M^{n,\delta} + A^{n,\delta}$ *and stopping times* $T^{n,\alpha}$ *such that* $P(\{T^{n,\alpha} \leq \alpha\}) \leq \frac{1}{\alpha}$ *and furthermore*

[5] The predictable σ-algebra on $\Omega \times \mathbf{R}_+$ is $\mathcal{R} = \sigma(\mathbf{L})$.

$$\sup_n E^n\left\{[M^{n,\delta}, M^{n,\delta}]_{t\wedge T^{n,\alpha}} + \int_0^{t\wedge T^{n,\alpha}} |dA_s^{n,\delta}|\right\} < \infty. \tag{7.8}$$

Note that it is implicit in Definition 7.5 that each semimartingale X^n (and hence also $X^{n,\delta}$, $M^{n,\delta}$, $A^{n,\delta}$ and $T^{n,\alpha}$) can be defined on a different probability space Θ^n. Definition 7.5 is taken from [10].

Theorem 7.6. *Let $(X^n)_{n\geq}$ be a sequence of semimartingales, $X \in \mathbf{D}$, and suppose $X^n \Rightarrow X$. Then X^n satisfies UT if and only if it satisfies UCV.*

Proof. Suppose first $(X^n)_{n\geq 1}$ satisfies UT. By considering stopping times of the form $T = \inf\{t > 0 : |X_t^n| \geq c\}$, and then H of the form $H = 1_{[0,T]}(t)$ (which is in $\mathbf{S}^n$), it follows that $\{\sup_{s\leq t}|X_s^n|; n \geq 1\}$ is stochastically bounded. Using this and Theorem 3.2 (ii) which approximates $[X, X]$ as a limit of sums of squared increments of X in ucp), we see that $[X^n, X^n]$ is also stochastically bounded. Therefore, the number of jumps of X^n bigger than $\delta(n \geq 1)$ is stochastically bounded, whence $(X^{n,\delta})$ is stochastically bounded too. Apply the preceding again to deduce that $[X^{n,\delta}, X^{n,\delta}]$ is also stochastically bounded.

Let $X^{n,\delta} = M^{n,\delta} + A^{n,\delta}$ be the decomposition of $X^{n,\delta}$ where $A^{n,\delta}$ is taken to be natural (as mentioned in (7.6)). Given $\varepsilon > 0$, one can find K such that $P^n([X^{n,\delta}, X^{n,\delta}] > K) < \varepsilon$. Let $T^{n,k} = \inf\{t > 0 : [X^{n,\delta}, X^{n,\delta}]_t > K\} \wedge t$. Then $P(T^{n,k} < t) < \varepsilon$, and moreover

$$E\{[X^{n,\delta}, X^{n,\delta}]_{T^{n,k}}\} \leq K + 4\delta^2,$$

using $(a+b)^2 \leq 2a^2 + 2b^2$ and also (7.5). Since $A^{n,\delta}$ is natural one has that $E^n\{[M^{n,\delta}, A^{n,\delta}]_R\} = 0$ for stopping times R such that $M^{n,\delta}$ is bounded. Since $[X^{n,\delta}, X^{n,\delta}] = [M^{n,\delta}, M^{n,\delta}] + 2[M^{n,\delta}, A^{n,\delta}] + [A^{n,\delta}, A^{n,\delta}]$, we deduce

$$E^n\{[M^{n,\delta}, M^{n,\delta}]_{T^{n,k}}\} \leq E^n\{[X^{n,\delta}, X^{n,\delta}]_{T^{n,k}}\} \leq K + 4\delta^2. \tag{7.9}$$

If $H^n \in \mathbf{S}^n$, $|H^n| \leq 1$, then by Doob's maximal quadratic inequality we have

$$\begin{aligned} E\left\{\left(\sup_{r\leq T^{n,k}} \int_0^r H_s^n dM_s^{n,\delta}\right)^2\right\} &\leq 4E\left\{\int_0^{T^{n,k}} (H_s^n)^2 d[M^{n,\delta}, M^{n,\delta}]_s\right\} \\ &\leq 4E\{[M^{n,\delta}, M^{n,\delta}]_{T^{n,k}}\} \leq K + 4\delta^2 \end{aligned}$$

by (7.9), and combining this with the UT property of $X^{n,\delta}$, and since

$$P([M^{n,\delta}, M^{n,\delta}]_{T^{n,k}} > K) \leq \varepsilon + \frac{K + 4\delta^2}{K^2},$$

we have that

$$\left\{\int_0^r H_s^n dA_s^{n,\delta}; H^n \in \mathbf{S}^n, |H^n| \leq 1, n \geq 1\right\}$$

is stochastically bounded. Note that

$$\frac{|dA_s^{n,\delta}|}{dA_s^{n,\delta}} = H_s^{n,\delta} \in \mathbf{D} \text{ has } |H_s^{n,\delta}| = 1,$$

and since $A^{n,\delta}$ is natural we can take $H^{n,\delta} \in \mathbf{L}$ without loss. Therefore we have that

$$\lim_{k\to\infty} \sum_{t_i \in \pi^k} H_{t_i}^{n,\delta,k}(A_{t_i+1}^{n,\delta} - A_{t_i}^{n,\delta}) = \int |dA_s^{n,\delta}|,$$

and we deduce $\int_0^{t\wedge T^{n,k}} |dA_s^{n,\delta}|$ is stochastically bounded for each K. Since the jumps of $A^{n,\delta}$ (and hence also of $|A^{n,\delta}|$) are bounded by 2δ, it follows that we have UCV.

Next suppose $(X^n)_{n\geq 1}$ satisfies UCV. Since $X^n \Rightarrow X$, there exists $\delta > 0$ such that $J_\delta(X^n)$ is stochastically bounded. By UCV we also have that

$$\left\{ \int_0^t |dA_s^{n,\delta}|, [M^{n,\delta}, M^{n,\delta}]_t; n \geq 1 \right\}$$

is stochastically bounded. This implies that $X^n - M^{n,\delta}$ satisfies the property:

$$\left\{ \int K_{s-}^n d(X^n - M^{n,\delta})_s; K^n \in \mathbf{D}; |K^n| \leq 1; n \geq 1 \right\}$$

is stochastically bounded. Now let $\varepsilon > 0$. There exists K such that $P^n([M^{n,\delta}, M^{n,\delta}]_t > K) < \varepsilon$ for all n. Define

$$T^n = \inf\{s : [M^{n,\delta}, M^{n,\delta}]_s > K\} \wedge t.$$

Then $P^n(T^n < t) < \varepsilon$. Next let $H^n \in \mathbf{D}$, $|H^n| \leq 1$. Then we have

$$\begin{aligned}
P^n(|(H_-^n \cdot M^{n,\delta})_t| > K) &\leq P^n\left(\sup_{s\leq t} |(H_-^n \cdot M^{n,\delta})_s| > K\right) \\
&\leq P^n(\sup_{s\leq T^n} |(H_-^n \cdot M^{n,\delta})_s| > K) + \varepsilon \\
&\leq \frac{1}{K^2} E\{\sup_{s\leq T^n} ((H_-^n \cdot M^{n,\delta})_s)^2\} + \varepsilon \\
&\leq \frac{1}{K^2} E\{[M^{n,\delta}, M^{n,\delta}]_{T^n}\} + \varepsilon \leq \frac{K + 4\varepsilon^2}{K^2} + \varepsilon.
\end{aligned}$$

This last quantity can be made arbitrarily small, and thus $M^{n,\delta}$ satisfies UT as well. □

We note that without the hypothesis that $X^n \Rightarrow X$, we have that if (X^n) satisfies UT, then it satisfies UCV, but if it satisfies UCV we need the extra hypotheses that $J_\delta(X^n)$ is stochastically bounded to prove it satisfies UT.

