The structure of hyperfinite stochastic integrals

by

Tom L. Lindstrøm,
University of Oslo.

1. Introduction.

The hyperfinite theory for stochastic integration goes back to R.M. Anderson [2], who constructed a Brownian motion as the standard part of a hyperfinite random walk, and defined the stochastic integral with respect to this random walk as a pathwise Stieltjes integral. The theory was further developed by H.J. Keisler [7], and extended to more general classes of martingales by Lindstrøm [9], and Hoover and Perkins [6], independently (confer also the work of K.D. Stroyan). A further extension to the infinite dimensional case was given in [10]. The papers by Keisler and Hoover-Perkins effectfully demonstrated the power of the nonstandard approach by proving new strong existence results for stochastic differential equations.

A central issue in the first papers was to show that what could be obtained by the standard theory could also be obtained by the hyperfinite theory, e.g. it was shown in [9] that if ${}^\circ M^+$ is the "right standard part" of a hyperfinite SL^2-martingale, and X is a process standard integrable with respect to ${}^\circ M^+$, then there exists a hyperfinite process Y - called a 2-lifting of X - which is integrable with respect to M, and such that ${}^\circ(\int YdM)^+ = \int Xd{}^\circ M^+$. Moreover, it was shown that all local L^2-martingales could in a natural way be represented as right standard parts of hyperfinite martingales, and it was argued that these two results implied that the standard theory could be derived from the nonstandard theory.

But not all hyperfinite integrable processes are liftings, and thus the answer above urges us to consider the opposite question; can the hyperfinite theory be "richer" than the standard one? Or - on the contrary - is it true, that given a hyperfinite stochastic integral $Y = \int XdM$, we can obtain the standard part of Y as a stochastic integral of a process in a natural way connected to the standard part of M? These are the questions we shall consider in this paper. Unluckily, we do not have many positive results (- the paper is almost a collection of counter-examples -), but we shall try to argue that the one result we do have, has so nice consequences that the study should be continued. To see this, let us consider what consequences different answers to our question would have: If the nonstandard theory really is "richer", we have the possibility that it can be used to express new connections and to obtain new results. But we also have the possibility that since the class of integrals is larger, fewer results may hold for it, e.g. an inequality which is true for a class of standard integrals may fail for the corresponding nonstandard class. A problem of this kind was encountered by Keisler in the proof of his existence theorem for solutions of stochastic differential equations: He wanted to use a standard inequality of Krylov, concerning processes of the form

$$x(t,\omega) = x_0 + \int_0^t f(s,\omega)ds + \int_0^t g(s,\omega)db(s,\omega),$$

to show that a process X was a lifting. To complete this argument he had to replace the term $\int g(s,\omega)db(s,\omega)$ with a nonstandard term $\int G(s,\omega)d\chi(s,\omega)$, where χ was a hyperfinite random walk having the Brownian motion b as standard part. Using the lifting theorems he could do this if G was a lifting, but in his problem G depended on X in such a way that it was a lifting only if X was a lifting! Keisler avoided this circularity using an approximation argument, but no doubt his proof had been much simpler if he had had a nonstandard Krylov inequality without a lifting condition. Applying the representation theorems of the first parts of this paper, we shall in sections 5 and 6 prove such an inequality and use it to simplify Keisler's proof. We hope this will convince the reader of the importance of a better knowledge and control of hyperfinite stochastic integrals.

In the next section of this paper we give some examples of what we can and can not hope to obtain in representing nonstandard stochastic integrals by standard stochastic integrals. In the third and fourth section we prove our main result (Theorem 5), saying that for a class of martingales M we may obtain the standard parts of nonstandard stochastic integrals $\int XdM$ as standard integrals of processes having the same finite dimensional distributions as $^{\circ}M^{+}$. This is the result we use to prove the Krylov-inequality and Keisler's theorem. In the final section we try to show by an example that the extra power of hyperfinite stochastic integration is significant, and that it should be possible to put it to good use.

We shall use the terminology and notation of [7] and [9]. A suitable reference for nonstandard analysis in general is the book by Stroyan and Luxemburg [13], and for nonstandard probability theory in particular the survey paper by Loeb [11]. For the standard theory for stochastic integration, see Métivier [12]. We shall assume that our nonstandard models have the necessary saturation properties (see [13]).

An earlier (unpublished) version of sections 5 and 6 was referred to by Fenstad in [5] under the title "Hyperfinite stochastic integration and stochastic differential equations".

2. Two examples.

It is not hard to find examples which show that if M is an SL^2-martingale and X is integrable with respect to M, we can not always find a Y integrable with respect to ${}^\circ M^+$, such that ${}^\circ(\int XdM)^+ = \int Yd{}^\circ M^+$. The example we shall give shows that it is not true even when M is as nice and regular as a Brownian motion:

Example 1: Let $\eta \in {}^*\mathbb{N}-\mathbb{N}$; we shall use the hyperfinite timeline $T = \{\frac{k}{\eta} : 0 \leq k \leq \eta\}$. Let $\Omega = \{-1,1\}^T$, and let P be the uniform probability measure on Ω; $P\{\omega\} = \frac{1}{2^\eta}$.

Let $\chi : T \times \Omega \to {}^*\mathbb{R}$ be the Anderson process

$$\chi(t,\omega) = \sum_{s=0}^{t} \frac{\omega(s)}{\sqrt{\eta}};$$

then $\beta = {}^\circ\chi^+$ is a Brownian motion. Let $X : T \times \Omega \to {}^*\mathbb{R}$ be defined by

$$X(\tfrac{k}{\eta},\omega) = 1 \quad \text{if } k \text{ is even},$$

$$X(\tfrac{k}{\eta},\omega) = 0 \quad \text{if } k \text{ is odd}.$$

We shall show that there is no process Y such that $\int Y d\beta = {}^\circ(\int X d\chi)^+$: By Theorem II-21 of [9], $[{}^\circ(\int X d\chi)^+](t) = {}^\circ[\int X d\chi]^+(t) = \frac{1}{2}t$ since $\int X d\chi$ is S-continuous and hence well-behaved. But $[\int Y d\beta](t) = \int_0^t Y^2 dt$, and thus if $\int Y d\beta = \int X d\chi$, we must have $Y^2 = \frac{1}{2}$ a.e. Hence we may find a 2-lifting Z of Y such that $Z^2 = \frac{1}{2}$. But now

$$E((\int_0^t (X-Z)d\chi)^2) = E([\int_0^t (X-Z)d\chi]) \geq (1 - \frac{1}{\sqrt{2}})^2 t.$$

On the other hand, by construction of Z we should have ${}^\circ(\int X d\chi)^+ = \int Y d\beta = {}^\circ(\int Z d\chi)^+$, and we have got a contradiction.

