PAPER • OPEN ACCESS

On the interpretation of Stratonovich calculus

View the article online for updates and enhancements.

On the interpretation of Stratonovich calculus

W Moon[1,3] and J S Wettlaufer[2,3]

[1] Institute of Theoretical Geophysics, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, CB3 0WA, UK
[2] Mathematical Institute University of Oxford, Oxford OX2 6GG, UK
[3] Yale University, New Haven, CT 06520-8109, USA
E-mail: wm275@damtp.cam.ac.uk and wettlaufer@maths.ox.ac.uk

Abstract

The Itô–Stratonovich dilemma is revisited from the perspective of the interpretation of Stratonovich calculus using shot noise. Over the long time scales of the displacement of an observable, the principal issue is how to deal with finite/zero autocorrelation of the stochastic noise. The former (non-zero) noise autocorrelation structure preserves the normal chain rule using a mid-point selection scheme, which is the basis Stratonovich calculus, whereas the instantaneous autocorrelation structure of Itô's approach does not. By considering the finite decay of the noise correlations on time scales very short relative to the overall displacement times of the observable, we suggest a generalization of the integral Taylor expansion criterion of Wong and Zakai (1965 *Ann. Math. Stat.* **36** 1560–4) for the validity of the Stratonovich approach.

Keywords: stochastic differential equations, Stratonovich calculus, noise

1. Introduction

Stochastic dynamical models are basic to the understanding of the role of random forcing in a wide range of scientific and engineering systems (e.g. [2–4]). The central theoretical approaches arise from the Einstein and Langevin studies of Brownian motion [5, 6], which provide mathematically different but physically equivalent and complimentary descriptions of the fate of a body (a pollen particle in water observed under a microscope) under the influence of

random non-deterministic collisions (water molecules). Einstein determined the time evolution of the probability density of particles by solving the Fokker–Planck equation whereas Langevin wrote down an explicit deterministic momentum equation for a particle augmented by a Gaussian white noise forcing which perturbs the particle trajectory. The approach of Langevin now constitutes a canonical 'stochastic differential equation' (SDE) for a daunting scope of systems, but since its introduction the mathematical and physical interpretation of the noise term has been discussed and debated.

Whether viewing the problem from configuration space through the solution of the Fokker–Planck equation, solving a Langevin equation, or generating statistical realizations by performing Monte-Carlo simulations of the particles under consideration (e.g. [7]), a consistent interpretation and calculational scheme of the noise structure is needed. The approach developed by Itô [8] rests upon the Markovian and Martingale properties; the former captures the concept of a 'memoryless' process, wherein the conditional probability distribution of future states depends solely on the present state and the latter that, given all prior events, the expectation value of future stochastic events equals the present value. These properties have the advantage of simplifying many complicated time integrals but the disadvantage of requiring a new calculus which does not obey the traditional chain rule. In contrast, the approach of Stratonovich [9] does not invoke the Martingale property and preserves the chain rule and allows white noise to be treated as a regular derivative of a Brownian (or Weiner) process, W_t.

A model SDE (to which we shall return later) for a variable $x(t)$ is

$$\frac{dx}{dt} = a(x, t) + b(x, t)\frac{dW_t}{dt}, \tag{1}$$

wherein the first term is 'deterministic' and the second term is 'stochastic'. The case in which $b(x, t)$ is constant is referred to as additive noise and when it depends on $x = x(t)$ it is multiplicative noise. Now, despite being able to integrate equation (1) in a formal sense as

$$x(t) - x(t_0) = \int_{t_0}^{t} a(x, t')\,dt' + \int_{t_0}^{t} b(x, t')\,dW_t(t'), \tag{2}$$

the crux of the 'Itô–Stratonovich dilemma' resides in the integral over the Brownian process W_t. The issue is laid bare by recalling the definition of the definite integral of a real valued function $f(t)$ in terms of the Riemann sum viz.,

$$\int_{t_0}^{t} f(t')\,dt' \equiv \lim_{n\to\infty}\sum_{k=0}^{n-1} f(\tilde{t}_k)(t_{k+1} - t_k), \tag{3}$$

where importantly $\tilde{t}_k \in [t_k, t_{k+1}]$. The issues are (i) the Brownian process W_t is nowhere differentiable *and* because $\frac{dW_t}{dt}$ is δ-autocorrelated then in any interval on the real line the white noise it represents fluctuates an infinite number of times with an infinite variance, and (ii) the defining limit of the integral in equation (3) depends on the place in $[t_k, t_{k+1}]$ where $\tilde{t}_k$ is chosen; the choice that provides the seed of the dilemma and the origin of the two different calculi.

In the Itô approach the choice is $\tilde{t}_k = t_k$, which maintains the Martingale property due to the fact that t_k is the present value in the integrand, thereby forcing the expectation value in $[t_k, t_{k+1}]$ to take the present value. In contrast, the Stratonovich approach defines $\tilde{t}_k = (t_k + t_{k+1})/2$, which abandons the Martingale property and hence maintains the normal

rules of calculus. Stratonovich referred to his choice as a 'symmetrization' between past and future. Despite its wide usage, we have not found a physical interpretation of the basis of the past/future symmetrization of Stratonovich calculus.

