Numerical Generation of Stochastic Differential Equations in Climate Models

Brian Ewald

Department of Mathematics, Florida State University, Tallahassee, Florida, USA

Cécile Penland

Physical Science Division/ESRL/NOAA 325 Broadway, Boulder, CO 80305-3328
November 2007

1. Introduction

It is, at present, impossible for a computer model to simulate weather and climate at all spatial and temporal scales. This would be no problem if the dynamical processes at widely different scales were independent of each other, but they are not. The modeler is therefore faced with the decisions of what physical processes to model explicitly, what processes to ignore, and what processes to parameterize. The word "parameterize" is used to indicate an approximation of a special kind; phenomena that are not explicitly treated in a numerical model are "parameterized" in a model when they are approximated by expressions involving phenomena that are. For example, the barotropic vorticity equation describes horizontal rotational motions only. The effects of divergent flow and vertical mass transport on these motions are parameterized by means of damping and forcing terms proportional to the horizontal rotational motions themselves. These parameterizations are not "ad hoc," as has sometimes been suggested, but are based on knowledge of data and physical principles.

Whether or not they are physically based, parameterizations are approximations and, as such, some are better than others. The more characteristics of a real process a parameterization can mimic, the better the approximation is expected to be. However, there are virtues in minimalism; resources expended in reproducing a marginally important effect are probably spent better elsewhere. Thus, returning to our example, we find that

Computational Methods for the Atmosphere and the Oceans
Special Volume (Roger M. Temam and Joseph J. Tribbia, Guest Editors) of
HANDBOOK OF NUMERICAL ANALYSIS, VOL. XIV
P.G. Ciarlet (Editor)

ISSN 1570-8659
DOI 10.1016/S1570-8659(08)00206-8

interactions between resolved and unresolved scales of vorticity are usually parameterized by hyperdiffusion in general circulation models with finite resolution in order to dissipate the numerical build-up of enstrophy at the smallest scales when these interactions are completely ignored. Accounting for the fact that these interactions can add enstrophy to the resolved scales as well as bleed it out from them, or that such interactions are more likely to be chaotically varying than temporally smooth, has been considered too expensive and difficult to be worth the trouble.

Accounting for the rapidly varying, chaotic nature of multiscale dynamical interactions in terms of random functions is the purview of stochastic theory. However, facing the importance of stochastic, subscale variability to weather and climate could happen only as computers became powerful enough to show that increasingly fine resolution in models never seemed to account for this variability completely. Further, powerful computers are needed to allow the rigorous application of stochastic theory in weather/climate models since numerical algorithms designed to optimize deterministic models are often not applicable to stochastic models. The art of numerical prediction and the theory of stochastic differential equations (SDEs) developed more or less contemporaneously, and certainly independently, over several decades so that the cultural shock of the two disciplines' discovery of each other has been severe. Happily, the last few years have seen the development of fruitful collaboration between geoscientists and applied mathematicians specializing in the development of numerical schemes for generating SDEs.

It is the purpose of this chapter to summarize numerical procedures for evaluating solutions of classical SDEs in climate prediction and research. Even this narrow focus is broad enough that we are likely to have overlooked relevant research, and we apologize to those researchers whose work falls into that category. The next section of this chapter discusses the central limit theorem, which directs how a system with scale separation may be approximated as an SDE. We do not provide a proof or even a rigorous statement of it, but rather set notation and examine some of its practical implications. The third section reviews the stochastic Taylor expansion and relates it to the development of stochastic numerical integration methods. The fourth section gives an overview of stochastic numerical methods as used in climate research, and the fifth presents examples. We conclude the chapter with a discussion of current directions in this field.

2. The central limit theorem

In the following, we discuss an extension of the traditional central limit theorem (e.g., DOOB [1953], WILKS [1995]) usually employed by geoscientists to justify the use of Gaussian distributions. Informally, the classical central limit theorem states that the sum of independently sampled quantities is approximately Gaussian. In the version described below, we consider dynamical systems described by a slow timescale and faster timescales. The equations are averaged over a temporal interval large enough that the fast timescales collectively act as Gaussian stochastic forcing of the slow, coarse-grained system. In the mathematical literature (e.g., FELLER [1966]), the fact that fine details of how the fast processes are distributed do not strongly affect the coarse-grained behavior of the slower dynamics is often called "the invariance principle." As the proof

of this theorem is far outside the scope of this review article, we state a commonly used form of it and refer the interested reader to the literature for details. GARDINER [1985] gives an heuristic description; we prefer PAPANICOLAOU and KOHLER [1974] for a technical statement of the theorem.

A dynamical system consisting of separated timescales may be written in the following manner:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \varepsilon \boldsymbol{G}(\boldsymbol{x}, t) + \varepsilon^2 \boldsymbol{F}(\boldsymbol{x}, t), \tag{2.1}$$

where $\boldsymbol{x}$ is an N-dimensional vector. In Eq. (2.1), the smallness parameter ε should not be taken as a measure of importance. It does measure the rapidity with which the terms on the right-hand side of Eq. (2.1) vary relative to each other; one can think of ε^2 as the ratio of the characteristic timescale of the first term to the characteristic timescale of the second. If we now cast Eq. (2.1) in terms of a scaled time coordinate,

$$\Delta s = \varepsilon^2 \Delta t, \tag{2.2}$$

Eq. (2.1) becomes

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}s} = \frac{1}{\varepsilon}\boldsymbol{G}(\boldsymbol{x}, s/\varepsilon^2) + \boldsymbol{F}(\boldsymbol{x}, s/\varepsilon^2). \tag{2.3}$$

We further assume that the first term in Eq. (2.1) decays quickly "enough" in the time interval Δt. In the limit $\varepsilon \to 0$, $\Delta t \to \infty$ with $\varepsilon^2 \Delta t$ remaining finite, the central limit theorem states that Eq. (2.3) converges weakly to a "Stratonovich" (see below) SDE in the scaled coordinates:

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{F}'(\boldsymbol{x}, s)\mathrm{d}s + \mathbf{G}'(\boldsymbol{x}, s) \cdot \mathrm{d}\boldsymbol{W}(s). \tag{2.4}$$

The symbol $\boldsymbol{W}$ in Eq. (2.4) is a K-dimensional vector, each component of which is an independent Wiener process, or Brownian motion, and the "$\cdot$" symbol indicates that it is to be interpreted in the sense of Stratonovich, discussed below. The symbol $\mathbf{G}'(\boldsymbol{x}, s)$ is a matrix, the first index of which corresponds to a component of x, and the second index of which corresponds to a component of $\mathrm{d}\boldsymbol{W}$. Using angle brackets to denote expectation values, we state the following properties of the vector Wiener process $\boldsymbol{W}$:

$$\boldsymbol{W}(s) \text{ is a vector of Gaussian random variables,} \tag{2.5a}$$

$$<\boldsymbol{W}(s)> = \mathbf{0}, \tag{2.5b}$$

$$<\boldsymbol{W}(s)\boldsymbol{W}^{\mathrm{T}}(t)> = \mathbf{I}\min(s, t), \tag{2.5c}$$

$$<\mathrm{d}\boldsymbol{W}(s)\mathrm{d}\boldsymbol{W}^{\mathrm{T}}(t)> = \mathbf{I}\,\delta(s - t). \tag{2.5d}$$

In Eq. (2.5), the symbol $\mathbf{I}$ denotes the identity matrix. The Wiener process is continuous, but is only differentiable in a generalized sense:

$$\mathrm{d}W_k = \xi_k \mathrm{d}t, \tag{2.6}$$

where "white noise" ξ_k is really a concept that can only be approximated in nature. As its name suggests, white noise has a flat spectrum and, therefore, infinite power. The Wiener process, on the other hand, has a power spectrum that decreases everywhere

with the square of the frequency. For more detailed descriptions of the properties of Wiener processes and white noise, we refer the interested reader to ARNOLD [1974], Chapter 3.

The white noise limit leading to Eq. (2.4) essentially represents the combined effect of the weakly correlated, rapid variations in the first term of Eq. (2.1) that occur within time Δt as the Gaussian variable in Eq. (2.4). This is why it is called the central limit theorem. For details and proof of the central limit theorem, we recommend, for example, the articles by WONG and ZAKAI [1965], KHASMINSKII [1966], and PAPANICOLAOU and KOHLER [1974]. Examples of geophysical applications may be found in KOHLER and PAPANICOLAOU [1977], PENLAND [1985], MAJDA et al. [1999], and SARDESHMUKH, PENLAND and NEWMAN [2001].

Let us examine the special case where the ith component of the rapidly varying term in Eq. (2.3) can be written as

$$G_i(\boldsymbol{x}, s/\varepsilon^2) = \sum_{k=1}^{K} G_{ik}(\boldsymbol{x}, s)\eta_k(s/\varepsilon^2), \tag{2.7}$$

where $\eta_k(s/\varepsilon^2)$ is a stationary, centered, and bounded random function. The integrated lagged covariance matrix of η is defined to be

$$C_{km} \equiv \int_0^\infty < \eta_k(t)\eta_m(t+t') > \mathrm{d}t', \quad k, m = 1, 2, \ldots, K, \tag{2.8}$$

where angle brackets denote expectation value. With these restrictions, the central limit theorem states that in the limit of long times ($t \to \infty$) and small $\varepsilon(\varepsilon \to 0)$, taken so that $s = \varepsilon^2 t$ remains fixed, the conditional probability density function (*pdf*) for $\boldsymbol{x}$ at time s given an initial condition $\boldsymbol{x}_o(s_o)$ satisfies the backward Kolmogorov equation (e.g., HORSTHEMKE and LÉFÈVER [1984], BHATTACHARYA and WAYMIRE [1990])

$$\frac{\partial p(\boldsymbol{x}, s|\boldsymbol{x}_o, s_o)}{\partial s_o} = \mathcal{L}p(\boldsymbol{x}, s|\boldsymbol{x}_o, s_o), \tag{2.9}$$

where

$$\mathcal{L} = \sum_{i,j}^{N} a_{ij}(\boldsymbol{x}_o, s_o)\frac{\partial^2}{\partial x_{oi}\partial x_{oj}} + \sum_{i}^{N} b_i(\boldsymbol{x}_o, s_o)\frac{\partial}{\partial x_{oi}} \tag{2.10}$$

and

$$a_{ij}(\boldsymbol{x}, s) = \sum_{k,m=1}^{K} C_{km}G_{ik}(\boldsymbol{x}, s)G_{jm}(\boldsymbol{x}, s), \tag{2.11a}$$

$$b_i(\boldsymbol{x}, s) = \sum_{k,m=1}^{K} C_{km}\sum_{j=1}^{N} G_{jk}(\boldsymbol{x}, s)\frac{\partial G_{im}(\boldsymbol{x}, s)}{\partial x_j} + F_i(\boldsymbol{x}, s). \tag{2.11b}$$