Next we give some general conditions that imply UT:

If $(X^n)_{n\geq 1}$ is a sequence of supermartingales such that

$$\inf_n(\inf_s X^n_s) \geq b \qquad (b \in \mathbf{R})$$

then (X^n) satisfies UT (cf [9]); (7.10)

If $(X^n)_{n\geq 1}$ is a sequence of local martingales and if for each $t < \infty$ one has

$$\sup_n E^n\{\sup_{s\leq t} |\Delta X^n_s|\} < \infty$$

then (X^n) satisfies UT (cf [9]). (7.11)

Clearly the condition UCV gives conditions which are easy to verify in practice. We give two examples:

– Let $(X^n)_{n\geq 1}$ be a sequence of semimartingales with decompositions

$$X^n = M^n + A^n \text{ such that } \sup_n\{E^n\{[M^n, M^n]_t\} + E^n\{\int_0^t |dA^n_s|\}\} < \infty, \tag{7.12}$$

each $t > 0$. Then $(X^n)_{n\geq 1}$ satisfies UCV.

– Let $(X^n)_{n\geq 1}$ be a sequence of semimartingales with decompositions

$$X^n = M^n + A^n \text{ such that } \sup_n\{\operatorname{Var}(M^n_t) + E^n\{\int_0^t |dA^n_s|\}\} < \infty, \tag{7.13}$$

each $t > 0$. Then $(X^n)_{n\geq 1}$ satisfies UCV. (Here $\operatorname{Var}(M^n_t)$ refers to the variance of the random variable M^n_t).

Note that (7.12) and (7.13) are trivially equivalent. Combining (7.11) and (7.12) we get: let $X^n \Rightarrow X$ and suppose X^n has decompositions

$$X^n = M^n + A^n$$

such that

$$\sup_n \left\{ E^n \left\{ \sup_{s\leq t} |\Delta M^n_s| \right\} + E^n \left\{ \int_0^t |dA^n_s| \right\} \right\} < \infty. \tag{7.14}$$

Then $(X^n)_{n\geq 1}$ satisfies UT and UCV.

Theorem 7.7. *Let $(X^n)_{n\geq 1}$ be a sequence of vector valued semimartingales, X a vector valued process in $\mathbf{D}$, and assume $X^n \Rightarrow X$ and that $(X^n)_{n\geq 1}$ is a good sequence. Then $(X^n)_{n\geq 1}$ satisfies UT and UCV.*

Proof. We treat the scalar case. By Theorem 7.6 it suffices to show that UT holds. Suppose $(X^n)_{n\geq 1}$ is good but UT does not hold. Then there must exist $H^n \in \mathbf{S}$, $|H^n| \leq 1$, and constants c_n increasing to ∞ such that for some $\varepsilon > 0$,

$$\liminf_{n\to\infty} P^n\left\{ \int H^n_{s-} dX^n_s \geq c_n \right\} \geq \varepsilon.$$

But this implies that

$$\liminf_{n\to\infty} P^n\left\{ \int \frac{1}{c_n} H^n_{s-} dX^n_s \geq 1 \right\} \geq \varepsilon \tag{7.15}$$

as well. Since $|H^n| \leq 1$ we have that $\frac{1}{c_n}H^n$ converges uniformly in distribution to the zero process. The goodness of $(X^n)_{n\geq 1}$ then implies that $\int \frac{1}{c_n} H^n_{s-} dX^n_s$ converges in distribution to 0, which contradicts (7.15). □

The next theorem is a key step to showing that each of UT and UCV imply goodness. For a sequence of vector processes (H^n, X^n) each defined on a space Θ^n, we let $\mathbf{H}^n = (\mathcal{H}^n_t)_{t\geq 0}$ denote the smallest filtration making (H^n, X^n) adapted and also satisfying the usual hypotheses. $\mathbf{H} = (\mathcal{H}_t)_{t\geq 0}$ is analogous for the limiting process (H, X).

Theorem 7.8. *Let (H^n, X^n), (H, X) be vector processes, each in* $\mathbf{D}$, *and suppose $(H^n, X^n) \Rightarrow (H, X)$. Assume UT or equivalently UCV holds. Then X is an* $\mathbf{H}$ *semimartingale.*

Proof. $X \in \mathbf{D}$ by hypothesis. If $H^m \in \mathbf{S}(\mathbf{H})$, $|H^n| \leq 1$, and $\lim_{m\to\infty} H^m = 0$ uniformly, we need to show $\lim_{m\to\infty} H^m \cdot X = 0$ in probability. Note that for $H^m \in \mathbf{S}$, $H^m \cdot X$ is well defined for any $X \in \mathbf{D}$. Since the limit is 0 — a constant — it suffices to show that $H^m \cdot X$ converges to 0 in distribution. Thus it suffices to show that

$$\{H^n \cdot X, H^m \in \mathbf{S}(\mathbf{H}), |H^m| \leq 1\} \tag{7.16}$$

is stochastically bounded.

Let

$$j(H, X) = \{s \geq 0 : P(\Delta H_s \neq 0 \text{ or } \Delta X_s \neq 0) > 0\}. \tag{7.17}$$

One can check fairly easily (cf, eg, [8, p. 313]) that $j(H, X)$ is at most countable. Therefore, $Q = \mathbf{R}_+ \setminus j(H, X)$ is dense. Since $(H^n, X^n) \Rightarrow (H, X)$, we have that the finite dimensional distributions, restricted to Q-valued tuples, of (H^n, X^n) converge to (H, X). (Typically this is denoted $(H^n, X^n) \overset{\mathcal{L}(Q)}{\Rightarrow} (H, X)$). This fact, together with the UT property of $(X^n)_{n\geq 1}$, is enough to conclude, using simple approximation arguments, that (7.16) is stochastically bounded. □

The next theorem (Theorem 7.10) is the key result in the theory of weak convergence of stochastic integrals. One can prove it using the UT approach (see [9]) or the UCV approach (see [10]). The UT approach is fairly intuitive given our definition of a semimartingale as a good integrator, but it is a little complicated to execute. The UCV approach is intuitively very simple, as is the proof. The main disadvantage is that we need a technical result concerning the Skorohod topology. Note that we will generalize this approach in Section 4,5,6 of Part II of these notes (at the end of this volume). To motivate the argument of the proof let us first make an observation.