The outline of this paper is as follows. In the next section we summarize our approach and discuss its connection to the dilemma generally and to other studies. We then outline the conventional physical science viewpoint and the perspective of stochastic calculus, before coming to our main point. Finally, we discuss the relationship to other approaches before concluding.

2. A brief comparison and contrast

Wong and Zakai [1] argued that in any real world system perfect white noise does not exist and that Brownian motion $x(t)$ approximates a description $x_n(t)$ that is continuous with at least a piece-wise continuous derivative. By showing that $x_n(t) \to x(t)$ as $n \to \infty$ they recovered Stratonovich calculus. Accordingly, the choice of stochastic calculus resides in the characteristics of the noise and continuity arguments. In finance, it is argued that short time scale processes are truly discontinuous and thus Itô calculus is preferred (e.g. [10]), thereby maintaining the Martingale property. However, in physics the continuous motion of Brownian particles influenced by high frequency white noise has long been considered within the framework of normal calculus [11]. Conceptually, there is no clear distinction between the statistics of water molecules colliding pollen grains and trading options or stocks. Hence, it remains an open exercise to answer the questions of when, how and whether to use continuity considerations as a core criterion to choose either of the calculi being discussed here.

As noted in the introduction, the complimentary approaches of Einstein and Langevin, with a reliance on δ-autocorrelated noise, provide consistent testable predictions of Brownian motion. The combination of this and the additional consistency with Stratonovich calculus led to the suggestion that physical scientists should avoid Itô calculus [12]. Although Van Kampen [13] cautiously suggested that the Langevin equation is intrinsically insufficient for representing systems with internal noise, it is still the case that Stratonovich calculus is appropriate for both internal and external noise and there are an enormous number of problems in which one cannot make this distinction.

Here we focus on an ostensibly physical argument to discuss the origin of Stratonovich calculus. We generalize the theorem Wong and Zakai [1] using the integral Taylor expansion and appealing solely to the L^2 integrability of a function, thereby avoiding discussions of the regularity of a stochastic process. Because the principal difference between the Itô–Langevin and the Stratonovich–Langevin formulations lies in the drift term, we focus on this in our approach and examine the intrinsic nature of short time scale processes approximated by white noise to decide which calculus is more appropriate. As opposed to Turelli [14], who took as a continuous deterministic model an approximation of a discretized model (in population dynamics), we begin with a well defined deterministic model.

3. The conventional physical science perspective

The term 'white noise' refers to a random signal $\Gamma(t)$ in which all frequencies contribute equally to the power spectral density. For a general stationary random process the autocorrelation $\mathcal{R}(t' - t) = \langle \Gamma(t')\Gamma(t) \rangle$, where $\langle \cdot \rangle$ is the ensemble average, is by definition independent of time and symmetric/even $\mathcal{R}(t' - t) = \mathcal{R}(t - t')$. If $t - t' \equiv \tau$, then let τ_c be the de-correlation time such that $\mathcal{R}(\tau_c) = e^{-1}\mathcal{R}(0)$. Typically, as $\tau_c \to 0$ and $\int_{-\infty}^{\infty} \mathcal{R}(\tau_c)\,d\tau_c$ remains finite, we say that $\mathcal{R}(\tau_c) = \delta(0)$ and the white noise has a δ-autocorrelation structure.

    In science and engineering the traditional form of the Langevin equation, slightly different than that in equation (1), is written as

$$\frac{dx}{dt} = a(x, t) + b(x, t)\Gamma(t), \tag{4}$$

where the noise forcing is δ-autocorrelated as $\langle \Gamma(t')\Gamma(t) \rangle = \delta(t' - t)$. Based on this, we calculate $\langle \frac{dx}{dt} \rangle$ starting with the first integral

$$x(t + \tau) - x(t) = \int_{t}^{t+\tau} \left(a\left[x(t'), t' \right] + b\left[x(t'), t' \right]\Gamma(t') \right) dt'. \tag{5}$$

We assume that τ is small relative to the time over which the macroscopic system evolves, but large relative to the time scale of fluctuations in the stochastic process $\Gamma(t)$. When $a(x, t)$ and $b(x, t)$ are analytic, we can Taylor expand them about a time t as

$$a(x(t'), t') = a(x(t), t) + \frac{\partial a}{\partial x}(x(t') - x(t)) + \frac{\partial a}{\partial t}(t' - t) + \ldots$$

$$b(x(t'), t') = b(x(t), t) + \frac{\partial b}{\partial x}(x(t') - x(t)) + \frac{\partial b}{\partial t}(t' - t) + \ldots. \tag{6}$$

following Risken [15]. Inserting (6) into (5) leads to

$$x(t + \tau) - x(t) = a(x(t), t)\tau + \frac{\partial a}{\partial x}\int_{t}^{t+\tau}(x(t') - x(t))\,dt' + \frac{\partial a}{\partial t}\frac{1}{2}\tau^2$$

$$+ b(x(t), t)\int_{t}^{t+\tau}\Gamma(t')\,dt' + \frac{\partial b}{\partial x}\int_{t}^{t+\tau}(x(t') - x(t))\Gamma(t')\,dt'$$

$$+ \frac{\partial b}{\partial t}\int_{t}^{t+\tau}(t' - t)\Gamma(t')\,dt' + h.o.t. \tag{7}$$