In this limit, the conditional *pdf* also satisfies a forward Kolmogorov equation (called a "Fokker–Planck equation" in the scientific literature) in the scaled coordinates:

$$\frac{\partial p(\boldsymbol{x}, s|\boldsymbol{x}_o, s_o)}{\partial s} = \mathcal{L}^\dagger p(\boldsymbol{x}, s|\boldsymbol{x}_o, s_o), \tag{2.12}$$

where

$$\mathcal{L}^\dagger p = \sum_{i,j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j}\left[a_{ij}(\boldsymbol{x}, s)p\right] - \sum_{i}^{N} \frac{\partial}{\partial x_i}\left[b_i(\boldsymbol{x}, s)p\right]. \tag{2.13}$$

In short, the *moments* of $\boldsymbol{x}$ can be approximated by the moments of the solution to the "Stratonovich" SDE

$$\mathrm{d}x = \boldsymbol{F}(\boldsymbol{x}, s)\mathrm{d}s + \mathbf{G}(\boldsymbol{x}, s)\mathbf{S} \cdot \mathrm{d}\boldsymbol{W}. \tag{2.14}$$

In Eq. (2.14), $\mathbf{G}(\boldsymbol{x}, s)$ is the matrix whose (i, j)th element is $G_j^i(\boldsymbol{x}, s)$, and $\mathbf{S}$ is a matrix where the (k, m)th element of $\mathbf{S}\mathbf{S}^{\mathrm{T}}$ is C_{km}. Note that $\mathbf{S}$ is only unique up to its product with an arbitrary orthogonal matrix. Notice also that the usual factor of one-half found in most formulations of the Fokker–Planck equation has been absorbed into the definition of C_{km}.

Before proceeding, it may be well to discuss the physical systems described by Eq. (2.4). We have presented a sparse outline of how a dynamical system with two timescales, both finite and with one timescale much shorter than the other, may be described as a SDE we denoted with the name Stratonovich. This class of systems obeys the rules of classical Riemannian calculus and is appropriate for dynamical systems that we assume to be continuous from the outset, before we start thinking about treating the system as stochastic. That is, when constructing the calculus to describe the dynamics, the continuous limit is taken before the white noise limit is taken. Mathematically, these systems are described by integrals defined by limits of Riemann sums, where contributions from both the beginning and the end of each time interval are taken into account with equal weight. Details may be found in, for example, KLOEDEN and PLATEN [1992].

There is another class of systems, called "Ito systems," that arises when these limits are taken in opposite order. Physically, one has a discrete system at the smaller timescale, and the disrupting influence that is to be treated as white noise is uncorrelated between the discrete time steps. At a longer timescale, the discrete time steps are approximately continuous, and we construct a continuous calculus to handle the dynamical evolution. These Ito systems also involve Brownian motion, but they do not follow the classical Riemannian rules of calculus. In contrast to Stratonovich integrals, Ito integrals are defined by limits of Riemann sums involving contributions from only the beginning of each time interval.

The difference between Ito and Stratonovich calculus is important for scientists to understand because most of the physical phenomena they deal with are Stratonovich, while most mathematical references on stochastic numerical techniques are primarily interested in Ito schemes. As we shall see, there is a formal equivalence between Ito and Stratonovich descriptions of reality, so a theorem about an Ito process can generally be

carried over to the corresponding Stratonovich process. Unfortunately, the transformation from one description to the other can be prohibitively difficult.

Roughly speaking, the Wiener process (Eqs. (2.5) and (2.6)) wiggles around enough that infinitely many ways of integrating this function can be defined, each giving a different answer. The names "Ito" and "Stratonovich" label two of these calculi, each of which is a result of applying a limiting procedure to equations describing a physical system. Let us say we wish to evaluate the integral

$$I = \int_{t_0}^{T} W(t)\mathrm{d}W. \tag{2.15a}$$

Classically trained scientists might immediately write the solution to this equation as

$$I_S = [W^2(T) - W^2(t_0)]/2. \tag{2.15b}$$

However, this result comes from dividing the interval $(t_0, T]$ into n partitions

$$S_n = \sum_{i=1}^{n} W(\tau_i)[W(t_i) - W(t_{i-1})], \tag{2.16}$$

with $t_n = T$ and τ_i chosen arbitrarily in the interval $(t_{i-1}, t_i]$. As $n \to \infty$ and the partitions become progressively finer, S_n converges to I (Eq. (2.15a)). For regular deterministic processes, the answer does not depend on *where* τ_i is chosen in the interval $(t_{i-1}, t_i]$. However, for integrals over the Wiener process, it *does* matter (e.g., ARNOLD [1974]). The Stratonovich solution (Eq. (2.15b)) results as the mean square convergence of Eq. (2.16) when τ_i is chosen to be the midpoint between t_{i-1} and t_i. The Ito solution,

$$I_I = [W^2(T) - W^2(t_0)]/2 - (T - t_0)/2, \tag{2.17}$$

results as the mean square convergence of Eq. (2.16) when τ_i is chosen equal to the beginning of the interval t_{i-1}.

Which is the "correct" solution? Each can be, as can any of the other calculi corresponding to an infinite number of choices for where one chooses τ_i. Here, we concentrate on the two calculi that have been associated with naturally occurring phenomena: Ito and Stratonovich.

Looking at Eqs. (2.15b) and (2.17), one might note that there is an easy transformation between I_S and I_I. In fact, there is always a transformation between the solution to a Stratonovich SDE and an Ito SDE. The solution to the Stratonovich SDE (written in component form)

$$\mathrm{d}x_i = F_i(\boldsymbol{x}, t)\mathrm{d}t + \sum_{\alpha} G_{i\alpha}(\boldsymbol{x}, t) \cdot \mathrm{d}W_\alpha \tag{2.18a}$$

is equivalent to the Ito SDE

$$\mathrm{d}x_i = \left[F_i(\boldsymbol{x}, t) + \frac{1}{2} \sum_{\alpha j} G_{j\alpha}(\boldsymbol{x}, t) \frac{\partial G_{i\alpha}(\boldsymbol{x}, t)}{\partial x_j} \right] \mathrm{d}t + \sum_{\alpha} G_{i\alpha}(\boldsymbol{x}, t)\mathrm{d}W_\alpha. \tag{2.18b}$$

By "equivalent" it is meant that solving Eq. (2.18a) using Stratonovich calculus gives the same solution as solving Eq. (2.18b) using Ito calculus. That is, each equation evaluated for $\boldsymbol{x}$ using its appropriate calculus would describe an experimental outcome equally well as the other; the statistics of $\boldsymbol{x}$ in each case are the same. In Eq. (2.18), we have omitted the symbol "$\circ$" in keeping with standard mathematical notation of an Ito SDE.

As for Stratonovich systems, the transition *pdf* for an Ito system satisfies a Fokker–Planck equation (Eqs. (2.12) and (2.13)). However, the drift and diffusion terms, $\boldsymbol{b}$ and $\mathbf{a}$, are somewhat different. For the Ito SDE

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, t)\mathrm{d}t + \mathbf{G}\mathrm{d}\boldsymbol{W}, \tag{2.19}$$

we have simply

$$\boldsymbol{b} = \boldsymbol{F}(\boldsymbol{x}, t) \tag{2.20a}$$

and

$$\mathbf{a} = \mathbf{G}\mathbf{G}^{\mathrm{T}}/2. \tag{2.20b}$$

For longer discussions on the difference between Ito and Stratonovich systems, we refer the reader to ARNOLD [1974], HORSTHEMKE and LÉFÈVER [1984], and KLOEDEN and PLATEN [1992]. In the following, we shall assume that the analytical work leading to an SDE has been done, the correct calculus has been identified, and that the scientist is ready to perform a numerical simulation of it.

3. The stochastic Taylor series and application to numerical schemes

3.1. Basic structure of the stochastic Taylor series

Stochastic Taylor series is an important theoretical and analytical tool in the study of SDEs and their numerical approximation. These are described in great detail in KLOEDEN and PLATEN [1992]; we repeat only their most salient properties here. To see how they are derived, we will first consider the derivation of a similar deterministic formula. For this, consider the following ordinary differential equation (ODE):

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(x). \tag{3.1}$$

We can then use the chain rule to write:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}t} = f'(x)\frac{\mathrm{d}x}{\mathrm{d}t} = f'(x)a(x). \tag{3.2}$$

In Eq. (3.2) and in all equations of this section, prime denotes differentiation with respect to the argument. We can write all of this in a form analogous to that which we will use for the stochastic Taylor series, using the differential operator $L = a\dfrac{\mathrm{d}}{\mathrm{d}t}$, and we see that, if $\mathrm{d}x = a(x)\mathrm{d}t$, then

$$\mathrm{d}f(x) = Lf(x)\mathrm{d}t. \tag{3.3}$$

That is,

$$f(x(T)) = f(x(0)) + \int_0^T Lf(x(t))\mathrm{d}t. \tag{3.4}$$

If we apply this formula with Lf for f, we will have that

$$Lf(x(T)) = Lf(x(0)) + \int_0^T L^2 f(x(t))\mathrm{d}t. \tag{3.5}$$

We substitute this into the integral in the previous equation to arrive at

$$\begin{aligned} f(x(T)) &= f(x(0)) + \int_0^T Lf(x(0))\mathrm{d}t + \int_0^T \int_0^t L^2 f(x(s))\mathrm{d}s\mathrm{d}t, \\ &= f(x(0)) + Lf(x(0))T + \int_0^T \int_0^t L^2 f(x(s))\mathrm{d}s\mathrm{d}t. \end{aligned} \tag{3.6}$$

If we then use the change of variables formula (Eq. (3.4)) with $L^2 f$ for f, we get that

$$\begin{aligned} f(x(T)) &= f(x(0)) + Lf(x(0))T + \int_0^T \int_0^t L^2 f(x(0))\mathrm{d}s\mathrm{d}t \\ &\quad + \int_0^T \int_0^t \int_0^s L^3(x(u))\mathrm{d}u\mathrm{d}s\mathrm{d}t, \\ &= f(x(0)) + Lf(x(0))T + \frac{1}{2}L^2 f(x(0))T^2 + \int_0^T \int_0^t \int_0^s L^3(x(u))\mathrm{d}u\mathrm{d}s\mathrm{d}t. \end{aligned} \tag{3.7}$$

We can then continue this for as long as desired, yielding a Taylor series-like formula with an integral remainder term.

