

PAPER • OPEN ACCESS

Analytical approximation to the multidimensional Fokker–Planck equation with steady state

To cite this article: R J Martin *et al* 2019 *J. Phys. A: Math. Theor.* **52** 085002

View the [article online](#) for updates and enhancements.

Recent citations

- [Modeling biological refractory periods and synaptic depression in an artificial neuron](#)
Eric Hamel and Richard Labib



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Analytical approximation to the multidimensional Fokker–Planck equation with steady state

R J Martin^{1,3} , R V Craster¹ , A Pannier¹  and M J Kearney² 

¹ Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom

² Senate House, University of Surrey, Guildford, GU2 7XH, United Kingdom

E-mail: richard.martin1@imperial.ac.uk

Received 4 August 2018, revised 15 December 2018

Accepted for publication 15 January 2019

Published 29 January 2019



CrossMark

Abstract

The Fokker–Planck equation is a key ingredient of many models in physics, and related subjects, and arises in a diverse array of settings. Analytical solutions are limited to special cases, and resorting to numerical simulation is often the only route available; in high dimensions, or for parametric studies, this can become unwieldy. Using asymptotic techniques, that draw upon the known Ornstein–Uhlenbeck (OU) case, we consider a mean-reverting system and obtain its representation as a product of terms, representing short-term, long-term, and medium-term behaviour. A further reduction yields a simple explicit formula, both intuitive in terms of its physical origin and fast to evaluate. We illustrate a breadth of cases, some of which are ‘far’ from the OU model, such as double-well potentials, and even then, perhaps surprisingly, the approximation still gives very good results when compared with numerical simulations. Both one- and two-dimensional examples are considered.

Keywords: stochastic processes, mean reversion, Fokker–Planck equation

(Some figures may appear in colour only in the online journal)

1. Introduction

The Fokker–Planck equation (FPE) arises in a broad range of problems from physics, engineering science, economics and mathematical modelling. Part of this breadth of application

³ Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

can be traced back to its success in modelling generic transport processes provided the dynamics can be represented by a Hamiltonian or Lagrangian with random components [19]. Alternatively, a time series of data can be analysed as a Markov process to create an effective FPE that captures the statistics of the observed process. This has led to modelling based upon extracting FPEs from experimental or observable data for instance in turbulent cascades [22], fractal-generated turbulence [45], modelling the beat fluctuations in heart-rate [24], electronic noise and kinetics [11], electronic circuits with nonlinear resistance [16], systems with over-damped Langevin dynamics [17], climate modelling [14], or from nonlinear friction [31, 47]. Financial modelling yields a wealth of further applications [23]. The modelling of market behaviour, where deviation from equilibrium is likely to be accompanied by higher volatility and/or slower mean-reversion, means that the invariant distribution is fat-tailed or leptokurtic [36]. Agent-based models [1] for the herd behaviour of interactions between traders, and the influence of rumours, social interactions and external information [7], take ideas from Kirman’s stochastic models of information transmission [29] to arrive at FPEs. Rather than study a single specialised case we consider generic FPEs with the main restriction being that we consider mean reverting processes.

Another important application for FPEs is in Kalman filtering, which in addition to physics also impacts upon control theory, optimisation, and time series analysis. Examples include: subatomic particle tracks [9]; movements in the ionosphere [43]; chemical reactions [41]; and extensive use in econometrics for making predictions about financial variables in systems dominated by stochastic behaviour, e.g. [15]. The basic operation of Kalman-type filters is long-established (see e.g. [26, 27, 37, 46]), and a key ingredient is the state transition density. Initial work used linear assumptions, i.e. an underlying Gaussian model. When the drift and covariance are nonlinear functions of the state vector, the extended Kalman filter [18] is often used, but in essence this employs a linearisation so that the Gaussian is used locally. This is usually acceptable for short time periods, but may not be over longer ones. A better approximation to the transition density for non-Gaussian processes, such as that we provide, is therefore highly desirable.

The normalised form of the FPE, in one dimension, that we investigate is

$$\frac{\partial f}{\partial \tau} = -\frac{\partial}{\partial y} [A(y)f] + \frac{\partial^2 f}{\partial y^2} \equiv \mathcal{L}^\dagger f, \quad f(0, y) = \delta(y - y_0) \quad (1)$$

where f is the probability density function, τ is nondimensional time and y is a spatial variable; the initial condition is a delta-function centred at y_0 . There is a general drift function $A(y)$ for which we will take various choices as examples in later sections. The steady-state solution $f(\infty, y) \propto \exp \int^y A(z) dz$, is a normalisable probability density: this can be ensured by the condition $\limsup_{|y| \rightarrow \infty} -yA(y) > 1 + \varepsilon$ for some $\varepsilon > 0$. As alluded to earlier this partial differential equation, the FPE, is connected to a stochastic differential equation (SDE):

$$dY_t = \kappa A(Y_t) dt + \sqrt{2\kappa} dW_t, \quad (2)$$

with $\tau = \kappa t$. More generally an SDE having both spatially varying mean and variance is

$$dX_t = \mu_X(X_t) dt + \sigma_X(X_t) dW_t; \quad (3)$$

however, provided σ_X is bounded away from zero we can make the substitution (also known as the Lamperti transformation) from X to Y defined by $dy/dx = \sqrt{2\kappa}/\sigma_X(x)$, which places (3) in the normalised form (2); therefore the analysis we present for the normalised form is more general than it first appears. In higher dimensions we take A to be the gradient of a potential, i.e. conservative, and note that under some common assumptions the Lamperti transformation

can be appropriately generalised [39]. It is worth noting that in one particular case, the *affine* model in which μ and σ^2 are both linear in X_t , the FPE associated with (3) is exactly solvable [25, 32, 34]. In general, though, the full solution of the FPE is considerably more difficult than the stationary solution.

The best-known example with a steady state is the OU process given by $A(y) = -\theta y$ (θ is a parameter), for which there is a well-known explicit Gaussian solution [40]. However, one often wants to deviate from this model because away from equilibrium the force field A cannot be expected to rise without limit, but instead be bounded; equivalently, the equation describes diffusion in the presence of a potential which cannot be expected to be quadratic in general. In these situations the steady state will no longer be Gaussian. Departing from the OU, whilst gaining closer connection to the physical model under consideration, loses analytical tractability and typically numerical methods are required. Naturally, one would like the best of both worlds: analytical tractability and physical relevance.

Given the connection with SDEs, there is a choice between the deterministic or stochastic approaches in terms of which is more practical to tackle numerically. Choosing the latter naturally leads to Monte-Carlo methods: evaluating the density of a stochastic process requires not only a large number of simulations, but also kernel density estimation (KDE) at each point in time to produce a smooth estimate from the simulated data points: For literature on KDE, see, e.g. [4, 42, 44]. The quality of the estimation depends on the kernel width and the number of simulations. The optimal choice of width is not straightforward [13, 28]⁴. In the work we have done, if $f(\infty, y)$ is very fat-tailed, for example student- t , then simulation of $\gtrsim 100\,000$ paths is required to get an L^1 error (integrated absolute error) of $4 \cdot 10^{-2}$. In two dimensions, $\gtrsim 100\,000$ paths only yield an L^1 error of $2 \cdot 10^{-1}$. Furthermore, for every evaluation point one typically needs to make use of every path simulated, so the computation time of estimating the density is much higher than the simulation time and suffers considerably from the curse of dimensionality; see [10] for background and numerical algorithms that aim to lift the curse. This impediment means that numerical methods for the FPE are preferred and literature combining Monte-Carlo and KDE methods is rare and specific, e.g. [5]. Thus, when we perform numerical calculations we do so upon the FPE directly; for efficiency and accuracy we use spectral methods and detail these in section 2.1. Even so, the numerical simulations in two dimensions become very time-consuming, particularly if one is not looking for the solution at short time. Furthermore, if one desires to sweep over different parameter values to analyse the effect on the results, the problem gets even worse.

All this points to the need for a simple, fast, approximation. Further, it is desirable to have a modus operandi that gives results that are intuitive and offer direct insight into the problem at hand. Ideally, we would like to address the metaphysical question of ‘how diffusions think about solving themselves’. In this respect, what we are going to describe—in the first instance (10)—does give clear intuition, in that the various terms in the equation make it clear how the solution behaves; also, the result at some basic level *looks* like an expression for the evolution of a probability density, in a way that an infinite series of eigenfunctions does not.

Most analytical approximations are based on the summation of eigenfunctions [40] that are often orthogonal polynomials and special functions, which is unsurprising given the linearity of the FPE; the OU can also be approached this way using Hermite polynomials [38]. Alternative asymptotic approaches, such as WKB methods [8], are useful for analysing the approach to equilibrium but are limited to studying specific regimes involving small or large diffusivities. A somewhat different approach, upon which the first steps were made in [36],

⁴The bandwidth that minimizes the mean integrated square error, Scott’s rule, is of order $N^{-1/(d+4)}$ where N is the number of simulations and d the number of dimensions and leads to an accuracy of $O(N^{-4/(d+4)})$.

consists in expressing the solution as a product of terms, rather than the more usual sum. Intuitively, with a sum it is difficult to represent the initial delta-function without generating oscillatory artefacts, whereas with a product this is simple: a narrow Gaussian, of width tending to zero as $\tau \rightarrow 0$, will capture that *regardless of what the other terms in the product are*; another term can capture the steady state; and a series of correction terms ‘patches-up’ the mid-term behaviour. Intuitively, we are expanding around an OU model, in the sense of finding the characteristics of a mean-reverting solution as exemplified by the OU case and capturing these characteristics for the general case, while reproducing the OU case exactly. There is a passing similarity with the WKB approximation, mainly in the use of a logarithmic transformation, but WKB expands around a zero-volatility (deterministic) model and is singular in the zero-volatility limit.

There are some important consequences of using products. The logarithm of the density is represented by a sum, and so: (i) positivity is guaranteed, in a way that it is typically not using linear methods; (ii) from the theoretical perspective there is a clear relation to entropy, and calculation of that from an approximated density is virtually impossible if the approximated density is anywhere negative. Based on these points, we therefore choose to develop this approach.

We begin by introducing our approach in the one-dimensional setting (section 2), giving the key results and the main technical route to them. A range of examples demonstrates the efficiency of the results as we move away from the OU process. We then move on to the higher-dimensional case in section 3 and show numerical simulations. Section 4 discusses extensions and potential limitations of the method, with concluding remarks drawn together in section 5.

