Smooth random functions, random ODEs, and Gaussian processes

Silviu-Ioan Filip, Aurya Javeed, Lloyd Nicholas Trefethen

SMOOTH RANDOM FUNCTIONS, RANDOM ODEs, AND GAUSSIAN PROCESSES

SILVIU FILIP*, AURYA JAVEED†, AND LLOYD N. TREFETHEN‡

Abstract. The usual way that mathematicians work with randomness is by a rigorous formulation of the idea of Brownian motion, which is the limit of a random walk as the step length goes to zero. A Brownian path is continuous but nowhere differentiable, and this non-smoothness is associated with technical complications that can be daunting. However, there is another approach to random processes that is more elementary, involving *smooth random functions* defined by finite Fourier series with random coefficients, or equivalently, by trigonometric polynomial interpolation through random data values. We show here how smooth random functions can provide a very practical way to explore random effects. For example, one can solve smooth random ordinary differential equations using standard mathematical definitions and numerical algorithms, rather than having to develop new definitions and algorithms of stochastic differential equations. In the limit as the number of Fourier coefficients defining a smooth random function goes to ∞, one obtains the usual stochastic objects in what is known as their Stratonovich interpretation.

Key words. band-limited white noise, Brownian motion, Chebfun, Dirichlet kernel, energy landscape, Fourier–Wiener series, Gaussian process, stochastic DE, white noise

AMS subject classifications. 42A16, 60G15, 62M40

1. Introduction.

This paper is about random functions of the kind illustrated in Figure 1.1. These functions are smooth, in fact band-limited, though they approach non-smooth form as the wavelength parameter λ approaches 0. They are defined both in a standard normalization, where it is function values themselves that are of interest, and in what we call the *big* normalization, of amplitude $\sqrt{2/\lambda}$ times larger, where it is integrals that are of interest to drive smooth analogues of Brownian motion and stochastic differential equations (SDEs).

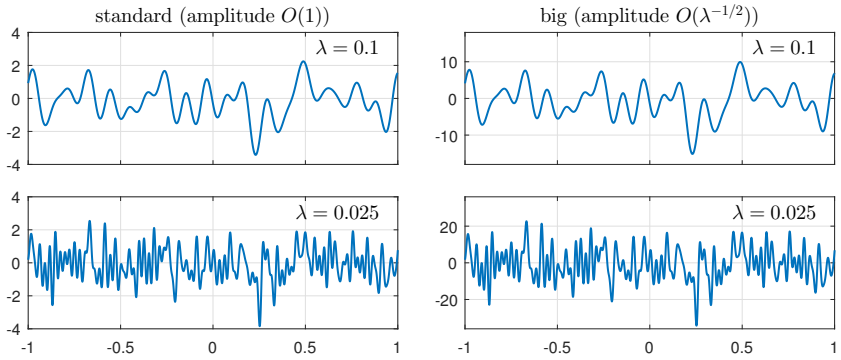

FIG. 1.1. *Smooth random functions on* $[-1, 1]$ *with wavelength parameters* $\lambda = 0.1$ *and* 0.025. *The two columns look the same at first glance, but their vertical scales are different, illustrating that two different choices of normalization are appropriate depending on whether one is concerned with function values (left column) or their integrals (right). In Chebfun, these functions can be generated by the commands* `randnfun(lambda)` *and* `randnfun(lambda,'big')`, *respectively.*

*`filips@maths.ox.ac.uk`, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK.

†`aj463@cornell.edu`, Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA.

‡`trefethen@maths.ox.ac.uk`, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK.

We developed smooth random functions for practical computation in the software package Chebfun [11, 46], whose aim is to enable MATLAB-style computing with continuous functions and operators as opposed to discrete vectors and matrices. Indeed, the starting point of this study was the question, what might be a continuous analogue of a random vector generated by the MATLAB command `randn(n,1)`? Our definitions allow the domain to be any interval $[a, b]$, and extensions to multiple dimensions are presented in Section 7.

Nothing in this article is mathematically or scientifically new. The idea of Fourier series with random coefficients ("Fourier–Wiener series") goes back to Wiener in 1924 [48], with generalizations by Paley, Wiener, and Zygmund in 1933–34 [36]. A leader in this subject for many years was Jean-Pierre Kahane [21, 22], and there is an advanced monograph by Marcus and Pisier [30]. Random series have been used extensively for time series analysis [17], computational simulations [54], and theoretical applications in a range of fields, both in one and in higher dimensions (see Section 7). Our aim in this paper is to present this idea in the form of a very concrete, easily accessible tool that may be useful for those learning or exploring random phenomena. Rather than starting with technicalities of Gaussian processes or stochastic differential equations, we go directly to the smooth random functions, while giving references to published sources for fuller and more rigorous treatments.

A key feature of this paper is a sequence of Chebfun-based examples illustrating how smooth random functions can be used in applications. Further examples can be found in Chapter 12 of [47], where these functions were first introduced, and in the online examples collection at `www.chebfun.org`.

2. Periodic smooth random functions. Let $\lambda > 0$ be a wavelength parameter, and suppose we are interested in periodic random functions on the interval $[-L/2, L/2]$ for some $L > 0$; an interval $[a, b]$ is handled by the obvious translation. Our definitions based in the Fourier domain are as follows. In a moment we shall give equivalent definitions based in the space domain.

DEFINITION 2.1. *A* COMPLEX PERIODIC SMOOTH RANDOM FUNCTION *for given $\lambda, L > 0$ is a function*

$$f(x) = \sum_{j=-m}^{m} c_j \exp\left(\frac{2\pi i j x}{L}\right), \quad m = \lfloor L/\lambda \rfloor, \tag{2.1}$$

where each c_j is an independent sample from $N(0, 1/(2m+1)) + iN(0, 1/(2m+1))$. A REAL PERIODIC SMOOTH RANDOM FUNCTION *is the real part of a complex one. Equivalently, it is a function*

$$f(x) = a_0 + \sqrt{2} \sum_{j=1}^{m} \left[a_j \cos\left(\frac{2\pi j x}{L}\right) + b_j \sin\left(\frac{2\pi j x}{L}\right) \right], \quad m = \lfloor L/\lambda \rfloor, \tag{2.2}$$

where each a_j and b_j is an independent sample from $N(0, 1/(2m+1))$.

As usual, $N(\mu, V)$ denotes the real normal distribution of mean μ and variance V, and $\lfloor \cdot \rfloor$ is the floor function. According to standard terminology, f is a *trigonometric polynomial of degree m* [52]. To verify that (2.2) is equivalent to the real part of (2.1), we can write $c_j = \alpha_j + i\beta_j$, where α_j and β_j are independent samples from $N(0, 1/(2m+1))$, and note that $\exp(2\pi i j x/L) = \cos(2\pi j x/L) + i \sin(2\pi j x/L)$.

Grouping together real terms, we find that the real part of (2.1) can be expanded as

$$\alpha_0 + \sum_{j=1}^{m} \left[(\alpha_j + \alpha_{-j}) \cos\left(\frac{2\pi j x}{L}\right) + (-\beta_j + \beta_{-j}) \sin\left(\frac{2\pi j x}{L}\right) \right].$$

Since α_j and α_{-j} are independent samples from $N(0, 1/(2m+1))$ for $j \geq 1$, their sum is a sample from $N(0, 2/(2m+1))$, hence equivalent to $\sqrt{2}a_j$ with a_j from $N(0, 1/(2m+1))$; similarly for the terms involving β_j and β_{-j}.

A theorem summarizes some of the properties of these functions. We say that a periodic function is *k-band-limited* if it can be written as a Fourier series with wave numbers confined to $[-k, k]$.