We will now derive stochastic Taylor series for the SDE

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, t)\mathrm{d}t + \mathbf{G}(\boldsymbol{x}, t)(\cdot)\mathrm{d}\boldsymbol{W}, \tag{3.8}$$

where, as above, $\boldsymbol{x}$ and $\boldsymbol{F}$ are N-vectors, $\boldsymbol{W}$ is a K-dimensional Brownian motion, and $\mathbf{G}$ is an $N \times K$ matrix. The symbol "$(\cdot)$" is to be interpreted as "$\cdot$" if the Wiener process is to be integrated in the sense of Stratonovich, and is to be ignored if the noise is to be integrated in the sense of Ito. Instead of the chain rule, we have the Ito formula (also called the "stochastic chain rule:" GARDINER [1985], KLOEDEN and PLATEN [1992]) for any sufficiently smooth function $\boldsymbol{f}(\boldsymbol{x}, t)$:

$$\mathrm{d}\boldsymbol{f}(\boldsymbol{x}, t) = L^0\boldsymbol{f}(\boldsymbol{x}, t)\mathrm{d}t + \sum_{j=1}^{K} L^j\boldsymbol{f}(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{W}_j(t). \tag{3.9}$$

Here, if Eq. (3.8) is an Ito equation,

$$L^0 = \frac{\partial}{\partial t} + \sum_{n=1}^{N} F_n \frac{\partial}{\partial x_n} + \frac{1}{2}\sum_{i,n=1}^{N}\sum_{j=1}^{K} G_{ij}G_{nj}\frac{\partial^2}{\partial x_i \partial x_n} \tag{3.10a}$$

and

$$L^j = \sum_{n=1}^{N} G_{nj} \frac{\partial}{\partial x_n}. \tag{3.10b}$$

Note that Eq. (3.10b) operated on the ith component f_i of $\boldsymbol{f}$ will be the ijth component of an $N \times K$ matrix. If the equation is Stratonovich, then Eq. (3.10b) is as in the Ito case, but

$$L^0 = \frac{\partial}{\partial t} + \sum_{n=1}^{N} F_n \frac{\partial}{\partial x_n}. \tag{3.10c}$$

We will now rewrite the SDE in integral form:

$$\boldsymbol{x}(T) = \boldsymbol{x}(0) + \int_0^T \boldsymbol{F}(\boldsymbol{x}, t)\mathrm{d}t + \int_0^T \mathbf{G}(\boldsymbol{x}, t)(\cdot)\mathrm{d}\boldsymbol{W}(t), \tag{3.11}$$

and the integral form of the Ito formula:

$$f(x(T), T) = \boldsymbol{f}(\boldsymbol{x}(0), 0) + \int_0^T L^0 \boldsymbol{f}(\boldsymbol{x}(t), t)\mathrm{d}t + \sum_j \int_0^T L^j \boldsymbol{f}(\boldsymbol{x}(t), t)(\cdot)\mathrm{d}W^j(t). \tag{3.12}$$

Then, as in the deterministic case above, we can use one of $L^0\boldsymbol{f}$ or $L^j\boldsymbol{f}$ for $\boldsymbol{f}$ in the Ito formula and substitute it back in. Note that now we have a choice as to what substitution to make, and that every time we make such a substitution, we replace one integral with two iterated integrals, and that we iterate different differential operators (L^0 and the L^js) rather than just one. As such, the formulas can quickly become unwieldy, and it is useful to have a compact notation.

For this purpose, we introduce the multi-indices $\alpha = (j_1, j_2, \ldots, j_l)$. Each j_k is either 0 or a number from 1 to K. If $j_k = 0$, j_k refers to L^0 if it is a superscript of L, or to $\mathrm{d}t$ if it is a superscript of $\mathrm{d}\boldsymbol{W}$. In other words, $\mathrm{d}t = \mathrm{d}W^0(t)$ by convention. If j_k is a number from 1 to K, it refers to L^{j_k} or $\mathrm{d}W^{j_k}(t)$. We also have the iterated differential operators

$$L^\alpha = L^{j_1} L^{j_2} \ldots L^{j_l} \tag{3.13}$$

and the iterated integrals

$$I_\alpha[\boldsymbol{f}(\boldsymbol{x}, t)]_{T_1}^{T_2} \equiv \int_{T_1}^{T_2} \int_{T_1}^{t} \cdots \int_{T_1}^{s} \boldsymbol{f}(\boldsymbol{x}(u), u)(\cdot)\mathrm{d}W^{j_1}(u)\mathrm{d}W^{j_2}(s) \cdots \mathrm{d}W^{j_l}(t). \tag{3.14}$$

If the integrand $\boldsymbol{f}(\boldsymbol{x}, t)$ is omitted from the iterated integral, it is assumed to be 1. Estimations of I_α for Ito integrals are generally different from those for Stratonovich integrals; we shall introduce those approximations when we get to examples of different schemes. Meanwhile, with this notation in hand, we can write any stochastic Taylor series in the form

$$\boldsymbol{f}(\boldsymbol{x}(T), T) = \boldsymbol{f}(\boldsymbol{x}(0), 0) + \sum_{\alpha \in A} L^\alpha \boldsymbol{f}(\boldsymbol{x}(0), 0) I_\alpha + \sum_{\beta \in B} I_\beta[\boldsymbol{f}(\boldsymbol{x}, t)], \tag{3.15}$$

where the integrals go from 0 to T. Note that here, B contains those multi-indices β for which $L^{\beta}\boldsymbol{f}$ could be the next substitution for $\boldsymbol{f}$ in Ito's formula, and A contains the αs which must have come earlier. When we use Eq. (3.15) to develop a numerical scheme, the approximation itself comprises multi-indices belonging to A, and the truncation error comprises the multi-indices belonging to B.

Before going on, we will clarify this notation with a one-dimensional example; that is, both $\boldsymbol{F}$ and $\mathbf{G}$ are scalars F and G, implying that the arbitrary function f and the Brownian motion W are scalars as well. If we apply Ito's formula (Eq. (3.12)) twice, with $L^0 f$ and then $L^1 f$ for f, we arrive at the expansion

$$\begin{aligned} f(x(T),T) = f(x(0),0) &+ \int_0^T L^0 f(x(0),0)\mathrm{d}t + \int_0^T L^1 f(x(0),0)(\cdot)\mathrm{d}W(t) \\ &+ \int_0^T \int_0^t L^0 L^0 f(x(s),s)\mathrm{d}s\mathrm{d}t + \int_0^T \int_0^t L^1 L^0 f(x(s),s)(\cdot)\mathrm{d}W(s)\mathrm{d}t \\ &+ \int_0^T \int_0^t L^0 L^1 f(x(s),s)\mathrm{d}s(\cdot)\mathrm{d}W(t) + \int_0^T \int_0^t L^1 L^1 f(x(s),s)(\cdot)\mathrm{d}W(s)\mathrm{d}W(t) + \cdots \end{aligned} \tag{3.16}$$

This is more compactly written as

$$\begin{aligned} f(x(T),T) = f(x(0),0) &+ I_{(0)} L^0 f(x(0),0) + I_{(1)} L^1 f(x(0),0) \\ &+ I_{(0,0)}\Big[L^{(0,0)} f(x(t),t)\Big]_0^T + I_{(1,0)}\Big[L^{(1,0)} f(x(t),t)\Big]_0^T \\ &+ I_{(0,1)}\Big[L^{(0,1)} f(x(t),t)\Big]_0^T + I_{(1,1)}\Big[L^{(1,1)} f(x(t),t)\Big]_0^T . \end{aligned} \tag{3.17}$$

In Eq. (3.17), $I_0 = T$ and $I_1 = (W(T) - W(0))$. If we approximate our integral with terms involving only the initial condition, I_0 and I_1, then $A = \{(0), (1)\}$ and $B = \{(0,0), (1,0), (0,1), (1,1)\}$.

3.2. *Application to numerical schemes*

To create a numerical scheme for an SDE, we can expand differences using the stochastic Taylor series and drop the high-order terms. Of course, we break up the integration interval $(0, T]$ into discrete time steps so that the ith time step consists of an expansion analogous to Eq. (3.17) with $\boldsymbol{f}(\boldsymbol{x}(0), 0)$ replaced by $\boldsymbol{f}(\boldsymbol{x}(t_{i-1}), t_{i-1})$, $\boldsymbol{f}(\boldsymbol{x}(T), T)$ replaced by $\boldsymbol{f}(\boldsymbol{x}(t_i), t_i)$, T replaced by the time step Δ, and $(\boldsymbol{W}(T) - \boldsymbol{W}(0))$ replaced by a vector $\Delta\boldsymbol{W}$ of Gaussian random variables with each component having *variance* Δ. For example, consider Eq. (3.17), replacing $f(x(t), t)$ with $x(t)$ and dropping the remainder terms. Since $L^0 x = F$ and $L^1 x = G$, we are left with the stochastic Euler scheme (revisited in Section 4):

$$x(t_{i+1}) = x(t_i) + F(x(t_i), t_i)\Delta + G(x(t_i), t_i)\Delta W(t_i) + \text{Remainder}. \tag{3.18}$$

In the so-called strong schemes (see below), the estimation of ΔW is z_i, a Gaussian random deviate with mean zero and variance Δ. That is, roughly speaking, the deterministic part of the equation is updated with the time step, and the stochastic part is updated with the square root of the time step.