2. Theory in one dimension

We introduce the methodology and results in one dimension. A key step is the decision to work with the normalised density g , and the derivative, h , of the log-density⁵ defined as $h = -(\partial/\partial y) \log g$, with $g(\tau, y) = f(\tau, y)/f(\infty, y)$. Thereby g solves the backward equation

$$\frac{\partial g}{\partial \tau} = A(y) \frac{\partial g}{\partial y} + \frac{\partial^2 g}{\partial y^2} \equiv \mathcal{L}g \tag{4}$$

with initial condition a delta-function of strength $1/f(\infty, y_0)$ at y_0 , and h solves the nonlinear partial differential equation (PDE)

$$\frac{\partial h}{\partial \tau} = \frac{\partial}{\partial y} \left\{ A(y) h + \frac{\partial h}{\partial y} - h^2 \right\} \tag{5}$$

which has a singular initial condition, in the sense that $h \sim (y - y_0)/2\tau$ as $\tau \rightarrow 0$. This singular behaviour can also be seen by dominant balance in (5).

Next we observe that for the OU model $A(y) = \theta(y_\infty - y)$, with the constant y_∞ denoting the long-term mean, we have exactly

$$\text{(OU)} \quad h(\tau, y) = \frac{\theta\sqrt{q}(y - y_0)}{1 - q} + \frac{\theta\sqrt{q}(y_\infty - y)}{1 + \sqrt{q}}, \quad q = e^{-2\theta\tau}, \tag{6}$$

as is easily verified by substituting it into (5) and ploughing through the algebra. Finally, from its definition, $g(\infty, y)$ must always equal unity, and so $h(\infty, y) = 0$. This inspires the ansatz

⁵ f, g, h are understood to mean f_Y, g_Y, h_Y respectively.

$$h(\tau, y) = \frac{\theta\sqrt{q}(y - y_0)}{1 - q} + \frac{\sqrt{q}}{1 + \sqrt{q}}A(y) + \sqrt{q}o(1)_{q=1} \quad (7)$$

for arbitrary A . When (7) is inserted into the PDE for h , (5), and a Laurent expansion performed around $\tau = 0$, the lhs and rhs agree at $O(\tau^{-2})$ and $O(\tau^{-1})$, explaining why we are writing the error term in (7) as $o(1)$ in the short-time limit. This error can then in principle be approximated as a Taylor series around $\tau = 0$, which we will discuss later.

Another matter presents itself when this Laurent expansion is done: the development of (7) is

$$h(\tau, y) = \frac{y - y_0}{2\tau} + \frac{A(y)}{2} + o(1), \quad \tau \rightarrow 0,$$

and we observe that θ is absent from both the first two terms. Accordingly, all θ 's give the same leading-order behaviour, and so we cannot say anything about θ simply by looking at the first two terms in the short-time expansion of the solution. In this sense, therefore, θ is now arbitrary, representing an estimate of the mean reversion speed of the force-field A , or, equivalently, the reversion speed of the OU process 'about which' we are expanding the given model A . Given that the leading-order asymptotics do not tell us what θ to use, some other method of inference is necessary, and we return to this later. In this context we call (7) the leading-order approximation for h .

To deduce g and f we integrate (7) from y_0 to y , giving (again for arbitrary A)

$$g(\tau, y) \sim (\dots)_{\tau, y_0} \exp\left(\frac{-\frac{1}{2}\theta\sqrt{q}(y - y_0)^2}{1 - q}\right) f(\infty, y)^{\frac{-\sqrt{q}}{1+\sqrt{q}}} f(\infty, y_0)^{\frac{\sqrt{q}}{1+\sqrt{q}}}$$

where $(\dots)_{\tau, y_0}$ generically denotes a function of τ, y_0 . By means of the reciprocity condition ⁶

$$g(\tau, y | y_0) = g(\tau, y_0 | y) \quad (8)$$

that is obeyed by the exact solution, we can infer the dependence of the prefactor on y_0 , to obtain

$$g(\tau, y) \sim (\dots)_\tau \exp\left(\frac{-\frac{1}{2}\theta\sqrt{q}(y - y_0)^2}{1 - q}\right) f(\infty, y)^{\frac{-\sqrt{q}}{1+\sqrt{q}}} f(\infty, y_0)^{\frac{-\sqrt{q}}{1+\sqrt{q}}}.$$

The OU case requires the prefactor to be

$$\frac{(\theta/2\pi)^{\frac{\sqrt{q}}{1+\sqrt{q}}}}{\sqrt{1 - q}}.$$

(Another way of deriving the prefactor is to write

$$g(\tau, y) = n(\tau) e^{-\int_{y_0}^y h(\tau, z) dz},$$

substitute into (1), and solve for the function n , which obeys a first-order linear differential equation; we shall refer to this technique later on.) Thence

⁶This can also be written $f(\tau, y | y_0)/f(\infty, y) = f(\tau, y_0 | y)/f(\infty, y_0)$. Viewed as a function of y (and τ), the lhs obeys the adjoint forward equation, whereas the rhs obeys the backward equation. However, those are the same PDE, with the same initial condition; alternatively, we could invoke the Kolmogorov criterion, e.g. [30, section 1.5].

$$g(\tau, y) \sim \frac{1}{\sqrt{1-q}} \exp\left(\frac{-\frac{1}{2}\theta\sqrt{q}(y-y_0)^2}{1-q}\right) \left(\frac{\theta/2\pi}{f(\infty, y)f(\infty, y_0)}\right)^{\frac{\sqrt{q}}{1+\sqrt{q}}} \tag{9}$$

and

$$f(\tau, y) \sim \frac{(\theta/2\pi)^{\frac{\sqrt{q}}{1+\sqrt{q}}}}{\sqrt{1-q}} \exp\left(\frac{-\frac{1}{2}\theta\sqrt{q}(y-y_0)^2}{1-q}\right) f(\infty, y)^{\frac{1}{1+\sqrt{q}}} f(\infty, y_0)^{\frac{-\sqrt{q}}{1+\sqrt{q}}} \tag{10}$$

which are the lowest-order approximations.

Equation (7), and its consequences, have several facets worthy of comment. First, they are exact for any OU model of reversion speed θ regardless of the reversion level, i.e. for $A(y) = \theta(y_\infty - y)$. Another way of putting this is to say that the correction terms to (7) will be expressible as functions of $(d/dy)(A(y) + \theta y)$. Secondly, the approximations (9) and (10) are necessarily positive, and correct in both the short- and long-time limits, regardless of $A(y)$. Thirdly, the approximations (9) and (10) obey the reciprocity condition.

We can proceed to determine higher-order terms in the representations for, say, h by continuing (7) and introducing an expansion for the remainder so

$$h(\tau, y) = \frac{p}{1+p} \left(\frac{\theta(y-y_0)}{1-p} + A(y) + \sum_{r=1}^{\infty} (1-p)^r b_r(y) \right) \tag{11}$$

where $p = \sqrt{q} = e^{-\theta\tau}$, and consider the remainder terms, (b_r) , that also depend parametrically on the starting-point y_0 . An alternative expansion is in powers of $(1-q)$, giving a different series, but with similar convergence properties and we do not pursue this further here. The power of \sqrt{q} in (11) ensures $h(\infty, y) = 0$ for any truncation of the series (i.e. sum as far as $r = N$). By inserting (11) into (5) and comparing coefficients in powers of $(1-p)^{r+1}$ we find that

$$\frac{d}{dy} [(y-y_0)^{r+2} b_{r+1}(y)] = (y-y_0)^{r+1} \mathfrak{F}_r(y)$$

where \mathfrak{F}_r is a complicated quadratic expression invoking A, b_1, \dots, b_r and their derivatives. So although (5) is second-order nonlinear, successive terms in the expansion can be recursively obtained by solving a first-order linear differential equation, which can be integrated immediately to give⁷

$$b_{r+1}(y) = \frac{1}{(y-y_0)^{r+2}} \int_{y_0}^y (z-y_0)^{r+1} \mathfrak{F}_r(z) dz.$$

In the special case $r = 0$, we have

$$b_1(y) = \frac{1}{\theta(y-y_0)^2} \int_{y_0}^y (z-y_0) \left(\frac{d}{dz} + A(z) + \frac{\theta}{2}(z-y_0) \right) \frac{d}{dz} (A(z) + \theta z) dz$$

which vanishes whenever $A(y) + \theta y$ is a constant, as it should. As an aside, the term $b_1(y)$ also vanishes in another special case, i.e. when

$$\tilde{A}' + \frac{1}{2}\tilde{A}^2 = \theta(y^2/4 + y_0/2 + \text{const}), \quad \tilde{A}(y) \equiv A(y) + \theta y;$$

⁷The lower limit has to be y_0 , as otherwise b_{r+1} will be singular at $y = y_0$.

which has \tilde{A} as the logarithmic derivative of a parabolic cylinder function. This solution is sporadic in that it also depends on the starting-point, that is, for this choice of $A(y)$ the function $b_1 = 0$ only if y_0 is chosen correctly.

As (7) is a sum, the expressions for f, g will be infinite products, and so we refer to the method as an infinite product expansion. The focus here is not on the extraction of higher-order terms: rather, it is on the leading-order term, which is probably the most applicable. But it is notable that, should it be desired, one can also treat the correction term by spectral methods: that is to say, write $\tilde{f}(\tau, y)$ for (10), derive the PDE that it satisfies (this is another parabolic PDE), and solve it approximately by means of a Galerkin or collocation expansion [2]. As the initial spike and also the long-term behaviour have already been accounted for in \tilde{f} , the unknown function f/\tilde{f} is unity in both limits $\tau \rightarrow 0, \infty$, and therefore well approximated by spectral methods, which are ideally suited to smooth problems.

The first steps towards a product expansion were made in [36], which carries out a power series expansion on broadly similar lines and provides an expansion of the form

$$h(\tau, y) = \frac{\theta(qy - \sqrt{q}y_0)}{1 - q} + \sqrt{q} \sum_{r=0}^{\infty} (1 - q)^r b_r^{\text{old}}(y). \tag{12}$$

The development here confers several advantages, besides greater compactness and elegance, over [36]: (i) the reciprocity condition (8) is enforced, whereas it was not in [36]; (ii) it is more readily adaptable to higher dimensions; (iii) it is more accurate over shorter time-scales; (iv) the leading-order approximation in [36] exhibits instability when $|y_0|$ is large. What is shown here can be obtained from (12) by taking the initial term and $b_0^{\text{old}}(y)$ as the new leading-order term, with minor alterations, and modifying the prefactor in the derivation of g so as to enforce (8). Otherwise, however, it is neither more nor less convergent, being in effect a rearrangement of the terms.