Let $(x_n)_{n\geq 1}$, x, $(y_n)_{n\geq 1}$, y be functions in D where $(x_n, y_n) \to (x, y)$ in the Skorohod topology (that is $D_{\mathbf{R}^2}[0,\infty)$, *not* $D_{\mathbf{R}}[0,\infty) \times D_{\mathbf{R}}[0,\infty)$!). Assume further that y_n is piecewise constant and the number of discontinuities of y_n in a bounded time interval is uniformly bounded in n. Assuming all terms make sense, we then have

$$\left(x_n, y_n, \int x_n(s-)dy_n(s), \int y_n(s-)dx_n(s)\right) \to \left(x, y, \int x_{s-}dy_s, \int y_{s-}dx_s\right) \tag{7.18}$$

in the Skorohod topology $D_{\mathbf{R}^4}[0,\infty)$. In view of (7.18), it makes sense to try to approximate the processes involved by piecewise constant processes, but in such a way that they converge along with the approximating processes. Before giving Theorem 7.10 we establish a lemma that plays an essential role in its proof.

Suppose (E,ρ) is a metric space, and let $(\theta_k)_{k\geq 1}$ be a sequence of i.i.d. random variables, uniform on $[\frac{1}{2}, 1]$. Fix $z \in D$, $\varepsilon > 0$, and define inductively

$$\begin{aligned} &T_0 = 0,\\ &T_{k+1} = \inf\{t > T_k : \rho(z_t, z_{T_k}) \vee \rho(z_{t-}, z_{T_k}) \geq \varepsilon\theta_k\} \end{aligned}$$

and let $y_k(z) = z_{T_k}$. (Note that $T_k = T_k(z)$; that is for each z we get a different sequence of times T_k.) We define

$$I^\varepsilon(z)_t = y_k(z) \text{ if } T_k \leq t < T_{k+1}. \tag{7.19}$$

Then $\rho(z_t, I^\varepsilon(z)_t) \leq \varepsilon$, for all t. The role of the $(\theta_k)_{k\geq 1}$ is to "spread" ε over an interval which then ensures the almost sure convergence of the $T_k^n = T_k(z^n)$ when z^n converges to z.

Lemma 7.9. *For I^ε defined as in (7.19), if $\lim_{n\to\infty} z_n = z$ in the Skorohod topology $D_E[0,\infty)$, then $(z_n, I^\varepsilon(z_n)) \to (z, I^\varepsilon(z))$ a.s. in the Skorohod topology $D_{E^2}[0,\infty)$.*

We refer the reader to [10, p. 1067] for a proof.

Theorem 7.10. *If (H^n, X^n) defined on Θ^n converges in distribution in the Skorohod topology to (H, X) and if $(X^n)_{n\geq 1}$ are semimartingales satisfying UCV (or equivalently UT), there exists a filtration* H *such that X is an* H *semimartingale and moreover*

$$(H^n, X^n, H_-^n \cdot X^n) \Rightarrow (H, X, H_- \cdot X). \tag{7.20}$$

That is, the sequence $(X^n)_{n\geq 1}$ is good.

Proof. That (UT) and (UCV) are equivalent under these hypotheses is Theorem 7.6. That X is an H semimartingale is Theorem 7.8. Thus it remains to establish (7.20).

Recall that for $x \in D$ and $\delta > 0$, we defined $J_\delta(x)$ in (7.2) as an operator that is used to shrink the large jumps to the size δ. Then $X^{n,\delta} = X^n - J_\delta(X^n)$ is a semimartingale with jumps bounded by δ. We define

$$Z^n = (H^n, X^n, J_\delta(X^n), X^{n,\delta}).$$

Let I^ε be as defined in (7.19). Then $I^\varepsilon(Z^n)$ is adapted to a filtration $\mathbf{K}^n = \mathbf{H}^n \vee \mathbf{U}$, where $\mathbf{U}$ is independent of $\mathbf{H}^n$. (By the independence, we note that X^n remains a semimartingale for the larger filtration $\mathbf{K}^n$.) Let $H^{n,\varepsilon}$ denote the first component of Z^n (which is M^{kd} valued in general, where M^{kd} represents $k \times d$ matrices). Then $|H^n - H^{n,\varepsilon}| \le \varepsilon$ and moreover

$$(H^n, X^n, J_\delta(X^n), X^{n,\delta}, H^{n,\varepsilon}) \Rightarrow (H, X, J_\delta(X), X^\delta, H^\varepsilon).$$

Next define:

$$\begin{aligned}
U^n &= \int H^n_{s-} dX^n_s; \\
U^{n,\varepsilon} &= \int H^{n,\varepsilon}_{s-} dX^{n,\delta}_s + \int H^n_{s-} dJ_\delta(X^n)_s; \\
U &= \int H_{s-} dX_s; \\
U^\varepsilon &= \int H^\varepsilon_{s-} dX^\delta_s + \int H_{s-} dJ_\delta(X)_s.
\end{aligned}$$

Then it follows as in (7.18) that

$$(H^n, X^n, U^{n,\varepsilon}) \Rightarrow (H, X, U^\varepsilon).$$

Finally we let

$$R^{n,\varepsilon} = U^n - U^{n,\varepsilon}.$$

Then

$$\begin{aligned}
R^{n,\varepsilon} &= \int (H^n_{s-} - H^{n,\varepsilon}_{s-}) dX^{n,\delta}_s \\
&= \int (H^n_{s-} - H^{n,\varepsilon}_{s-}) dM^{n,\delta}_s + \int (H^n_{s-} - H^{n,\varepsilon}_{s-}) dA^{n,\varepsilon}_s
\end{aligned}$$

where $X^{n,\delta} = M^{n,\delta} + A^{n,\delta}$ is a decomposition of $X^{n,\delta}$ into the sum of a local martingale and a finite variation process. Using Doob's maximal quadratic inequality we have that for any stopping time T,

$$E\left\{\sup_{s \le t\wedge T} |R^{n,\varepsilon}_s|\right\} \le \varepsilon \left\{2E\left\{[M^{n,\delta}, M^{n,\delta}]^{\frac{1}{2}}_{t\wedge T}\right\} + 2E\left\{\int_0^{t\wedge T} |dA^{n,\delta}_s|\right\}\right\}.$$

An analogous estimate holds for $U - U^\varepsilon$. We now apply the UCV hypothesis to conclude that $(H^n, X^n, U^n) \Rightarrow (H, X, U)$. □

We wish to make several remarks. First note that if we combine Theorems 7.7 and 7.10, we have that if $(H^n, X^n) \Rightarrow (H, X)$, then $(X^n)_{n\ge 1}$ is *good* *if and only if* UCV (or equivalently UT) holds. Second, if convergence in distribution is replaced by convergence in probability in the hypothesis of Theorem 7.10 (in this case of course all processes are defined on the same space), then convergence in probability will also hold in the conclusion.

Third, we can use Theorem 7.10 to prove some nice properties of goodness (Theorems 7.11 through 7.13). The first theorem shows that goodness is inherited via stochastic integration. The proof is similar to the proof of Theorem 7.7.