As another example, consider Eq. (3.15) with $A = \{(0), (1), (1, 1)\}$ and $B = \{(0, 0), (1, 0), (0, 1), (0, 1, 1), (1, 1, 1)\}$. That is, we expand the term

$$\int_0^T \int_0^t L^1 L^1 f(x(s), s)(\cdot)\mathrm{d}W(s)\mathrm{d}W(t) = \int_0^T \int_0^t L^1 L^1 f(x(0), 0)(\cdot)\mathrm{d}W(s)\mathrm{d}W(t)$$
$$+ \int_0^T \int_0^t \int_0^s L^0 L^1 L^1 f(x(u), u)\mathrm{d}u(\cdot)\mathrm{d}W(s)\mathrm{d}W(t)$$
$$+ \int_0^T \int_0^t \int_0^s L^1 L^1 L^1 f(x(u), u)(\cdot)\mathrm{d}W(u)\mathrm{d}W(s)\mathrm{d}W(t) \tag{3.19}$$

and again write in compact form

$$I_{(1,1)}\Big[L^{(1,1)} f(x(t), t)\Big]_0^T = I_{(1,1)} L^{(1,1)} f(x(0), 0)$$
$$+ I_{(0,1,1)}\Big[L^{(0,1,1)} f(x(t), t)\Big]_0^T + I_{(1,1,1)}\Big[L^{(1,1,1)} f(x(t), t)\Big]_0^T . \tag{3.20}$$

Again taking $f(x(t), t) = x(t)$

$$L^{(1,1)} x(t) = \boldsymbol{G}(x(t), t)\frac{\partial \boldsymbol{G}(x(t), t)}{\partial x} \tag{3.21a}$$

and either the Ito expression (KLOEDEN and PLATEN [1992])

$$I_{(1,1)} = (\Delta W^2 - \Delta^2)/2 \quad \text{(Ito)} \tag{3.21b}$$

or the Stratonovich expression

$$I_{(1,1)} = \Delta W^2/2 \quad \text{(Stratonovich)}, \tag{3.21c}$$

we arrive at the so-called Milsteyn scheme

$$x(t_{i+1}) = x(t_i) + F(x(t_i), t_i)\Delta + \boldsymbol{G}(x(t_i), t_i) z_i$$
$$+ \boldsymbol{G}(x(t_i), t_i)\frac{\partial \boldsymbol{G}(x(t_i), t_i)}{\partial x} I_{(1,1)} + \text{Remainder}. \tag{3.22}$$

In Eq. (3.22), $I_{(1,1)}$ is appropriately estimated by either Eq. (3.21b) or (3.21c), again with ΔW represented by z_i.

3.3. *Order of approximation*

To obtain the order of the approximation, we need to know what order various terms are. It turns out that this depends on the sense, i.e., strong or weak, in which we want to approximately solve the SDE. The so-called strong schemes approximate the entire path of the solution process in an L^2-sense (see below). Weak schemes, in contrast, yield paths whose *moments* approximate all of the moments of the SDE; the paths themselves, however, may not be pointwise solutions to the SDE.

We shall clarify strong solutions first. If $\boldsymbol{x}(t)$ is the actual solution process and $\boldsymbol{y}(t)$ is our strong numerical approximation, then

$$\sqrt{< \sup_{0 \le t \le T} |\boldsymbol{x}(t) - \boldsymbol{y}(t)|^2 >} = \boldsymbol{O}(\Delta t^{\gamma}), \tag{3.23}$$

where γ is the (strong) order of convergence. The vertical bars denote an appropriate norm, the Euclidean norm, for example. Note that this is indeed a strong convergence condition, usually much stronger than is needed or even necessarily desired. Strong schemes will, on average, approximate the path of $\boldsymbol{x}(t)$ that corresponds to the generated realization of the noise $\boldsymbol{W}(t)$. If we have no particular reason to believe that we are approximating the realization of the unresolved chaotic process that occurs in nature (and we generally call it noise because we do not have this belief), the added complexity and computational expense of strong schemes over weak schemes (see below) may not be warranted.

If one desires a strong scheme, then the correct order of an increment involving the iterated integral I_α depends on whether the integral is Ito or Stratonovich. For Ito integrals, the order is the number of deterministic terms (i.e., $j_k = 0$) plus half the number of stochastic terms (i.e., j_k between 1 and K), *except* that increments that are entirely deterministic (i.e., all the j_ks are 0) have effective order one-half smaller than this. For example, an increment with $\alpha = (0, 0, 1, 1, 0)$ has (strong) order 4, with $\alpha = (1, 0, 0)$ has order 2.5, and with $\alpha = (0, 0, 0, 0, 0)$ has order 4.5. It is understood that the iterated stochastic integrals are approximated using expressions appropriate to the Ito scheme, and that strong schemes can be developed for every integer and half-integer order.

For a particular example, consider Eq. (3.18). The first term on the right-hand side is, of course, order 0. The next two terms, with $\alpha = (0)$ and $\alpha = (1)$, are each of strong order 0.5. Thus, Eq. (3.18) represents an Ito scheme of strong order 0.5. The three next higher terms shown in Eq. (3.17) possess multi-indices $\alpha = (0, 0)$, $\alpha = (0, 1)$, and $\alpha = (1, 0)$. They are therefore of strong order 1.5. The last term, with $\alpha = (1, 1)$, is of strong order 1.

By the same reasoning, the Milsteyn scheme (Eq. (3.22)) is of strong order 1. Its next higher order terms in the remainder are of strong order 2 and 1.5. Hence, the Milsteyn scheme is an algorithm of strong order 1.

Similar counting of order is valid for strong Stratonovich integrals. However, only whole integer orders are valid, so all the half-order integers get bumped up to the next whole-order integer order. Note that the stochastic Euler scheme is not valid for Stratonovich integrals.

It is worth pointing out that it can be difficult to generate the necessary stochastic increments I_α for strong schemes (KLOEDEN and PLATEN [1992]) while still maintaining the formal accuracy to which the scheme was developed. For a one-dimensional noise, for example, the increment $I_{(0,1,1)}$, which appears in second-order schemes, is difficult to generate, since it is not Gaussian and has a complicated joint distribution with smaller order terms. For a multi-dimensional noise, the first-order increment $I_{(j,k)}$ is already difficult to generate.

We now turn to weak schemes. For so-called weak schemes, instead of approximating paths of solutions, we merely attempt to approximate the moments of solutions well. That is, we try to get the statistics correct. So, if $\boldsymbol{x}(t)$ is the actual solution process and $\boldsymbol{y}(t)$ is our numerical approximation, then

$$\left| < \boldsymbol{x}(t)^p > - < \boldsymbol{y}(t)^p > \right| = \boldsymbol{O}(\Delta t^\gamma) \tag{3.24}$$

for every $p = 1, 2, \ldots$, and where γ is the (weak) order of convergence. Note that weak schemes approximate every moment of the solution well.

It is important to note that what is said below will be valid *only* for Ito equations, and *not* for Stratonovich equations. This is not a major difficulty for sufficiently simple systems where it is possible to convert a Stratonovich equation to an Ito one (see Eq. 2.18).

If a weak scheme is desired, then the correct order of an increment involving the iterated integral I_α is simply the length of α. That is, both deterministic and stochastic terms in the increment have effective order 1. For high-order schemes this represents a considerable savings over strong schemes since many fewer increments are necessary. Weak schemes can be generated for any integral order. Note that all strong schemes of some order are also weak schemes, generally of a higher order. For example, the Euler scheme (Eq. (3.18)) is a strong scheme of order 0.5 and also a weak scheme of order 1.

A further savings can be realized with weak schemes since it is not necessary to simulate fully the stochastic increments. It is only necessary to simulate the first few moments (depending on the order of the scheme) of the increments.

For example, for a first-order weak scheme, such as the Euler scheme, it is only necessary that the simulated increments agree with the true stochastic increments up to the third moments, and these only need to agree up to first order in Δt (KLOEDEN and PLATEN [1992]). So, instead of generating a Gaussian random variable with mean 0 and variance Δt for the increment $\Delta W(t)$, it is sufficient to generate a coin-flip and generate a random variable $\Delta \widehat{W}$ such that

$$P(\Delta \widehat{W} = \sqrt{\Delta t}) = P(\Delta \widehat{W} = -\sqrt{\Delta t}) = \frac{1}{2}. \tag{3.25}$$

Note that even though we only approximate this increment correctly up through the third moment (and of these, only the second moment is nonzero), the scheme still approximates all moments of the solution well.

Second-order weak schemes need to get the first five moments of the stochastic increments correct, third-order schemes the first seven moments, etc.

To generate ΔW for a second-order scheme with a one-dimensional noise, the random variable $\Delta\widehat{W}$ with

$$P(\Delta\widehat{W} = \pm\sqrt{3\Delta t}) = \frac{1}{6}, \quad P(\Delta\widehat{W} = 0) = \frac{2}{3} \tag{3.26}$$

suffices, and for the iterated integral $I_{(1,0)}$, we can use $\Delta Z = \frac{1}{2}\Delta\widehat{W}\Delta t$ (KLOEDEN and PLATEN [1992]). If the scheme has a multi-dimensional noise, we will also need to generate the increments $I_{(j,k)}$ for $j \neq k$. For this, we can independently generate $V_{j,k}$ for $j < k$ such that

$$P(V_{j,k} = \pm\Delta t) = \frac{1}{2} \tag{3.27}$$

and $V_{k,j} = -V_{j,k}$, and approximate $I_{(j,k)}$ by $\Delta\widehat{W}^j\Delta\widehat{W}^k + V_{j,k}$. Generating all of this for just one time-step is faster than generating even one Gaussian random variable.

What can we do if we want a weak scheme for a Stratonovich SDE? We cannot use stochastic Taylor series to generate a weak scheme from scratch for a Stratonovich system since the weak order of a Stratonovich increment is not straightforward to estimate. However, we do have two possible options. First, as suggested above, we may be able to use Eq. (2.18) to convert from the Stratonovich system to the equivalent Ito system, thereby allowing use of weak Ito schemes. The other possibility is to make use of the fact that every strong scheme is also a weak scheme of the same order or higher. Thus, we can simply use a strong Stratonovich scheme and use it weakly, even to the extent of employing the simplifications of the stochastic increments as indicated for Ito weak schemes above.

4. Popular methods

The following summary of popular numerical algorithms is devoted to numerically integrating SDEs of the form

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, t)\mathrm{d}t + \mathbf{G}(\boldsymbol{x}, s)(\cdot)\mathrm{d}\boldsymbol{W}(s). \tag{4.1}$$

Many definitions introduced in the preceding sections are reproduced for the convenience of the reader. This is not intended to be an exhaustive list, nor do we weigh the relative advantages and disadvantages of the various schemes. Our motivation in listing the schemes we have chosen is twofold. 1) Scientists often wish to "stochastify" an existing numerical model of a deterministic process. Several of the schemes discussed below (e.g., Ewald–Témam and the stochastic Runge–Kutta schemes) have been shown to converge to appropriate stochastic integrals by minimally adjusting well-known deterministic integration schemes. 2) If a stochastic model is to be written from scratch, one may wish to use a scheme that was developed specifically for SDEs. We have included schemes that we, the authors, have found satisfy one or the other need. As before, the first subscript of the matrix $\mathbf{G}$ corresponds to a component of $\boldsymbol{x}$; the second subscript corresponds to a component of $\mathrm{d}\boldsymbol{W}$. It will be made clear if an algorithm can only support one interpretation. In all cases, time step is denoted Δ. The symbol z_α is used to denote a centered (i.e., zero-mean) Gaussian random variable with *variance* Δ. If one chooses

to estimate a weak Ito integral, then z_α will be replaced by a random variable sampled as described in Section 3. However, in what follows, we shall assume that Gaussian estimations for the increment of the Wiener process are being used. The subscript on z_α is used to indicate that this random variable is used as the αth component of a vector Wiener process. The double stochastic integral over first the αth, and then the βth component of the vector Wiener process is denoted $I_{(\alpha,\beta)}$, and it will be clear whether that symbol represents Ito or Stratonovich integration. Finally, the scheme introduced by EWALD and TÉMAM [2003, 2005] has become known as the Ewald–Témam scheme in the meteorological literature (e.g., HANSEN and PENLAND [2006, 2007]) and that is the convention we follow here.