The approach of expanding around an OU process leaves a free parameter, θ , that controls the intermediate-time behaviour, and the remainder series implicitly depends upon it. As we said earlier, we cannot infer θ from the $O(\tau^{-1})$ or $O(\tau^0)$ terms in the short-time expansion. We can argue that θ should be chosen to minimise $b_1(y)$ as given earlier, and as that function vanishes when $A'(y) + \theta$ is identically zero, it seems reasonable to choose θ so as to minimise $A'(y) + \theta$ ‘on average’. This motivates the choice

$$\hat{\theta} = \langle -A' \rangle_{\infty} = \langle A^2 \rangle_{\infty} \tag{13}$$

where $\langle \cdot \rangle_{\infty}$ means an average over the invariant density $f(\infty, \cdot)$. This was proposed in [36], albeit using a different line of reasoning based on a sort of Rayleigh–Ritz argument to identify the least negative eigenvalue of \mathcal{L}^\dagger , the differential operator of the FPE⁸. Clearly (13) is correct in the OU case, and it guarantees $\hat{\theta} > 0$, all of which make it a pragmatic choice, but there is another compelling reason based on a connection with entropy and information theory, which runs as follows. Consider, for some p.d.f. ψ , the family of distributions $\psi(y - \mu)$ indexed by the parameter $\mu \in \mathbb{R}$. It is desired to estimate μ (from data), and the standard way of doing this is the maximum likelihood estimator. Writing

$$f(y | \mu) = \psi(y - \mu)$$

we seek to maximise $\log f(y | \mu)$ w.r.t. μ . The Fisher information [33, section 2.5] is the expectation of the square of the μ -derivative of the log-likelihood, and hence is

⁸ It was, however, reliant on the operator \mathcal{L}^\dagger having a discrete spectrum, which it may not.

$$\begin{aligned} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \mu} \log f(y | \mu) \right)^2 f(y | \mu) dy &= \int_{-\infty}^{\infty} \left(\frac{\psi'(y - \mu)}{\psi(y - \mu)} \right)^2 \psi(y - \mu) dy \\ &= \int_{-\infty}^{\infty} \left(\frac{\psi'(y)}{\psi(y)} \right)^2 \psi(y) dy; \end{aligned}$$

if we set $\psi(y) = f(\infty, y)$ then this is exactly the definition of $\hat{\theta}$. In broad terms, the higher the Fisher information, the more certain we are about the estimation of the parameter in question, and indeed the reciprocal of the Fisher information furnishes the Cramér–Rao lower bound for the variance of any unbiased estimator. This has an interpretation in terms of the mean reversion: the higher the average speed of mean reversion, the more certain we are about our estimate of the mean from a given dataset, and vice versa. Using the Fisher information as an estimator of reversion speed is therefore natural. As will be seen later, the method extends in a natural way to higher dimensions, with $\partial/\partial\mu$ replaced by ∇ , and then the Fisher information is a positive definite symmetric matrix rather than just a positive number.

2.1. Results and discussion in one dimension

We consider a range of one dimensional examples deviating from the OU by different degrees and compare the leading-order approximation (10) to numerical solutions from a PDE solver. The numerical solution is computed by means of Fourier spectral collocation in the spatial direction coupled with a fourth-order Runge–Kutta algorithm in time and uses a narrow Gaussian as initial condition; a sufficiently large spatial domain is taken such that the FPE does not interact with the edge, and convergence is checked by mode-doubling. We take advantage of the linear diffusion by using an integrating-factor, and also the fast Fourier transform, to design a highly efficient solver; such methods are standard in scientific computing [6, 48] and the fully-converged numerical simulations act as the ‘gold standard’ against which we compare the approximations. Even then, the PDE solver is far slower than the leading-order approximations. Our solution is a truncation from an infinite product, and so it is of central importance to demonstrate its efficiency. The following examples explore the domain of validity of the leading-order approximation in which it successfully replicates the time-evolution of the numerical FPE solutions. We mainly choose cases from fat-tailed invariant distributions, which were our main motivation; the last case exhibits the adaptability of the approximation in places where it was unexpected.

2.1.1. Sech-power model. One way of moving away from the linear force field (quadratic potential well) of the OU model is to make the force field grow less rapidly away from equilibrium by stipulating

$$A(y) = -\frac{\hat{\delta}}{\hat{\gamma}} \tanh \hat{\gamma} y. \quad (14)$$

This is an example, of the well-known Pearson diffusions [21], which is obtained from the local volatility model

$$dX_t = -\kappa X_t dt + \sigma \sqrt{1 + \gamma^2 X_t^2} dW_t, \quad (15)$$

in which volatility increases away from equilibrium, using the transformation $\gamma X = \sinh \hat{\gamma} Y$; see [36] for more details and some applications in mathematical finance. The associated potential is also known in mathematical physics as the Pöschl–Teller potential [20] in the special

case $\hat{\delta}/\hat{\gamma}^2 = 2$. The resulting Schrödinger equation is solvable in terms of special functions and its link with the FPE has been used to derive analytical solutions, see [3]. The steady state is a sech-power:

$$f(\infty, y) = \frac{\hat{\gamma}(\cosh \hat{\gamma}y)^{-\hat{\delta}/\hat{\gamma}^2}}{\mathbf{B}\left(\frac{\hat{\delta}}{2\hat{\gamma}^2}, \frac{1}{2}\right)}, \quad \langle -A' \rangle_{\infty} = \frac{\hat{\delta}^2}{\hat{\delta} + \hat{\gamma}^2} \quad (16)$$

with \mathbf{B} denoting the Beta function; the explicit solutions mean that this is an attractive model that has been well-studied, for instance, for systems with nonlinear random vibrations (16) occurs, see [35].

In the limit where $\hat{\gamma} \rightarrow 0$ we recover the OU model, so $\hat{\gamma}$ is a measure of the deviation from OU. We have looked at many parameter sets, the comparison of the leading order solution to full numerical simulations is consistently qualitatively pleasing, and a typical example in figure 1 shows the comparison.

2.1.2. Dry-friction. Dry-friction [47] is the limit of the sech-power model obtained when $\hat{\delta} = \hat{\gamma} \rightarrow \infty$, and we then have a discontinuous $A(y)$ as

$$A(y) = -\text{sgn } y, \quad f(\infty, y) = \frac{e^{-|y|}}{2}, \quad \langle -A' \rangle_{\infty} = 1.$$

This case is interesting in terms of the physics it describes, but also as additionally the transition density is available in closed form [47], e.g. by the usual route of Laplace transforming the Fokker–Planck equation:

$$f(\tau, y | y_0) = \frac{e^{-(y-y_0)^2/4\tau}}{\sqrt{4\pi\tau}} e^{-\tau/4} e^{(|y_0|-|y|)/2} + \frac{e^{-|y|}}{2} \Phi\left(\frac{\tau - |y| - |y_0|}{\sqrt{2\tau}}\right) \quad (17)$$

or

$$g(\tau, y | y_0) = \frac{e^{-(y-y_0)^2/4\tau}}{\sqrt{\pi\tau}} e^{-\tau/4} e^{(|y_0|+|y|)/2} + \Phi\left(\frac{\tau - |y| - |y_0|}{\sqrt{2\tau}}\right) \quad (18)$$

and this exact solution provides a convenient benchmark against which to test our theory.

In this example, f arises as a sum of two pieces, and so h , rather than being a simpler function than g (as it is for example with the OU), is more complicated. Nonetheless, comparing to numerical simulation the accuracy is excellent when starting at the origin (not shown) or fairly near the origin ($y_0 = -2$, figure 2(a)), but worse when it is much further away ($y_0 = -5$, figure 2(b)). This points to a separate development of the theory that deals with far-field expansions, and that we pursue later in section 4: as we will show by consideration of (5), the first term in either of the above expressions corresponds to an approximation in which we start a long way from equilibrium and the drift is small, as here.

2.1.3. Student- t . Another popular deviation, and as noted in the introduction important across many fields, from the OU is that associated with fat-tailed distributions. We use, as in [36], a distribution that conveniently has student- t as its steady state in the Y coordinates:

$$A(y) = -\frac{y}{1 + \hat{\gamma}^2 y^2}, \quad f(\infty, y) = \frac{\hat{\gamma}(1 + \hat{\gamma}^2 y^2)^{-(\nu-1)/2}}{\mathbf{B}\left(\frac{\nu-2}{2}, \frac{1}{2}\right)}, \quad \langle -A' \rangle_{\infty} = \frac{\nu-2}{\nu+1}.$$

This model gives rise to fatter tails than the sech-power example, because the force-field decays to zero as $|y| \rightarrow \infty$. The distinction is further accentuated by noting that the sech-power

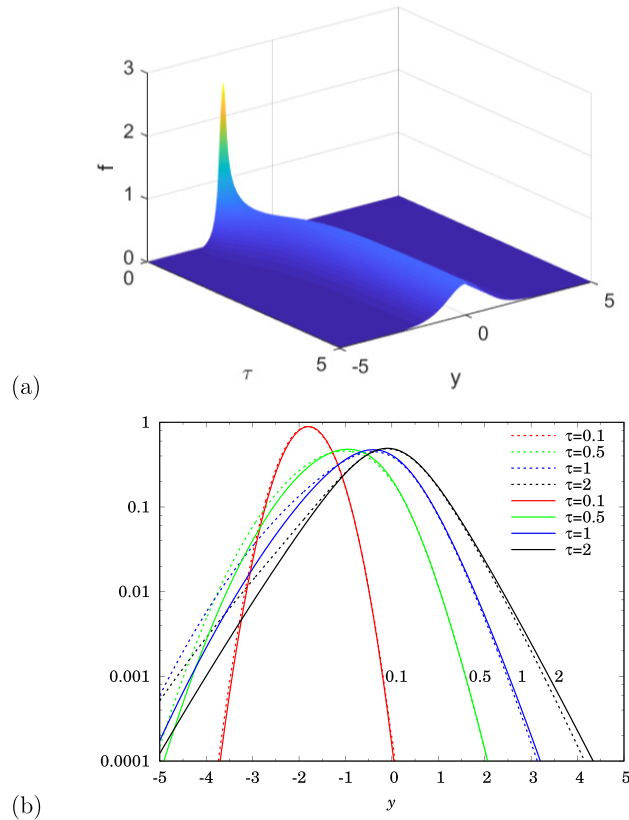


Figure 1. Density as a function of position (sech^2 , $\hat{\gamma} = 1$, $\hat{\delta} = 2$), with $y_0 = -2$, (a) the profile of f versus times from the numerical simulation and (b) comparing the leading-order approximation (solid) with the numerics (dotted).

case can be obtained, for certain parameter values, from transforming a model in which the steady state is student- t in X coordinates; we demonstrate the efficacy of the leading-order approximation in figure 3.