THEOREM 2.2. *A periodic smooth random function f (whether real or complex) is L-periodic, entire, and $(2\pi/\lambda)$-band-limited. The stochastic process from which f is a sample is stationary (i.e., it has a distribution that is translation-invariant), with values $f(x)$ at each x distributed according to $N(0,1) + iN(0,1)$ in the complex case and $N(0,1)$ in the real case.*

Proof. The periodicity is immediate from (2.1) or (2.2), and f is entire (i.e., analytic throughout the complex x-plane) since it is a finite sum of complex exponentials or sines and cosines. Since $|j| \leq m \leq L/\lambda$, the maximum value of the coefficients $|2\pi j/L|$ in (2.1) or (2.2) is bounded by $2\pi/\lambda$, so f is $(2\pi/\lambda)$-band-limited. Stationarity of the stochastic process follows from (2.1) since translating a function $c_j \exp(2\pi i j x/L)$ amounts to changing the argument but not the modulus of c_j and the distribution $N(0, 1/(2m+1)) + iN(0, 1/(2m+1))$ is argument-invariant. Hence any translated process has an identical distribution to the original. The same argument-invariance of $N(0, 1/(2m+1)) + iN(0, 1/(2m+1))$ also ensures that the sum in (2.1) is distributed according to $N(0,1) + iN(0,1)$ at each point x, and its real part accordingly has the distribution $N(0,1)$. □

These definitions of random functions are rooted in the Fourier domain: they describe a random function as a sum of Fourier modes with random amplitudes. (We may think either of random complex amplitudes or of random real amplitudes coupled with random phases.) Equivalently, as suggested in Figure 2.1, we can construct random functions in the spatial domain. The essential point here is that there is an equivalence between the $2m+1$ Fourier series coefficients $\{c_j\}$ of (2.1) and the $2m+1$ function values

$$d_j = f(x_j), \quad -m \leq j \leq m,$$

where the equispaced gridpoints x_j are defined by

$$x_j = jh, \quad h = \frac{L}{2m+1}, \quad -m \leq j \leq m. \tag{2.3}$$

If $\mathbf{c} = (c_0, \ldots, c_m, c_{-m}, \ldots, c_{-1})^T$ and $\mathbf{d} = (d_0, \ldots, d_m, d_{-m}, \ldots, d_{-1})^T$, then $\mathbf{d} = F\mathbf{c}$, where F is the $(2m+1) \times (2m+1)$ *discrete Fourier transform matrix*

$$F = \begin{pmatrix} 1 & 1 & 1 & \cdots \\ 1 & \omega & \omega^2 & \cdots \\ 1 & \omega^2 & \omega^4 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}, \quad \omega = \exp\left(\frac{2\pi i}{2m+1}\right).$$

This matrix, mapping coefficients to function values, is $\sqrt{2m+1}$ times a unitary matrix. In other words its inverse, mapping function values to coefficients, is $F^{-1} = (2m+1)^{-1}F^*$, where * denotes the conjugate transpose.

Thus there are two equivalent ways to specify a $(2\pi/\lambda)$-band-limited L-periodic function: as a linear combination of $2m+1$ Fourier modes, or as the unique trigonometric interpolant through $2m+1$ data values. The idea of Lagrange interpolation gives an explicit representation of such interpolants. The function

$$D(x) = \frac{\sin((2m+1)\pi x/L)}{(2m+1)\sin(\pi x/L)} \tag{2.4}$$

is the trigonometric interpolant through the data values 1 at $x = 0$ and 0 at the other $2m$ points x_j, $j \neq 0$. (It is equal to the sum (2.1) in the case $c_j = (2m+1)^{-1}$ for all j.) Thus D is the *cardinal function* associated with interpolation through a single data point, also known as a *periodic sinc function* or the *Dirichlet kernel*. From here we see that any $(2\pi/\lambda)$-band-limited L-periodic function can be specified in the form

$$f(x) = \sum_{j=-m}^{m} d_j\, D(x - jh), \quad m = \lfloor L/\lambda \rfloor.$$

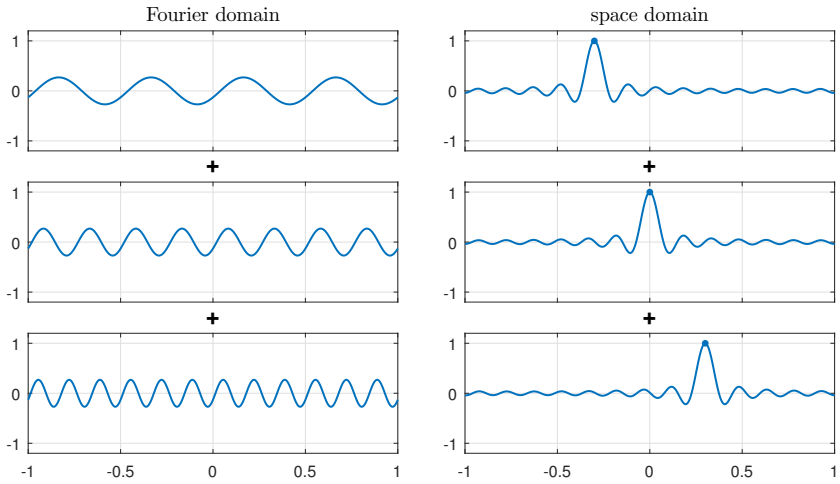

FIG. 2.1. *Two ways of viewing a smooth random function f. The Fourier domain view is that f is a linear combination of a finite collection of sine waves with different wave numbers and phases. The space domain view is that it is a trigonometric interpolant through random data values at equally spaced points, or equivalently, a linear combination of translates of the Dirichlet kernel (the periodic sinc function). Each column of this figure is intended to suggest how the random function is obtained by adding up such pieces.*

The observations above depend only on the fact that F is nonsingular. The equivalence we need for smooth random functions follows from the further fact that F is a multiple of a unitary matrix. In the definition (2.1), the variables c_j are independent samples from $N(0, 1/(2m+1)) + iN(0, 1/(2m+1))$. By a standard result of multivariate statistics, this is the same as saying that their joint probability distribution is

$$p(\mathbf{c}) = C \exp(-(2m+1)\|\mathbf{c}\|^2/2),$$

where $\|\cdot\|$ is the 2-norm and the constant $C > 0$ normalizes the total probability to 1. Since $\mathbf{d} = F\mathbf{c}$ and F is $\sqrt{2m+1}$ times a unitary matrix, this is equivalent to saying that the joint probability distribution of the values $\{d_j\}$ is

$$p(\mathbf{d}) = C' \exp(-\|\mathbf{d}\|^2/2),$$

where C' is again a normalization constant. Therefore the values d_j are independent samples from $N(0,1) + iN(0,1)$. This observation establishes that the following definition is equivalent to the earlier one.

DEFINITION 2.3. *A* REAL OR COMPLEX PERIODIC SMOOTH RANDOM FUNCTION *for given $\lambda, L > 0$ is a function*

$$f(x) = \sum_{j=-m}^{m} d_j D(x - jh), \quad m = \lfloor L/\lambda \rfloor \tag{2.5}$$

with h defined by (2.3)*, where each d_j is an independent sample from $N(0,1)$ or $N(0,1) + iN(0,1)$, respectively.*

It is worth emphasizing this equivalence. *Fourier series with random coefficients are the same as trigonometric interpolants through random data values.* Though we have made use of the particular choice (2.3) of grid points x_j, it follows from translation-invariance that this choice does not matter. Translation to any other equispaced grid will produce the same distribution of smooth random functions.

Periodic smooth random functions define a Gaussian process with mean zero, as we shall discuss in Section 6. As a taste of this interpretation, we note here that translation-invariance reveals how values of a smooth random function depend on one another *between* grid points. The covariance of the stochastic process is the function $C(x, y)$ defined as the expected value of the product $f(x)f(y)$, or $\overline{f(x)}f(y)$ in the complex case. Because of stationarity, this is equal to the expected value of $\overline{f(x-y)}f(0)$. By (2.5), $f(0)$ reduces to the random number d_0, and since the coefficients d_j are uncorrelated with d_0 for $j \neq 0$, it follows from (2.5) that the expected value of $\overline{f(x-y)}f(0)$ reduces to the expected value of $\overline{d_0}d_0 D(x-y)$, that is, $D(x-y)$ in the real case and $2D(x-y)$ in the complex case.