With this example in mind there seems to be no reason to look for classes of martingales M such that given X, we can always find a Y such that $\int Y d{}^\circ M^+ = {}^\circ(\int X dM)^+$. But often we are not interested in the process itself, only in its distribution. Perhaps we should weaken our statement above by replacing ${}^\circ M^+$ with a process N having the same finite dimensional distributions. Hence we could ask if given X and M, we can find a martingale N with the same finite dimensional distributions as ${}^\circ M^+$ and a process Y, such that $\int Y dN = {}^\circ(\int X dM)^+$. Again it is not difficult to find examples which show that this is not true. Our example shows that it does not even hold for the "nicest" kind of discontinuous martingales, the well-behaved ones:

Example 2: We use the same time-line as in Example 1. A martingale $Z : T \times \Omega \to {}^*\mathbb{R}$ is described informally as follows: $Z(0) = 0$. If $0 \leq t < \frac{1}{\eta^{1/4}}$, $\Delta Z(t)$ is $-\eta^{-3/4}$, $(1 - \eta^{-3/4})$ or 0 according to the following rules: If $\Delta Z(s,\omega) = 1 - \eta^{-3/4}$ for some $s < t$, then $\Delta Z(t,\omega) = 0$. If not, then $\Delta Z(t)$ is $-\eta^{-3/4}$ with probability $(1-\eta^{-3/4})$ and $(1-\eta^{-3/4})$ with probability $\eta^{-3/4}$. For $t \geq \frac{1}{\eta^{1/4}}$, $\Delta Z = 0$.

Let now $X : T \times \Omega \to {}^*\mathbb{R}$ be defined by

$$X(\tfrac{k}{\eta},\omega) = (-1)^k,$$

and let $M = \int X dZ$. Then M is a well-behaved martingale and $Z = \int X dM$.

Now ${}^\circ M^+$ is constant zero on a set of measure $(1-\eta^{-3/4})^{\eta^{3/4}} \approx e^{-1}$, while ${}^\circ Z^+$ is different from zero on a set of Loeb-measure one. Hence ${}^\circ Z^+$ can not be a stochastic integral of a process with the same distributions as ${}^\circ M^+$.

Things are even worse than this; we shall see in Example 8 that the statement is not true in general for S-continuous processes. All the same, this is the concept we shall work with in the following two sections.

Example 2 has been constructed independently by Hoover and Perkins [6] to show that the stochastic integral of a well-behaved process is not necessarily well-behaved, which was also our original purpose.

3. The μ-Brownian motions and their integrals.

Let μ be a measure on $[0,1]$ such that $\mu([0,1]) < \infty$, and the cumulative distribution $g : [0,1] \to \mathbb{R}_+$ is a continuous function. If $\langle Z, \{\mathcal{F}_t\}, \nu \rangle$ is a stochastic basis, an n-dimensional

μ-Brownian motion with respect to this basis is an n-dimensional martingale M such that

$$(g(t)-g(s))I = E([M](t)-[M](s) \mid \mathcal{F}_s)$$

for all $s<t$. Here I is the identity $n\times n$-matrix and $[M]_{ij} = [M_i, M_j]$. By a well-known characterization of Brownian motions it follows that if μ is the Lebesgue-measure, then M is a Brownian motion.

If M is a μ-Brownian motion, we can define a new process $\tilde{M}: [0,g(1)]\times\Omega \to \mathbb{R}_+$ by $\tilde{M}(t,\omega) = M(g^{-1}(t),\omega)$. $\tilde{M}$ is well-defined since if $g(t_1) = g(t_2)$, then $M(t_1) = M(t_2)$ a.e. Then $\tilde{M}$ is a martingale, and

$$(t-s)I = E([\tilde{M}](t)-[\tilde{M}](s) \mid \bigcup_{g(r)=t} \mathcal{F}_r)$$

By the characterization above, $\tilde{M}$ is Brownian motion. Since all Brownian motions have the same finite dimensional distributions, we have proved:

Lemma 3: Let M and N be two μ-Brownian motions; then M and N have the same finite dimensional distributions.

If N is a real-valued L^2-martingale adapted to $\langle Z, \{\mathcal{F}_t\}, \nu\rangle$, the Doleans-measure of $[N]$ is the measure $\nu_{[N]}$ on the predictable sets defined by

$$\nu_{[N]}(\langle s,t]\times A_s) = E(1_{A_s}([N](t)-[N](s)))$$

for $A_s \in \mathcal{F}_s$, $s<t$.

So if f is predictable

$$\int_{]s,t]\times A_s} f \, d\nu_{[N]} = E(1_{A_s} \int_s^t f d[N])$$

Assume that M is another L^2-martingale, and assume that the paths of $[N]$ are absolutely continuous with respect to the corresponding paths of $[M]$ a.s. Then $\nu_{[N]}$ is absolutely continuous with respect to $\nu_{[M]}$: Since

$$\nu_{[M]}(A) = \int 1_A d\nu_{[M]} = E(\int 1_A d[M])$$

and

$$\nu_{[N]}(A) = \int 1_A d\nu_{[N]} = E(\int 1_A d[N]),$$

we have

$$\nu_{[M]}(A) = 0 \Rightarrow \int 1_A d[M] = 0 \text{ a.e.} \Rightarrow \int 1_A d[N] = 0 \text{ a.e.} \Rightarrow \nu_{[N]}(A) = 0.$$