2.1.4. Double-well potentials. A gross deviation from OU is that of double-well potentials, and quite remarkably we find that the leading-order approximation still performs well capturing both the quantitative features and the qualitative behaviour, see figures 4 and 5. We take a very general form

$$f_Y(\infty, y) = Ke^{-y^2/2} \frac{y^2 + \gamma^2}{((y - \alpha_1)^2 + \beta_1^2)((y - \alpha_2)^2 + \beta_2^2)}$$

with zeros at $y = \pm i\gamma$ and poles at $y = \alpha_j \pm i\beta_j$ for $j = 1, 2$; in [36] a limited case with just the quadratic in the numerator, but no denominator, that does not allow the flexibility to explore the parameters. These act as follows: $\gamma \rightarrow 0$ makes the two wells disjoint, so that it becomes progressively less easy to transit from one well to the other; α controls their location; $\beta \rightarrow 0$ makes them deeper. The force-field is

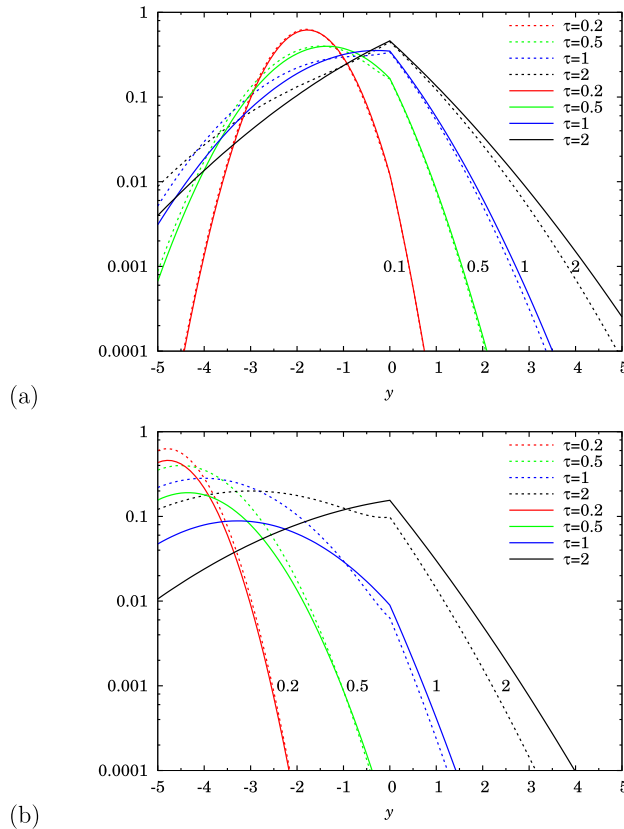


Figure 2. Density as a function of position (dry-friction), with (a) $y_0 = -2$, (b) $y_0 = -5$ at several different time points and comparing the leading-order approximation (solid) with the numerical simulation (dotted).

$$A(y) = -y + \frac{2y}{y^2 + \gamma^2} - \frac{2(y - \alpha_1)}{(y - \alpha_1)^2 + \beta_1^2} - \frac{2(y - \alpha_2)}{(y - \alpha_2)^2 + \beta_2^2}$$

and there are explicit forms for $\langle -A' \rangle_\infty$, and the normalising constant K , that can be obtained from the complex error function, or which can simply be evaluated numerically. We have evaluated several parameter sets and the results are shown in figures 4 and 5 are typical. To be specific, the parameters are: poles at $\pm 2 \pm i$, zeros at $\pm i/\sqrt{2}$, and for this we have $\langle -A' \rangle_\infty \approx 1.557$. It is evident that starting from the equilibrium point (figure 4) produces different results from starting in one of the wells (figure 5). In the former, the approximation is excellent whilst in the latter, the approximation overestimates the rate at which the process ‘finds out about’ the other well, with the density being shared between the two wells at too early a time, though of course as $\tau \rightarrow \infty$ the results must again agree. Nonetheless even when starting in one of the wells the leading-order approximation gives qualitative insight. It is perhaps surprising that the approximation works at all, and that it does is suggestive that our philosophy of building an approximation based upon ‘how diffusions think about themselves’ and using the slightly counter-intuitive approach of using a product, rather than a sum, expansion is of value.

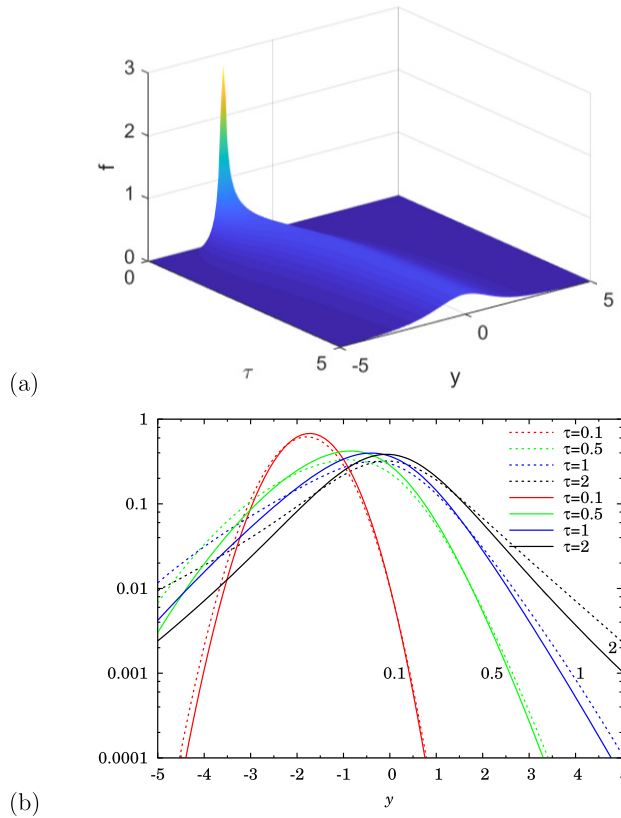


Figure 3. Density as a function of position (student- t) with (a) the profile of f versus times from the leading-order approximation and (b) comparing the leading-order approximation (solid) with the numerics (dotted).

3. Multivariate theory

Buoyed by the success of the one-dimensional theory, and the exemplars that demonstrate the viability of the approach we employ across a range of illustrative cases, we move to higher dimensions where the availability of a fast accurate approximation to partially cure the ‘curse of dimensionality’ is attractive. In extending to the multivariate case we use two approaches: we can consider the case of the multivariate OU process as a guideline, and we can attempt to glue together components from the general one-dimensional case. There are, however, some preliminaries before we embark upon this. The general form (noting the discussion in [39] regarding assumptions required for the multivariate Lamperti transformation to move from X_t to Y_t) is

$$dY_t = \kappa A(Y_t) dt + \sqrt{2\kappa} dW_t \tag{19}$$

where, in m dimensions, $W_t, Y_t \in \mathbb{R}^m$ and $A : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is the force field and the corresponding FPE is (with $\tau = \kappa t$ as before)

$$\frac{\partial f}{\partial \tau} = -\nabla \cdot (Af) + \nabla^2 f. \tag{20}$$

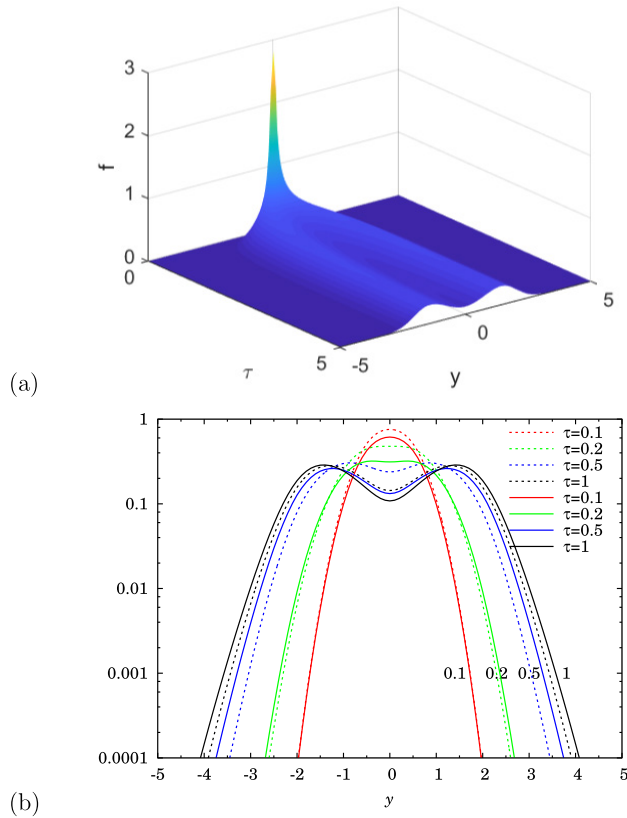


Figure 4. Density as a function of position (Double-well Vsn 1, poles $\pm 2 \pm i$, zeros $\pm i/\sqrt{2}$, $(\alpha, \beta, \gamma) = (2, 1, \frac{1}{\sqrt{2}})$) with $y_0 = 0$. (a) Shows the field profile in time, taken from the leading-order approximation and (b) comparison of the leading-order approximation (solid) versus numerics (dotted).

It is natural to take A , the force field, conservative: that is to say it is the gradient of a potential. Indeed, $A = \nabla \log f_\infty$, where f_∞ is the steady-state solution and then $g = f/f_\infty$ obeys the adjoint equation

$$\frac{\partial g}{\partial \tau} = A \cdot \nabla g + \nabla^2 g \tag{21}$$

so that if $H = -g^{-1} \nabla g$, we have the vector equation

$$\frac{\partial H}{\partial \tau} = \nabla(A \cdot H + \nabla \cdot H - H \cdot H). \tag{22}$$

If A were considered non-conservative then the equations (21) and (22) are no longer correct.