Figure 2.2 shows an example of one kind of use of smooth random functions. In many applications one would like to take random initial data to explore the typical behavior of a system. The figure shows a computation in which a periodic smooth random function has been taken as the initial condition for the Cahn–Hilliard equation,

$$u_t = -10^{-2} u_{xx} - 10^{-5} u_{xxxx} + 10^{-2} (u^3)_{xx}, \tag{2.6}$$

which models phase separation in binary alloys and fluids [8]. We take periodic boundary conditions on the interval $x \in [-1, 1]$ for $t \in [0, 3000]$, and the simulation is carried out with the "spin" stiff PDE integrator in Chebfun [31] in about 15 seconds of laptop time using essentially this code:

```
S = spinop([-1,1],[0 3000]);
rng(6), S.init = -.5*randnfun(.2,'trig');
S.lin = @(u) -1e-2*diff(u,2) - 1e-5*diff(u,4);
S.nonlin = @(u) 1e-2*diff(u.^3,2);
spin(S,96,.04,'iterplot',250)
```

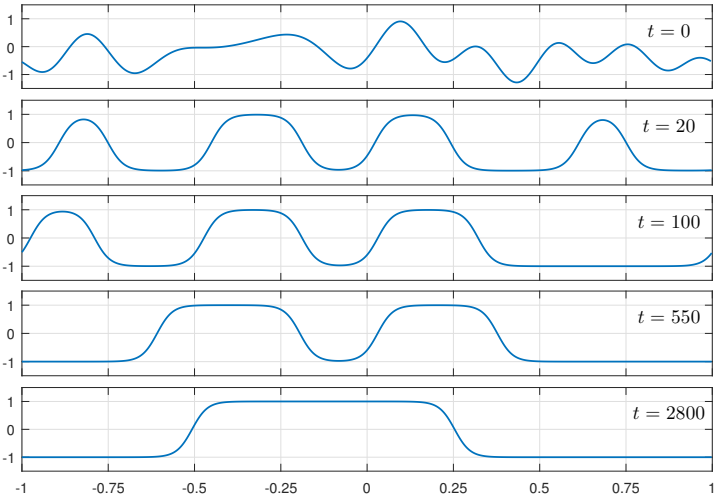

FIG. 2.2. *Solution of the Cahn–Hilliard equation* (2.6) *with a smooth periodic random function with* $\lambda = 0.2$ *as the initial condition* $u(x,0)$*. As* t *increases, the solution coalesces into fewer and fewer regions with values* $\approx \pm 1$*, always conserving the overall integral, until eventually a steady state is reached with just one region of each sign in this periodic domain. Smooth random functions are employed in applications for exploring the typical behavior of a dynamical system. As we shall see in Section* 7*, sometimes that system may be the universe itself.*

3. Nonperiodic smooth random functions. The random functions just discussed are periodic, but applications usually do not call for periodicity. As a practical matter, we construct nonperiodic smooth random functions by forming periodic ones on a longer interval $[-L'/2, L'/2]$ or $[0, L']$ with $L' > L$ and then truncating. In principle one should take $L' \to \infty$, so that no trace of periodicity remains in the original interval. A mathematically precise treatment of this limit would involve random Fourier transforms as opposed to series, but we shall not pursue this. Accordingly, in the remainder of this paper, when we speak of smooth random functions without specifying periodicity, we refer to a construction based on a finite value $L' \geq L$. Informally, one should imagine $L' \gg L$, or more precisely $L' - L \gg \lambda$, but formally, our statements apply irrespective of the choice of any finite value $L' \geq L$. The Chebfun `randnfun` function takes $L' \approx 1.2L$. Any fixed ratio $L'/L > 1$ is enough to ensure that effects of periodicity go away as $\lambda \to 0$.

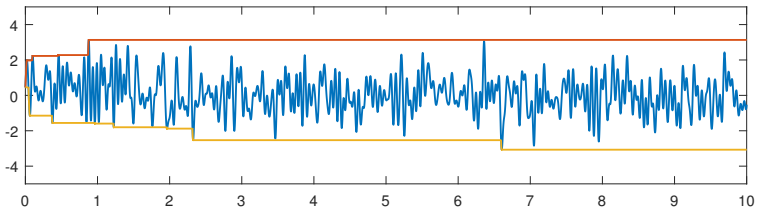

FIG. 3.1. *A smooth random function with* $\lambda = 0.05$ *on* $[0, 10]$ *together with its cumulative maximum and minimum functions. With probability* 1*, these widen at a rate proportional to* $(\log L)^{1/2}$ *on* $[0, L]$ *as* $L \to \infty$*.*

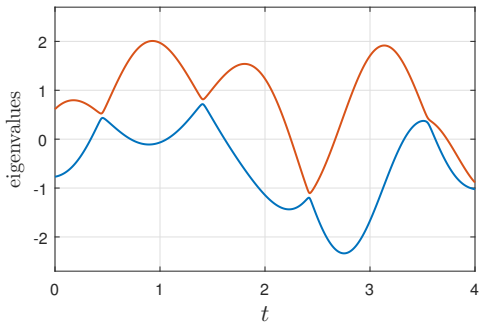

FIG. 3.2. *Eigenvalues of a* 2×2 *real symmetric matrix function* (3.1) *with smooth random entries, related to the process known as Dyson Brownian motion. When the values of* $f(t)$ *and* $g(t)$ *cross, the eigenvalues come close together but do not cross.*

Smooth random functions have many properties that mimic those of the random vectors `randn(n,1)` mentioned in the introduction. For example, Figure 3.1 shows a smooth random function with $\lambda = 0.05$ on the domain $[0, 10]$ produced by the Chebfun command `f = randnfun(0.05,[0,10])`. What are its maximum and minimum? Approximately speaking, both numbers will be of order 1, and more precisely, for any fixed λ, according to the theory of extreme value statistics, one can expect them to grow at a rate proportional to $(\log L)^{1/2}$ on $[0, L]$ as $L \to \infty$ because of the square-exponential tail of the normal distribution. (A key mathematical result in this area is the *Borell–TIS inequality* [1].) The figure gives some hint of this behavior by including cumulative minimum and maximum curves drawn by the Chebfun command `plot([f; cummax(f); cummin(f)])`. One could investigate precise formulations of such observations, and that would be an interesting subject for research. As is customary in probability theory, properties of smooth random functions will hold not with certainty, but with probability 1. For example, a smooth random function for any fixed L and $\lambda \le L$ is nonconstant—with probability 1.

Another example of the kind of exploration that is readily carried out with non-periodic smooth random functions is illustrated in Figure 3.2. If f and g are real functions of t, then for any $\varepsilon > 0$ and any t, the symmetric matrix

$$A(t) = \begin{pmatrix} f(t) & \varepsilon \\ \varepsilon & g(t) \end{pmatrix} \tag{3.1}$$

has two distinct real eigenvalues (separated by at least 2ε). If ε is small, however, the two eigenvalues will have a near-crossing at points where f and g cross. The figure illustrates this effect for a case where f and g are smooth random functions on $[0, 4]$ with $\lambda = 1$ and $\varepsilon = 0.05$. A generalization of this example, which goes by the name of *Dyson Brownian motion* [12], is the effect that real symmetric matrices with Brownian path entries also show eigenvalue level avoidance (with probability 1).

4. Big smooth random functions, white noise, and Brownian paths. The tempting question is always, what happens as $\lambda \to 0$? This is the white noise limit, but one must be careful.

Smooth random functions as we have defined them oscillate faster and faster as $\lambda \to 0$, always with amplitude $O(1)$. They do not converge pointwise. One might imagine that a reasonable notion of a random function corresponding to $\lambda = 0$ would be a function taking independent values from $N(0, 1)$ at each point x. Such functions

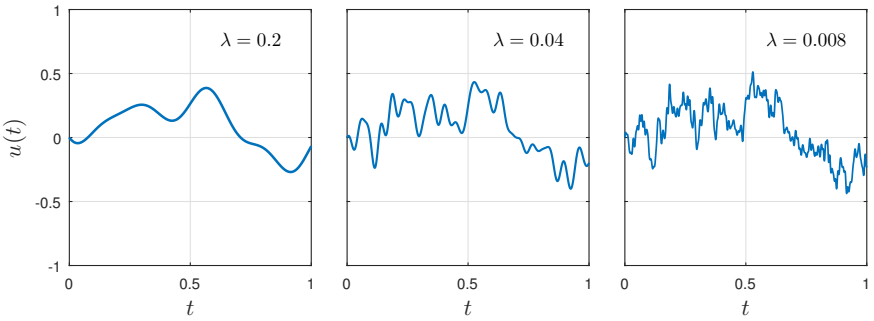

FIG. 4.1. *Indefinite integrals of big smooth random functions give smooth random walks, which converge to Brownian paths as* $\lambda \to 0$.

would not be Lebesgue measurable, however, and hence would not even be integrable; it is not clear what use they would be.

Another idea for $\lambda \to 0$ comes from the observation that integrals of smooth random functions converge to zero in this limit because of sign cancellation.[1] So one could also speak of a limit function for $\lambda = 0$ in the form of a *distribution.* It would be the zero distribution, however, which is not very interesting.

The mathematical and scientific substance for the limit $\lambda \to 0$ appears when the functions are rescaled by $O(\lambda^{-1/2})$. The precise definition we make is that a smooth random function in the "big" normalization is the same as before, but with (2.1), (2.2), and (2.5) multiplied by $\sqrt{2/\lambda}$. Here are the definitions followed by the appropriate restatement of Theorem 2.2.

DEFINITION 4.1. *A* REAL OR COMPLEX BIG PERIODIC SMOOTH RANDOM FUNCTION *is defined as in Definitions* 2.1 *and* 2.3*, except with the variances* $1/(2m+1)$ *of Definition* 2.1 *increased to* $2/((2m+1)\lambda)$ *and the variances* 1 *of Definition* 2.3 *increased to* $2/\lambda$.