This implies that there is a predictable Radon-Nikodym derivative h such that

$$\nu_{[N]}(A) = \int_A h \, d\nu_{[M]}$$

for all predictable A. Also notice that if $A_s \in \mathcal{F}_s$, $s < t$:

$$E(1_{A_s}([N](t)-[N](s))) = \nu_{[N]}(]s,t]\times A_s) = \int_{]s,t]\times A_s} h \, d\nu_{[M]} = E(1_{A_s} \int_s^t h \, d[M])$$

and hence

$$E([N](t) - [N](s) | \mathcal{F}_s) = E(\int_s^t h \, d[M] | \mathcal{F}_s)$$

A special case is when $\nu_{[M]}$ is the restriction to the predictable sets of a product measure $\mu \times \nu$ on $[0,1] \times Z$, where almost all the paths of $[N]$ are absolutely continuous with respect to μ. In this case the formula becomes

$$E([N](t) - [N](s) | \mathcal{F}_s) = E(\int_s^t h \, d\mu | \mathcal{F}_s)$$

We now consider the case where N is an n-dimensional martingale, and we assume that there is a measure μ such that for all $i \leq n$ almost all paths of $[N_i]$ are absolutely continuous with respect to μ. By what we have just seen, there exists a predictable $n \otimes n$-process H such that

$$E([N](t) - [N](s) | \mathcal{F}_s) = E(\int_s^t H \, d\mu | \mathcal{F}_s).$$

Let X be a predictable $n \otimes n$-process such that $H = X^t X$ – where ${}^t X$ denotes the transpose of X –, and let Y be a predictable process such that $YX = P_{(\mathrm{Ker}\, X)^\perp}$ and $Y \upharpoonright (\mathrm{Im}\, X)^\perp = 0$ (here $P_{(\mathrm{Ker}\, X)^\perp}$ is the projection on the orthogonal complement of the kernel of X, and $(\mathrm{Im}\, X)^\perp$ is the orthogonal complement to the image of X.) Then $X \cdot Y = P_{\mathrm{Im}\, X}$, and Y is a kind of partial inverse of X. If the measure μ is finite, it follows from the definition of Y that Y is integrable with respect to N.

We shall say that the probability space $\langle Z, \mathcal{F}, \nu \rangle$ is μ-<u>large</u> with respect to the basis $\{\mathcal{F}_t\}$, if there exists a μ-Brownian motion χ° adapted to a basis $\{\mathcal{F}_t'\}$ such that $\mathcal{F}_1$ and $\mathcal{F}_1'$ are independent. If this is the case, we may define the process

$$\chi = \int Y dN + \int P_{\mathrm{Ker}\, X} \, d\chi^\circ.$$

If $\mathcal{H}_t$ is the σ-algebra generated by $\mathcal{F}_t$ and $\mathcal{F}_t'$, we shall show that χ is a μ-Brownian motion adapted to the family $\{\mathcal{H}_t\}$. Obviously χ is a martingale w.r.t. $\{\mathcal{H}_t\}$, and hence it suffices to calculate the quadratic variation:

$$E([\chi](t)-[\chi](s)|\mathcal{H}_s)_{ij} =$$

$$E(\int_s^t \sum_{k,l} Y_{ik}Y_{jl}\, d[N_k,N_l] +$$

$$+\int_s^t \sum_{k,l} Y_{ik}(P_{\mathrm{Ker}\,X})_{jl}\, d[N_k,\chi_l^\circ] +$$

$$+\int_s^t \sum_{k,l} (P_{\mathrm{Ker}\,X})_{ik}\, Y_{jl}\, d[\chi_k^\circ,N_l] +$$

$$+\int_s^t \sum_{k,l} (P_{\mathrm{Ker}\,X})_{ik}\,(P_{\mathrm{Ker}\,X})_{jl}\, d[\chi_k^\circ,\chi_l^\circ]|\mathcal{H}_s) =$$

$$= E(\int_s^t \sum_{k,l} Y_{ik}Y_{jl} \sum_m X_{km}X_{lm}\, d\mu|\mathcal{H}_s)$$

$$+ E(\int_s^t \sum_k (P_{\mathrm{Ker}\,X})_{ik}\,(P_{\mathrm{Ker}\,X})_{jk}\, d\mu|\mathcal{H}_s)$$

$$= E(\int_s^t (YX)^t(YX)d\mu|\mathcal{H}_s)_{ij} + E(\int_s^t P_{\mathrm{Ker}\,X}\,{}^tP_{\mathrm{Ker}\,X}\, d\mu|\mathcal{H}_s)_{ij}.$$

Since $YX = P_{(\mathrm{Ker}\,X)^\perp}$, this proves that $E([\chi](t)-[\chi](s)|\mathcal{H}_s) = (g(t)-g(s))I$, and hence χ is a μ-Brownian motion.

We may now define

$$Z = \int X\,d\chi = \int X\cdot Y\,dN + \int X P_{\mathrm{Ker}\,X}\,d\chi^\circ = \int XY\,dN,$$

and we shall prove that $Z = N-N_o$. Since $XY = P_{\mathrm{Im}\,X}$ is a projection, $\sum_{i=1}^n [Z_i] \leq \sum_{i=1}^n [N_i]$; and if $Z \neq N-N_o$, we must have inequality. But we have

$$E([Z](t)-[Z](s)|\mathcal{H}_s) = E(\int_s^t XY\,d[N]^t(XY)|\mathcal{H}_s) =$$

$$= E(\int_s^t XY\,X^tX^tY^tX\,d\mu|\mathcal{H}_s) = E(\int_s^t X\,P_{(\mathrm{Ker}\,X)^\perp}\,{}^t(X\cdot P_{(\mathrm{Ker}\,X)^\perp})\,d\mu|\mathcal{H}_s)$$

$$= E(\int_s^t X^t X\,d\mu|\mathcal{H}_s) = E([N](t)-[N](s)|\mathcal{H}_s)$$

and hence $Z = N - N_o$.

We have hence proved:

Theorem 4: Let $\langle Z, \mathcal{F}, \nu \rangle$ be a probability space which is μ-large with respect to the basis $\langle Z, \{\mathcal{F}_t\}, \nu \rangle$, and let N be an L^2-martingale adapted to this basis. Assume that the measure μ is finite, and that almost all paths of each $[N_i]$ are absolutely continuous with respect to μ. Let X, Y and χ° be as defined above. Then

$$\chi = \int Y \, dN + \int P_{\mathrm{Ker}\, X} \, d\chi^\circ$$

is a μ-Brownian motion, and

$$N = \int X \, d\chi + N_o .$$

Hence N can be written as a stochastic integral of a μ-Brownian motion.

The proof above is not new; the idea goes back to Doob [4]. But since we have not been able to find exactly the version we need in the literature, and a knowledge of the proof will be useful in the sequel, we have repeated it here.

4. The representation theorem.

From now on Ω shall be a hyperfinite probability space of the kind considered in Keisler [7]; i.e. Ω is of the form Ω_o^T for some hyperfinite set Ω_o, and a hyperfinite time-line T. Let P be the uniform, internal probability measure on Ω, and let $L(P)$ be its Loeb-measure.