3.1. Symmetric OU

By a symmetric OU process, we mean (in normal form) that

$$dY_t = -\kappa a Y_t dt + \sqrt{2\kappa} dW_t \tag{23}$$

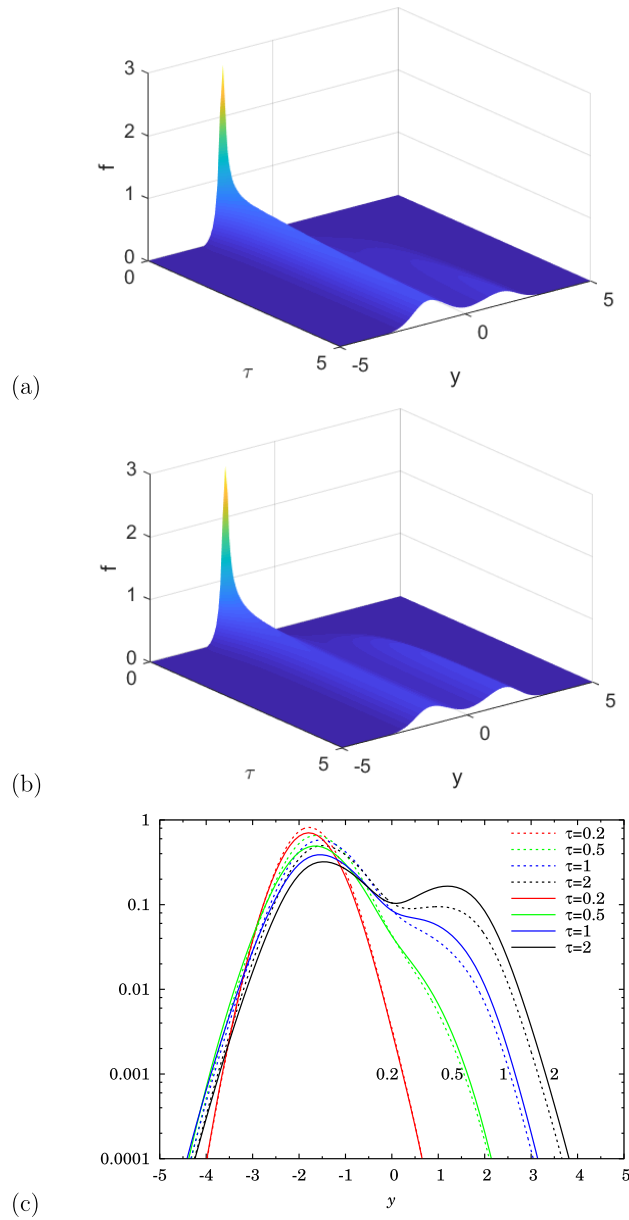


Figure 5. Density as a function of position (Double-well $V_{sn} 1$, poles $\pm 2 \pm i$, zeros $\pm i/\sqrt{2}$, $(\alpha, \beta, \gamma) = (2, 1, \frac{1}{\sqrt{2}})$) with $y_0 = -2$, at several different time points: (a) PDE solver, (b) leading-order approximation, (c) comparison at several different time points (dotted line is PDE solver, solid is leading-order approximation).

where italic bold letters are square matrices and W_t is an m -dimensional standard Brownian motion, i.e. different coordinates are independent and so $\mathbf{E}[dW_t dW_t^\dagger] = \mathbf{I} dt$ where \dagger denotes the transpose. The force field is conservative if, and only if, the matrix \mathbf{a} , which we call the generator, is symmetric, and this is assumed henceforth. The steady-state is a Gaussian

distribution of mean 0 and with covariance matrix $\sigma_\infty = \mathbf{a}^{-1}$. Further, writing \mathbf{q} as the matrix exponential

$$\mathbf{q} = \exp(-2\mathbf{a}\tau),$$

which is evaluated by diagonalising \mathbf{a} , this is eased by noting that \mathbf{a} is symmetric, we find the following

$$\begin{aligned} \sigma(\tau) &= \mathbf{a}^{-1}(\mathbf{I} - \mathbf{q}) \\ f(\infty, y) &= |\mathbf{a}/2\pi|^{1/2} \exp(-\frac{1}{2}y^\dagger \mathbf{a} y) \\ g(\tau, y | y_0) &= \frac{1}{|\mathbf{I} - \mathbf{q}|^{1/2}} \exp\left(-\frac{1}{2}y^\dagger \frac{\mathbf{a}\mathbf{q}}{\mathbf{I} - \mathbf{q}} y - \frac{1}{2}y_0^\dagger \frac{\mathbf{a}\mathbf{q}}{\mathbf{I} - \mathbf{q}} y_0 + y_0^\dagger \frac{\mathbf{a}\sqrt{\mathbf{q}}}{\mathbf{I} - \mathbf{q}} y\right) \\ &= \frac{1}{|\mathbf{I} - \mathbf{q}|^{1/2}} \exp\left(-\frac{1}{2}(y - y_0)^\dagger \frac{\mathbf{a}\sqrt{\mathbf{q}}}{\mathbf{I} - \mathbf{q}} (y - y_0)\right) \\ &\quad \times \exp\left(\frac{1}{2}y^\dagger \frac{\mathbf{a}\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}} y\right) \exp\left(\frac{1}{2}y_0^\dagger \frac{\mathbf{a}\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}} y_0\right) \\ H(\tau, y | y_0) &= \frac{\mathbf{a}\sqrt{\mathbf{q}}}{\mathbf{I} - \mathbf{q}}(y - y_0) + \frac{\mathbf{a}\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}}y \end{aligned} \tag{24}$$

where the notation $|\cdot|$ denotes the determinant of the matrices. A technical, and notational, point is that as \mathbf{q} lies in the commutative matrix ring $\mathbb{R}[\mathbf{a}]$, we can legitimately write rational functions of \mathbf{a}, \mathbf{q} as if they were scalar indeterminates, providing the denominator is an invertible matrix. Thus $\mathbf{a}/(\mathbf{I} - \mathbf{q}), \mathbf{a}(\mathbf{I} - \mathbf{q})^{-1}, (\mathbf{I} - \mathbf{q})^{-1}\mathbf{a}$ are all equivalent. It is notable that the equation for H is, as in one dimension (6), a simple sum of two terms.

3.2. General theory

With the OU case in hand we now proceed to the non-OU case, and emphasise that although A is not linear, it has to be a conservative field; broadly we follow the approach of the univariate case, but there are technicalities associated with higher dimension. Comparing with the univariate case, the next step is to define a matrix analogue of q and also of θ . Recalling the previous discussion on Fisher information, we use the following for the multivariate analogue:

$$\mathbf{q} = \exp(-2\boldsymbol{\theta}\tau), \quad \text{and} \quad \boldsymbol{\theta} = \langle -\nabla A \rangle = \langle -\nabla \nabla \log f_\infty \rangle. \tag{25}$$

As A is conservative, its matrix $\boldsymbol{\theta}$ of partial derivatives is symmetric and is also equal the Hessian of $-\log f_\infty$. It is then immediate that $\boldsymbol{\theta}$ is positive definite because of the identity

$$-\nabla \nabla (\log \psi) = \frac{\nabla \psi}{\psi} \frac{\nabla \psi}{\psi} - \frac{\nabla \nabla \psi}{\psi};$$

setting $\psi = f_\infty$, multiplying by f_∞ and integrating over \mathbb{R}^m causes the second term to vanish, while the first is the integral of a tensor square. Analogously to the univariate case, and also (24), we adopt the ansatz

$$H(\tau, y) = \frac{\boldsymbol{\theta}\sqrt{\mathbf{q}}}{\mathbf{I} - \mathbf{q}}(y - y_0) + \frac{\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}}A(y) + \sqrt{\mathbf{q}} o(1). \tag{26}$$

The first term may be integrated immediately to give a Gaussian, but the second is not in general a conservative field, except in certain cases: (i) any one-dimensional model or composite

of one-dimensional models (by multiplying the marginals); (ii) any spherical model, i.e. one in which $f(\infty, y)$ is a function of $(y - \mu)^\dagger (y - \mu)$ for some constant vector μ . This seems to be a difficulty as, no well-defined antiderivative exists in general. Upon further investigation the ‘non-conservative’ part of this term is $O(\tau)$ because, if we write

$$B(y) = \frac{\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}} A(y),$$

then

$$\frac{\partial B_i}{\partial y_j} - \frac{\partial B_j}{\partial y_i} = O(\tau)$$

which at leading order is ignorable; it also disappears as $\tau \rightarrow \infty$.

We are then at liberty to define a function $\Omega(\tau, y)$ as

$$\Omega(\tau, y) = \exp \int_{\mu_\infty}^y dx \cdot \left(\frac{\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}} A(x) \right), \quad (27)$$

where μ_∞ is the long-term mean $\langle Y \rangle_\infty$, and the path of integration, which needs to be specified whenever the integrand is non-conservative, is a straight line. As the integral in (27) is taken along a line, its computation does not present greater difficulties as the dimension is raised; for all the examples considered later it can be evaluated in closed form. This is the main technical hurdle and we can now proceed in much the same way as the univariate case to find that

$$g(\tau, y) \sim (\dots)_{\tau, y_0} \exp \left(-\frac{1}{2} (y - y_0)^\dagger \frac{\boldsymbol{\theta} \sqrt{\mathbf{q}}}{\mathbf{I} - \mathbf{q}} (y - y_0) \right) \cdot \frac{\Omega(\tau, y_0)}{\Omega(\tau, y)}$$

and, from reciprocity, as g must be symmetric in y, y_0 we must have

$$g(\tau, y) = (\dots)_\tau \exp \left(-\frac{1}{2} (y - y_0)^\dagger \frac{\boldsymbol{\theta} \sqrt{\mathbf{q}}}{\mathbf{I} - \mathbf{q}} (y - y_0) \right) \cdot \frac{1}{\Omega(\tau, y) \Omega(\tau, y_0)}$$

so that the prefactor now depends on τ only. As $\tau \rightarrow 0$ we must have, as the density initially grows as a Gaussian, that

$$g(\tau, y_0) \sim \frac{|\boldsymbol{\theta}/2\pi|^{1/2}}{|\mathbf{I} - \mathbf{q}|^{1/2}} \cdot \frac{1}{f(\infty, y_0)}, \quad \tau \rightarrow 0,$$

while as $\tau \rightarrow \infty$ we have $g(\tau, y) \rightarrow 1$. This demands a prefactor of the form

$$\frac{1}{|\mathbf{I} - \mathbf{q}|^{1/2}} \left(\frac{|\boldsymbol{\theta}/2\pi|}{f(\infty, \mu_\infty)^2} \right)^{\rho(\tau)},$$

where $\rho(0) = \frac{1}{2}$, and $\rho(\infty) = 0$, in view of the work leading up to (9) we write

$$\rho(\tau) = \frac{1}{m} \text{tr} \frac{\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}}. \quad (28)$$