Theorem 7.11. *Suppose* $(H^n, X^n) \Rightarrow (H, X)$, *and* $(X^n)_{n\geq 1}$ *is a good sequence of semimartingales. Then* $Y^n = H^n_- \cdot X^n$ *is also a good sequence of semimartingales.*

Proof. We treat the scalar case. By Theorem 7.10 it suffices to show $(Y^n)_{n\geq 1}$ satisfies UT. Suppose it does not. Then as in the proof of Theorem 7.7 there exists a sequence $K^n \in \mathbf{S}$, $|K^n| \leq 1$, and constants c_n such that for some $\varepsilon > 0$,

$$\liminf_{n\to\infty} P^n\left\{ \int K^n_{s-} dY^n_s \geq c_n \right\} \geq \varepsilon,$$

or equivalently

$$\liminf_{n\to\infty} P^n\left\{ \int \frac{1}{c_n} K^n_{s-} H^n_s dX^n_s \geq 1 \right\} \geq \varepsilon. \tag{7.21}$$

The hypothesis that $(X^n)_{n\geq 1}$ is good implies that the integrals in (7.21) converge to 0, which is a contradiction. □

Theorem 7.12. *Let* $(X^n, Y^n)_{n\geq 1}$ *be a sequence of semimartingales such that* $(X^n, Y^n) \Rightarrow (X, Y)$, *and both* $(X^n)_{n\geq 1}$ *and* $(Y^n)_{n\geq 1}$ *are good. Then*

$$(X^n, Y^n, [X^n, Y^n]) \Rightarrow (X, Y, [X, Y]) \tag{7.22}$$

and also $[X^n, Y^n]$ *and* X^nY^n *are good.*

Proof. Integration by parts yields

$$X^nY^n = \int X^n_{s-} dY^n_s + \int Y^n_{s-} dX^n_s + [X^n, Y^n],$$

and (7.22) follows trivially. By Theorem 7.11 it suffices to show that $[X^n, Y^n]$ is good. But goodness implies UCV by Theorem 7.7, and the goodness of $[X^n, Y^n]$ follows easily. □

Theorem 7.13. *Let* $(X^n)_{n\geq 1}$ *be good, and suppose* $f : \mathbf{R}^d \times \mathbf{R}_+ \to \mathbf{R}$ *is* C^2 *on* $\mathbf{R}^d$ *and* C^1 *on* $\mathbf{R}_+$. *Let* $Y^n_t = f(X^n_t, t)$. *Then* $(Y^n)_{n\geq 1}$ *is also a good sequence.*

Proof. One need only apply the change of variables formula, and Theorems 7.12 and 7.11. □

We remark that the convergence in (7.22) is not as robust as it might seem. The next example, due to Jacod [6, p. 395], shows that one can have $X^n \Rightarrow X$, but $[X^n, X^n] \not\Rightarrow [X, X]$; thus a condition such as goodness is truly needed for the convergence of the quadratic variations.

Example 7.14. Let X^n be the non-random process $X_t^n = \sum_{i=1}^{[n^2t]}(-1)^i \frac{1}{n}$. Then $|X^n| \le \frac{1}{n}$, whence $X^n \Rightarrow 0$. On the other hand,

$$[X^n, X^n]_t = \sum_{i=1}^{[n^2t]} \left(\frac{1}{n}\right)^2 = \frac{[n^2t]}{n^2}$$

which converges to t. Since $[X, X] = 0$ trivially, we have $[X^n, X^n] \not\Rightarrow [X, X]$.

One of the primary uses of Theorem 7.10 is to the study of stochastic differential equations, which is the topic of §8..

8. Weak Convergence of Stochastic Differential Equations

In this section we consider stochastic differential equations in a form similar to those of § 5.. Let $(X^n)_{n\ge1}$ be a sequence of semimartingales, and $(J^n)_{n\ge1}$ be a sequence of processes in $\mathbf{D}$. Let

$$F, F^n : D_{\mathbf{R}^k}[0,\infty) \to D_{\mathbf{M}^{km}}[0,\infty)$$

have property (i) of Definition 5.1: that is, for $t > 0$ we have $F^n(X)_t = F^n(X^t)_t$ and $F(X)_t = F(X^t)_t$, which is a non-anticipation requirement. Note that we do not make the Lipschitz hypothesis ((5.1)(ii)). We will study equations of the type

$$X_t^n = J_t^n + \int_0^t F^n(X^n)_{s-} dZ_s^n \tag{8.1}$$

and give conditions that imply $X^n \Rightarrow X$, where X is a solution of the limiting equation

$$X_t = J_t + \int_0^t F(X)_{s-} dZ_s. \tag{8.2}$$

Note that without the Lipschitz assumption on $(F^n)_{n\ge1}$ nor F we do not have uniqueness of solutions either for (8.1) or for (8.2). If we are willing to assume that *a priori* the solutions (8.1) are relatively compact, we have the following simple result:

Theorem 8.1. *Suppose that (J^n, X^n, Z^n) satisfies equation (8.1), that (J^n, X^n, Z^n) is relatively compact in the Skorohod topology for $D_{\mathbf{R}^{2k+m}}[0,\infty)$, and that $(J^n, Z^n) \Rightarrow (J, Z)$ and that $(Z^n)_{n\geq 1}$ is good. Assume further that F^n, F satisfy*

$$\begin{aligned}&\text{if } (x_n, y_n) \to (x, y) \text{ in the Skorohod topology, then}\\ &(x_n, y_n, F^n(x_n)) \to (x, y, F(x)) \text{ in the Skorohod topology.}\end{aligned} \tag{8.3}$$

Then any limit point of the sequence $(X^n)_{n\geq 1}$ satisfies (8.2).

Proof. Suppose a subsequence of $(X^n)_{n\geq 1}$ converges in distribution. Then along a further subsequence, the triple (J^n, X^n, Z^n) will converge in distribution, to a process (J, X, Z). Theorem 7.10 then gives that X satisfies (8.2). □

The assumption (8.3) is that F^n and F are Skorohod continuous. Some examples of such are the following:

Example 8.2. Let $g : \mathbf{R}^k \times \mathbf{R}_+ \to \mathbf{M}^{km}$ and $h : \mathbf{R}_+ \to \mathbf{R}_+$ be continuous. The following functionals are non-anticipating and Skorohod continuous.

(i) $F(x)_t = g(x_t, t)$

(ii) $F(x)_t = \int_0^t h(t-s)g(x_s, s)ds$

If $k = m = 1$, then:

(iii) $F(x)_t = \sup_{s\leq t} h(t-s)g(x_s, s)$

(iv) $F(x)_t = \sup_{s\leq t} h(t-s)g(\Delta x_s, s)$.

Before stating our main result we need to make some definitions.

Definition 8.3. *(X, T) is a* local solution *of (8.2) if there exists a filtration $\mathbf{F}$ for which X, J, and Z are adapted, Z is a semimartingale, T is a stopping time, and such that*

$$X_{t\wedge T} = J_{t\wedge T} + \int_0^{t\wedge T} F(X)_{s-}dZ_s. \tag{8.4}$$

We say that we have strong local uniqueness *if any two solutions (X^1, T^1), (X^2, T^2) satisfy $X_t^1 = X_t^2$, $0 \leq t \leq T^1 \wedge T^2$.*

To define *weak* local uniqueness (that is, local uniqueness in the sense of distributions), we need the stopping times to be functions of the solutions.