If $\omega \in \Omega$ and $t \in T$, let

$$\omega \upharpoonright t = \langle \omega(s) : s \leq t \rangle,$$

and let $\mathcal{G}_t$ be the internal algebra generated by the equivalence relation $\omega \equiv_t \omega' \Leftrightarrow \omega \upharpoonright t = \omega' \upharpoonright t$. Let the stochastic basis $\langle \Omega, \{\mathcal{F}_t\}, L(P) \rangle$ be the one constructed from the internal basis $\langle \Omega, \{\mathcal{G}_t\}, P \rangle$ as in [6], [7], and [9]. Define

$$(\omega \upharpoonright t) = \{\omega' \in \Omega : \omega \upharpoonright t = \omega' \upharpoonright t\}.$$

To be sure that our space $\langle \Omega, L(\mathcal{G}_1), L(P) \rangle$ is μ-large, we shall change it a little: Let Ω_0' be a hyperfinite - but not finite - set, and define

$$\Omega' = \Omega_0^T \times \Omega_0'^T.$$

If $\omega \in \Omega'$, denote its components by ω_1, ω_2, and let $\pi : \Omega' \to \Omega$ be the projection $\pi(\omega) = \omega_1$. Define $\mathcal{G}_t' = \pi^{-1}(\mathcal{G}_t)$, $\mathcal{F}_t' = \pi^{-1}(\mathcal{F}_t)$, and $(\omega \upharpoonright t)' = \pi^{-1}(\omega \upharpoonright t)$. Having done this, we shall forget about the original space Ω; we shall delete the prime and write Ω for Ω', and $\mathcal{G}_t$, $\mathcal{F}_t$ and $(\omega \upharpoonright t)$ will be the objects obtained by applying the definitions above to the new Ω. (The trick of enlarging the probability space is probably unnecessary anyhow, since our space is so enormous; but we don't want to get too far afield by showing it.)

By these definitions it is not hard to see that $\langle \Omega, L(\mathcal{G}_1), L(P) \rangle$ is μ-large with respect to $\{\mathcal{F}_t'\}$.

We shall assume that our time-line is of the form

$$T = \{\tfrac{k}{\eta} : k \in {}^*\mathbb{N},\ 0 \leq k \leq \eta\}$$

for some infinite $\eta = (\gamma!)$. If $H \in {}^*\mathbb{N} \setminus \mathbb{N}, H \leq \gamma$, let

$$T_H = \{0, \frac{1}{H}, \frac{2}{H}, \ldots\ldots, 1\} \subset T.$$

An internal <u>transformation of</u> Ω of <u>mesh</u> $1/H$ is an internal bijection $h : \Omega \to \Omega$ such that if $t \in T_H$, h maps $(\omega|t)$ onto $(h\omega|t)$.

H.J. Keisler has proved (Theorem 9.4 of [7]) that two continuous $\mathcal{F}_t$-Markov processes x and y have the same finite dimensional distributions if and only if there is an internal transformation h of infinitesimal mesh such that

$$y(\cdot,\omega) = x(\cdot,h\omega) \text{ a.e.}$$

Let $M : T \times \Omega \to {}^*\mathbb{R}$ be an S-continuous SL^2-martingale, let $U \in SL^2(M)$, and assume that μ is a finite measure on $[0,1]$ such that almost all paths of each ${}^\circ[M_i]^+ = [{}^\circ M_i^+]$ are absolutely continuous with respect to μ. If $Z = \int U \, dM$, it follows that almost all paths of all ${}^\circ[Z_i]^+ = [{}^\circ Z_i^+]$ are absolutely continuous with respect to μ.

By Theorem 4, there exists μ-Brownian motions χ°, χ, χ' and processes Y, Y', X, X' such that

$$\chi = \int Y \, d\,{}^\circ M^+ + \int P_{\mathrm{Ker}\, X} \, d\chi^\circ$$

$$\chi' = \int Y' \, d\,{}^\circ Z^+ + \int P_{\mathrm{Ker}\, X'} \, d\chi^\circ$$

and

$${}^\circ Z^+ = \int X' \, d\chi'.$$

By Lemma 3 and the theorem by Keisler mentioned above, there is an internal transformation h of infinitesimal mesh such that

$$\chi'(\cdot,\omega) = \chi(\cdot,h\omega) \text{ a.e.}$$

Let $K : T \times \Omega \to T \times \Omega$ be $\mathrm{id} \times h$, then $\chi' = \chi \circ K$, and we get

$${}^\circ Z^+ = \int X' \, d\chi' = \int X' \, d\chi \circ K = \int X' Y \circ K \, d({}^\circ M^+ \circ K) + \int X' P_{\mathrm{Ker}\, X} \circ K \, d(\chi^\circ \circ K).$$

Obviously $°M^+ \circ K$ has the same finite dimensional distributions as $°M^+$, and hence we see that if the last integral above is zero, we can write $°Z^+ = °(\int U\,dM)^+$ as a standard stochastic integral with respect to a process which has the same finite dimensional distributions as M.

One way of making the integral $\int X' P_{\mathrm{Ker}\,X} \circ K\, d(X^{\circ} \circ K)$ zero, is to let $P_{\mathrm{Ker}\,X} = 0$. Hence we say that a martingale $N : [0,1] \times \Omega \to \mathbb{R}^n$ is <u>nondegenerate</u> if there is a finite, continuous measure μ on $[0,1]$ such that almost all paths of each $[N_i]$ are absolutely continuous with respect to μ, and the Radon-Nikodym derivative H defined in the last section has $\det H \neq 0$ almost everywhere. By the theory developed in the last section, this is equivalent to that N can be written as a stochastic integral of μ-Brownian motion, where the integrand is nondegenerate.

Thus we have proved the following theorem:

<u>Theorem 5</u>: Let $M : T \times \Omega \to {}^*\mathbb{R}^n$ be an S-continuous SL^2-martingale adapted to $\langle \Omega, \{\mathcal{G}_t'\}, P \rangle$ such that $°M^+$ is nondegenerate, and let $U \in SL^2(M)$. Then there are a martingale $N : [0,1] \times \Omega \to \mathbb{R}^n$ adapted to $\langle \Omega, \{\mathcal{F}_t\}, L(P) \rangle$ having the same finite dimensional distributions as $°M^+$, and a process $V \in \Lambda^2(N)$ such that

$$°(\int U\,dM)^+ = \int V\,dN.$$

Example 1 shows that we can not in general have $N = °M^+$.