THEOREM 4.2. *A big periodic smooth random function* f *(whether real or complex) is* L*-periodic, entire, and* $(2\pi/\lambda)$*-band-limited. The stochastic process from which* f *is a sample is stationary, with values* $f(x)$ *at each* x *distributed according to* $N(0, 2/\lambda) + iN(0, 2/\lambda)$ *in the complex case and* $N(0, 2/\lambda)$ *in the real case.*

Note that since $m \approx L/\lambda$, we have $2/((2m+1)\lambda) \approx 1/L$. Thus in the big normalization, the random coefficients of the series (2.1) and (2.2) have variances essentially independent of λ as $\lambda \to 0$.

The point of the rescaling emerges when we look at integrals. Figure 4.1 plots indefinite integrals of three big smooth random functions on $[0, 1]$ with parameters $\lambda = 1/5$, $1/25$, and $1/125$. The seed used to initialize the random number generator is set in the same way for each case, so these are successively finer approximations of the same random curve.[2] One sees smaller-scale features appearing as λ decreases and more terms are included in the series (2.2). We call such paths *smooth random walks,* and as we shall state in Theorem 4.3 below, they converge to Brownian paths as $\lambda \to 0$. One of our favorite references on mathematical Brownian motion is [32].

[1] This statement holds with probability 1. In the remainder of the paper, we shall not always mention this qualification.

[2] To ensure that setting the random number seed has the desired effect, Chebfun picks random coefficients in the order $c_0, c_1, c_{-1}, c_2, c_{-2}, \ldots$, i.e., starting at the low wave numbers and then increasing.

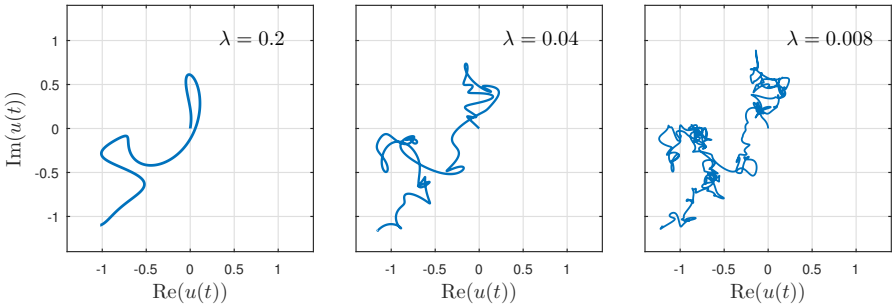

FIG. 4.2. *A complex analogue of Figure* 4.1 *shows indefinite integrals of big smooth complex random functions, that is, smooth random walks in* 2*D.*

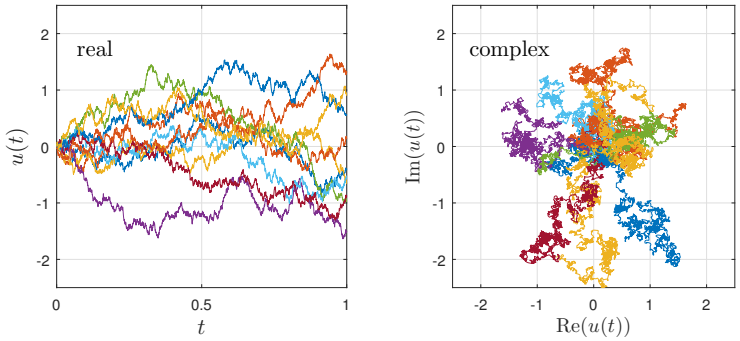

FIG. 4.3. *Real and complex smooth random walks, ten samples each with* $\lambda = 0.001$. *This value is small enough that the curves are Brownian paths roughly to plotting accuracy.*

Figure 4.2 presents an analogous trio of images for indefinite integrals of complex smooth random functions. These we call *complex smooth random walks,* again converging to a familiar form of complex (or simply 2D) Brownian paths as $\lambda \to 0$.

Each sample path of a random process looks different (and it is striking how the human eye is wired to see personalities in them!). Figure 4.3 shows ten examples each of smooth real and complex random walks, all with $\lambda = 0.001$. Taking smaller values of λ would have little visible effect. In Chebfun, one can generate a figure like this with the command `plot(cumsum(randnfun(.001,[0 1],10)))`.

The mathematics of Brownian paths began to be worked out by Einstein, Smoluchowski and others in the first decade of the 20th century. The core of this subject is the idea that Brownian paths are the integral of white noise, i.e., of a signal with equal energy at all wave numbers. The paradox is that the notion of white noise does not make sense, because for noise to be truly white, it would have to have infinite amplitude and infinite energy. In a physical application, noise must be colored. There must be a cutoff, a minimal space scale, and in the case of physical Brownian motion of the kind observed in microscopes, this is provided by the finite size of the molecules that randomly impact a small particle suspended in a liquid.

In the 1920s, nevertheless, Wiener found a way to make the notion of Brownian paths rigorous without a small space scale cutoff: in our terms, to set $\lambda = 0$. (Other mathematicians important in the early history include Kolmogorov and Levy.) The essential idea is to take as the primary object not noise but its integral—the Brownian

path. Wiener showed that one can define such paths and the associated probability measure in a mathematically rigorous way, giving what is now known as the *Wiener process*. A Brownian path $W(t)$ is continuous (with probability 1), but it is not smooth. Its derivative exists nowhere (again with probability 1), which is to be expected since the derivative would have to be white noise.

A remarkable property of Brownian paths is that, although the details of a path from A to B are infinitely complicated, it is possible to get from A to B without tracking those details. A random walk with finite steps taken from a normal distribution can be regarded as a sample of a Brownian path at discrete times, and if one needs values at points in-between, one can calculate them later (the *Brownian bridge*). Using a term from computer science, we may say that the infinite complexity of Brownian paths is not an obstacle because they can be computed by "lazy evaluation."

The most fundamental virtue of the pointwise approach to stochastic analysis initiated by Wiener is that, scientifically speaking, it is just the right idealization. By way of analogy, the subject of continuum mechanics builds on the fact that although the air in a pump, say, is composed of a vast number of discrete molecules, it can be modeled as a continuum. We know that air is not really a continuum, yet most of us would feel that interpreting it that way is not merely convenient, but in some sense intellectually the right thing to do for many purposes. The pointwise, $\lambda = 0$ approach to stochastic analysis has the same quality. Truly white noise may be a physical impossibility, but we can make sense of its integral, and this seems intellectually right. And just as with continuum mechanics, this model of stochastics has the particular benefit that it connects the subject with partial differential equations. For example, the theory of harmonic measure associates exit probabilities of Brownian paths with solutions of the Laplace equation. The basis of such connections is the fact that the density of an ensemble of Brownian motions obeys a diffusion PDE, with spreading at a rate characterized by $\sqrt{t}$. This observation goes all the way back to Einstein.

The difficulty with the pointwise approach to stochasticity, however, is that it is highly technical. In continuum mechanics we can write down the gas laws or the Navier–Stokes equations without discussing the underlying molecules, but it is not possible to state the principles of stochastic analysis so easily. Stochastic analysis requires special foundations, and they are technically advanced. These in turn require special notations and special numerical methods, which are different from the familiar methods of nonstochastic numerical computation [24]. This becomes particularly an issue in the context of differential equations, the subject of the next section.

Ultimately, anyone working in this field will need to deal with the technicalities. Smooth random functions, however, provide an elementary way to get started. Since $\lambda > 0$ always, they build just on ordinary calculus and ordinary numerical methods (quadrature in this section, solution of ODEs in the next).

The possibility of defining Brownian paths via Fourier series with random coefficients goes back a long way. As discussed by Kahane [21, 22], Wiener considered such series in [48], and the discussion was generalized by him and Paley and Zygmund in several papers including [36]. As these authors noted, the integral of the series (2.1) or (2.2) contains coefficients mollified by the factor $1/j$. Thus, for example, integration of (2.1) gives

$$\int_0^x f(s)ds = c_0 x + \frac{L}{2\pi i} \sum_{\substack{j=-m \\ j\neq 0}}^{m} \frac{c_j}{j} \left[\exp\left(\frac{2\pi i j x}{L}\right) - 1 \right].$$

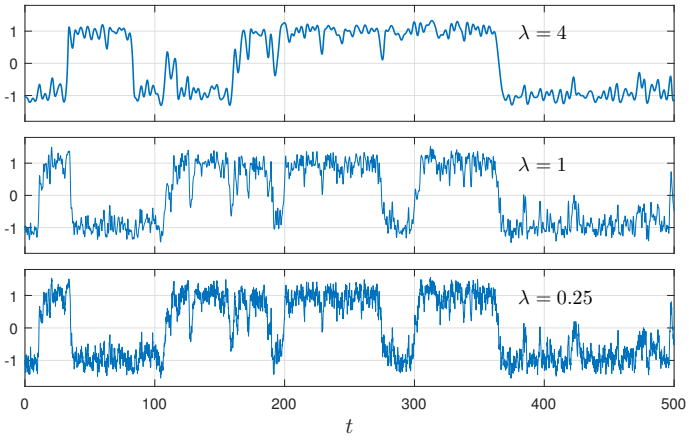

FIG. 5.1. *Solutions to the bistable ODE with noise* (5.1) *for three values of* λ*, all starting from the same random number seed. The solution alternates randomly between one state and the other.*

With $m = \infty$ this becomes an infinite series whose convergence is not guaranteed, since $O(j^{-1})$ coefficients do not decrease quite fast enough, but if the coefficients are random, that ensures convergence with probability 1.