For $x(t') - x(t)$, we can treat equation (5) iteratively and then take the ensemble average, which is

$$\langle x(t + \tau) - x(t) \rangle = a(x, t)\tau + b\frac{\partial b}{\partial x}\int_{t}^{t+\tau}dt'\int_{t}^{t'}\langle \Gamma(t'')\Gamma(t') \rangle\,dt'' + O(\tau^2). \tag{8}$$

Therefore,

$$\lim_{\tau \to 0}\frac{1}{\tau}\langle x(t + \tau) - x(t) \rangle = a(x, t) + b\frac{\partial b}{\partial x}\frac{1}{\tau}\int_{t}^{t+\tau}dt'\int_{t}^{t'}\langle \Gamma(t'')\Gamma(t') \rangle\,dt''. \tag{9}$$

By the definition of white noise discussed above, $\left\langle \Gamma\left(t'\right)\Gamma\left(t''\right)\right\rangle = \delta\left(t'-t''\right)$, we have

$$\int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \left\langle \Gamma\left(t''\right)\Gamma\left(t'\right)\right\rangle \mathrm{d}t'' = \int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \delta\left(t''-t'\right)\mathrm{d}t''. \tag{10}$$

In the integral $\int_t^{t'} \delta\left(t''-t'\right)\mathrm{d}t''$, half of the contribution of the delta function is considered such that the value of the integral becomes $\frac{1}{2}$. Therefore, $\int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \left\langle \Gamma\left(t''\right)\Gamma\left(t'\right)\right\rangle \mathrm{d}t''$ is equal to $\frac{1}{2}\tau$. What is the physical meaning of the half contribution of the delta function? The entire interval $\left[t, t'\right]$ does not contribute to the integral $\int_t^{t'} \delta\left(t''-t'\right)\mathrm{d}t''$. We shrink the interval near t' as $\left[t'-\epsilon, t'\right]$, where ϵ can be interpreted as the size of a subinterval in the Riemann summation discussed above in distinguishing the stochastic calculi. Because the delta function is even $\delta\left(t''-t'\right) = \delta\left(t'-t''\right)$, the half contribution can be written as

$$\int_t^{t'} \delta\left(t''-t'\right)\mathrm{d}t'' = \frac{1}{2}\int_{t'-\epsilon}^{t'+\epsilon} \delta\left(t''-t'\right)\mathrm{d}t''. \tag{11}$$

Despite this being just a simple mathematical modification, we can interpret it as a representation that the weighting from past is equal to that from future in the neighbourhood of t'. The origin of this symmetry comes from the time symmetry of the autocorrelation function $\mathcal{R}\left(\tau\right)$. We will connect this half contribution of the autocorrelation to the Stratonovich integral below, but we note now the above development yields

$$\left\langle \frac{\mathrm{d}x}{\mathrm{d}t} \right\rangle = a\left(x, t\right) + \frac{1}{2}b\left(x, t\right)\frac{\partial}{\partial x}b\left(x, t\right), \tag{12}$$

which is exactly same as that obtained from Stratonovich calculus.

4. The stochastic calculus perspective

Mathematicians have long questioned the validity of the conventional definition of white noise because Brownian motion W_t is nowhere differentiable. Namely, defining $\Gamma\left(t\right)$ as $\frac{\mathrm{d}W_t}{\mathrm{d}t}$ is poorly grounded in normal calculus and we must revisit the interpretation of equation (4) in light of the mathematical definition of Brownian motion W_t;

(i) $W_0 = 0$

(ii) W_t is almost surely continuous (or sample-continuous) with respect to t

(iii) W_t has independent increments and $W_t - W_s$, with $s < t$, is governed by a normal distribution, $N\left(0, t-s\right)$, with zero mean and standard deviation $\sqrt{t-s}$.

Therefore, equation (4) can be rewritten as

$$\mathrm{d}x = a\left(x, t\right)\mathrm{d}t + b\left(x, t\right)\mathrm{d}W_t. \tag{13}$$

On the one hand, using this integral form avoids the issue of the differentiability of W_t, but this comes at the cost of having to introduce a different sort of calculus. However, before coming to the issue of the implementation of a particular stochastic calculus, we perform the same expansion as discussed previously, in (7) and (8), which leads to

$$\langle x(t+\tau) - x(t) \rangle = a(x,t)\tau + b(x,t)\frac{\partial}{\partial x}b(x,t)\left\langle \int_t^{t+\tau}\int_t^{t'} dW_t'' dW_t' \right\rangle + O(\tau^2). \quad (14)$$

The first term $a(x,t)$ can be interpreted as the deterministic response to macro-scale forcing and the second term embodies the interaction between a macroscopic quantity $b(x,t)$ and the microscopic accumulation of random noise forcing. On a time τ that is short from the macroscopic perspective, the interaction between two scales manifests itself multiplicatively. The contribution of the micro-scale to the change of the macroscopic quantity x appears as a double integral of the Brownian motion. It is at this point that we must decide upon a specific calculus, and this choice determines the value of the double integral in equation (14).