4.1. Euler method

This method converges only to Ito calculus. The order of its convergence was described in Section 3. Given Eq. (4.1), the algorithm is

$$x_i(t+\Delta) = x_i(t) + F_i(\boldsymbol{x}, t)\Delta + \sum_\alpha \boldsymbol{G}_{i\alpha}(\boldsymbol{x}, t) z_\alpha(t). \tag{4.2}$$

One generates the vector $\boldsymbol{z}(t)$ once at each time step and uses that same vector in updating every component of $\boldsymbol{x}$.

4.2. Heun method

This method (MCSHANE [1974], RÜMELIN [1982]) *does* converge, but only to Stratonovich calculus. The order of this convergence is of strong order 0.5 and weak order 1. As in the Euler scheme, we generate a vector $\boldsymbol{z}(t)$ once at each time step. Here, we use the Euler predictor as an intermediate variable

$$x_i'(t+\Delta) = x_i(t) + F_i(\boldsymbol{x}, t)\Delta + \sum_\alpha \boldsymbol{G}_{i\alpha}(\boldsymbol{x}, t) z_\alpha(t), \tag{4.3}$$

and then update $\boldsymbol{x}$ as follows:

$$\begin{aligned} x_i(t+\Delta) = x_i(t) &+ \frac{1}{2}\{F_i(\boldsymbol{x}, t)\Delta + F_i(\boldsymbol{x}', t+\Delta)\}\Delta + \cdots \\ \cdots &+ \frac{1}{2}\sum_\alpha \{\boldsymbol{G}_{i\alpha}(\boldsymbol{x}, t) + \boldsymbol{G}_{i\alpha}(\boldsymbol{x}', t+\Delta)\} z_\alpha(t). \end{aligned} \tag{4.4}$$

The Heun method is a special case of the stochastic Runge–Kutta methods.

4.3. Runge–Kutta methods

The general $(m+1)$th order stochastic Runge–Kutta scheme has been published in RÜMELIN [1982], where its convergence properties are discussed, and is also examined in KLOEDEN and PLATEN [1992]. The reader should be aware that when KLOEDEN and PLATEN [1992] state that the Heun scheme, a second-order Runge–Kutta algorithm, "does not converge," they mean that the scheme does not converge to an Ito solution.

RÜMELIN [1982] provides explicit criteria by which the $(m+1)$th-order Runge–Kutta scheme may be judged to converge to Ito or Stratonovich calculus.

HANSEN and PENLAND [2006] considered a special case of the fourth-order scheme since that is the order most commonly used by scientists using Runge–Kutta algorithms in deterministic modeling. They also used RÜMELIN's [1982] criterion to show that this scheme converges to Stratonovich calculus. The stochastic version of this algorithm involves several intermediate steps:

$$\boldsymbol{K}_o = \boldsymbol{F}(\boldsymbol{x}(t), t), \tag{4.5a}$$

$$\mathbf{M}_o = \mathbf{G}(\boldsymbol{x}(t), t), \tag{4.5b}$$

$$\boldsymbol{x}' = \boldsymbol{x}(t) + \frac{1}{2}\boldsymbol{K}_o\Delta + \frac{1}{2}\mathbf{M}_o z(t), \tag{4.5c}$$

$$\boldsymbol{K}_1 = \boldsymbol{F}(\boldsymbol{x}', t + \frac{1}{2}\Delta), \tag{4.5d}$$

$$\mathbf{M}_1 = \mathbf{G}(\boldsymbol{x}', t + \frac{1}{2}\Delta), \tag{4.5e}$$

$$\boldsymbol{x}'' = \boldsymbol{x}(t) + \frac{1}{2}\boldsymbol{K}_1\Delta + \frac{1}{2}\mathbf{M}_1 z(t), \tag{4.5f}$$

$$\boldsymbol{K}_2 = \boldsymbol{F}(\boldsymbol{x}'', t + \frac{1}{2}\Delta), \tag{4.5g}$$

$$\mathbf{M}_2 = \mathbf{G}(\boldsymbol{x}'', t + \frac{1}{2}\Delta), \tag{4.5h}$$

$$\boldsymbol{x}''' = \boldsymbol{x}(t) + \boldsymbol{K}_2\Delta + \mathbf{M}_2 z(t), \tag{4.5i}$$

$$\boldsymbol{K}_3 = \boldsymbol{F}(\boldsymbol{x}''', t + \Delta), \tag{4.5j}$$

$$\mathbf{M}_3 = \mathbf{G}(\boldsymbol{x}''', t + \Delta), \tag{4.5k}$$

$$\boldsymbol{x}(t+\Delta) = \boldsymbol{x}(t) + \frac{1}{6}(\boldsymbol{K}_o + 2\boldsymbol{K}_1 + 2\boldsymbol{K}_2 + \boldsymbol{K}_3)\Delta + \frac{1}{6}(\mathbf{M}_o + 2\mathbf{M}_1 + 2\mathbf{M}_2 + \mathbf{M}_3)z(t). \tag{4.5l}$$

As before, the same vector of random numbers $z(t)$ is used during the entire updating process.

4.4. Milstein method: explicit version

This method has versions corresponding to either Ito or Stratonovich calculus. Its strong order of convergence is of order 1, as is its weak order. The updating expression is

$$x_i(t+\Delta) = x_i(t) + F_i(\boldsymbol{x}, t)\Delta + \sum_{\alpha} G_{i\alpha}(\boldsymbol{x}, t) z_\alpha + \sum_{j,\alpha,\beta} G_{j\alpha}\frac{\partial G_{i\beta}}{\partial x_j} I_{(\alpha,\beta)}. \tag{4.6}$$

In Eq. (2.12), $I_{(\alpha,\beta)}$ represents the double integral

$$I_{(\alpha,\beta)} = \int_t^{t+\Delta} \mathrm{d}W_\beta(t') \int_t^{t'} \mathrm{d}W_\alpha(t''). \tag{4.7}$$

When $\alpha = \beta$, we estimate $I_{(\alpha,\alpha)}$ as

$$I_{(\alpha,\alpha)} = (z_\alpha^2 - \Delta)/2 \tag{4.8a}$$

for Ito calculus and

$$I_{(\alpha,\alpha)} = z_\alpha^2/2 \tag{4.8b}$$

for Stratonovich calculus. When $\alpha \neq \beta$, $I_{(\alpha,\beta)}$ is the same for both Ito and Stratonovich calculus and must be approximated (KLOEDEN and PLATEN [1992]). In *no* case should it be ignored. In order to achieve a strong convergence order of unity, one chooses an integer $n \geq C/\Delta$, with C some positive constant, and estimates $I_{(\alpha,\beta)}$ as follows:

$$I_{(\alpha,\beta)} = z_\alpha z_\beta/2 + \sqrt{\rho_n}(w_{\alpha,n} z_\beta - w_{\beta,n} z_\alpha) \ldots$$
$$\ldots + \frac{1}{2\pi} \sum_{m=1}^{n} \frac{1}{m} (v_{\alpha,m}(\sqrt{2} z_\beta + r_{\beta,m}) - v_{\beta,m}(\sqrt{2} z_\alpha + r_{\alpha,m})), \tag{4.9a}$$

where

$$\rho_n = \frac{1}{12} - \frac{1}{2\pi^2} \sum_{m=1}^{n} \frac{1}{m^2}, \tag{4.9b}$$

and where, similar to the components of z, the quantities $w_{\alpha,n}$, $v_{\alpha,m}$, $r_{\alpha,m}$, $w_{\beta,n}$, $v_{\beta,m}$, and $r_{\beta,m}$ are all independent Gaussian random variables having mean zero and variance Δ. Each of these random variables is also independent from the components of z. The accuracy of the estimation increases with n, but this estimation can be very expensive computationally. In a large class of special cases, one loses no accuracy by approximating

$$I_{(\alpha,\beta)} = z_\alpha z_\beta/2. \tag{4.10}$$

This class of special cases is known as "commutative noise" and obtains when

$$\sum_k G_{k\alpha}(\boldsymbol{x}, t) \frac{\partial G_{i\beta}(\boldsymbol{x}, t)}{\partial x_k} = \sum_k G_{k\beta}(\boldsymbol{x}, t) \frac{\partial G_{i\alpha}(\boldsymbol{x}, t)}{\partial x_k}. \tag{4.11}$$

If the commutativity relation Eq. (3.5) does not hold, one must ensure that the approximation of $I_{(\alpha,\beta)}$ yields sufficient accuracy. The simplicity of Eq. (4.10) usually makes checking the commutativity relation worthwhile.

If accuracy of order 1 is necessary, the commutativity relation does not hold, and one is trying to estimate an Ito integral, then one of the weak schemes discussed in Section 3 may be appropriate.

4.5. *Milsteyn method: implicit version*

In this and in all implicit algorithms, remember this principle: *Never, ever, put the stochastic term in the part to be inverted.* Depending on your random number generator, you may get lucky for a particular simulation, but the scheme will eventually either blow up or give a spurious contribution so large that it will contaminate the rest of the solution. A cautionary tale is provided when we give examples. That being said, the implicit Milsteyn scheme is recognizable from the explicit version. That is, one chooses the level α_i of implicitness for the ith component of $\boldsymbol{x}$ and uses the updating equation

$$x_i(t+\Delta) = x_i(t) + \{\alpha_i F_i(\boldsymbol{x}, t+\Delta) + (1-\alpha_i)F_i(\boldsymbol{x}, t)\}\Delta \ldots$$

$$\ldots + \sum_{\beta} G_{i\beta}(\boldsymbol{x}, t) z_\beta + \sum_{k\beta\gamma} G_{k\beta}(\boldsymbol{x}, t) \frac{\partial G_{i\gamma}(\boldsymbol{x}, t)}{\partial x_k} I_{(\beta,\gamma)}. \tag{4.12a}$$

In Eq. (4.12a), one estimates the multiple stochastic integral as

$$I_{(\alpha,\beta)} = (z_\alpha z_\beta - \delta_{\alpha,\beta}\Delta)/2 \qquad \text{(Ito)} \tag{4.12b}$$

$$I_{(\alpha,\beta)} = z_\alpha z_\beta/2 \qquad \text{(Stratonovich)} \tag{4.12c}$$

if the noise is commutative. Otherwise, one either employs the more accurate approximations described in the explicit scheme or chooses a weak version. The order of convergence for the implicit Milsteyn scheme is the same as that for the explicit Milsteyn scheme.