The central property of big smooth random functions is that their integrals converge to standard Brownian paths as $\lambda \to 0$. (The term *standard* refers to a normalization. The variance of the distribution of a Brownian path $W(t)$ is equal to Ct for some constant C, and the standard choices are $C = 1$ for the real case and $C = 2$ for the complex case.) We designate this as a theorem for clarity, but the statement below is not really precise, and indeed this paper does not present any of the definitions and details needed for rigorous stochastic analysis. For a full treatment, we recommend Chapter 16 of Kahane's book [21], particularly Theorem 2 on p. 236. Ultimately this result is due to Wiener.

THEOREM 4.3. *As* $\lambda \to 0$*, indefinite integrals of big smooth random functions (whether real or complex, periodic or nonperiodic) converge with probability* 1 *to standard Brownian paths.*

For a fascinating presentation of the mathematical properties of Brownian motion, see the book [32] mentioned earlier, and a more advanced treatment can be found in [39]. The physical side of the subject is presented in [15], also with a discussion of applications in finance.

5. Smooth random ODEs. Having defined smooth random functions, we can use them as forcing functions, or as coefficients, in ordinary differential equations (ODEs). Sometimes this is interesting for fixed λ, typically with the standard normalization, when one is interested in systems with a macroscopic random character, as illustrated in a PDE context by the Cahn–Hilliard example of Figure 2.2. Other times the motivation is noise, and then the right choice will be small values of λ with the big normalization. For example, the smooth random walks of the last section are solutions of the trivial ODE $u' = f$, where f is a big smooth random function. It is equally easy on a computer to incorporate smooth random functions in less trivial differential equations.

First, Figure 5.1 shows three solutions to a bistable equation with noise,

$$u' = u - u^3 + 0.7f, \quad t \in [0, 500], \ u(0) = -1, \tag{5.1}$$

where f is a big smooth random function with wavelength parameter λ. Without the noise term, this ODE has stable steady states at $u = -1$ and $+1$. With the noise, solutions tend to linger near one steady state before eventually making a transition to the other (essentially a Poisson process), switching back and forth infinitely often (with probability 1) as $t \to \infty$. In Chebfun, a suitable code is

```
f = randnfun(lambda,[0 500],'big');
N = chebop(0,500);
N.op = @(u) diff(u) - u + u^3 + .5*f; N.lbc = -1;
u = N\0; plot(u)
```

The figure plots solutions for three values of λ, revealing modest changes as λ decreases. In a scientific application one might be interested, for example, in the dependence of the mean switching time on the noise amplitude.

The next example, in Figure 5.2, is a nonlinear pendulum equation with noise,

$$\theta'' = -\sin(\theta) + 0.05f, \quad t \in [0, 200], \ \theta(0) = 3, \ \theta'(0) = 0, \tag{5.2}$$

where f is a big smooth random function with $\lambda = 0.2$,

```
N = chebop(0,200);
N.op = @(theta) diff(theta,2) + sin(theta); N.lbc = [3;0];
f = .05*randnfun(0.2,[0 200],'big');
theta = N\f; plot(theta)
```

Without the noise, the trajectory would oscillate forever around 0, but the noise has the effect of increasing the energy so that θ increases steadily up to around $t = 60$; the pendulum is swinging over and over. Then the energy happens to diminish a bit, giving a couple of bound oscillations, before at around $t = 90$ it increases again and the pendulum starts swinging over in the other direction. As in all such experiments, a new choice of f would change the details completely, but it would not change the qualitative behavior.

Our third example concerns a Hopf bifurcation in a two-variable ODE system that can be found in a number of references; we adapted this from Example 79 of Appendix B of [47]. The equations are

$$u' = -v + u(t/T - u^2 - v^2) + \varepsilon f, \quad v' = u + v(t/T - u^2 - v^2), \tag{5.3}$$

where f is a big smooth random function with $\lambda = 1$. With $\varepsilon = 0.01$, a bifurcation occurs near $t = 0$, as one would expect from a standard analysis, but the larger value $\varepsilon = 0.1$ advances the bifurcation point noticeably.

```
T = 100; dom = [-T,T];
N = chebop(@(t,u,v) [diff(u)+v-u*(t/T-u^2-v^2); ...
                     diff(v)-u-v*(t/T-u^2-v^2)],dom);
N.lbc = [0;0]; rng(0), f = 0.01*randnfun(1,dom,'big');
[u,v] = N\[f;0]; t = chebfun('t',dom); plot3(t,u,v,LW,lw)
```

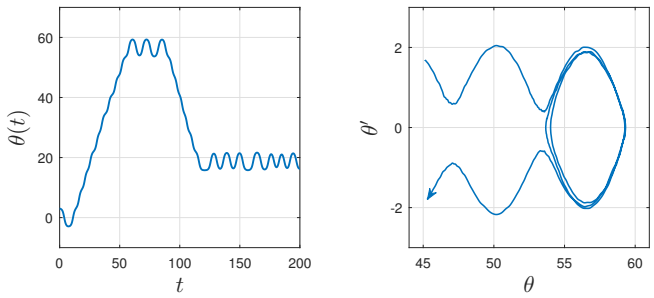

FIG. 5.2. *The nonlinear pendulum with noise* (5.2)*, showing a function of time* t *on the left and the corresponding phase plane trajectory for* $47 \le t \le 97$ *on the right. The noise induces transitions between bound states, in which* $\theta(t)$ *oscillates around a multiple of* 2π*, and unbound states, where the pendulum swings over and* $\theta(t)$ *increases or decreases steadily.*

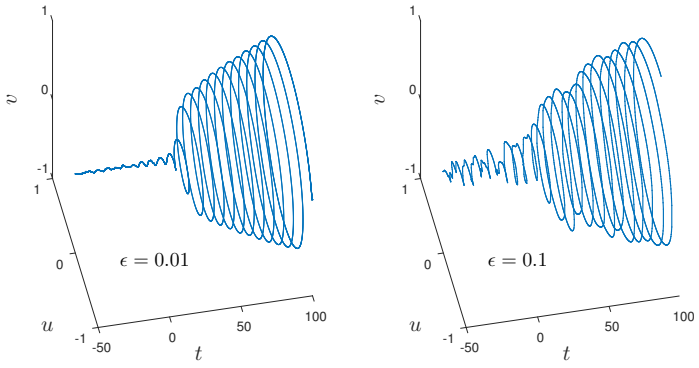

FIG. 5.3. *Hopf bifurcation with noise in equation* (5.3)*. With small noise (left), the bifurcation point is near* $t = 0$*, but larger noise (right) brings it forward.*

All these examples involve what is known as *additive noise,* where a random term is added to an ODE as a forcing function. Smooth random functions can also be used to approximate *multiplicative noise.* In the simplest case the ODE is $u' = fu$, where f is a big smooth random function, and this leads to a smooth approximation of *geometric Brownian motion.* For a more substantive example we follow Horsthemke and Lefever [20, p. 123] and consider the equation

$$u' = (1 + \sigma f)u - u^2, \quad t \in [0, 20], \ u(0) = 1, \tag{5.4}$$

where f is a big random function. With $\sigma = 0$, this system has a stable fixed point at $u = 1$ and an unstable one at $u = 0$. With $\sigma \neq 0$, as analyzed in [20], the trajectories stay mainly near $u = 1$ when σ is small but are often near $u = 0$ when σ is large. Figure 5.4 illustrates this difference.