The double integral can be written as $\int_t^{t+\tau} W' dW'$ and approximated by a Riemann sum as

$$\int_t^{t+\tau} W' dW' = \sum_{k=0}^{N-1} W_{t_k'}\Delta W_{t_k}, \quad (15)$$

where $\Delta W_{t_k} \equiv W_{t_{k+1}} - W_{t_k'}$. As should be intuitive from the previous discussion, the confusion lies in the choice of t_k' for $W_{t_k'}$. As can be seen below, this choice determines the average of the integral viz.,

$$\left\langle \sum_{k=0}^{k=N-1} W_{t_k'}\Delta W_{t_k} \right\rangle = \sum_{k=0}^{k=N-1} \left\langle \left(W_{t_k'} - W_{t_k} + W_{t_k} \right)\left(W_{t_{k+1}} - W_{t_k} \right) \right\rangle$$

$$= \sum_{k=0}^{k=N-1} \left\langle \left(W_{t_k'} - W_{t_k} \right)\left(W_{t_{k+1}} - W_{t_k} \right) \right\rangle$$

$$= \sum_{k=0}^{k=N-1} \left(t_k' - t_k \right). \quad (16)$$

Itô [8] argued that by choosing $t_k' = t_k$ the most important characteristic preserved is the Martingale property. This implies that the expectation value of a future event is equivalent to that of the present value in the subinterval $[t_k, t_{k+1}]$. Hence, the expectation value of the double integral is zero. The disadvantage of Itô calculus is that the chain rule of normal calculus is lost; one needs additional terms to preserve the Martingale property during integration. On the other hand, Stratonovich [9] chose $t_k' = \frac{t_k + t_{k+1}}{2}$, and the integral has a value of $\frac{1}{2}\tau$. This approach preserves the chain rule, implying that Stratonovich calculus must be used when we approximate noise processes that are continuous and have piece-wise continuous derivatives (e.g. [1]). When we approximate the rapid fluctuations of a stochastic process as white noise, then we must preserve the normal rules of calculus if we are to capture the dynamics of Brownian motion.

5. Autocorrelation structure and stochastic calculi: a simple example

In a classical physical world where the continuity is more or less guaranteed, Stratonovich calculus is evidently the most appropriate. However, when we consider the colliding of particles or the buying or selling of stocks or options, these short time-scale phenomena *appear* to be discontinuous. Such circumstances lead us to question the validity of the continuity of a

stochastic process. That said, it may be impossible to find events that have zero autocorrelation. Thus it is a natural question to examine the effect of a non-zero autocorrelation in a multiplicative noise process by investigating limiting behaviors, which may provide some insight into the nature of the appropriate stochastic calculus. Hence we consider the simplest case of multiplicitive behavior; $dx/dt = xF(t)$, where $F(t)$ is a stochastic process with a finite decorrelation time. For this example, using the same procedure used to arrive at equation (8), we find that

$$\left\langle x\left(t + \frac{\tau}{n}\right) - x(t) \right\rangle = x(t) \int_t^{t+\frac{\tau}{n}} \int_t^{t'} \left\langle F(t'')F(t') \right\rangle dt''dt'. \tag{17}$$

An important feature of our argument involves the time τ/n, with integer n, which is a subdivision of the total domain $[t, t + \tau]$. For example, consider the commonly used autocorrelation $\left\langle F(t'')F(t') \right\rangle = \frac{\sigma^2}{2\tau_c}e^{-|t''-t'|/\tau_c}$, where τ_c is the decorrelation time of the amplitude of the stochastic noise $F(t)$ satisfying $\tau_c \ll \tau$. Now, let $\tau_m \equiv \tau/n$ be a minimum time over which we can observe the decay of the stochastic noise.

Whereas it is always the case that $\tau \gg \tau_c$, the value of τ_m depends on the fineness n of the subdivision, with τ_m *increasing* as n *decreases*; coarsening the temporal resolution. So long as $\tau_m \gg \tau_c$ we can still justify the use of the stochastic differential equation but once $\tau_m \leqslant \tau_c$ this is no longer the case. Therefore, increasing τ_m to values above τ_c but still small relative to τ one reaches a value of $\tau_m \equiv \tau_\star$ that is sufficiently short that the decay of the noise can be observed but sufficiently large so that the stochastic differential equation is valid. This is achieved as follows. Consider n such that $\tau \gg \tau_m = \tau_c$, which we rewrite as $1 \gg \frac{1}{n} = \frac{\tau_c}{\tau}$. Now, as n decreases from $\frac{\tau}{\tau_c}$ it will reach a value $n = n_\star$ such that $1 \gg \frac{1}{n_\star} \gg \frac{\tau_c}{\tau}$ and our analysis is valid for $n_\star \geqslant n \gg 1$ as described presently.