In the calculations leading up to Theorem 5, we have proved the formula

$$°Z^+ = \int X\, Y \circ K\, d(°M^+ \circ K)$$

if $°M^+$ is nondegenerate. Applying K^{-1} on both sides, we get

$$°Z^+ \circ K^{-1} = \int X \circ K^{-1}\, Y\, d°M^+,$$

and we have proved

Theorem 6: Let M and U be as in Theorem 5. Then there is a process $W \in \Lambda^2({}^\circ M^+)$ such that $\int W\, d\,{}^\circ M^+$ has the same finite dimensional distributions as ${}^\circ(\int U\, dM)^+$.

The condition that M should be nondegenerate is not very satisfactory, but in dimension one we can at least make it look a little nicer. In this case the only way it can degenerate is that the Radon-Nikodym derivative becomes zero, and this can not happen if μ is absolutely continuous with respect to almost all the paths, as well as the other way around. Thus we have

Corollary 7: Let $M : T \times \Omega \to {}^*\mathbb{R}$ be an S-continuous SL^2-martingale adapted to $\langle \Omega, \{\mathcal{G}_t'\}, P \rangle$, and suppose that almost all the paths of $[M]$ are mutually absolutely continuous. Let $U \in SL^2(M)$. Then there are a martingale $N : [0,1] \times \Omega \to \mathbb{R}$ adapted to $\langle \Omega, \{\mathcal{F}_t\}, L(P) \rangle$ having the same finite dimensional distributions as ${}^\circ M^+$, and a process $V \in \Lambda^2(N)$ such that

$$ {}^\circ\left(\int U\, dM\right)^+ = \int V\, dN. $$

Moreover, there is a $W \in \Lambda^2({}^\circ M^+)$ such that $\int W\, d\,{}^\circ M^+$ and ${}^\circ(\int U\, dM)^+$ have the same finite dimensional distributions.

But what if the nondegeneracy condition isn't satisfied? The formula

$$ {}^\circ Z^+ = \int X' Y \circ K\, d({}^\circ M^+ \circ K) + \int X' P_{\mathrm{Ker}\, X} \circ K\, d(\chi^\circ \circ K) $$

provides an idea what to look for: Either we must try to show that $\mathrm{Ker}\, X \circ K \subset \mathrm{Ker}\, X'$, or we must find an example where this doesn't hold. Keeping to the latter strategy, we produce

Example 8: Let T, Ω, χ, and X be as in Example 1. Let A be the set

$$A = \{\omega \in \Omega : \sum_{0}^{\frac{1}{2}} X(s,\omega)\Delta\chi(s,\omega) > 0\}.$$

Obviously $L(P)(A) = \frac{1}{2}$. We now define a process $\beta : T \times \Omega \to {}^*\mathbb{R}$ by:

$$\beta(t,\omega) = \chi(t,\omega) \quad \text{for} \quad t \leq \tfrac{1}{2},$$

and for $t \geq \frac{1}{2}$,

$$\Delta\beta(t,\omega) = \Delta\chi(t,\omega) \quad \text{if} \quad \omega \notin A$$

$$\Delta\beta(t,\omega) = 0 \qquad\qquad \text{if} \quad \omega \in A.$$

Let

$$Z = \int X\, d\beta,$$

and assume - for contradiction - that there are an internal transformation K and a process Y such that

$$^{\circ}Z^{+} \circ K = \int Y\, d\,^{\circ}\beta^{+}.$$

Since $t \to {}^{\circ}Z^{+}(t,\omega)$ is constant on $[\frac{1}{2},1]$ if and only if $\omega \in A$ a.s., and the same holds for $^{\circ}\beta^{+}$, we must have $L(P)(K^{-1}(A)\Delta A) = 0$.

By definition of A, the distribution of $^{\circ}Z^{+}(\frac{1}{2})$ on A is the distribution of the absolute value of a gaussian variable with variance $\frac{1}{4}$, and by what we have just seen, this is also the distribution of $^{\circ}Z^{+} \circ K$ on A.

Let us find this last distribution in another way: Define $\beta' = \beta - Z = \int (1-X)d\beta$, and let $\tilde{Y}$ be a 2-lifting of Y:

$$(*) \qquad {}^{\circ}Z^{+} \circ K = \int Y\, d\,^{\circ}\beta^{+} = {}^{\circ}\left(\int \tilde{Y}\, d(Z+\beta')\right)^{+} = {}^{\circ}\left(\int \tilde{Y}\, dZ\right)^{+} + {}^{\circ}\left(\int \tilde{Y}\, d\beta'\right)^{+}.$$

Since Z and β' are independent, the quadratic variation of the process on the right is

$$°(\int_0^t \tilde{Y}^2 \, d[Z])^+ + °(\int_0^t \tilde{Y}^2 \, d[\beta'])^+ = 2 \cdot \tfrac{1}{2} \, °(\int_0^t \tilde{Y}^2(s)ds)$$

and since $[°Z^+ \circ K](t) = \frac{1}{2}t$ for $t \leq \frac{1}{2}$, it follows that

$$|\tilde{Y}(t,\omega)| = \frac{1}{\sqrt{2}} \quad \text{for} \quad t \leq \tfrac{1}{2}.$$

Again since Z and β' are independent, β' is a martingale on A, and hence the expectation of $°(\int \tilde{Y} \, d\beta)^+(\frac{1}{2})$ is zero.

The distribution of $°(\int Y \, dZ)^+(\frac{1}{2})$ on A is part of the distribution of a gaussian random variable with variance $\frac{1}{8}$. Thus

$$E(1_A °(\int Y \, dZ)^+(\tfrac{1}{2})) \leq \int_0^\infty \frac{x}{\sqrt{2\pi \cdot 1/8}} \, e^{-\frac{x^2}{2 \cdot 1/8}} \, dx = \frac{1}{4\sqrt{\pi}}$$

while

$$E(1_A °Z^+ \circ K(\tfrac{1}{2})) = \int^\infty \frac{x}{\sqrt{2\pi \cdot 1/4}} \, e^{-\frac{x^2}{2 \cdot 1/4}} \, dx = \frac{1}{2\sqrt{2n}}.$$

Hence the expectation of the left hand side of (*) over A is $\frac{1}{2\sqrt{2\pi}}$, while the expectation of the right hand side over A is $\frac{1}{4\sqrt{n}}$, contradiction. Thus the internal transformation K which we postulated can not exist, and hence there is no Y', K' such that

$$°Z^+ = \int Y' \, d°\beta^+ \circ K'.$$

But it is not immediately clear that this is a counterexample to Theorem 5 without the nondegeneracy condition, since $°\beta^+$ is not a Markov-process. However, it is not difficult to use Keisler's result to show that if two processes have the same finite dimensional distribution as $°\beta^+$, then one can be brought over into the other by an internal transformation. Hence Theorem 5 is false without the degeneracy condition.