None of what we have done in this section has made use of the theorems, notations, or algorithms of stochastic calculus and SDEs. Everything has involved ODEs of the usual kind computed by the usual numerical methods, so no technical issues have arisen. But of course, it is necessary to know how computations like these relate to stochastic analysis. There are two standard formulations, originating with Itô (in the

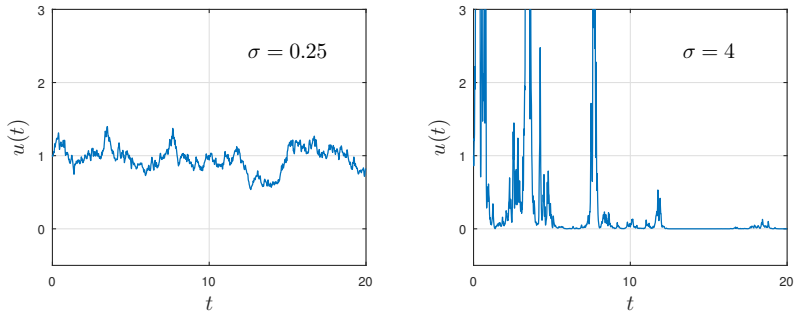

FIG. 5.4. *Solutions to the multiplicative noise equation* (5.4) *for small and large values of* σ*, both with* $\lambda = 0.05$*. As analyzed in* [20]*, the two behaviors are qualitatively different.*

1940s) and Stratonovich (in the 1960s). The mathematical relationships between the Itô and Stratonovich formulations are fully understood, and an equation written in either form can be converted into an equivalent equation in the other; when there is only additive noise, they are identical. Details can be found in many texts on stochastic analysis, such as [15, 24, 35]. In a word, smooth random ODEs as we have set them up correspond to the Stratonovich formulation. Equation (5.1), for example, is a smooth approximation to this SDE written in the standard notation:

$$dX_t = (X_t - X_t^3)dt + 0.5dW_t.$$

In this case, as there is only additive noise, the Itô and Stratonovich settings coincide. An example with multiplicative noise is (5.4), and in this case the convention is to include a "$\circ$" symbol to indicate that the formulation is Stratonovich. Equation (5.4) is a smooth approximation to this SDE:

$$dX_t = (X_t - X_t^2)dt + \sigma X_t \circ dW_t.$$

The central property of solutions of ODEs containing a big smooth random function is that they converge to solutions of Stratonovich SDEs as $\lambda \to 0$. As with Theorem 4.3, we state this as a theorem, but the statement is not precise, for that would require details of stochastic analysis. The result is essentially due to Wong and Zakai in a pair of papers in the mid-1960s [50, 51], of which an account is given in the book by Wong and Hajek [49]. "Wong–Zakai theory" is more general than this result, however, not requiring smoothness in the random functions, a property that would be regarded as needlessly restrictive by many mathematicians. Our recommended reference on this material is the paper [43] by Sussmann and its short summary [42], which connect SDEs and ODEs at the level of individual solutions paths (see also [25]). A popular reference for SDEs is [35], in which these matters are briefly discussed. For recent developments, see [18] and [54]. A fundamental generalization of some of these ideas is the theory of *rough paths* introduced by Lyons [14, 28].

THEOREM 5.1. *As* $\lambda \to 0$*, solutions to random ODEs containing a big smooth random function converge with probability* 1 *to solutions of Stratonovich SDEs.*

Note that the theorem allows for just one big smooth random function, not several. When more than one random variable is involved, the relationship with the theory of SDEs is not as simple. New issues also arise when there is more than one independent

variable, i.e., with stochastic PDEs. For information on these and many other matters in the approximation of stochastic systems by differential equations, see [9, 10, 23, 37, 43, 54].

6. Smooth random functions and Gaussian processes. A smooth random function is a sample path from a particular Gaussian process. Informally, a Gaussian process is a stochastic process depending on a continuous variable t in which for each fixed t, the value is a Gaussian random variable, and moreover, for any finite set $t_1, \dots, t_k$, the joint distribution is multivariate Gaussian. Gaussian processes are an old idea, but in recent decades interest in them has increased greatly with the advance of Bayesian reasoning in general and machine learning in particular [19, 29, 34, 38, 41].

A Gaussian process is determined by a mean function $\mu(t)$ (here, just the zero function) and a covariance function $C(t,t')$, defined (when the mean is zero) as the expected value of $\overline{f(t')}f(t)$. If $C(t,t')$ depends just on $t'-t$, the process is stationary (another term is *homogeneous*). For smooth random functions, we showed in Section 2 that $C(t,t')$ is the Dirichlet kernel (2.4): $C(t,t') = D(t'-t)$, or $2D(t'-t)$ in the complex case. Other choices of covariance function also make sense, and indeed, they have advantages. A particularly attractive choice is a Gaussian kernel, which is positive rather than oscillatory and decays square-exponentially rather than just inverse-linearly. We could have defined smooth random functions in terms of a Gaussian kernel, or with various other choices, both in Chebfun and in this article, and it would not have made much difference. The disadvantage is that it would have entailed the use of infinite Fourier series rather than finite ones. Ultimately we have been swayed by simplicity: the idea of a finite Fourier series with random coefficients all from the same distribution could hardly be more elementary.

Instead of the covariance function, engineers and physicists often speak equivalently of its Fourier transform, the *power spectrum* of a Gaussian process, that is, the function describing the rate of decay of the variances of the random coefficients as the wave number k increases. For our smooth random functions, the power spectrum is a step function taking a constant value for $|k| \le 2\pi/\lambda$ and the value 0 for $|k| > 2\pi/\lambda$.

Gaussian processes have a link with data-fitting that is one of the starting points of machine learning. (Related ideas were introduced in the geostatistics literature under the name of *kriging* in the 1950s and 1960s and are also connected with *radial basis functions* and *smoothing splines* [16].) Traditionally, we may think of data-fitting as beginning with the method of least-squares approximation going back to Gauss and Legendre. What is distinctive with the Gaussian process point of view is the central role of *probability*. To explain this we now outline how one might interpolate a set of data using smooth random functions.

Suppose that for some $\mu \ge 0$ we have $2\mu+1$ data values at points equispaced in the interval $[-L/2, L/2]$ and we know that these data come from an L-periodic function. Then there is a unique trigonometric interpolant of degree μ through these data, that is, of the form (2.1) or (2.2). The same holds if the points are not equispaced, so long as they are distinct.

A more flexible alternative, however, is to consider interpolants of higher degrees $m > \mu$. For example, suppose we take $m = 2\mu$. There are infinitely many interpolants of degree m; which one shall we pick? Here is where probability enters the discussion. In the Gaussian process framework, each interpolant is associated with a probability (more precisely a probability density), and a natural choice is to take the mean of all the interpolants with respect to these probabilities. Roughly speaking, this means we choose as our interpolant the trigonometric polynomial of degree m that is *most*

likely—or as one may also think of it, *least surprising*.

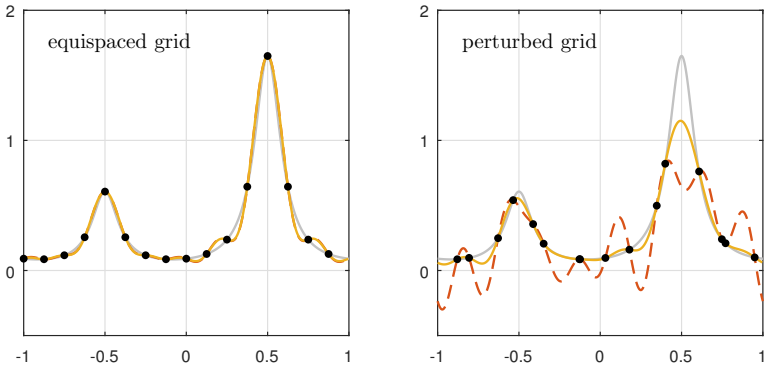

FIG. 6.1. *Gaussian process interpolation of* $2\mu+1$ *data values with* $\mu = 8$ *by a trigonometric polynomial of degree* $m = 16$*: equispaced sample points on the left, perturbed sample points on the right. (The data are samples of* $f(x) = \exp(\sin(\pi x))/(1+10\cos^2(\pi x))$*, shown in gray.) The dashed curves correspond to the Dirichlet kernel that is the basis of smooth random functions as defined in this article, and the continuous curves to a periodic Gaussian kernel. The latter is better for this kind of application because it is biased toward low wave numbers rather than treating all wave numbers below a certain cutoff equally.*

What does it mean to be least surprising? For a Gaussian process with mean zero, it means that the Fourier coefficients are as small as possible, and the choice of covariance function or power spectrum comes into play in determining the balance between smallness of coefficients at difference wave numbers. For our definition of smooth random functions, all the Fourier coefficients have the same variance, so small coefficients are favored no more strongly in the high wave numbers than in the low wave numbers. This means that interpolants based on smooth random functions as we have defined them tend to have a lot of energy in the high wave numbers, making them prone to oscillations. A Gaussian kernel, on the other hand, puts more pressure on the high wave number coefficients than on the low wave number coefficients to be small. This introduces a bias toward low wave numbers that makes interpolants smoother. Figure 6.1 illustrates this effect, which generalizes also to data fitting in the presence of noise.