With the above form of the autocorrelation, equation (17) becomes

$$\left\langle x\left(t + \frac{\tau}{n}\right) - x(t) \right\rangle = x(t) \times \frac{1}{2}\sigma^2 \frac{\tau}{n}\left[1 + \frac{\tau_c}{\tau/n}\left(e^{-\frac{\tau}{n\tau_c}} - 1\right)\right], \tag{18}$$

which for $n_\star \geqslant n \gg 1$ can be approximated as $\left\langle x\left(t + \frac{\tau}{n}\right) \right\rangle = x(t)\left(1 + \frac{1}{2}\sigma^2\frac{\tau}{n}\right)$ and then rewritten as

$$\left\langle x(t + \tau) \right\rangle = x(t)\left(1 + \frac{1}{2}\sigma^2\frac{\tau}{n}\right)^n. \tag{19}$$

The limit $n_\star \geqslant n \gg 1$ insures that the right hand side converges to $x(t)e^{\frac{1}{2}\sigma^2\tau}$ and hence

$$\left\langle x(t + \tau) \right\rangle = x(t)e^{\frac{1}{2}\sigma^2\tau} \approx x(t)\left(1 + \frac{1}{2}\sigma^2\tau\right), \tag{20}$$

consistent with the result from Stratonovich calculus, in which $\frac{1}{2}\sigma^2\tau$ is analogous to $\left\langle W_t\Delta W_t \right\rangle$ in the Riemann sum with $W_t = W_{(t_k+t_{k+1})/2}$ and $\Delta W_t = W_{t_{k+1}} - W_{t_k}$. This is the essence of Stratonovich calculus; choosing the mid-point in a subinterval is equivalent to including only *half of the effect of a finite autocorrelation*. Clearly, in this example, the magnitude of τ_c does not play a role unless $\tau_m \searrow \tau_c$ and the neccessary separation of time scales starts to be violated. Therefore, in

this case the commonly-used definition of white noise $\langle F(t'')F(t')\rangle \propto \delta(t''-t')$ becomes an increasingly better approximation, which we can understand as follows.

Recall that

$$\int_t^{t'} \frac{\sigma^2}{2\tau_c} e^{-\frac{|t'-t'|}{\tau_c}} dt' \approx \int_t^{t'} \sigma^2 \delta(t''-t') dt'' \sim O(\sigma^2). \tag{21}$$

We can proceed along the same lines as above but without the need to invoke a specific form of the autocorrelation. Rather, we only need to assume that $F(t)$ is stationary,

$$\int_t^{t+\frac{\tau}{n}} \int_t^{t'} \langle F(t'')F(t')\rangle dt'' dt' = \int_{-\tau/n}^0 ds \int_{t-s}^{t+\tau/n} K(s) dt' = \int_{-\tau/n}^0 K(s)(\tau+s) ds$$
$$\approx \frac{1}{2}\frac{\tau}{n} \int_{-\infty}^{\infty} K(s) ds, \tag{22}$$

where $K(s) = \langle F(t)F(t+s)\rangle$ and we exploited $\tau/n \gg \tau_c$, thereby neglecting s compared to τ/n (see e.g. chapter 15 of [16]). Here, $\frac{1}{2}\frac{\tau}{n}\int_{-\infty}^{\infty} K(s) ds$ is an approximation of $\langle W_t \Delta W_t \rangle$ in the Riemann sum, where the appropriate form of stochastic calculus is determined by the magnitude of $\int_{-\infty}^{\infty} K(s) ds$. Namely, if

$$\int_{-\infty}^{\infty} K(s) ds \sim O(\sigma^2), \tag{23}$$

where $\langle F^2(t)\rangle = \sigma^2$, the determination of $\langle W_t \Delta W_t \rangle$ follows Stratonovich calculus and hence choosing a δ-function for $K(s)$ is appropriate. However, if

$$\int_{-\infty}^{\infty} K(s) ds \ll \sigma^2, \tag{24}$$

then $\frac{1}{2}\frac{\tau}{n}\int_{-\infty}^{\infty} K(s) ds$ is negligible and the determination of $\langle W_t \Delta W_t \rangle$ follows Itô-calculus. Here, $K(s)$ can be considered approximately as a zero-measure function in L^2 space such that a δ-function cannot be used for the approximation of the autocorrelation function $K(s)$. This argument is similar to Morita's [17], who also considered three time scales and his choice of stochastic calculus also relies on the condition for the autocorrelation function, but as manifests itself in the form of a generalized Fokker–Planck equation.