This gives our final results, the multivariate counterparts of (9) and (10), as

$$g(\tau, y) \sim \frac{1}{|\mathbf{I} - \mathbf{q}|^{1/2}} \exp \left(-\frac{1}{2} (y - y_0)^\dagger \frac{\boldsymbol{\theta} \sqrt{\mathbf{q}}}{\mathbf{I} - \mathbf{q}} (y - y_0) \right) \frac{\left(\frac{|\boldsymbol{\theta}/2\pi|}{f(\infty, \mu_\infty)^2} \right)^{\rho(\tau)}}{\Omega(\tau, y) \Omega(\tau, y_0)} \quad (29)$$

and

$$f(\tau, y) \sim \frac{1}{|\mathbf{I} - \mathbf{q}|^{1/2}} \exp\left(-\frac{1}{2}(y - y_0)^\dagger \frac{\boldsymbol{\theta}\sqrt{\mathbf{q}}}{\mathbf{I} - \mathbf{q}}(y - y_0)\right) \frac{\left(\frac{|\boldsymbol{\theta}/2\pi|}{f(\infty, \mu_\infty)^2}\right)^{\rho(\tau)} f(\infty, y)}{\Omega(\tau, y)\Omega(\tau, y_0)}. \tag{30}$$

We consider the two special cases mentioned above. First, one-dimensional models: (27) reduces to

$$\Omega(\tau, y) = \left(\frac{f(\infty, y)}{f(\infty, \mu_\infty)}\right)^{\frac{\sqrt{q}}{1+\sqrt{q}}},$$

and as A is the logarithmic derivative of the invariant density we recover (9); μ_∞ does not enter the final expression. An extension is to consider a product of independent processes, so that $\nabla A(y)$, and hence $\boldsymbol{\theta}$ and \mathbf{q} , are diagonal. This also leads to simplifications particularly if we take the reversion speeds (θ_i) to be identical, then $\boldsymbol{\theta}$ and \mathbf{q} are multiples of the identity matrix, and the multivariate approximation (30) becomes simply the product of the univariate approximations (10). Secondly, the symmetric OU case which has $A(y) = -\mathbf{a}y$, $\boldsymbol{\theta} = \mathbf{a}$, $f(\infty, \mu_\infty)^2 = |\boldsymbol{\theta}/2\pi|$, and

$$\Omega(\tau, y) = \exp\left(-\frac{1}{2}y^\dagger \frac{\mathbf{a}\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}}y\right);$$

so we end up with (24), as we should. We now take illustrative examples and evaluate the leading-order approximations comparing to numerical simulation.

3.3. Fat-tailed and double-well results

Two cases of interest are a student- t model, as a generalisation of that in section 2.1.3, and a bivariate double-well model. For the bivariate student- t model we take a general form⁹

$$A(y) = -\frac{\nu + 2}{\nu} \left(1 + \frac{a_1 y_1^2 + a_2 y_2^2}{\nu}\right)^{-1} \begin{bmatrix} a_1 y_1 \\ a_2 y_2 \end{bmatrix}$$

for which

$$f(\infty, y) = \frac{\sqrt{a_1 a_2}}{2\pi} \left(1 + \frac{a_1 y_1^2 + a_2 y_2^2}{\nu}\right)^{-\frac{\nu+2}{2}}, \quad \langle -\nabla A \rangle_\infty = \frac{\nu + 2}{\nu + 4} \begin{bmatrix} \sqrt{a_1/a_2} & 0 \\ 0 & \sqrt{a_2/a_1} \end{bmatrix}.$$

Note that, writing $\boldsymbol{\theta}$ for $\langle -\nabla A \rangle_\infty$,

$$\frac{\sqrt{\mathbf{q}}}{\mathbf{I} + \sqrt{\mathbf{q}}} = \begin{bmatrix} \frac{\exp(-\theta_{11}\tau)}{1+\exp(-\theta_{11}\tau)} & 0 \\ 0 & \frac{\exp(-\theta_{22}\tau)}{1+\exp(-\theta_{22}\tau)} \end{bmatrix} = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} \text{ say,}$$

and since $\mu_\infty = 0$ we have

$$\Omega(\tau, y) = \left(\frac{f_\infty(y)}{f_\infty(0)}\right)^{\frac{a_1 Q_1 y_1^2 + a_2 Q_2 y_2^2}{a_1 y_1^2 + a_2 y_2^2}}$$

so (30) is explicit.

⁹ While y_0 continues to denote $y(0)$ and is a vector, y_1 and y_2 are being used to denote the components of y ; this should not cause confusion.

This model is an extension in two dimensions of the student- t model shown previously, so it has the same characteristics, including the fat tails. We take a typical example, the model with $a_1 = 1, a_2 = 3$, so that the density starts off circularly-symmetric, and ends up elliptical. Two sub-cases are displayed, one with $y_0 = (-2, 2)$ (figure 6) and the other with $y_0 = (3, 1)$ (figure 7) as contour plots. It is pleasing to see that the features are accurately captured both quantitatively and qualitatively.

The bivariate double-well model is particularly challenging, and very far from the simple OU model, and we take a general model

$$A(y) = -\mathbf{a}y + \frac{2y}{\|y\|^2 + \gamma^2} - \frac{2(y - \alpha_1)}{\|y - \alpha_1\|^2 + \beta_1^2} - \frac{2(y - \alpha_2)}{\|y - \alpha_2\|^2 + \beta_2^2}$$

where $a_1, a_2 \in \mathbb{R}^2, b_1, b_2, \gamma \in \mathbb{R}$ and \mathbf{a} is a 2-by-2 symmetric matrix. Then

$$f(\infty, y) = K e^{-y^\dagger \mathbf{a} y / 2} \frac{\|y\|^2 + \gamma^2}{(\|y - \alpha_1\|^2 + \beta_1^2)(\|y - \alpha_2\|^2 + \beta_2^2)}.$$

We present cases illustrating different features:

(a) The first uses the following parameters

$$\mathbf{a} = \mathbf{I}, \quad \alpha_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \beta_1 = 1, \quad \alpha_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \beta_2 = 1, \quad \gamma = \frac{1}{2}$$

for which we compute numerically

$$\boldsymbol{\theta} \approx \begin{bmatrix} 1.2633 & 0 \\ 0 & 1.2774 \end{bmatrix}, \quad K \approx 2.5352.$$

Similarly to the 1D case, α controls the position, β the depth and γ makes the wells more disjoint as it goes towards 0. Hence the wells are quite disjoint and located on the y_1 -axis; we consider two sub-cases with different starting points.

In the first one, $y_0 = (0, 0.5)$ is equidistant from the two wells and the resulting field evolution is illustrated in figure 8. The diffusion takes place at the same pace towards both wells and, even though we notice that the numerical solution converges slightly faster than the approximation, the agreement is reassuring.

The second one is the extreme case when it starts at $y_0 = (-1.5, 0)$ corresponding to the bottom of a well and its evolution is shown in figure 9. As could be expected since the wells are well-separated, the approximation struggles to capture the medium-term behaviour.

(b) Finally we consider

$$\mathbf{a} = \mathbf{I}, \quad \alpha_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \beta_1 = 1, \quad \alpha_2 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \quad \beta_2 = 0.7, \quad \gamma = 1$$

for which

$$\boldsymbol{\theta} \approx \begin{bmatrix} 0.9957 & -0.1990 \\ -0.1990 & 0.9957 \end{bmatrix}, \quad K \approx 4.0767.$$

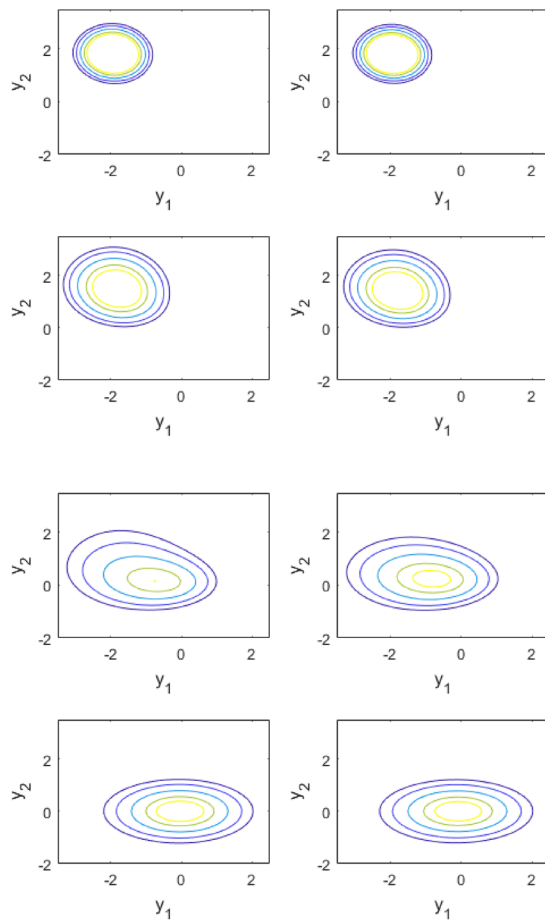


Figure 6. Bivariate student- t example (see text). PDE solver on the left, leading-order approximation on the right, $y_0 = (-2, 2)$. Densities at several different time points, from top to bottom: $\tau = 0.1, 0.25, 1, 5$.

In this example, due to the symmetry of α , the wells are located on the $y_1 = y_2$ line, and they are less separated because γ is bigger. However the main difference is that $\beta_1 \neq \beta_2$ which implies that the wells have different depths. Again we present two sub-cases.

The first one starts equidistant from the wells with $y_0 = (-1, 1)$ and is illustrated in figure 10. Once more we observe that the rate of convergence is the same towards both wells and that the exact and leading-order solutions have a very similar behaviour even though the starting point is further away from the wells than in case (a).

The second sub-case starts at the bottom of a well which is represented by $y_0 = (1.3, 1.3)$ and is shown in figure 11. As with case (a), the medium-term approximation diverges from the exact solution.

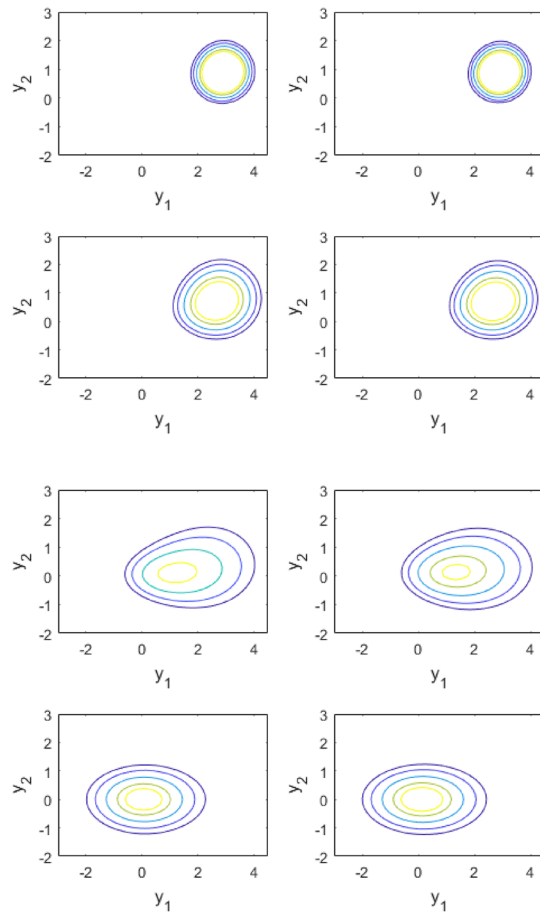


Figure 7. Bivariate student- t example (see text). PDE solver on the left, leading-order approximation on the right, $y_0 = (3, 1)$. Densities at several different time points, from top to bottom: $\tau = 0.1, 0.25, 1, 5$.