Thus when it comes to data fitting, more standard covariance functions such as the Gaussian will be preferable to the Dirichlet kernel of our smooth random functions. In Chebfun, there is a `gpr` command for Gaussian process regression (with or without noise), and it uses a Gaussian kernel by default. In the periodic case, the kernel is made periodic in a fashion proposed in [29].

7. Smooth random functions in multiple dimensions. Smooth random functions are readily generalized to multiple dimensions; we focus on the 2D case for concreteness. The new issue that arises here is that as well as being stationary, one would like the distribution to be *isotropic*. We achieve this by taking a finite bivariate Fourier series with random coefficients in a ball, not a square, of wave numbers:

$$f(x,y) = \sum_{k=-m}^{m} \sum_{j=-m_k}^{m_k} c_{jk} \exp\left(\frac{2\pi i(jx+ky)}{L}\right), \quad m_k = \sqrt{m^2-k^2}. \tag{7.1}$$

(For random functions on a rectangle, the ball becomes an ellipse.) This provides approximate isotropy for finite m, improving as $m \to \infty$. The analogue of a Gaussian process in multiple dimensions is called a *Gaussian random field,* and stationarity together with isotropy amount to the condition that the covariance function $C(\mathbf{x}, \mathbf{x}')$ depends only on $\|\mathbf{x} - \mathbf{x}'\|$.

Random functions have been employed for a wide variety of scientific applications, and there has been great interest in elucidating their properties. Early work on the 1D case was due to Steve Rice [40] during World War II, motivated by applications such as shot noise in signal processing, and 2D random functions were investigated a decade later by Longuet-Higgins in an analysis of ocean waves [27]. In cosmology, 3D random functions have been investigated to shed light on the distribution of galaxies in the universe and the structure of the cosmic microwave background; a celebrated paper in this area is that of Bardeen, et al. [3]. "Random energy landscapes" are a basic notion in fields including condensed matter physics [7], and string theorists are considering random functions in a higher-dimensional parameter space as a model to explain how a universe such as our own may have arisen [7, 13]. Bowen, et al. have investigated random functions in 2D as models of the fractal geometry of Arctic melt ponds [6], and for random functions in biology, see [44]. Mathematicians have also investigated properties of random functions extensively, and a leader in this area has been R. Adler [1, 2]. In this context random functions can be considered not just in Euclidean space but on manifolds.

A particularly down-to-earth example of a manifold is the unit sphere, and to construct smooth random functions on this domain, one can use a spherical harmonic series with random coefficients. For isotropy, it is appropriate to use coefficients in a triangular array:

$$f(\varphi, \theta) = \sum_{\ell=0}^{m} \sum_{j=-\ell}^{\ell} c_{\ell,j} Y_{\ell,j}(\varphi, \theta). \tag{7.2}$$

Here φ is longitude, θ is co-latitude (i.e., $\pi/2$ minus the latitude), and $Y_{\ell,j}$ denotes the spherical harmonic of degree ℓ and order j,

$$Y_{\ell,j}(\varphi, \theta) = P_\ell^j(\cos(\theta)) e^{ij\varphi},$$

where P_ℓ^j is the (normalized) associated Legendre function. Since the circumference of the unit sphere is $L = 2\pi$, we take $m = \lfloor 2\pi/\lambda \rfloor$ in analogy to (2.1) [26]. In Chebfun, smooth random functions on the sphere have been implemented by Grady Wright using Spherefun [45].

We shall not give further details of multidimensional smooth random functions but illustrate the subject with Figures 7.1 and 7.2. Incidentally, for dimensions greater than 1, there is an alternative notion of smooth random function of interest to physicists: a Fourier series in which all wave number vectors are $= \lambda$ rather than just $\le \lambda$ in magnitude for some $\lambda > 0$; the orientations of the waves, however, are not fixed. We call these *monochromatic smooth random functions,* and they arise in the study of quantum chaos as models of random high energy eigenfunctions of the Laplace operator [4, 5, 33]. In Chebfun, one can write e.g. `randnfun2(lambda,'mono')`.

8. Discussion. Our "standard" smooth random functions are Gaussian processes (or Gaussian random fields in multiple dimensions), but for simplicity, they are very simple ones: defined by a finite Fourier series with random coefficients, all

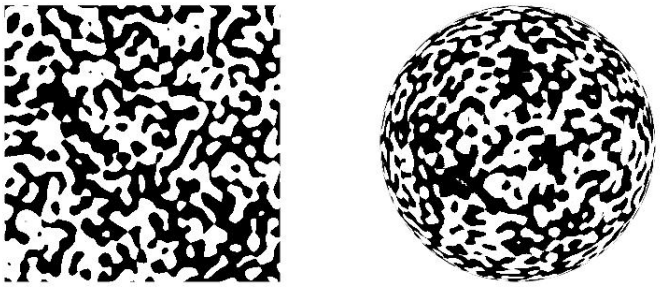

FIG. 7.1. *Random functions with* $\lambda = 0.1$ *on the square* $[-1,1]^2$ *and the unit sphere computed with* `randnfun2` *and* `randnfunsphere` *in Chebfun. The "zebra" plotting mode shows positive and negative values as white and black, respectively.*

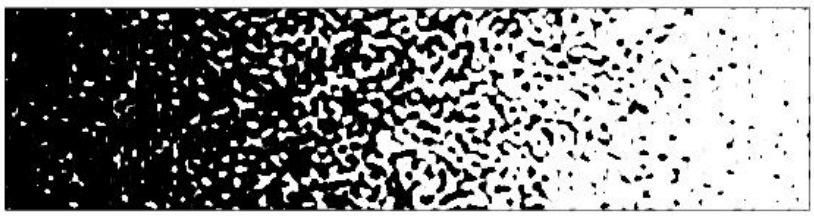

FIG. 7.2. *This image, which reminds the authors of the engravings of M. C. Escher, is a zebra plot of the function* $x + f$ *on the rectangle* $-2 \le x \le 2$, $0 \le y \le 1$, *where* f *is a 2D random function with* $\lambda = 0.05$. *The properties of surfaces like these have been investigated for applications in fields including cosmology, condensed matter physics, oceanography, climate modelling, image processing, and pure mathematics.*

from the same normal distribution. This is certainly not the only reasonable choice, but we offer it for simplicity and concreteness, to emphasize that useful explorations can be carried out without requiring the user first to confront the possibly daunting question, "what is your covariance function?" Many applications will involve some smoothing—the Cahn–Hilliard equation of Figure 2.2, for example—and in such cases the choice of covariance function may not matter much anyway.

Our "big" smooth random functions are the same, except with a different normalization appropriate to taking integrals to model stochastic effects. Again, our aim has been simplicity. The usual foundation of stochastic analysis among mathematicians is a pointwise conception of noise, as discussed in Section 4, even though, if white noise truly took values at points, they would have to be discontinuous and of infinite amplitude. The use of smooth random functions follows the alternative approach of conceiving noise as smooth and finite, with a small wavelength parameter $\lambda > 0$. This offers a way to explore stochastic effects without employing the special definitions of stochastic calculus (Itô, Stratonovich) or the associated special SDE algorithms (Euler–Maruyama, Milstein, ...). As recorded in Theorem 5.1, this corresponds to the Stratonovich calculus as $\lambda \to 0$, at least in cases with a single random variable.

As we stated at the outset, none of the ingredients we have put together are new. Fourier series with random coefficients have been investigated for a century, the idea

of band-limited white noise is a familiar one, and many variations on these themes have been exploited for a wide range of scientific and engineering problems.

We emphasize that we do not advocate smooth random functions for reasons of computational efficiency. On the contrary, existing algorithms of stochastic computation will be faster in many cases, and in particular, working with smooth random functions with $L/\lambda \gg 1000$ is unwieldy in Chebfun. The point of smooth random functions is conceptual and computational simplicity, not speed. Indeed, even in the realm of approximating Brownian paths and SDEs via random ODEs with smooth coefficients, more accurate approximations are possible with the use of alternative random Fourier series derived from the Karhunen–Loève expansion [53].

Although this presentation has avoided the details of stochastic analysis, there is no doubt that these are indispensable for a full understanding of stochastic effects—starting with the fundamental distinction between the Itô and Stratonovich integrals. The perspective of this paper is that, given the associated technical challenges, perhaps there is a place for simpler tools too. Imagine if we told students they had to learn measure theory before they could talk about integrals!