4.6. *Platen method: explicit version*

The classic MILSTEYN [1974, 1978] method requires knowledge of how $\mathbf{G}(\boldsymbol{x}, t)$ changes with each component of $\boldsymbol{x}$. For very complicated multiplicative stochastic terms, analytical expressions may be difficult or impossible to evaluate. In this case, there is a two-step process introduced by Platen (KLOEDEN and PLATEN [1992]). For each component of the noise, one evaluates an intermediate value using no random numbers,

$$x'_{i,\alpha}(t+\Delta) = x_i(t) + F_i(\boldsymbol{x}, t)\Delta + G_{i\alpha}(\boldsymbol{x}, t)\sqrt{\Delta}. \tag{4.13a}$$

One then updates $\boldsymbol{x}$ as follows:

$$x_i(t+\Delta) = x_i(t) + F_i(\boldsymbol{x}, t)\Delta + \sum_{\alpha} G_{i\alpha}(\boldsymbol{x}, t) z_\alpha$$

$$+ \frac{1}{\sqrt{\Delta}} \sum_{\alpha,\beta} \{G_{i\beta}(\boldsymbol{x}'_\alpha, t) - G_{i\beta}(\boldsymbol{x}, t)\} I_{(\alpha,\beta)}. \tag{4.13b}$$

As in the Milsteyn method, the difference between Ito and Stratonovich calculus is made in the estimation of $I_{(\alpha,\beta)}$.

4.7. Implicit strong Runge–Kutta scheme

KLOEDEN and PLATEN [1992] call this algorithm the implicit order 1.0 strong Runge–Kutta scheme, although it is not a simple heuristic adaptation of the deterministic scheme, because it is derived using a strong stochastic Taylor expansion and involves intermediate variables. Again, one chooses the level α_i of implicitness for the ith component of $\boldsymbol{x}$. The supporting values are evaluated as in Eq. (3.7), and the updating equation is

$$x_i(t+\Delta) = x_i(t) + \{\alpha_i F_i(\boldsymbol{x}(t+\Delta), t+\Delta) + (1-\alpha_i)F_i(\boldsymbol{x}, t)\}\Delta \ldots$$

$$\cdots + \sum_{\beta} G_{i\beta}(\boldsymbol{x}, t) z_\beta + \frac{1}{\sqrt{\Delta}} \sum_{\alpha,\beta} \{G_{i\beta}(\boldsymbol{x}'_\alpha, t) - G_{i\beta}(\boldsymbol{x}, t)\} I_{(\alpha,\beta)}. \tag{4.14}$$

4.8. Ewald–Témam: explicit scheme

The explicit and implicit versions of the Ewald–Téman scheme have both strong and weak convergence order of one. The explicit Ewald–Témam is a modification of the Milsteyn scheme where the explicit tendency is replaced with a discretization (EWALD and TÉMAM [2003, 2005]):

$$x_i(t+\Delta) = x_i(t) + F_i(\boldsymbol{x}, t)\Delta + \sum_{\alpha} G_{i\alpha}(\boldsymbol{x}, t) z_\alpha \ldots$$

$$\cdots + \sum_{j,\alpha,\beta} G_{j\alpha}(\boldsymbol{x}, t) \frac{\{G_{i\beta}(\boldsymbol{x} + \varepsilon_j \sqrt{\Delta}\hat{\boldsymbol{e}}_j, t) - G_{i\beta}(\boldsymbol{x}, t)}{\varepsilon_j \sqrt{\Delta}} I_{(\alpha,\beta)}. \tag{4.15}$$

In Eq. (4.15), $\hat{\boldsymbol{e}}_j$ is a unit vector corresponding to the component x_j. The vector ε has components less than or equal to unity, in units of $\boldsymbol{x}/\sqrt{t}$, and allows the modeler to adjust the discretized derivatives to the problem at hand. The comments already made concerning the double stochastic integral also apply here.

4.9. Ewald–Témam: implicit scheme

This scheme was devised to accommodate the architecture of extant climate models (EWALD and TÉMAM [2003, 2005]), including barotropic vorticity models (e.g., SARDESHMUKH and HOSKINS [1988]) and full general circulation models (e.g., SAHA, NADIGA, THIAW, WANG, WANG, ZHANG, VAN DEN DOOL, PAN, MOORTHI, BEHRINGER, STOKES, PENA, LORD, WHITE, EBISUZAKI, PENG and XIE [2006]). As such, this scheme is obviously different from any of the previous schemes described in this article. Deterministic climate models usually integrate the state vector first using a leapfrog step, followed by an implicit step. To implement the stochastic analog of this procedure, we rewrite Eq. (4.2) as

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{F}_1(\boldsymbol{x}, t)\mathrm{d}t + \boldsymbol{F}_2(\boldsymbol{x}, t)\mathrm{d}t + \mathbf{G}(\boldsymbol{x}, t)(\cdot)\mathrm{d}\boldsymbol{W}. \tag{4.16}$$

In Eq. (3.10), $\boldsymbol{F}_1(\boldsymbol{x}, t)$ is the explicit part of the model and $\boldsymbol{F}_2(\boldsymbol{x}, t)$ is the implicit part. The implicit leapfrog scheme of EWALD and TÉMAM [2003, 2005] is as follows:

$$\boldsymbol{x}'(t+2\Delta) = \boldsymbol{x}(t) + 2\boldsymbol{F}_1(\boldsymbol{x}(t+\Delta), t+\Delta)\Delta + \boldsymbol{M}(\boldsymbol{x}(t), t) + \boldsymbol{M}(\boldsymbol{x}(t+\Delta), t+\Delta), \tag{4.17a}$$

$$\boldsymbol{x}(t+2\Delta) = \boldsymbol{x}'(t+2\Delta) + 2\boldsymbol{F}_2(\boldsymbol{x}(t+2\Delta), t+2\Delta)\Delta. \tag{4.17b}$$

In the updating expressions Eq. (4.17a), the ith component of the vector $\boldsymbol{M}(\boldsymbol{x}(t), t)$ is

$$M_i(\boldsymbol{x}, t) = \sum_{j\alpha\beta} G_{j\alpha}(\boldsymbol{x}, t)\frac{\partial G_{i\beta}(\boldsymbol{x}, t)}{\partial x_j} I_{(\alpha,\beta)} + \sum_{\alpha} G_{i\alpha}(\boldsymbol{x}, t) z_\alpha. \tag{4.17c}$$

Again, we approximate the derivative in Eq. (4.17c) as

$$\frac{\partial G_{i\beta}(\boldsymbol{x}, t)}{\partial x_j} = \frac{G_{i\beta}(\boldsymbol{x} + \varepsilon_j\sqrt{\Delta}\hat{\boldsymbol{e}}_j, t) - G_{i\beta}(\boldsymbol{x}, t)}{\varepsilon_j\sqrt{\Delta}}, \tag{4.17d}$$

with all symbols defined as in the explicit case.

As written, this scheme is of strong order 1 for both Ito and Stratonovich systems. If a weak order 1 Ito scheme is sufficient, the $I_{(\alpha,\beta)}$ terms in Eq. (4.17c) may be neglected.

5. Random number generators

All of the stochastic numerical schemes require random deviates. The Gaussian deviates required by strong schemes may be obtained by applying the Box–Müller (PRESS et al. [1992]) technique to output from pseudorandom number generators, which provide numbers evenly distributed on the interval (0,1), and the non-Gaussian deviates required by weak schemes also employ output from pseudorandom generators. Many computer languages have intrinsic random number generators that claim to fill this need, and some of them are dangerous for our purposes. According to GARCÍA [2000], C++ has a random number generator in the <stdlib.h> that is *not* intended for scientific programming; it repeats the same sequence of numbers after 33,000 calls. For comparison, a computer model with a triangular truncation of T382 in the horizontal and 64 levels in the vertical has more than 5,500,000 grid points. Perhaps the best random number generator is called the Mersenne Twister (MATSUMOTO and NISHIMURA [1998]), which is freely available from the website *http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html*. The period of this generator is an enormous 10^{6000}. We are unaware of any problems associated with this random number generator and would be grateful if anyone who finds any were to inform us.

6. Frequently asked questions

Before we begin exploring some common mistakes, let us note that when the noise is purely additive, many of the more complicated schemes are equivalent to the stochastic Euler scheme. Some of the examples below use additive noise for simplicity; multiplicative noise only exacerbates the problems discussed in this section.

6.1. Why do we need the square roots in the noise terms?

Without treating the system as a legitimately generated Wiener process, we do not have dynamical consistency. Consider the simple Ornstein–Uhlenbeck (OU) process below:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\gamma x + \eta\xi. \tag{6.1}$$

In Eq. (6.1), white noise is denoted ξ, and η is a constant. It is easy to show that the stationary probability density is a centered Gaussian with variance $< x^2 > = \eta^2/2\gamma$. With values of $\gamma = 0.05$ and $\eta = 1$, $< x^2 > = 10$. Now we use the Euler method to numerically integrate Eq. (3.2), employing the relation $\mathrm{d}W = \xi \mathrm{d}t$. We shall also treat Eq. (6.1) naively, by treating $\eta\xi$ as a discrete white noise with unit variance at every time step and using the deterministic Euler scheme $x(t + \Delta) = x(t) + (-\gamma x(t) + \eta\xi)\Delta$. Figure 6.1 shows estimates of $< x^2 >$ for 10,000 samples from a time series sampled every 2.4 time units, using time steps of 0.2, 0.4, 0.6, and 0.8 time units. It is clear that the variance of the naively generated system increases monotonically with the time step, even for time steps an order of magnitude smaller than the standard deviation of the system to be generated. In fact, perusal of the schemes indicates that the variance for this system integrated naively should increase linearly. The stochastic Euler system, on the other hand, gives a reasonably accurate estimate of the variance. The moral of this story, of course, is that throwing random numbers into a deterministic integration scheme will provide noise with a variance dependent upon the time step.