Definition 8.4. *A tuple $(\widehat{J}, \widehat{Z}, \widehat{X}, \widehat{T})$ is a* weak local solution *if $(\widehat{J}, \widehat{Z})$ is a version of (J, Z), and (8.4) holds with $(\widehat{J}, \widehat{Z}, \widehat{X}, \widehat{T})$ replacing (J, Z, X, T). We say that* weak local uniqueness *holds for (8.2) if for any two weak local solutions (J^1, Z^1, X^1, T^1) and (J^2, Z^2, X^2, T^2) with $T^1 = h_1(X^1)$ and $T^2 = h_2(X^2)$ for measurable h_1, h_2, then $(X^1, (h_1 \wedge h_2)(X^1))$ and $(X^2, (h_1 \wedge h_2)(X^2))$ have the same distribution.*

We need to make some technical assumptions on the functional coefficients which are stronger than simple Skorohod continuity. Nevertheless Examples 8.2 can be shown to satisfy Condition 8.5 below.

Condition 8.5. Let Λ denote the collection of increasing maps λ of $\mathbf{R}_+$ to $\mathbf{R}_+$ with $\lambda_0 = 0$ and $\lambda_{t+h} - \lambda_t \leq h$, all t, $h \geq 0$. Assume that there exist mappings G^n, $G : D_{\mathbf{R}^k}[0,\infty) \times \Lambda \to D_{\mathbf{M}^{km}}[0,\infty)$ such that $F^n(x) \circ \lambda = G^n(x \circ \lambda, \lambda)$, and $F(x) \circ \lambda = G(x \circ \lambda, \lambda)$. Assume further

(a) For each compact subset $\mathcal{H} \subset D_{\mathbf{R}^k}[0,\infty) \times \Lambda$ and $t > 0$, (8.5)
$\lim_{n\to\infty} \sup_{(x,\lambda)\in\mathcal{H}} \sup_{s\leq t} |G^n(x,\lambda)_s - G(x,\lambda)_s| = 0,$

(b) For $(x_n, \lambda_n) \in D_{\mathbf{R}^k}[0,\infty) \times \Lambda$, if $\lim_{n\to\infty} \sup_{s\leq t} |x_n(s) - x(s)| = 0$ and $\lim_{n\to\infty} \sup_{s\leq t} |\lambda_n(s) - \lambda(s)| = 0$ for each $t > 0$,

then $\lim_{n\to\infty} \sup_{s\leq t} |G(x_n, \lambda_n)_s - G(x,\lambda)_s| = 0.$

Theorem 8.6. *Suppose $(J^n, Z^n) \Rightarrow (J, Z)$ where J^n, $J \in \mathbf{D}$; Z^n are semimartingales, and $(Z^n)_{n\geq 1}$ is good. Suppose F^n, F have representations in terms of G^n, G satisfying Condition 8.5. For $b > 0$, let*

$$\eta_n^b = \inf\{t > 0 : |F^n(X^n)_t| \vee |F^n(X^n)_{t-}| \geq b\}$$

and let $X^{n,b}$ denote the solution of

$$X_t^{n,b} = J_t^n + \int_0^t 1_{[0,\eta_n^b)}(s-) F^n(X^{n,b})_{s-} dZ_s^n$$

that agrees with X^n on $[0, \eta_n^b)$. Then $(J^n, X^{n,b}, Z^n)_{n\geq 1}$ is relatively compact and any limit point (J, X^b, Z) gives a local solution (X^b, T) of (8.2) with $T = \eta^c$, for any $c < b$.

Moreover if there exists a global solution X of (8.2) and weak local uniqueness holds, then $(J^n, X^n, Z^n) \Rightarrow (J, X, Z)$.

The proof of Theorem 8.6 involves some technical points, and we refer the reader to [10].

We next give two examples to show how Theorem 8.6 can be used.

Example 8.7 (Duffie-Protter). We can use Theorem 8.6 to help to derive and to justify models in continuous time stochastic finance theory as limiting cases of discrete models. As an example, let $(Y_i^n)_{i\geq 1}$ be the periodic rate of return on a security (such as a stock) with initial price S_0. After k periods the price of the security will be

$$S_k^n = S_0^n \prod_{i=1}^{k} (1 + Y_i^n).$$

Let $Z_t^n = \sum_{i=1}^{[nt]} Y_i^n$ and $S_t^n = S_{[nt]}^n$. Since $S_{k+1}^n - S_k^n = S_k^n Y_k^n$, we can write

$$S_t^n = S_0^n + \int_0^t S_{s-}^n dZ_s^n.$$

If $(Z^n)_{n\geq 1}$ is good with $Z^n \Rightarrow Z$ then the limiting equation is

$$S_t = S_0 + \int_0^t S_{s-} dZ_s$$

which is the stochastic exponential and has a unique global solution, and thus if $S_0^n \Rightarrow S_0$, by Theorem 8.6, $S^n \Rightarrow S$. Moreover by Theorem 7.11 we also thus know that $(X^n)_{n\geq 1}$ is good, hence if $(\theta_k^n)_{k\geq 1}$ represents a trading strategy and

$$G_t^n = \int_0^t \theta_{s-}^n dS_s^n$$

where $\theta_t^n = \theta_{[nt]}^n$ represents the resulting "gain" from the strategy θ^n, and if $(Z^n, \theta^n, S_0^n) \Rightarrow (Z, \theta, S_0)$ with $(Z^n)_{n\geq 1}$ still assumed to be good, we have $G^n \Rightarrow G$, with G given by

$$G_t = \int_0^t \theta_{s-} dS_s.$$

Many naturally occurring models have the property that $(Z^n)_{n\geq 1}$ is good.

Example 8.8. Emery [4] has discovered a class of martingales that have the Chaos Representation Property (CRP). A necessary condition to have this property, if $\langle M, M\rangle_t = t$, is that the local martingale M satisfy an equation of the type $[M,M]_t = t + \int_0^t \varphi_s dM_s$. A special case is:

$$d[X,X]_t = dt + f(X_{t-})dX_t, \tag{8.6}$$

where $f : \mathbf{R} \to \mathbf{R}$ is continuous. Therefore it is of interest to know when solutions of (8.6) exist. One can show existence for any such f by defining a sequence of discrete time martingales and then showing the sequence is relatively compact and that the limit satisfies (8.6). If one sets $\Delta Y_k^n = Y_{k+1}^n - Y_k^n$ and assumes $Y_0^n = 0$, then the discrete time analogue of (8.6) becomes

$$(\Delta Y_k^n)^2 = \frac{1}{n} + f(Y_k^n)\Delta Y_k^n. \tag{8.7}$$

One then solves (8.7) for ΔY_k^n:

$$\Delta Y_k^n = \frac{f(Y_k^n) \pm \sqrt{f(Y_k^n)^2 + 4/n}}{2}.$$

Call the solutions Z^+ and Z^-. In order for Y^n to be a martingale we are forced to choose

$$P(\Delta Y_k^n = Z^+) = \frac{Z^-}{Z^- - Z^+}$$

and

$$P(\Delta Y_k^n = Z^-) = 1 - P(\Delta Y_k^n = Z^+).$$

We then define $X_t^n = Y_{[nt]}^n$ and show it is relatively compact and that the limit satisfies (8.6). See [10, p. 1044] for details. The above argument also applies to the more general equation

$$d[X,X]_t = dt + F(X)_{t-}dX_t, \tag{8.8}$$

where F satisfies Condition 8.5. Note that a martingale is called a *normal martingale* if $\langle M, M\rangle_t = t$ and it has CRP, thus solving equations (8.6) or (8.7) is a way to generate candidates for normal martingales. Emery [4] has shown that if M satisfies

$$d[M,M]_t = dt + \beta M_{t-}dM_t; \quad M_0 = x,$$

then M is a normal martingale for $-2 \le \beta \le 0$. Note that $\beta = 0$ corresponds to Brownian motion, $\beta = -1$ is Azéma's martingale, and $\beta = -2$ is the parabolic martingale ($|M_t| = \sqrt{t}$).

In many situations the approximating semimartingales are *not* good. We give an example to illustrate this phenomenon.

Example 8.9. Let $W = (W_t)_{t\ge 0}$ be a standard Wiener process (that is, standard Brownian motion). Let us approximate W with an "approximate identity" as follows:

$$W_t^n = n\int_{t-\frac{1}{n}}^t W_s ds.$$

Then W^n is defined on the same space as W, W^n is adapted to the same filtration, and $\lim_{n\to\infty} W^n = W$ a.s., uniformly on compacts. However (W^n) is not good. Indeed, consider the equations

$$X_t^n = x + \int_0^t X_s^n dW_s^n. \tag{8.9}$$

Then $X_t^n = x\exp(W_t^n)$. But for the limiting equation

$$X_t = x + \int_0^t X_s dW_s$$

we have $X_t = x\exp(W_t - \frac{1}{2}t)$. Thus $W^n \Rightarrow W$, but $X^n \not\Rightarrow X$. This could not happen if $(W^n)_{n\ge 1}$ were good by Theorem 8.6.

If in Example 8.9 we rewrite W^n as $W^n = W + (W^n - W) = Y^n + Z^n$, then we have $Y^n \Rightarrow W$ and $(Y^n)_{n\geq 1}$ is good (in this case $Y^n = W$ for all n, so the convergence and goodness are trivial), and Z^n is a sequence of semimartingales converging to 0. Equation (8.8) can be re-written

$$X_t^n = X + \int_0^t X_s^n dY_s^n + \int_0^t X_s^n dZ_s^n.$$

This idea allows us to handle naturally arising situations where goodness does not apply. It generalizes a well-known approach due originally to E. Wong and M. Zakai.

Theorem 8.10. *Let Y^n, Z^n be semimartingales on Θ^n, and let $f : \mathbb{R}^d \to \mathbb{M}^{dm}$ be C^2 and bounded with bounded derivatives of first and second order. Define matrices of processes in* $\mathbb{D}$ *by*

$$H_t^n = \int_0^t Z_{s-}^n dZ_s^n \quad ^6$$

and

$$K_t^n = [Y^n, Z^n]_t.$$

Assume $(Y^n)_{n\geq 1}$ $(H^n)_{n\geq 1}$ are good, and $Z^n \Rightarrow 0$. Moreover assume

$$\mathbf{A}^n = (X_0^n, Y^n, Z^n, H^n, K^n) \Rightarrow \mathbf{A} = (X, Y, 0, H, K).$$

Let X^n be the solution of:

$$X_t^n = X_0^n + \int_0^t f(X_{s-}^n) dY_s^n + \int_0^t f(X_{s-}^n) dZ_s^n. \tag{8.10}$$

Then $(\mathbf{A}^n, X^n)$ is relatively compact and any limit point $(\mathbf{A}, X)$ satisfies

$$X_t = X_0 + \int_0^t f(X_{s-}) dY_s + \sum_{i,j,k} \int_0^t \partial_i f^j(X_{s-}) f^{ik}(X_{s-}) d(H_s^{kj} - K_s^{kj}), \tag{8.11}$$

where ∂_i denotes the partial derivative with respect to the i^{th} variable, f^j denotes the j^{th} column of f, etc.

Before we prove Theorem 8.10, we remark that the boundedness assumptions on f and its derivatives can be dropped, and X_0^n can be replaced with an exogenous process $(J_t^n)_{t\geq 0}$. See [10]. Also since $Z^n \Rightarrow 0$, it follows that H and K will *a fortiori* be continuous. Also, we have not assumed that $(K^n)_{n\geq 1}$ is a good sequence in Theorem 8.10; however the hypotheses will imply that $(K^n)_{n\geq 1}$ is also good (see (8.14) in the proof to follow).

Proof of Theorem 8.10. The proof will follow from the change of variables formula and integration by parts. First observe that

[6] Matrix entries: $(H_t^n)^{ij} = \int_0^t Z_{s-}^{n,i} dZ_s^{n,j}$; $(K_t^n)^{ij} = [Y^{n,i} Z^{n,j}]_t$

$$[Z^{n,i}, Z^{n,j}]_t = Z_t^{n,i} Z_t^{n,j} - Z_0^{n,i} Z_0^{n,j} - \int_0^t Z_{s-}^{n,i} dZ_s^{n,j} - \int_0^t Z_{s-}^{n,j} dZ_s^{n,i},$$

and therefore if $I_t^{n,i,j} = [Z^{n,i} Z^{n,j}]_t$, it follows that

$$I^{n,i,j} \Rightarrow -(H^{ij} + H^{ji}). \tag{8.12}$$

Since $I^{n,i,i}$ is non-decreasing and converges in distribution to a continuous process, it follows that $I^{n,i,i}$ is good. Moreover we can estimate the increments of $I^{n,i,j}$ by the increments of $I^{n,i,i}$ and $I^{n,j,j}$ to deduce that $(I^n)_{n\geq 1}$ itself is a good sequence.

Since f is assumed to be C^2, letting X^n be the solution of (8.9), by the change of variables formula we have

$$f^{ij}(X_t^n) = f^{ij}(X_0^n) + \sum_k \int_0^t \partial_k f^{ij}(X_{s-}^n) dX_s^k + R_t^{n,i,j},$$

where the increments of $R^{n,i,j}$ are dominated by a linear combination of the increments of $[Y^{n,k}, Y^{n,k}]$ and $[Z^{n,k}, Z^{n,k}]$, whence $(R^n)_{n\geq 1}$ is good.