4. Extensions

We have presented typical examples and demonstrated, via comparison with numerical simulation that the approximation is encouragingly capturing behaviour far from the OU case, despite being based, in some sense, around the OU. We now highlight three extensions: firstly, what happens when we big far from equilibrium, then half-line problems for which our existing theory is inconvenient, and then finally what happens if the field is non-conservative.

4.1. Far-field expansion

The method is less accurate when the starting-point is a long way from equilibrium, and more concretely this means

$$|A(y)| \ll \theta|y - \mu_\infty|$$

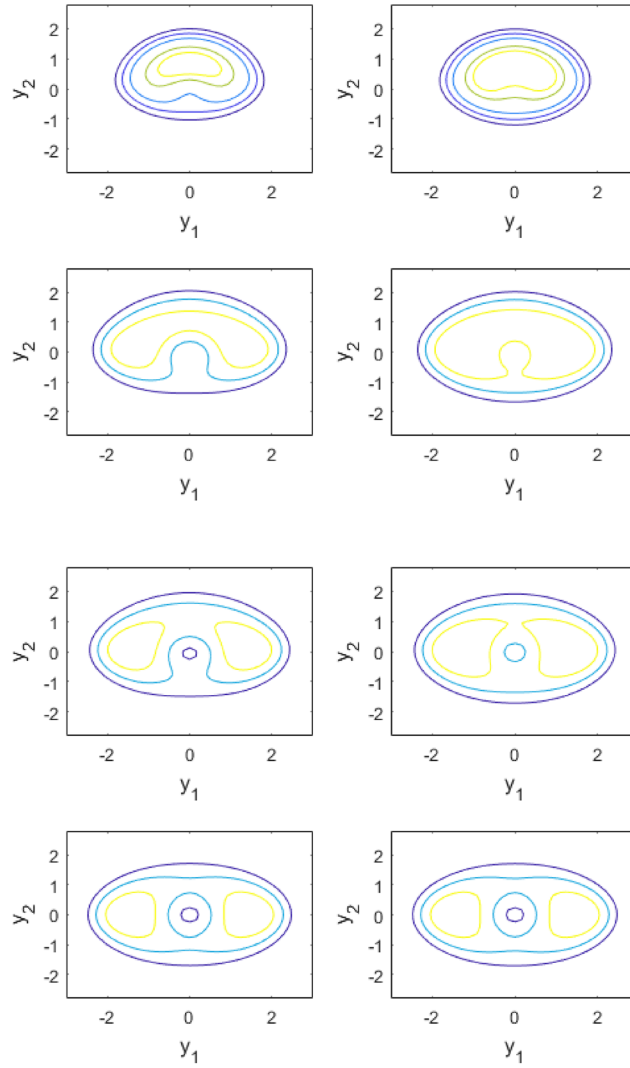


Figure 8. Two dimensional double-well example (a) (see text). PDE solver on the left, leading-order approximation on the right, $y_0 = (0, 0.5)$. Densities at several different time points, from top to bottom: $\tau = 0.3, 1, 1.5, 5$.

in one dimension. At one level this represents the deviation from the chosen OU model. However it can also be thought of as follows: $|y - \mu_\infty|/A(y)$ is, loosely, the time taken to get back to equilibrium, and if this is much larger than $1/\theta$, which is the reciprocal of the average speed of attraction, then we are in the far field—hence the above heuristic.

If we redevelop all the theory assuming no mean reversion (and there is little to be gained by staying in one dimension), we have, from (22),

$$H(\tau, y) \sim \frac{y - y_0}{2\tau} + \frac{A(y)}{2} + \nabla B_1(y)\tau + \dots, \quad \tau \rightarrow 0, \quad (31)$$

where we have replaced $b_1(y)$ by the gradient of a function B_1 . Equating the terms at $O(\tau^0)$ in (22) gives

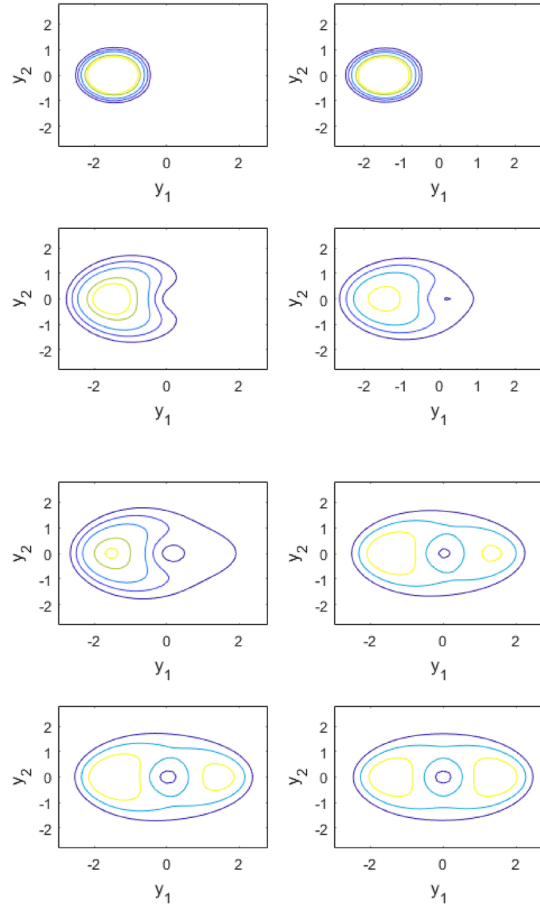


Figure 9. Two dimensional double-well example (a) (see text). PDE solver on the left, leading-order approximation on the right, $y_0 = (-1.5, 0)$. Densities at several different time points, from top to bottom: $\tau = 0.1, 0.8, 2, 5$.

$$B_1 + (y - y_0) \cdot \nabla B_1 = \frac{\|A\|^2}{4} + \frac{\nabla \cdot A}{2}.$$

In one dimension, this gives

$$B_1(y) = \frac{1}{y - y_0} \int_{y_0}^y \left(\frac{A(z)^2}{4} + \frac{A'(z)}{2} \right) dz$$

which is the average of the integrand, and hence approximated by the average of the values at the endpoints. If A is slowly-varying then, going back to the higher-dimensional case,

$$B_1(y) \approx \frac{1}{8} (\|A(y)\|^2 + \|A(y_0)\|^2).$$

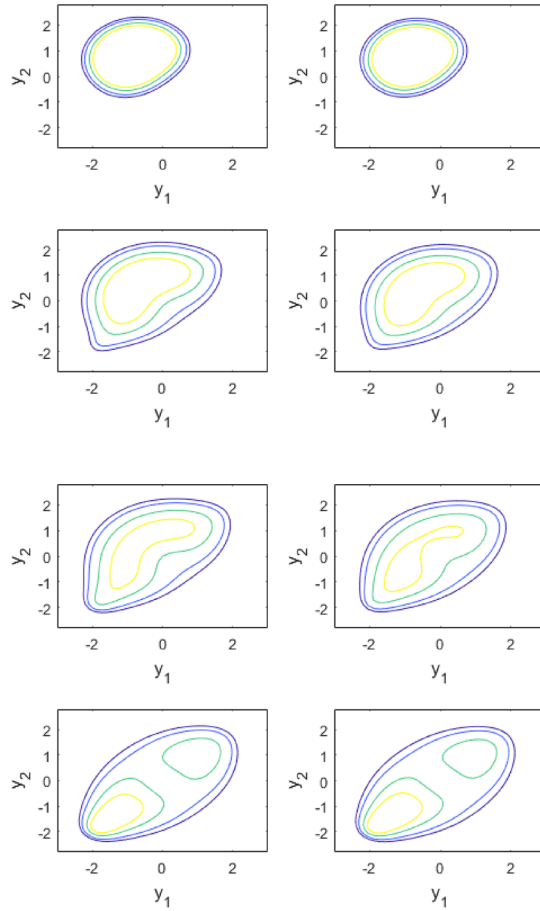


Figure 10. Two dimensional double-well example (b) (see text). PDE solver on the left, leading-order approximation on the right, $y_0 = (-1, 1)$. Densities at several different time points, from top to bottom: $\tau = 0.3, 1, 2, 5$.

Also, by the same arguments as in section 2, writing $g(\tau, y) = n(\tau)e^{-f h}$,

$$n(\tau) \sim \frac{e^{-(\|A(y_0)\|^2/4 + \nabla \cdot A(y_0)/2)\tau}}{\sqrt{4\pi\tau}}$$

and combining the results gives (again ignoring the $\nabla \cdot A$ term)

$$g(\tau, y) \approx \frac{e^{-\|y-y_0\|^2/4\tau} e^{-(\|A(y)\|^2 + \|A(y_0)\|^2)\tau/8}}{\sqrt{4\pi\tau f(\infty, y) f(\infty, y_0)}}, \quad \tau \rightarrow 0 \tag{32}$$

or

$$f(\tau, y) \approx \frac{e^{-\|y-y_0\|^2/4\tau} e^{-(\|A(y)\|^2 + \|A(y_0)\|^2)\tau/8}}{\sqrt{4\pi\tau}} \left(\frac{f(\infty, y)}{f(\infty, y_0)} \right)^{1/2}. \tag{33}$$

Now let us reconsider the one-dimensional dry-friction case, which we reproduce for convenience:

$$g(\tau, y | y_0) = \frac{e^{-(y-y_0)^2/4\tau}}{\sqrt{\pi\tau}} e^{-\tau/4} e^{(|y_0|+|y|)/2} + \Phi\left(\frac{\tau - |y| - |y_0|}{\sqrt{2\tau}}\right).$$

For $|y| + |y_0| > \tau$ (a V-shaped domain, in a plot of y versus τ) the first term predominates, and in fact is identical to what we have just derived in (32) as $\|A\| = 1$. This is understood as the far-field solution. Inside the ‘V’, i.e. $|y| + |y_0| < \tau$, the second term takes over; this is the term that governs the long-time limit, and the first term decays to zero then. It is elegant that these two halves combine to give the exact solution.