The choice of Stratonovich calculus for colored noise has been considered within the framework of a master equation by Horsthemke and Lefever [18]. Colored noise converges to Gaussian white noise through the application of the central limit theorem, which they argued through a rescaling of time (Δt) and space (Δx) assuming a diffusive process, viz., $\Delta t \to \Delta t/\epsilon^2$, $\Delta x \to \Delta x/\epsilon$. In the limit $\epsilon \to 0$ the master equation becomes the Fokker–Planck equation with Gaussian white noise interpreted using Stratonovich calculus. Horsthemke and Lefever note that the interpretation of Stratonovich calculus in the approach of Wong and Zakai [1] requires the strong regularity of colored noise (a C_1 function) for convergence to Gaussian white noise. However, Wong and Zakai [1] demonstrate the convergence directly from an SDE, whereas Horsthemke and Lefever [18] avoid the issue of regularity by beginning with an Uhlenbeck–Ornstein process, which is not C_1, and demonstrate convergence using a master

equation, but do not connect the finite autocorrelation of colored noise with a physical interpretation that underlies the construction of Stratonovich calculus.

6. Shot noise, white noise and the origin of Stratonovich calculus

It is the mid-point selection rule in the Riemann sum that is associated with the finite noise autocorrelation. However, traditionally this selection rule has been interpreted as the 'magical' choice required in order to recover the normal rules of calculus, but Stratonovich calculus itself does not provide any new dynamical insight that plays a role analogous to the Martingale property of Itô calculus. Hence, we seek a basic understanding of how the mid-point selection rule emerges as a *consequence* of applying the central limit theorem to colored noise. Our approach is to consider shot noise, which has an autocorrelation represented by a δ-function.

What is the origin of the δ-function? Brownian motion can be generated by a collection of independent random processes. In the Riemann-summation approximation of the integral $\int_t^{t+\tau} W' \mathrm{d}W'$, we first considered the subdivision of the time domain $[t, t+\tau]$ using τ/n with integer n. As discussed in the argument leading to equation (20), so long as it is large the freedom to choose n determined the time scale τ/n over which the decay of the noise forcing is observed; when $\tau/n \gg \tau_c$ then the noise signal appears as a discrete packet, appropriately simulated as shot noise (e.g. [19]).

Consider two 'square' signals with amplitude h and time duration w, occurring with probability αw. When both are contained in the same packet and have the same sign ($\pm h$) they are positively correlated. Otherwise they are independent and uncorrelated. In the probability domain Ω the autocorrelation $\mathcal{R}(\tau)$ is

$$\mathcal{R}(\tau) = \int_\Omega n_1 n_2 P(n_1, n_2)\, \mathrm{d}n_1 \mathrm{d}n_2, \tag{25}$$

where n_1 (n_2) is a random variable occurring at the time t ($t+\tau$) and $P(n_1, n_2)$ is the joint probability density function of n_1 and n_2. We need only consider the case when τ is smaller than w, which is

$$\begin{aligned}
\mathcal{R}(\tau) &= \int_\Omega n_1^2 P(n_1) P(n_2|n_1)\, \mathrm{d}n_1 \mathrm{d}n_2 \\
&= \left(1 - \frac{\tau}{w}\right) \alpha w \left(\frac{1}{2}\left(\frac{h}{w}\right)^2 + \frac{1}{2}\left(-\frac{h}{w}\right)^2\right) \\
&= \alpha h^2 \frac{1}{w}\left(1 - \frac{\tau}{w}\right),
\end{aligned} \tag{26}$$

from which we can readily calculate

$$\int_{-\infty}^{\infty} R(\tau)\, \mathrm{d}\tau \approx \alpha h^2 \int_{-w}^{w} \frac{1}{w}\left(1 - \frac{\tau}{w}\right) \mathrm{d}\tau = 2\alpha h^2. \tag{27}$$

Thus, the constant αh^2 corresponds to σ^2, h/w is analogous to $\mathrm{d}W/\mathrm{d}t$, and hence $\mathcal{R}(\tau) = 2\sigma^2 \delta(\tau)$. Therefore, despite our argument emerging from shot noise, it leads naturally to the definition of white noise in which there exists a finite time correlation that is too minute to be realized when viewed from a coarse time-scale, a coarseness which depends on the observer.

Clearly, a similar procedure applies directly to the Riemann-sum so long as the same constraints on the subdivision of the time domain are in place. When $\tau/n \simeq w$, two signals of the same sign are positively correlated in the interval. Hence, the random variables $n_1 n_2$ only have the value h^2 with probability $\int_\Omega P(n_1) P(n_2 | n_1) \, dn_1 dn_2$, where $P(n_1)$ is $\alpha\tau/n$ and $P(n_2 | n_1)$ is 1/2. Because the conditional probability $P(n_2 | n_1)$ is equal to the fraction of $n_2 > n_1$ in the domain $\left\{ (n_1, n_2) \middle| 0 < n_1 < \tau/n, \, 0 < n_2 < \tau/n \right\}$, $n_1 n_2$ take the value h^2 with probability $\frac{1}{2}\alpha\frac{\tau}{n}$ and the value 0 with the probability $1 - \frac{1}{2}\alpha\frac{\tau}{n}$. According to the central limit theorem, a very large collection of realizations $n_1 n_2$ is given by a Gaussian distribution with mean $\frac{1}{2}\alpha h^2\tau$. This result is analogous to $\frac{1}{2}\sigma^2 W_t^2$ which was obtained from the Stratonovich calculus integral of $\int_0^t W_t \, dW_t$. Moreover, the factor of 1/2 originates in the time order of n_1 and n_2 which is traced to the constraint $n_1 < n_2$ when $P(n_2 | n_1)$ is calculated. In other words, the present value n_1 is influenced in a time-symmetric manner and this factor of 1/2 reflects this symmetry through a properly ordered time integration.