Next we integrate by parts to obtain:

$$\begin{aligned}
\int_0^t f^{ij}(X_{s-}^n) dZ_s^j &= f^{ij}(X_t^n) Z_t^{n,j} - f^{ij}(X_0^n) Z_0^{n,j} \\
&\quad - \sum_k \int_0^t \partial_k f^{ij}(X_{s-}^n) Z_{s-}^{n,j} dX_s^{n,k} \\
&\quad - \int_0^t Z_{s-}^{n,j} dR_s^{n,i,j} \\
&\quad - \sum_k \int_0^t \partial_k f^{i,j}(X_{s-}^n) d[X^{n,k}, Z^{n,j}]_s \\
&\quad + [R^{n,i,j}, Z^{n,j}]_t \\
&= U_t^n - \sum_{k,\ell} \int_0^t \partial_k f^{i,j}(X_{s-}^n) f^{k,\ell}(X_{s-}^n) Z_{s-}^{n,j} dZ_s^{n,\ell} \\
&\quad - \sum_{k,\ell} \int_0^t \partial_k f^{i,j}(X_{s-}^n) f^{k,\ell}(X_{s-}^n) d([Y^{n,\ell}, Z^{n,j}]_s \\
&\quad + [Z^{n,\ell}, Z^{n,j}]_s),
\end{aligned}$$

where $U^n \Rightarrow 0$. Continuing:

$$= U_t^n - \sum_{k,\ell} \int_0^t \partial_k f^{ij}(X_{s-}^n) f^{k,\ell}(X_{s-}^n) d(H_s^{n,j,\ell} + K_s^{n,\ell,j} + I_s^{n,\ell,j}). \tag{8.13}$$

We have already seen that $(I^n)_{n\geq 1}$ is good, and we calculated its limit in (8.12). Note further that

$$\left|K_{t+h}^{n,j,k} - K_t^{n,j,k}\right| \leq \frac{1}{2}\left\{[Y^{n,j}, Y^{n,j}]_{t+h} + I_{t+h}^{n,k,k} - [Y^{n,j}, Y^{n,j}]_t - I_t^{n,k,k}\right\} \tag{8.14}$$

and it follows that $(K^n)_{n\geq 1}$ is good. Therefore it remains only to substitute (8.13) into (8.9) to complete the proof. □

9. Applications to Numerical Analysis of SDE's

Let us consider a simple stochastic differential equation driven by a semimartingale Y:

$$X_t = X_0 + \int_0^t f(X_{s-})dY_s \tag{9.1}$$

where f is a continuous function (not necessarily Lipschitz). One is often interested in estimating quantities of the form $E\{g(X_a)\}$ for a fixed time a. One could use a Monte Carlo method if the law of $g(X_a)$ or of X_a were known, but in general it is not. Therefore one uses the structure of the SDE (9.1) to estimate the law of X_a. The simplest method is the Euler method. (There are more complicated numerical schemes that converge faster, but we intend to combine our results with a Monte Carlo procedure, and since Monte Carlo convergence is slow, we do not consider them here.) A straightforward extension of the Euler method of ordinary differential equations leads to an Euler scheme of the type

$$\overline{X}^0_{t_{k+1}} = \overline{X}^0_{t_k} + f(\overline{X}^0_{t_k})(Y_{t_{k+1}} - Y_{t_k}) \tag{9.2}$$

where $\pi^n = \{0 = t_0 < t_1 < \ldots < t_{k_n} = a\}$ is a sequence of partitions of $[0,a]$ such that $\lim_{n\to\infty} \operatorname{mesh}(\pi^n) = 0$. We denote $\overline{X}^0$ to be the approximation (the dependence on n is implicit) to distinguish it from a solution X of (9.1).

It is convenient in this context to use a different scheme than the naive one (9.1). However note that for our scheme (9.4) below, the two schemes will agree on the partition points $(t_k)_{k\geq 1}$ of $\pi^n[0,a]$. We define

$$\eta(t) = t_k \text{ for } t_k \leq t < t_{k+1} \tag{9.3}$$

for a partition $\pi^n[0,a] = \{0 < t_0 < \ldots < t_{k_n} = a\}$ where again the dependence on n is implicit. Let $\overline{X}$ satisfy the equation:

$$\overline{X}_t = X_0 + \int_0^t f(\overline{X}_{\eta(s-)})dY_s, \tag{9.4}$$

so that the integrands in the stochastic integral are piecewise constant. Note that $\overline{X}_{t_k} = \overline{X}^0_{t_k}$ for partition points $(t_k)_{k\geq 1}$. We can put more general assumptions on η, as the next lemma shows.

Lemma 9.1. *Let (Y^n) be a sequence of semimartingales, $X^n \in \mathbf{D}$, and $\eta_n \in \mathbf{D}$, nondecreasing, $\eta_n(t) \le t$, and $\lim_{n\to\infty} \eta_n(t) = t$ for all $t \ge 0$. Assume also $(Y^n)_{n\ge 1}$ is good and that $(X^n, Y^n) \Rightarrow (X, Y)$. Then*

$$\int X^n_{\eta_n(s-)} dY^n_s \Rightarrow \int X_{s-} dY_s.$$

Proof. Recall the notation J_δ introduced in §7. (equation (7.2)). Then for $\delta > 0$ we have

$$\int \left(J_\delta(X^n)_{\eta^n(s-)} - J_\delta(X^n)_{s-}\right) dY^n_s \Rightarrow 0.$$

Therefore there exists a sequence $(\delta_n)_{n\ge 1}$ tending to 0 such that

$$\int \left(J_{\delta_n}(X^n)_{\eta^n(s-)} - J_{\delta_n}(X^n)_{s-}\right) dY^n_s \Rightarrow 0.$$

However the asymptotic continuity of X^{n,δ_n} implies that $(X^{n,\delta_n}_t - X^{n,\delta_n}_{\eta_n(t)})_{t\ge 0} \Rightarrow 0$, whence

$$\int \left(X^{n,\delta_n}_{s-} - X^{n,\delta_n}_{\eta_n(s-)}\right) dY^n_s \Rightarrow 0,$$

and therefore

$$\int (X^n_{s-} - X^n_{\eta_n(s-)}) dY^n_s \Rightarrow 0$$

and the lemma is proved. □

Theorem 9.2. *Let (Y^n) be a good sequence of semimartingales such that $Y^n \Rightarrow Y$ Let η_n be as in Lemma 9.1. Let $f : \mathbf{R}^d \to \mathbf{M}^{km}$ be bounded and continuous, and let $\overline{X}^n$ satisfy*

$$\overline{X}^n_t = X_0 + \int_0^t f\left(\overline{X}^n_{\eta_n(s-)}\right) dY^n_s. \tag{9.5}$$

Then $(\overline{X}^n, Y^n)$ is relatively compact and any limit point satisfies

$$X_t = X_0 + \int_0^t f(X_{s-}) dY_s. \tag{9.6}$$

If all Y^n are defined on the same sample space as Y and if Y^n converges to Y in ucp and pathwise uniqueness holds for (9.6), then $\overline{X}^n$ converges to X in ucp as well.

Proof. The relative compactness is complicated and we refer the reader to Kurtz and Protter [10].

The fact that any limit point satisfies (9.6) then follows from Lemma 9.1.