5. A nonstandard Krylov inequality.

Let us give an application of the results of the last section. In [8], N.V. Krylov proved the following inequality:

Proposition 9: For all positive reals M, and every positive integer n, there exists a real k with the following property: Suppose

$$f : [0,1] \times \Omega \to \mathbb{R}^n, \quad g : [0,1] \times \Omega \to \mathbb{R}^n \otimes \mathbb{R}^n$$

are progressively measurable, bounded functions with

$$\|f(t,\omega)\|, \ \|g(t,\omega)\|, \ (\det g(t,\omega))^{-2} \leq M.$$

Let b be an n-dimensional Brownian motion on Ω, and

$$x(t,\omega) = x_0 + \int_0^t f(s,\omega)ds + \int_0^t g(s,\omega)db(s,\omega).$$

Then for any L^{n+1}-function $h : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ with $h(t,x) \geq 0$:

$$E(\int_0^1 h(s,x(s,\omega))ds) \leq k\|h\|_{n+1}.$$

As we shall see in the next section, it would be very useful to have a nonstandard version of this result. Hence we may ask whether the following holds:

Theorem 10: For all positive reals M, and every positive integer n, there exists a real k with the following property:

Suppose

$$F : T \times \Omega \to {}^*\mathbb{R}^n, \quad G : T \times \Omega \to {}^*\mathbb{R}^n \otimes {}^*\mathbb{R}^n$$

are S-bounded, nonanticipating processes with

$$\|F(t,\omega)\|, \ \|G(t,\omega)\|, \ (\det G(t,\omega))^{-2} \leq M.$$

Let χ be an n-dimensional Anderson process, and let

$$X(t,\omega) = x_o + \int_0^t F(s,\omega)ds + \int_0^t G(s,\omega)d\chi(s,\omega).$$

Then for any positive L^{n+1}-function $h : [0,1] \times \mathbb{R}^n \to \mathbb{R}$:

$$E\left(\int_0^1 h(s, {}^\circ X(s,\omega)^+)ds\right) \leq k\|h\|_{n+1}.$$

The idea of the proof is obvious; we write ${}^\circ X^+$ on the form of the process X in Proposition 9 and apply that proposition. It is easy to see that there is a progressively measurable f such that $\int f(s,\omega)ds = {}^\circ(\int F(s,\omega)ds)^+$. By Theorem 5, there is an n-dimensional Brownian motion b, and a process g such that $\int g\,db = {}^\circ(\int G\,d\chi)^+$ (we are confusing the predictable and the progressively measurable sets, but there is no danger in this since we obtain the latter by adding all product measurable null-sets to the former.) Hence we can write

$${}^\circ X^+ = x_o + \int f\,ds + \int g\,db.$$

To apply Proposition 9, we just have to check that we have the right bounds on f and g. The only one that takes a little work is the bound on $|\det g|^{-1}$. We first show that it can be replaced by a notion which is easier to handle:

Lemma 11: Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a positive, symmetric linear map, and let

$$\|A\| = \sup\{\|A\xi\| : \|\xi\| = 1\}$$

$$l(A) = \inf\{\langle \xi, A\xi\rangle, \|\xi\| = 1\}$$

$M(A) = \sup\{m : m$ is a component in a matrix representation of A with respect to an orthonormal basis$\}$.

Then $M(A) = \|A\|$, $l(A) \geq \det(A)/\|A\|^{n-1}$ and $\det A \geq l(A)^n$.

Proof: $M(A)$ and $\|A\|$ are both equal to the largest eigenvalue of A, and $l(A)$ is equal to the smallest. Since $\det A$ is the product of all eigenvalues $l(A)\cdot\|A\|^{n-1} \geq \det A$, and $\det A \geq l(A)^n$.

The lemma shows that as long as we know that g is bounded, we may replace the condition that $\det g^{-1}$ is bounded by the condition that $l(g)^{-1}$ is bounded. This is useful since l satisfies the following superadditivity properly:

Lemma 12: If A and B are two nonnegative definite $n \times n$-matrices, then $l(A+B) \geq l(A) + l(B)$.

The proof is obvious.

Let us now study the properties of the integrand g in the expression $\int g\, db = {}^\circ(\int G\, d\chi)^+$. To find g, we first find a process H such that

$$(*) \qquad E([{}^\circ(\int G\, d\chi)^+](t) - [{}^\circ(\int G\, d\chi)^+](s) \mid \mathcal{F}_s) = E(\int_s^t H\, d\mu \mid \mathcal{F}_s)$$

Since each path of $[{}^\circ(\int G\, d\chi)^+] = {}^\circ[\int G\, d\chi]^+$ is absolutely continuous, we may take H to be the derivative

$$H(s,\omega) = ({}^{\circ}[\int G\,d\chi]^{+})'(s,\omega),$$

since this process obviously satisfies (*), and is progressively measurable. Since $H = g\,{}^{t}g$, it is enough to show that $\|H\|$ and $l(H)^{-1}$ are uniformly bounded. Using the independence of $\Delta\chi_k$ and $\Delta\chi_l$ when $k \neq l$, we get:

$$\begin{aligned}
&([\int G\,d\chi](t) - [\int G\,d\chi](s))_{ij} = \\
&= \sum_{r=s}^{t} \Big(\sum_{k=1}^{n} G_{ik}(r)\Delta\chi_k(r)\Big)\Big(\sum_{l=1}^{n} G_{jl}(r)\Delta\chi_l(r)\Big) \\
&= \sum_{r=s}^{t} \sum_{k,l} G_{ik}(r)G_{jl}(r)\Delta\chi_k(r)\Delta\chi_l(r) \\
&\approx \sum_{r=s}^{t} \sum_{k} G_{ik}(r)G_{jk}(r)\Delta t \\
&= \Big(\int_s^t G\,{}^{t}G\,dt\Big)_{ij}.
\end{aligned}$$

By assumption and Lemma 11, $l(G(r)^{t}G(r))^{-1}$ is uniformly bounded, and by Lemma 12 so is $l(\int_s^t G(r)\,{}^{t}G(r)dr/(t-s))^{-1}$. But by definition, this must also hold for H and g, and by applying Lemma 11 again, we see that $\det g^{-1}$ is uniformly bounded. This proves Theorem 10.