In this simple argument one sees that the δ-autocorrelated noise (used in the definition of white noise) and the Stratonovich calculus approach have the same origin. Both formalisms use different approaches to capture the accumulated influence of short time-scale correlations of a noise source that are not represented in the long time-scale dynamics. Importantly, temporal continuity of the noise is not a necessary condition. In order to demonstrate this, we used discontinuous autocorrelated shot noise to recover both the δ-function definition of white noise and the mid-point selection procedure of Stratonovich calculus. Moreover, one can make the simple case more realistic by treating h as a random variable and the same logic holds.

Finally, we note that despite our pedagogical example $dx/dt = xF(t)$, the results are general. Thus, when we replace x in the right hand side with $b(x, t)\frac{\partial}{\partial x}b(x, t)$, we can generalize our argument to any multiplicative noise case. There are of course a myriad of ways to decide which calculus is most appropriate to the problem at hand. Hence, whereas the integral Taylor expansion method enables one to only focus on the nature of the short time-scale processes, the comparison between the integral of autocorrelation and the variance of the short time-scale process focuses the choice on the specific mathematical model or scientific problem. Of particular interest to us is the question of how the stability of the non-autonomous SDE's of interest in climate dynamics are influenced by such considerations [20].

7. Related approaches

According to our result using shot noise, even an infinitesimal noise correlation in a stochastic differential equation can be interpreted using Stratonovich calculus. One might consider our analysis as a generalised version of the Wong and Zakai [1] approach because shot noise here is not a C_1 function. Considering the ubiquity of colored noise in real systems, Itô calculus might be interpreted as an idealized mathematical procedure, only applicable to true white noise processes that are never realized in nature. Hence, we ask whether there are situations in which Itô calculus can be used for colored noise?

The conceptual model we are dealing with in stochastic differential equations is that we have a principle deterministic process whose fate is influenced by short time-scale fluctuations, which are considered as noise. In building a mathematical model it is common to ignore the influence of the short time scale processes on the deterministic dynamics, but there are situations when this may be known to be a poor assumption such as is in the presence of inertial [21] or feedback [22] effects. Indeed, Kupferman and colleagues [21] studied systems with multiplicative *colored* noise and inertia to find that if the correlation time of the noise is faster (slower) than the relaxation time, this leads to the Itô (Stratonovich) calculus form of the limiting stochastic differential equation. Similarly, Itô calculus is invoked to interpret experiments wherein the time delay of the feedback is much larger than the noise correlation time [22].

These results may also be reinterpreted within the framework of our shot noise formalism. When the inertial or feedback time scales are much shorter than the noise decorrelation time, one can ignore the former and focus on the latter using the logic that lead us to the Stratonovich calculus interpretation. However, when the inertial or feedback time scales are much longer than the noise decorrelation time, the duration w of the shot noise is no longer equivalent to the decorrelation time scale of the noise. Rather, in order for the principle dynamics to be valid, we must interpret w as the inertial or feedback time scale. Hence, during a time increment w, we must assume that two consecutive random events n_1 and n_2 are independent, $P(n_1, n_2) = P(n_1) P(n_2)$, which is analogous to the equation (24). Therefore, despite the colored noise in this case, Itô calculus prevails. Furthermore, this description provides an unambiguous interpretation of the discrete nature of a system [18], which is believed to be a central criterion for the use of Itô calculus.

8. Conclusion

When a system is described as having multiplicative noise, the criterion for choosing which stochastic calculus is most appropriate—Itô or Stratonovich—has been a source of confusion and great discussion. In most areas of physical science where white noise is defined in terms of a δ-function autocorrelation, Stratonovich calculus is preferred, mainly due to the consistency with the results emerging from the Fokker–Planck equation and the fluctuation-dissipation theorem. In economics, where the Martingale property is considered as the most essential aspect of stochastic random variables, Itô calculus is widely accepted and used in the development of models. On the other hand, in population dynamics, Itô calculus is viewed as a proper continuous approximation of an underlying discrete model in which the Martingale property is guaranteed by construction. However, it would appear more prudent to choose the calculus depending on a clear set of objective considerations.