Under the assumptions of the final assertion, we can treat $(\overline{X}^n, X)$ as a solution of a single system. The uniqueness then gives us that $(\overline{X}^n, X) \Rightarrow (X, X)$, whence $\overline{X}^n - X \Rightarrow 0$, and the conclusion follows. □

We note that often in applications we will have $Y^n = Y$ for all n; such a sequence is of course trivially good.

In certain circumstances we can give an analysis of the error in the Euler scheme; that is, we can determine the asymptotic distribution of the normalized error. Here weak convergence is essential, as we will see by considering the Brownian case in Example 9.4.

Theorem 9.3. *Let Y be a given $\mathbf{F}$ semimartingale and let f be a C^1 $\mathbf{M}^{dm}$ matrix valued function. Let $0 = T_0^n < T_1^n < \dots$ be $\mathbf{F}$ stopping times and define*

$$\eta_n(t) = T_k^n \text{ if } T_k^n \leq t < T_{k+1}^n,$$

and let $\overline{X}^n$ satisfy (9.5). Let α_n be a sequence of positive constants tending to ∞, and set

$$U^n = \alpha_n(\overline{X}^n - X)$$

and define

$$Z_t^{n,i,j} = \alpha_n \int_0^t \left(Y_{s-}^i - Y_{\eta_n(s-)}^i\right) dY_s^j.$$

Assume that $(Z^n)_{n\geq 1}$ is good and that $(Y, Z^n) \Rightarrow (Y, Z)$. Then $U^n \Rightarrow U$, where U satisfies

$$U_t = \sum_i \int_0^t \nabla f_i(X_{s-})U_{s-}dY_s^i + \sum_{i,j} \int_0^t \sum_k \partial_k f^i(X_{s-})f^{j,k}(X_{s-})dZ_s^{i,j}. \quad (9.7)$$

Proof. The hypothesis imply that (9.6) has a unique solution, hence

$$(\overline{X}^n, X, Y, Z^n) \Rightarrow (X, X, Y, Z).$$

Let us treat only the scalar use ($d = m = 1$). Observe that

$$\overline{X}_{s-}^n - \overline{X}_{\eta_n(s-)}^n = f(\overline{X}_{\eta_n(s-)}^n)(Y_{s-} - Y_{\eta_n(s-)}).$$

Therefore

$$\begin{aligned} U_t^n &= \int_0^t \alpha_n \left(f(\overline{X}_{s-}^n) - f(X_{s-})\right) dY_s - \int_0^t \alpha_n \left(f(\overline{X}_{s-}^n) - f(\overline{X}_{\eta_n(s-)}^n)\right) dY_s \\ &= \int_0^t \frac{f(\overline{X}_{s-}^n) - f(X_{s-})}{\overline{X}_{s-}^n - X_{s-}} U_{s-}^n dY_s \\ &\quad - \int_0^t f\left(\overline{X}_{\eta_n(s-)}^n\right) + f\left(\overline{X}_{\eta_n(s-)}^n\right)\left(Y_{s-} - Y_{\eta_n(s-)}\right) \\ &\quad - f\left(\overline{X}_{\eta_n(s-)}^n\right)\left(Y_{s-} - Y_{\eta_n(s-)}\right)^{-1} dZ_s^n. \end{aligned}$$

Next let $T^{n,a} = \inf\{t > 0 : |U_t^n| > a\}$. Then $U_{t\wedge T^{n,a}}^n$ is relatively compact, and any limit point will satisfy (9.7) on $[0, T^a]$, where $T^a = \inf\{t > 0 : |U_t| > a\}$. But $\lim_{a\to\infty} T^a = \infty$ a.s., so $U^n \Rightarrow U$. □

Example 9.4. Let us take $Y_t = \begin{pmatrix} W_t \\ t \end{pmatrix}$ in Theorem 9.3, where W is an $n-1$ dimensional standard Wiener process (or Brownian motion). Let $\eta_n(t) = \frac{[nt]}{n}$. Then taking $\alpha_n = \sqrt{n}$, we have $(Y, Z^n) \Rightarrow (Y, Z)$, where Z is independent of Y. Moreover $Z^{im} = Z^{mi} = 0$, and since Z^{ij} are continuous local martingales with $[Z^{ij}, Z^{k\ell}]_t = \begin{cases} 0 & ij \neq k\ell \\ \frac{1}{2}t & ij = k\ell \end{cases}$ we conclude that Z is also a Brownian motion, independent of W. Note that since Z is independent of W, it need not "live" on the same space as W. Thus the limiting process U could appear only through weak convergence, in general.

References

1. Ahn, H., Protter, P., "A Remark on Stochastic Integration", *Sém. de Proba. XXVII*, LNM 1583, 312-315 (1994).
2. Dellacherie, C., *Capacités et Processus Stochastiques*, Springer, Berlin (1972).
3. Duffie, D., Protter, P., " From discrete to continuous time finance: weak convergence of the financial gain process", *J. Math. Finance* 2, 1-15 (1992).
4. Emery, M., "On the Azéma Martingales", *Sém. de Proba. XXIII*, LNM 1372, 66-87 (1989).
5. Ethier, S., Kurtz, T. G., *Markov Processes: Characterization and Convergence*, Wiley, New York (1986).
6. Jacod, J., "Theoremes Limites pour les Processus", *Ecole d'Eté de Proba. de St. Flour XIII*, LNM 1117, 299-409 (1985).
7. Jacod, J., Protter, P., "A Remark on the Weak Convergence of Processes in the Skorohod Topology", *J. Theoretical Probability*, 6, 463-472 (1993).
8. Jacod, J., Shiryaev, A. N., *Limit Theorems for Stochastic Processes*, Springer, Berlin (1987).
9. Jakubowski, A., Mémin, J., Pagès, G., "Convergence en Loi des Suites d'Intégrales Stochastiques sur l'Espace $\mathbf{D}^1$ de Skorohod", *Proba. Th. Rel. Fields* 81, 111-137 (1989).
10. Kurtz, T. G., Protter, P., "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations", *Ann. Proba.* 19, 1035-1070 (1991).
11. Kurtz, T. G., Protter, P., "Characterizing the Weak Convergence of Stochastic Integrals", in *Stochastic Analysis* (M. Barlow and N. Bingham, eds.), Cambridge U. P., 255-259 (1991).
12. Kurtz, T. G., Protter, P., "Wong-Zakai, Corrections, Random Evolutions, and Numerical Schemes for SDEs", in *Stochastic Analysis*, Academic Press, 331-346 (1991).
13. Mémin, J., Slominski, L., "Condition UT et Stabilité en Loi des Solutions d'Equations Différentielles Stochastiques", *Sém. de Proba. XXV*, LNM 1485, 162-177 (1991).
14. Pratelli, M., "La Classe des Semimartingales qui Permettent d'Intégrer les Processus Optionnels", *Sém. de Proba. XVII*, LNM 986, 311-320 (1983).
15. Protter, P., *Stochastic Integration and Differential Equations*, Springer, Berlin (1990).
16. Slominski, L., "Stability of Strong Solutions of Stochastic Differential Equations", *Stoch. Processes and Their Appl.* 31, 173-202 (1989).