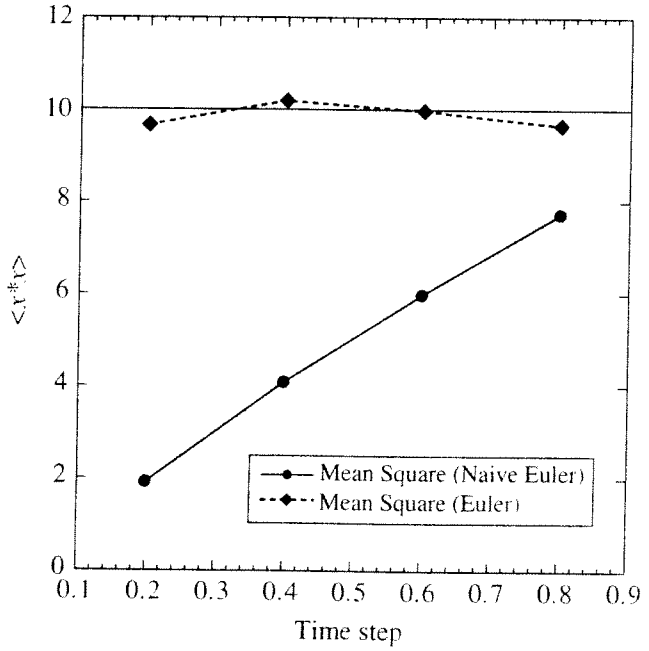

FIG. 6.1 Mean square of 10,000 samples of Eq. (6.1). Solid circles: Naive Euler scheme. Solid diamonds: stochastic Euler scheme, Eq. (4.1).

6.2. What happens if I use more than one random vector per time step?

There may be a temptation to hard-wire a random number generator into the code at every place the symbol z_α appears in the updating equation. If the system is multivariate, doing so is equivalent to using more than one random vector per time step. This is deadly if the correlations between variables are important. As an illustration, consider two damped oscillators coupled only through the driving noise:

$$d\boldsymbol{x} = \mathbf{L}\boldsymbol{x}dt + \mathbf{G}d\boldsymbol{W}, \tag{6.2a}$$

where

$$\mathbf{L} = \begin{pmatrix} -0.5 & 1 & 0 & 0 \\ -1 & -0.5 & 0 & 0 \\ 0 & 0 & -0.5 & 1 \\ 0 & 0 & -1 & -0.5 \end{pmatrix} \tag{6.2b}$$

and

$$\mathbf{G} = \begin{pmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & 0 & \sqrt{1.5} & 0 \\ 0 & 1/\sqrt{2} & 0 & \sqrt{1.5} \end{pmatrix}. \tag{6.2c}$$

From the fluctuation–dissipation relationship (e.g., PENLAND and SARDESHMUKH [1995], NEWMAN, SARDESHMUKH and PENLAND [1997]), this system yields a covariance matrix

$$<\boldsymbol{x}\boldsymbol{x}^T> = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}. \tag{6.2d}$$

Generating the vector of random numbers before employing it in the time stepping procedure, as one ought to do, yielded the following sample covariance matrix:

$$\mathbf{C}_{\text{Sample}} = \begin{pmatrix} 1.979 & -0.006 & 0.993 & -0.029 \\ -0.006 & 1.990 & 0.016 & 1.001 \\ 0.993 & 0.016 & 2.024 & -0.013 \\ 0.002 & 1.001 & -0.013 & 2.010 \end{pmatrix}, \tag{6.2e}$$

whereas hard-wiring the random number generator into the updating equation yielded

$$\mathbf{C}_{\text{Hard-wire}} = \begin{pmatrix} 2.002 & 0.014 & -0.013 & 0.002 \\ 0.014 & 2.020 & 0.012 & -0.005 \\ -0.013 & 0.012 & 2.014 & 0.002 \\ 0.002 & -0.005 & 0.002 & 2.016 \end{pmatrix}. \tag{6.2f}$$

The vector of random numbers must be preserved during the entire course of the time step, or the correlations between state vector components are lost. This occurs with additive noise; the consequences of being cavalier with multiplicative noise can be dire indeed.

6.3. What happens if I use a naive implicit scheme?

This cautionary tale is reproduced from EWALD, PENLAND and TÉMAM [2004]. Consider the following forced diffusion equation:

$$\frac{dx}{dt} = (k^2 - r)x + F, \tag{6.3a}$$

where $r = r_o + r_s\eta$, and where η is white noise. That is,

$$rdt = r_o dt + r_s dW. \tag{6.3b}$$

An analytical solution for the stationary probability distribution function (*pdf*) is derived from the Fokker–Planck equation (HORSTHEMKE and LÉFÈVER [1984]) and is shown in Fig. 6.2 (heavy solid line) for the parameters $k = 0.1$, $r_o = 0.51$, $r_s = 0.5$, and $F = 0.5$. Also shown is the sample *pdf* from an integration of Eq. (6.3) using the implicit Ewald–Témam (filled circles). In this implementation, we have defined $a_1 = F$ and $a_2 = (k^2 - r_o)x$; further details may be found in EWALD, PENLAND and TÉMAM [2004]. For comparison is the sample *pdf* estimated from a time series generated using a naive application of an implicit scheme as follows:

$$x'(t + 2\Delta) = x(t) + 2a_1(x(t + \Delta), t + \Delta)\Delta, \tag{6.4a}$$

$$x(t + 2\Delta) = x(t) + 2a_2(x'(t + 2\Delta), t + 2\Delta)\Delta. \tag{6.4b}$$

Eq. (6.4) was used to integrate Eq. (6.3a) with η estimated as a Gaussian random variable of unit variance. This exercise shows the danger of throwing random numbers into a deterministic model and expecting to get useful results.

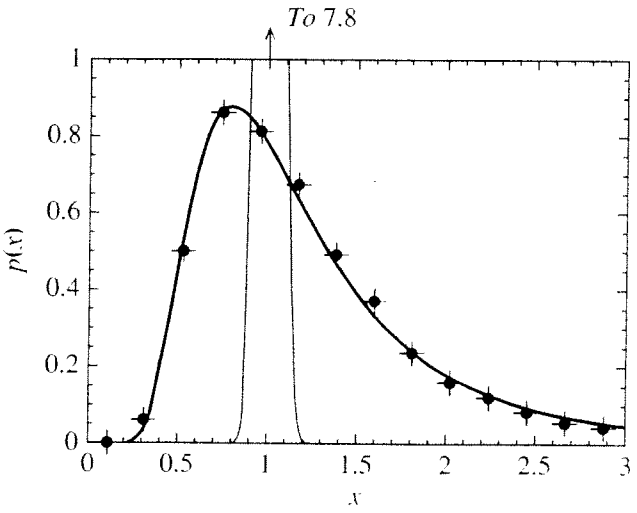

FIG. 6.2 Probability density function estimated from integrating Eq. (6.3). Heavy solid line: theoretically expected *pdf*. Filled circles with crosses: implicit Ewald–Témam. Light solid line: naive implicit scheme.

6.4. *How do I implement stochastic forcing in a dynamical model based on measured statistics?*

There is no single answer to this question. A stochastic parameterization cannot be implemented unless something is known about the timescales of the system to be parameterized. The common practice of simply augmenting a deterministic parameter in a model with an uncorrelated random component of specified, experimentally estimated, variance can cause most of the undesirable results discussed in this section (see, e.g., Fig. 6.2). Some of these effects can be ameliorated somewhat by using red noise, rather than white noise, but this is often extremely difficult in a high-dimensional, massively parallelized numerical model. Besides, using red noise rather than white noise does not guarantee the problems will go away. In the following, we consider two procedures that are appropriate for important special cases.

One solution to the quandary is possible when the timescale of the fast system is known at least to some approximation. In that case, the central limit theorem can be applied and the equivalent stochastic forcing identified. SARDESHMUKH, PENLAND and NEWMAN [2001], for example, applied this procedure to the linearized barotropic vorticity equation with stochastic fluctuations in the zonally symmetric mean velocity. After accounting for the effects of the annual cycle and El Niño, their four-times-daily measured velocities were found to have a spectrum varying roughly as the inverse square of the frequency. They, therefore, decided to approximate the fluctuations as an Ornstein–Uhlenbeck process with a timescale fast enough that the rest of the system would see those fluctuations as white. As is well known (e.g., HORSTHEMKE and LÉFÈVER [1984]) and as is readily verified by the formulas provided in Section 2, the white noise approximation of an Ornstein–Uhlenbeck process u with variance σ^2 and decay rate γ is

$$u\mathrm{d}t \approx (\sigma\sqrt{2/\gamma})\mathrm{d}W. \tag{6.5}$$

The larger σ and γ are, with σ^2/γ remaining constant, the better the approximation is.

It turned out that the decay time $\tau_s = 1/\gamma$ of the velocity fluctuations considered by SARDESHMUKH, PENLAND and NEWMAN [2001] was somewhat too large to present a legitimate candidate for a white noise parameterization (SARDESHMUKH, PENLAND and NEWMAN [2003]). However, when the decay time is small enough, Eq. (6.5) can be used to parameterize a physical quantity u stochastically.

A second procedure (HANSEN and PENLAND [2007]) can be used if the parameters of the stochasticity are diagnosed using data assimilation in, for example, a forecast model. That study was originally concerned with whether or not standard data assimilation procedures, such as various versions of the ensemble Kalman filter (ANDERSON [2001], BISHOP, ETHERTON and MAJUMDAT [2001], EVENSEN [1994], WHITAKER and HAMILL [2002]), could be applied to diagnosing the coefficients of random terms in a dynamical model. The answer was a reserved "yes," but the reservation had implications for stochastic integration.

The technique is best described using the example presented by HANSEN and PENLAND [2007], who considered a stochastic version of the chaotic Lorenz system:

$$\begin{aligned} dx &= -a_0(x - y)dt - a_s(x - y) \cdot dW, \\ dy &= (r_0 x - xz - y)dt, \\ dz &= (xy - b_0 z)dt. \end{aligned} \tag{6.6}$$

In Eq. (6.6), the typical chaotic parameter values are $a_0 = 10$, $r_0 = 28$, and $b_0 = 8/3$. In addition, there is a stochastic component to be integrated in the sense of Stratonovich, with an addition parameter $a_s = 0.1$. HANSEN and PENLAND [2007] integrated Eq. (6.6) using a stochastic fourth-order Runge–Kutta scheme (see Section 4c) using a time step (the details are important) of 0.00025 model units (mu). The resulting trajectory was taken as "truth." Employing the deterministic Lorenz model (i.e., Eq. (6.6) with $a_s = 0$) but augmented with the condition $da_0/dt = 0$, as the assimilation model, the ensemble Kalman filter was used to assimilate the state vector (x, y, z, a_0). The deterministic Lorenz model was integrated using a standard fourth-order Runge–Kutta method with a time step of 0.01mu, and data from the "true" trajectory were assimilated every fifth time step, i.e., every $\tau_{obs} = 0.05$ mu. Data assimilation resulted in an estimate of $a_0 = 10.2$ with a standard deviation of that estimate being 0.42.