The proof above shows how we can use the representation theorems of the last section to extend standard results to nonstandard situations. Notice that we need not extend the measure-space in this case, since the processes are nondegenerate. Also notice that we have proved the nonstandard result without getting involved in the standard proof; the advantage of this should be clear to anyone who has tried to understand the proof of Krylov's inequality.

6. Keisler's existence theorem.

In [7], H.J. Keisler proved a strong existence result for solution of stochastic differential equations using Krylov's inequality. As he only had the standard version of Proposition 9 at his disposal, Keisler had to rely on a long and rather complicated approximation argument to obtain his result. We shall now see how the nonstandard version of Theorem 10 can be used to simplify his proof.

We first review some notions from Keisler's paper: If $H \in {}^*\mathbb{N}$, an H-<u>element</u> is a subset of ${}^*\mathbb{R}^d$ of the form

$$\{z \in {}^*\mathbb{R}^d : y_k \leq z_k < y_k + \frac{1}{H}, \quad k = 1, \ldots, d\}$$

where $y = (y_1, \ldots, y_d)$ is a point in ${}^*\mathbb{R}^d$ such that each y_i is a multiple of $\frac{1}{H}$. Each H-element in ${}^*\mathbb{R}^d$ has $3^d - 1$ neighbours. The union of a hyperfinite set of H-elements is called an H-<u>set</u>. If $H \in \mathbb{N}$, A is an S-bounded H-set in $[0,1] \times \mathbb{R}^n$, and ω is in the set of measure one where the process X of Theorem 10 is S-continuous, then if ${}^\circ X({}^\circ t, \omega)^+ \in A({}^\circ t, \omega)$, $X(t, \omega)$ is in $A(t, \omega)$ or one of its neighbouring sets. Applying Theorem 10 to the characteristic function of A, we get

$$\Pr\{(t,\omega) \in T \times \Omega : X(t,\omega) \in A(t)\} \leq 3k \cdot {}^*\mu(A)^{1/n+1} + \frac{1}{H},$$

where $\Pr$ is the internal product measure on $T \times \Omega$, and μ is the Lebesgue-measure on $\mathbb{R}^{n+1}$.

For each $r \in \mathbb{N}$, define:

$\mathcal{H}_r = \{H \in {}^*\mathbb{N}$: For all F and G nonanticipating processes $\|F\|$, $\|G\|$, $(\det G)^{-2} \leq M$, we have for all H-sets A bounded by r, that if

$$X(t,\omega) = x_0 + \sum_0^t F(s,\omega)\Delta t + \sum_0^t G(s,\omega)\Delta\chi(s,\omega)$$

then

$$Pr\{(t,\omega) : X(t,\omega) \in A(t,\omega)\} \leq 3k^*\mu(A)^{1/n+1} + \frac{1}{H}\}.$$

By the internal definition principle there must for each $r \in \mathbb{N}$ be an infinite $H \in \mathcal{H}_r$, and by saturation we may find an infinite H which is in all of them.

We may now prove Keisler's theorem:

Theorem 13: Suppose $f : [0,1] \times \mathbb{R}^n \to \mathbb{R}^n$, $g : [0,1] \times \mathbb{R}^n \to \mathbb{R}^n \otimes \mathbb{R}^n$ are bounded, measurable functions, and that $|\det g|^{-2}$ is uniformly bounded. Let b be the standard part of an Anderson process with values in ${}^*\mathbb{R}^n$. Then the equation

$$x(t,\omega) = x_0 + \int_0^t f(s,x(s,\omega))ds + \int_0^t g(s,x(s,\omega))db(s)$$

has a continuous solution.

Proof: Let M be a bound on $\|f\|$, $\|g\|$ and $|\det g|^{-2}$, and let $H \in {}^*\mathbb{N}-\mathbb{N}$ be the hyperfinite number constructed above. Let $F : T \times {}^*\mathbb{R}^n \to {}^*\mathbb{R}^n$, $G : T \times {}^*\mathbb{R}^n \to {}^*\mathbb{R}^n \otimes {}^*\mathbb{R}^n$ be H-liftings of f and g respectively (i.e. liftings constant on H-elements).

Consider the process

$$X(t,\omega) = x_0 + \int_0^t F(s,X(s))ds + \int_0^t G(s,X(s,\omega))d\chi(s,\omega),$$

and let x be a progressively measurable, continuous process having X as a uniform lifting. To prove that x is a solution of our stochastic differential equation, it is enough to show that $F(s,X(s,\omega))$ and $G(s,X(s,\omega))$ are liftings of $f(s,x(s,\omega))$ and $g(s,x(s,\omega))$ respectively, since then

$$°\int F(s,X(s,\omega))ds = \int f(s,x(s,\omega))ds \quad \text{and} \quad °\int G(s,X(s,\omega))d\chi = \int g(s,x(s,\omega))db.$$

But $°F$ and $f \circ st$ differs only on a null-set in the Loeb-algebra generated by the H-sets, and the same holds for $°G$ and $g \circ st$. But the probability that X shall be in such a set is zero according to the definition of H. This proves the theorem.

Since the processes in the last two sections all are nondegenerate and have quadratic variations absolutely continuous with respect to the Lebesgue-measure, we need only a very simple version of the theory of sections 3 and 4. Hence rather short, direct proofs of Krylov's inequality and Keisler's theorem are possible (see the forthcoming book by Albeverio, Fenstad, and Høegh-Krohn [1]).

7. The power of nonstandard stochastic integration.

We have so far mainly been concerned with reducing hyperfinite stochastic integrals to standard stochastic integrals, but our examples have shown that this is not always possible, and when it <u>is</u> possible, only in a rather indirect way. In this section we shall try to show by an example why we believe that the extra power of the nonstandard theory will be of importance in the mathematical modeling of statistical phenomena.