The core difference between the approaches is seen through the presence of the Martingale property in the multiplicative noise term of the stochastic Langevin equation, and hence manifests itself in the drift term of the associated Fokker–Planck equation. Thus, the principal ambiguity is associated with the ensemble mean of the product of two Brownian motion processes and in the original form it is difficult to assess the validity of the Martingale property. Here we used integral Taylor expansions to pinpoint the source of the deviation between the two calculi, which resides in the characteristics of the short time-scale noise process, the nature of which is described by the autocorrelation. The approach allows one to see the origin of the

Stratonovich calculus, which resides in the mid-point selection scheme that is synonymous with including the effect of a finite autocorrelation.

It appears that most realistic signals simulated by white noise have non-Markovian structure, that is, a *finite decorrelation time* τ_c. When one writes down a model stochastic differential equation, one typically initially assumes that τ_c is very small relative to the characteristic deterministic dynamical time scales, in which case white noise is a proper approximation. Thus, in such a case, the integral of the noise autocorrelation function and the variance of the noise are of a similar order, and the δ-function autocorrelation is a good approximation that is consistent with Stratonovich's calculus. However, when the integral of the noise autocorrelation function is much smaller than the variance of the noise, the δ-function autocorrelation is not an appropriate definition of white noise and Itô's calculus is appropriate.

Focusing on the noise itself, in the microscopic limit it is no longer Brownian but is instead represented by discontinuous finite-time signals, or shot noise. A small but finite autocorrelation of shot noise can be characterized by a δ-function, and the accumulation over time of its autocorrelation is represented by the mid-point selection procedure in the Riemann sum of Stratonovich calculus. This demonstrates that the origin of δ-correlated noise and Stratonovich calculus is the *infinitesimal autocorrelation* of the stochastic noise. Thus, the consistency between the Fokker–Planck equation and Stratonovich calculus has the same physical source, the finite autocorrelation of the stochastic noise that is ignored on time scales long relative to the fluctuation time scales. Therefore, a finite noise autocorrelation is the key criterion for choosing a stochastic calculus, but we must take care in interpreting its origin generally as well as on a case by case basis.

Acknowledgments

References

[1] Wong E and Zakai M 1965 On the convergence of ordinary integrals to stochastic integrals *Ann. Math. Stat.* **36** 1560–4

[2] Van Kampen N G 1976 Stochastic differential equations *Phys. Rep.* **24** 171–228

[3] Hasselmann K 1976 Stochastic climate models part I. Theory *Tellus* **28** 473–85

[4] Benzi R, Sutera A and Vulpiani A 1981 The mechanism of stochastic resonance *J. Phys. A: Math. Gen.* **14** 453

[5] Einstein A 1905 On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat *Ann. Phys.* **17** 16

[6] Langevin M P 1908 Sur la théorie du mouvement Brownien *C. R. Acad. Sci., Paris* **146** 530–3

[7] Kloeden P E and Pearson R A 1977 The numerical solution of stochastic differential equations *J. Aust. Math. Soc.* B **20** 8–12

[8] Itô K 1944 Stochastic integral *Proc. Imperial Acad. Tokyo* **20** 519–24

[9] Stratonovich R L 1966 A new representation for stochastic integrals and equations *SIAM J. Control* **4** 362–71

[10] Shreve S E, Chalasani P and Jha S 2004 *Stochastic Calculus for Finance* vol 1 (New York: Springer)

[11] Uhlenbeck G E and Ornstein L S 1930 On the theory of the Brownian motion *Phys. Rev.* **36** 823–41

[12] West B J, Bulsara A R, Lindenberg K, Seshadri V and Shuler K E 1979 Stochastic processes with non-additive fluctuations: I. Itô and Stratonovich calculus and the effects of correlations *Physica* A **97** 211–33

[13] Van Kampen N G 1981 Itô versus Stratonovich *J. Stat. Phys.* **24** 175–87

[14] Turelli M 1977 Random environments and stochastic calculus *Theor. Pop. Biol.* **12** 140–78

[15] Risken H 1996 *The Fokker–Planck Equation: Methods of Solution and Applications* vol 18 (Berlin: Springer)

[16] Reif F 1965 *Fundamentals of Statistical and Thermal Physics* (New York: McGraw-Hill)

[17] Morita T 1981 On the interpretation of multiplicative white noise *Phys. Lett.* A **82** 215–7

[18] Horsthemke W and Lefever R 2006 *Noise Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology (Springer Series in Synergetics* vol 15*)* (Berlin: Springer)

[19] Lathi B P 1998 *Modern Digital and Analog Communication Systems (Oxford Series in Electrical and Computer Engineering)* (New York: Oxford University Press)

[20] Moon W and Wettlaufer J S 2013 A stochastic perturbation theory for non-autonomous systems *J. Math. Phys.* **54** 123303

[21] Kupferman R, Pavliotis G A and Stuart A M 2004 Itô versus Stratonovich white-noise limits for systems with inertia and colored multiplicative noise *Phys. Rev.* E **70** 036120

[22] Pesce G, McDaniel A, Hottovy S, Wehr J and Volpe G 2013 Stratonovich-to-Ito transition in noisy systems with multiplicative feedback *Nat. Commun.* **4** 2733