It is now that the central limit theorem (Section 2) becomes important. Our "truth" (Eq. 6.6) looks like Eq. (2.14). Our assimilation model looks like Eq. (2.1), with the x-component of $\varepsilon^2 \boldsymbol{F}$ identified as $a_0(x - y)$ and with $\varepsilon \boldsymbol{G}$ identified as $(x - y)$ times a centered (zero-mean) variable having a standard deviation of 0.42 in the same units as a_0. During the data assimilation, that variable is held fixed for $\tau_{obs} = 0.05$ mu after which another independent value is inserted into the assimilation model. Thus, the integrated lagged covariance (Eq. 2.8) of this variable is simply $(0.42)^2 \tau_{obs}$, yielding a coefficient of the Wiener process equal to $(x - y)$ times the square root of that, or 0.098 $(x - y)$. This is not far from the true expression 0.1 $(x - y)$.

All this may be interesting, but the relation to stochastic integration is not obvious until one asks, "If the central limit theorem worked in the data assimilation scheme, why won't it work in forward integration?" In fact, it does. If we go to all the trouble to assimilate the stochastic parameter into a forecast model, the Stratonovich SDE is estimated to a remarkably accurate extent by a piecewise deterministic model, where the stochastic parameter is drawn from a Gaussian distribution having the *observed* mean and standard deviation (10.2 and 0.42, respectively, in the Hansen–Penland example). It is then injected into the forward integration, after which it is *held constant for exactly the assimilation period* ($\tau_{obs} = 0.05$ mu in this example) before it is replaced by another draw from the same distribution. Of course, for this to work, the assimilation period still has to be much smaller than the timescales of the dynamics the forecast model is trying to predict, or the central limit theorem is not satisfied.

7. Discussion

As the popularity of stochastic climate modeling increases, so does the temptation to do a quick and dirty job of it. The traditional identity of uncertainty as representing the extent to which a scientist "fails" to eliminate uncontrolled experimental variables causes

many scientists to harbor a certain discomfort with dynamically generated stochasticity, resulting in a very human denial of its importance. Once this denial can no longer be maintained, one may try to account for stochasticity in a numerical model while minimizing the effort needed to do so by simply adding a random number generator to the deterministic model. Unfortunately, it has been shown many times that this attitude can lead to wrong answers that look reasonable. We have tried to summarize approaches that combine necessary rigor and ease.

Ease is nearly as important as rigor. The immense practical difficulties of implementing stochastic numerical techniques in state-of-the art general circulation models are real and should be approached with sympathy and respect. It is for this reason that we have included not only numerical algorithms commonly employed for their theoretical advantages but also those algorithms that are similar to those already employed by geoscientists in numerical models. For example, the dynamical core of many deterministic models in climate research consists of a semi-implicit leapfrog scheme; the implicit Ewald–Témam scheme is therefore ideally suited for such models.

In this chapter, we have tried to clarify the reasons for the existence of two calculi found in nature. We have summarized some of their basic properties, including the differences and connections between them as well as a rule of thumb for the use of each: if the system one wishes to model derives its stochasticity from unresolved continuous chaotic processes, Stratonovich calculus is appropriate. If the process being modeled possesses discrete, uncorrelated components that are treated as approximately continuous, the appropriate calculus is Ito (HORSTHEMKE and LÉFÈVER [1984]). Practically, the difference between these calculi is in the definition of the integral over a Wiener process (Brownian motion) and the numerical approximation of that integral.

Although it is possible for the reader to use the algorithms presented here without understanding the basic theory behind them, we cannot recommend such a course. We have therefore presented a review of the stochastic Taylor series and its role in developing the numerical algorithms for integrating SDEs. In particular, we have noted that one need use strong schemes only if the actual trajectory of the system through phase space is required. If one is interested only in the statistics of the system, and this is usually the case, one may employ the more efficient weak schemes. As we have seen, a strong scheme of one order of convergence is equivalent to a weak scheme of at least that order. Sometimes, order of convergence is not the primary consideration in choosing a scheme; however, the coefficient in front of the order estimate can also make a difference. For this reason, if at all possible, we encourage modelers to experiment with different schemes on simple systems for which the analytic solution is known.

Finally, we urge the readers to refrain from taking "short cuts" since most of the theorems pertaining to deterministic numerical integrations do not apply to the stochastic case. The numerical integration of SDEs may be somewhat unfamiliar to traditional climate modelers, but stochastic schemes are well defined, straightforward, and increasingly common. For us, it is indeed welcome to see widespread advantage taken of these useful and elegant techniques.

References

ANDERSON, J.L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2284–2903.

ARNOLD, L. (1974). *Stochastic Differential Equations: Theory and Applications* (John Wiley and Sons, New York), pp. 228.

BHATTACHARYA, R.N., WAYMIRE, E.C. (1990). *Stochastic Processes with Applications* (John Wiley and Sons, New York), pp. 672.

BISHOP, C.H., ETHERTON, B.J., MAJUMDAT S.J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Wea. Rev.* **129**, 420–436.

DOOB, J.L. (1953). *Stochastic Processes* (John Wiley and Sons, New York), pp. 654.

EVENSEN, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10143–10162.

EWALD, B.D., TÉMAM, R. (2003). Analysis of stochsastic numerical schemes for the evolution equations of geophysics. *Appl. Math. Lett.* **16**, 1223–1229.

EWALD, B.D., TÉMAM, R. (2005). Numerical analysis of stochastic schemes in geophysics. *SIAM J. Numer. Anal.* **42**, 2257–2276.

EWALD, B., PENLAND, C., TÉMAM, R. (2004). Accurate integration of stochastic climate models. *Mon. Wea. Rev.* **132**, 154–164.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications, vol. 2* (John Wiley and Sons, New York), pp. 626.

GARCÍA, A.L. (2000). *Numerical Methods for Physics*, second ed. (Prentice Hall, Upper Saddle River, New York), pp. 423.

GARDINER, C.W. (1985) *Handbook of Stochastic Methods*, second ed. (Springer-Verlag, Berlin, Germany), pp. 442.

HANSEN, J., PENLAND, C. (2006). Efficient approximation techniques for integrating stochastic differential equations. *Mon. Wea. Rev.* in press.

HANSEN, J., PENLAND, C. (2007). On stochastic parameter estimation using data assimilation. *Physica D* in press.

HORSTHEMKE, W., LÉFÈVER, R. (1984). *Noise-Induced Transitions. Theory and Applications in Physics, Chemistry, and Biology* (Springer-Verlag, Berlin, Germany), pp. 318.

KHASMINSKII, R.Z. (1966). A limit theorem for solutions of differential equations with random right-hand side. *Theory Prob. Appl.* **11**, 390–406.

KLOEDEN, P.E., PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, Germany), pp. 632.

KOHLER, W., PAPANICOLAOU, G.C. (1977). Wave propagation in a randomly inhomogeneous ocean. In: Keller, J.B., Papadakis, J.S. (eds.), *Chapter IV of Wave Propagation and Underwater Acoustics. Lecture Notes in Physics, vol. 70* (Springer Verlag, Berlin, Germany), pp. 153–223.

MAJDA, A.J., TIMOFEYEV, I., VANDEN EIJNDEN, E. (1999). Models for stochastic climate prediction. *Proc. Natl. Acad. Sci.* (USA), **96**, 14687–14691.

MATSUMOTO, M., NISHIMURA, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. Model. Comput. Simul.* **8**, 3–30.

MCSHANE, E.J. (1974). *Stochastic Calculus and Stochastic Models* (Academic Press, New York).

MILSTEYN, G.N. (1974). Approximate integration of stochastic differential equations. *Theor. Prob. Appl.* **19**, 557–562.

MILSTEYN, G.N. (1978). A method of second-order accuracy integration of stochastic differential equations. *Theor. Prob. Appl.* **23**, 396–401.

NEWMAN, M., SARDESHMUKH, P.D., PENLAND, C. (1997). Stochastic forcing of the wintertime extratropical flow. *J. Atmos. Sci.* **54**, 435–455.

PAPANICOLAOU, G.C., KOHLER, W. (1974). Asymptotic theory of mixing stochastic differential equations. *Commun. Pure Appl. Math.* **27**, 641–668.

PENLAND, C. (1985). Acoustic normal mode propagation through a three-dimensional internal wave field. *J. Acoust. Soc. Am.* **78**, 1356–1365.

PENLAND, C., SARDESHMUKH, P.D. (1995). The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999–2024.

PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T., FLANNERY, B.P. (1992). Numerical Recipes in Fortran, *The Art of Scientific Computing*, second ed. (Cambridge University Press, Cambridge), pp. 963.

RÜMELIN, W. (1982). Numerical treatment of stochastic differential equations. *SIAM J. Numer. Anal.* **19**, 605–613.

SAHA, S., NADIGA, S., THIAW, C., WANG, J., WANG, W., ZHANG, Q., VAN DEN DOOL, H.M., PAN, H.-L., MOORTHI, S., BEHRINGER, D., STOKES, D., PENA, M., LORD, S., WHITE, G., EBISUZAKI, W., PENG, P., XIE, P. (2006). The NCEP climate forecast system. *J. Clim.* **19**, 3483–3517.

SARDESHMUKH, P.D., HOSKINS, B.J. (1988). The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.* **45**, 1228–1251.

SARDESHMUKH, P.D., PENLAND, C., NEWMAN, M. (2001). Rossby waves in a fluctuating medium. In: Imkeller, P., von Storch, J.-S. (eds.), *Stochastic Climate Models*. In: Progress in Probability **49** (Birkhaueser Verlag, Basel, Switzerland).

SARDESHMUKH, P.D., PENLAND, C., NEWMAN, M. (2003). Drifts induced by multiplicative red noise with application to climate. *Europhys. Lett.* **63**, 498–504.

WHITAKER, J.S., HAMILL, T.M. (2002). Ensemble data assimilation withoug perturbed observations. *Mon. Wea. Rev.* **131**, 1485–1490.

WILKS, D.S. (1995). *Statistical Methods in the Atmospheric Sciences* (Academic Press, San Diego, CA), pp. 467.

WONG, E., ZAKAI, M. (1965). On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Stat.* **36**, 1560–1564.