We shall look at models for Brownian motion. A well-known way of modeling these phenomena, is to start with a sequence of independent random variables, and then use Donsker's theorem (see Billingsley [3], or - for a nice nonstandard proof - Anderson [2]) to obtain a Brownian motion process. Thus let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables on a probability space (Ω, P) taking each of the values ± 1 with probability $\frac{1}{2}$. Define the stochastic process $\chi_n : [0,1] \times \Omega \to \mathbb{R}$ by

$$\chi_n(t,\omega) = \sum_{k=1}^{[nt]} \frac{\xi_k(\omega)}{\sqrt{n}} + (nt - [nt]) \frac{1}{\sqrt{n}} \xi_{([nt]+1)}(\omega)$$

Then χ_n converges in distribution in $C([0,1])$ to a Brownian motion χ.

In considering the physical aspects of this model, it does not seem improbable that we should come across processes of the form

$$\tilde{\chi}_n = \int X_n \, d\chi_n$$

where

$$X_n(t,\omega) = 1 \quad \text{if} \quad t \in \left[\frac{j}{n}, \frac{j+1}{n}\right\rangle \quad \text{for some even} \quad j \in \mathbb{N},$$

and

$$X_n(t,\omega) = 0 \quad \text{if} \quad t \in \left[\frac{j}{n}, \frac{j+1}{n}\right\rangle \quad \text{for some odd} \quad j \in \mathbb{N}.$$

It is natural to represent these processes in the limit model as the limit in distribution of the $\tilde{\chi}_n$. We denote this limit process by $\tilde{\chi}$.

<u>Proposition 14</u>: Let Y be integrable with respect to χ, then

$$E\Big(\big(\tilde{\chi}(1) - \int_0^1 Y \, d\chi\big)^2\Big) \geq \frac{1}{4}.$$

Proof: By definition of the stochastic integral, there is for each given $\varepsilon > 0$ a process $\tilde{Y}$ of the form

$$\tilde{Y} = \sum_{i=1}^{n} a_i \cdot 1_{F_{s_i}} \chi\langle s_i, t_i]$$

such that

$$\left\| \int_0^1 Y \, d\chi - \int_0^1 \tilde{Y} \, d\chi \right\|_2 < \varepsilon .$$

Since χ_n and $\tilde{\chi}_n$ converges in distribution to χ and $\tilde{\chi}$ respectively, there must be an n such that

$$\left| \|\tilde{\chi}_n - \int \tilde{Y} \, d\chi_n\|_2 - \|\tilde{\chi} - \int \tilde{Y} \, d\chi\|_2 \right| < \varepsilon$$

for $n > n_o$.

But taking this n large enough, we can get X_n to change value as often as we wish without $\tilde{Y}$ changing. Since the least value of $x^2 + (1-x)^2$ is $\frac{1}{2}$ for $x = \frac{1}{2}$, we have for large enough n :

$$\|\tilde{\chi}_n - \int \tilde{Y} \, d\chi_n\|_2^2 = \|\int (X_n - \tilde{Y}) d\chi_n\|_2^2 \geq \|\int (X_n - \tilde{Y})^2 ds\|_1^2 - \varepsilon \geq \frac{1}{4} - 2\varepsilon .$$

Putting these results together, we have

$$E((\tilde{\chi}(1) - \int_0^1 \tilde{Y} \, d\chi)^2) \geq (\sqrt{\tfrac{1}{4} - 2\varepsilon} + 2\varepsilon)^2 ,$$

and since $\varepsilon > 0$ is arbitrary, the proposition follows.

This result tells us that it is impossible to obtain $\tilde{\chi}$ as a stochastic integral or as a limit of stochastic integrals of χ. What we <u>can</u> do, is to leave the limit-model, go back to the approximations, and there write $\tilde{\chi}_n$ as $\int X_n d\chi_n$. That this simple procecedure can not be reflected within the model, seems to point at a weakness of the limit construction.

The nonstandard model of this phenomenon would clearly be the one given in Example 1; the χ of that example corresponding to the χ above, and $\int X \, d\chi$ corresponding to $\tilde{\chi}$. The result of the example is of course just another version of Proposition 14; we can not obtain $^{\circ}(\int X \, d\chi)^{+}$ as a standard stochastic integral of χ. But by definition, $\int X \, d\chi$ is a <u>nonstandard</u> stochastic integral of χ. Thus the nonstandard model faithfully represents more properties of the approximations than does the standard model, and this reflects the extra power of the hyperfinite stochastic integral. Let us end by hoping for a more effective example than the one above; an example where the importance of the processes $\tilde{\chi}_n$ are not only postulated but shown, and where the extra expressive power is put to good use.

References

1. S. Albeverio, J.E. Fenstad, and R. Høegh-Krohn: Forthcoming book.

2. R.M. Anderson: A nonstandard representation of Brownian motion and Itô integration. Israel J. Math. 25 (1976), 15-46.

3. P. Billingsley: Convergence of probability measures, John Wiley and Sons, 1968.

4. J.L. Doob: Stochastic processes, John Wiley and Sons, 1953.

5. J.E. Fenstad: Nonstandard methods in stochastic analysis and mathematical physics, To appear, Jahr. Deut. Math. Verein.

6. D.N. Hoover and E. Perkins: Nonstandard construction of the stochastic integral and applications to stochastic differential equations I, II, Preprints, 1980.

7. H.J. Keisler: An infinitesimal approach to stochastic analysis, University of Wisconsin preprint, 1980.

8. N.V. Krylov: Some estimates of the probability density of a stochastic integral. Math. of the USSR-Izvestia, Vol 8 (1974), pp. 233-254.

9. T.L. Lindstrøm: Hyperfinite stochastic integration I, II, III, To appear, Math. Scand 46 (1980).

10. T.L. Lindstrøm: Stochastic integration in hyperfinite dimensional linear spaces, Oslo Preprint 1980.

11. P.A. Loeb: An introduction to nonstandard analysis and hyperfinite probability theory, in A.T. Bharucha-Reid (ed): Probabilistic analysis and related topics, Vol. 2, Academic Press 1979, 105-142.

12. M. Métivier: Reelle und vektorwertige Quasimartingale und die Theorie der stochastischen Integration, LNM 607, Springer-Verlag, 1977.

13. K.D. Stroyan and W.A.J. Luxemburg: Introduction to the theory of infinitesimals, Academic Press, 1976.