Acknowledgments. We are grateful for help and advice, among others, from David Allwright, Dmitry Belyaev, Teo Deveney, Guillaume Dubach, Richard Easther, Mike Giles, Jonathan Goodman, Abdul-Lateef Haji-Ali, Des Higham, Peter Kloeden, Andrew Stuart, and Grady Wright. It was Haji-Ali's arrival at Oxford that prompted the beginning of this project in late 2016. This article was written during a 2017–18 sabbatical visit by the third author to the Laboratoire de l'Informatique du Parallélisme at ENS Lyon hosted by Nicolas Brisebarre, Jean-Michel Muller, and Bruno Salvy.

REFERENCES

[1] R. J. Adler and J. E. Taylor, *Random Fields and Geometry,* Springer, 2009.
[2] R. J. Adler, J. E. Taylor, and K. J. Worsley, *Applications of Random Fields and Geometry,* book in preparation.
[3] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, *The statistics of peaks of Gaussian random fields,* Astrophys. J. 304 (1986), pp. 15–61.
[4] M. V. Berry, *Regular and irregular semiclassical wavefunctions,* J. Phys. A 10 (1977), pp. 2083–2091.
[5] E. Bogomolny and C. Schmit, *Random wave functions and percolation,* J. Phys. A 40 (2007), p. 14033.
[6] B. Bowen, C. Strong, and K. M. Golden, *Modeling the fractal geometry of Arctic melt ponds using the level sets of random surfaces,* J. Fractal Geom., to appear.
[7] A. J. Bray and D. S. Dean, *Statistics of critical points of Gaussian fields on large-dimensional spaces,* Phys. Rev. Lett. 98 (2007), 150201.
[8] J. W. Cahn and J. E. Hilliard, *Free energy of a nonuniform system. I. Interfacial free energy,* J. Chem. Phys. 28 (1958), pp. 258–267.
[9] F. Castell and J. Gaines, *The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations,* Annales Inst. Henri Poincaré et Statistiques, Gauthier-Villars, 1996, pp. 231–250.
[10] H. Doss, *Liens entre équations différentielles stochastiques et ordinaires,* Ann. Inst. H. Poincaré 13 (1977), pp. 99–125.
[11] T. A. Driscoll, N. Hale, and L. N. Trefethen, *Chebfun Guide,* Pafnuty Publications, Oxford, 2014; see also `www.chebfun.org`.
[12] F. J. Dyson, *A Brownian-motion model for the eigenvalues of a random matrix,* J. Math. Phys. 3 (1962), pp. 1191–1198.
[13] R. Easther, A. H. Guth, and A. Masoumi, *Counting vacua in random landcapes,* arXiv:1612.05224v1, 2016.
[14] P. K. Friz and M. Hairer, *A Course on Rough Paths,* Universitext, 2014.
[15] C. Gardiner, *Stochastic Methods,* 4th ed., Springer, 2009.

[16] P. J. GREEN AND B. W. SILVERMAN, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach,* CRC Press, 1993.
[17] J. D. HAMILTON, *Time Series Analysis,* Princeton U. Press, 1994.
[18] X. HAN AND P. E. KLOEDEN, *Random Ordinary Differential Equations and their Numerical Solution,* Springer, 2017.
[19] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* 2nd ed., Springer, 2009.
[20] W. HORSTHEMKE AND R. LEFEVER, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology,* Springer, 1984.
[21] J.-P. KAHANE, *Some Random Series of Functions,* 2nd ed., Cambridge U. Press, 1993.
[22] J.-P. KAHANE, *A century of interplay between Taylor series, Fourier series and Brownian motion,* Bull. London Math. Soc. 29 (1997), pp. 257–279.
[23] D. KELLY AND I. MELBOURNE, *Smooth approximation of stochastic differential equations,* Ann. Prob. 44 (2016), pp. 479–520.
[24] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations,* 2nd ed., Springer, 1999.
[25] J. LAMPERTI, *A simple construction of certain diffusion processes,* J. Math. Kyoto U. 4-1 (1964), pp. 161–170.
[26] R. LAPRISE, *The resolution of global spectral methods,* Bull. Amer. Meteor. Soc. 73 (1992), pp. 1453–1454.
[27] M. S. LONGUET-HIGGINS, *The statistical analysis of a random, moving surface,* Phil. Trans. Roy. Soc. Lond. A 249 (1957), pp. 321–387.
[28] T. LYONS, *Differential equations driven by rough signals,* Rev. Mat. Iberoamericana 14 (1998), pp. 215–310.
[29] D. J. MACKAY, *Introduction to Gaussian processes,* NATO ASI Series F Computer and Systems Sciences 168 (1998), pp. 133–166.
[30] M. B. MARCUS AND G. PISIER, *Random Fourier Series with Applications to Harmonic Analysis,* Princeton U. Press, 2016.
[31] H. MONTANELLI AND N. BOOTLAND, *Solving periodic semilinear stiff PDEs in 1D, 2D and 3D with exponential integrators,* arXiv:1604.08900, 2016.
[32] P. MÖRTERS AND Y. PERES, *Brownian Motion,* Cambridge U. Press, 2010.
[33] P. O'CONNOR, J. GEHLEN, AND E. J. HELLER, *Properties of random superpositions of plane waves,* Phys. Rev. Lett. 58 (1987), p. 1296.
[34] A. V. OPPENHEIM AND G. C. VERGHESE, *Signals, Systems and Inference,* Pearson, 2015.
[35] B. ØKSENDAL, *Stochastic Differential Equations: An Introduction with Applications,* Springer, 2003.
[36] R. E. A. C. PALEY, N. WIENER, AND A. ZYGMUND, *Notes on random functions,* Math. Zeit. 37 (1933), pp. 647–668.
[37] G. PAPANICOLAU AND KOHLER, *Asymptotic theory of mixing stochastic differential equations,* Comm. Pure Appl. Math. 27 (1974), pp. 641–668.
[38] C. E. RASMUSSEN AND C. K. I. WILLIAMS, *Gaussian Processes for Machine Learning,* MIT Press, 2006.
[39] D. REVUZ AND M. YOR, *Continuous Martingales and Brownian Motion,* Springer, 1991.
[40] S. O. RICE, *Mathematical analysis of random noise,* Bell Syst. Tech. J. 23 (1944), pp. 282–332 and 24 (1945), pp. 46–156.
[41] B. SCHÖLKOPF AND A. J. SMOLA, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond,* MIT Press, 2001.
[42] H. J. SUSSMANN, *An interpretation of stochastic differential equations as ordinary differential equations which depend on the sample point,* Bull. Amer. Math. Soc. 83 (1977), 296–298.
[43] H. J. SUSSMANN, *On the gap between deterministic and stochastic ordinary differential equations,* Ann. Prob. 6 (1978), pp. 19–41.
[44] A. SWISHCHUK AND J. WU, *Evolution of Biological Systems in Random Media: Limit Theorems and Stability,* Springer, 2013.
[45] A. TOWNSEND, H. WILBER, AND G., B. WRIGHT, *Computing with functions in spherical and polar geometries I. The sphere,* SIAM J. Sci. Comp., 38 (2016), pp. C403–C425.
[46] L. N. TREFETHEN, *Computing numerically with functions instead of numbers,* Comm. ACM 10 (2015), pp. 91–97.
[47] L. N. TREFETHEN, A. BIRKISSON AND T. A. DRISCOLL, *Exploring ODEs,* SIAM, 2018.
[48] N. WIENER, *Un problème de probabilités dénombrables,* Bull. Soc. Math. France 52 (1924), pp. 569–578.
[49] E. WONG AND B. HAJEK, *Stochastic Processes in Engineering Systems,* Springer, 1985.
[50] E. WONG AND M. ZAKAI, *On the convergence of ordinary integrals to stochastic integrals,* Ann.

Math. Stat. 36 (1965), pp. 1560–1564.
[51] E. WONG AND M. ZAKAI, *On the relation between ordinary and stochastic differential equations,* Int. J. Engng. Sci. 3 (1965), pp. 213–229.
[52] G. B. WRIGHT, M. JAVED, H. MONTANELLI, AND L. N. TREFETHEN, *Extension of Chebfun to periodic functions,* SIAM J. Sci. Comput. 37 (2015), C554–C573.
[53] D. XIU, *Numerical Methods for Stochastic Computations,* Princeton University Press, 2010.
[54] Z. ZHANG AND G. E. KARNIADAKIS, *Numerical Methods for Stochastic Partial Differential Equations with White Noise,* Springer, 2017.