The basic principle is that (32) is a generic way of understanding the far-field behaviour of equations of non-OU type in the far-field, where the drift is small: for example, when $A(y) = -a \tanh \hat{\gamma} y$ or $-y/(1 + \hat{\gamma}^2 y^2)$ as previously studied. However, extending (31) and (32) to a complete solution that works in all régimes does not seem to be straightforward—there is no exponential damping factor in (31), so the expression is clearly wrong as $\tau \rightarrow \infty$. This is a matter for further research.

4.2. Square-root process

Another new branch of the theory concerns processes that are bound to lie in the half-line by reason of the volatility decaying to zero at some point, without loss of generality $X = 0$. (This is as opposed to a reflecting boundary condition, which is another way of constraining a process; we do not consider that case here, but doubtless it could be.) A famous example of this is the so-called square root process, described in, for example, [12, 32]:

$$dX_t = a(b - X_t) dt + \sigma \sqrt{X_t} dW_t, \quad X \geq 0.$$

It is unhelpful to transform this into (2) because the drift A will become infinite at the origin: $1/\sigma_X$ becomes undefined. The problem can be rescaled by setting $Y = 2aX/\sigma^2$ and $\tau = at$, giving

$$f(\infty, y) = \frac{y^{\nu-1} e^{-y}}{\Gamma(\nu)}, \quad \nu = 2ab/\sigma^2$$

and

$$f(t, y | y_0) = \frac{e^{-y}}{1 - e^{-\tau}} \left(\frac{ye^\tau}{y_0}\right)^{(\nu-1)/2} \exp\left(-\frac{(y+y_0)e^{-\tau}}{1 - e^{-\tau}}\right) I_{\nu-1}\left(\frac{2\sqrt{yy_0}e^{-\tau}}{1 - e^{-\tau}}\right)$$

$$g(t, y | y_0) = \frac{\Gamma(\nu)}{1 - e^{-\tau}} \left(\frac{e^\tau}{yy_0}\right)^{(\nu-1)/2} \exp\left(-\frac{(y+y_0)e^{-\tau}}{1 - e^{-\tau}}\right) I_{\nu-1}\left(\frac{2\sqrt{yy_0}e^{-\tau}}{1 - e^{-\tau}}\right)$$

where $I_\nu(z)$ denotes the modified Bessel function of the first kind. We have for short time

$$h(\tau, y) = \frac{e^{-\tau/2}(1 - \sqrt{y_0/y})}{1 - e^{-\tau}} - \frac{e^{-\tau/2}}{1 + e^{-\tau/2}} + \frac{\nu - \frac{1}{2}}{2y} + o(1), \quad \tau \rightarrow 0$$

where we have used $I_{\nu-1}(z) \sim e^z/\sqrt{2\pi z}$ as $z \rightarrow \infty$. We can replace $\frac{\nu - \frac{1}{2}}{2y}$ with $\frac{e^{-\tau/2}}{1 + e^{-\tau/2}} \frac{\nu - \frac{1}{2}}{y}$, thereby keeping an error term of the same order, $O(\tau)$, but ensuring $h(\infty, y) = 0$. This gives

$$h(\tau, y) = \frac{e^{-\tau/2}(1 - \sqrt{y_0/y})}{1 - e^{-\tau}} + \frac{e^{-\tau/2}}{1 + e^{-\tau/2}} \left(-1 + \frac{\nu - \frac{1}{2}}{y}\right) + o(1), \quad (34)$$

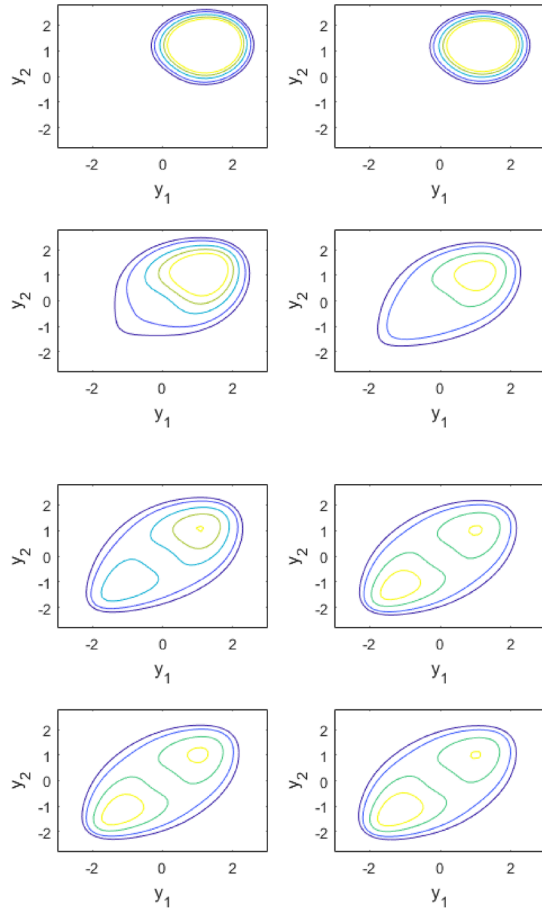


Figure 11. Two dimensional double-well example (b) (see text). PDE solver on the left, leading-order approximation on the right, $y_0 = (1.3, 1.3)$. Densities at several different time points, from top to bottom: $\tau = 0.3, 2, 5, 10$.

which is an analogue of (11) for generalised square-root processes. The y -dependence is different, but there are obvious parallels, notably the $\frac{e^{-\tau/2}}{1-e^{-\tau}}$ in the first term.

The expansion can also be derived directly, if we define the normal form of a process on the half-line as

$$dY_t = \kappa A(Y_t) dt + \sqrt{2\kappa Y_t} dW_t \tag{35}$$

from which

$$\frac{\partial h}{\partial \tau} = \frac{\partial}{\partial y} \left\{ A(y) h + y \frac{\partial h}{\partial y} - y h^2 \right\}. \tag{36}$$

In the expansion of h , the first term must, by dominant balance of the $\partial h / \partial \tau$ and $\partial(yh^2) / \partial y$ terms, look like

$$h(\tau, y) \sim \frac{1 - \sqrt{c/y}}{\tau}, \quad \tau \rightarrow 0;$$

the constant c must equal y_0 , by consideration of the initial behaviour, and to ensure $h(\infty, y) = 0$ we write

$$h(\tau, y) \sim \frac{2\theta\sqrt{q}(1 - \sqrt{y_0/y})}{1 - q} + \dots$$

as the first term, where $\theta > 0$ as before is arbitrary. The next term is obtained, as before, by requiring that the $O(\tau^{-1})$ terms balance in (36) and we find

$$h(\tau, y) \sim \frac{2\theta\sqrt{q}(1 - \sqrt{y_0/y})}{1 - q} + \frac{\sqrt{q}}{1 + \sqrt{q}} \left(\frac{A(y)}{y} - \frac{1}{2y} \right) + \dots \quad (37)$$

where the last term ($-1/2y$) is needed to balance the $(yh)'$ term in (36). (No such term appears in (7), because the relevant term in (5) is h'' , which vanishes at leading order in τ .) This gives the same result as the square-root process, for which $A(y) = \nu - y$ in normal form, provided we set $\theta = \frac{1}{2}$.

Similar considerations apply to doubly-bounded processes e.g.

$$dX_t = (a - bX_t) dt + \sigma\sqrt{1 - X_t^2} dW_t, \quad X \in [-1, 1].$$

Both of these are examples of processes identified by Wong [49], where the usual method of solution is orthogonal expansion using the Hermite, Laguerre or Jacobi polynomials.

There is also, as intimated above, the possibility of studying processes with one or two reflecting barriers.

4.3. Non-conservative problems

The theory of diffusion in the presence of a non-conservative force field is considerably more difficult. Even the OU case is not straightforward, but it can be related to its symmetric case as follows. To retain the general setting let us write the SDE as

$$dX_t = -\kappa \mathbf{a}_X dt + \mathbf{b}_X dW_t$$

with $\mathbf{a}_X, \mathbf{b}_X$ square matrices of dimension m . Writing

$$Y = \sqrt{2\kappa} \mathbf{b}_X^{-1} X, \quad \mathbf{a} = \kappa^{-1} \mathbf{b}_X^{-1} \mathbf{a}_X \mathbf{b}_X,$$

we convert the SDE to its normal form (23), but \mathbf{a} might not be symmetric. The invariant density is still multivariate Gaussian with zero mean but the covariance matrix σ_∞ is not \mathbf{a}^{-1} : instead it is given by the Lyapunov equation,

$$\mathbf{a}\sigma_\infty + \sigma_\infty\mathbf{a}^\dagger = 2\mathbf{I}$$

which is most easily solved for σ_∞ by writing it as a set of linear equations in its elements. Also

$$f(\tau, y | y_0) = \frac{1}{|2\pi\sigma(\tau)|^{1/2}} \exp\left(\frac{-(y - \mu(\tau))^\dagger \sigma(\tau)^{-1} (y - \mu(\tau))}{2}\right)$$

with

$$\mu(\tau) = e^{-a\tau} y_0, \quad \sigma(\tau) = 2 \int_0^\tau e^{-as} e^{-a^\dagger s} ds = \sigma_\infty - e^{-a\tau} \sigma_\infty e^{-a^\dagger \tau}.$$

From this, with $H = -\nabla \log(f/f_\infty)$ as before, we can see by direct calculation that

$$\begin{aligned} H(\tau, y | y_0) &= (\sigma(\tau)^{-1} - \sigma_\infty^{-1})y - \sigma(\tau)^{-1}\mu(\tau) \\ &= (e^{a\tau}\sigma_\infty - \sigma_\infty e^{-a^\dagger\tau})^{-1}(y - y_0) + (e^{a\tau}\sigma_\infty - \sigma_\infty e^{-a^\dagger\tau})^{-1}\sigma_\infty(e^{-a^\dagger\tau} - 1)\sigma_\infty^{-1}y. \end{aligned}$$

The first term, which is singular at $\tau = 0$, corresponds to the first term in (7) and (26), if we accept the substitutions

$$\frac{\theta\sqrt{q}}{1-q} \rightsquigarrow (e^{a\tau}\sigma_\infty - \sigma_\infty e^{-a^\dagger\tau})^{-1}, \quad \theta \rightsquigarrow \sigma_\infty^{-1}$$

(notionally, $\sqrt{q} \rightsquigarrow e^{-a\tau}$ or $e^{-a^\dagger\tau}$), while the second term corresponds similarly, from the substitutions

$$\frac{\sqrt{q}}{1+\sqrt{q}} = \frac{\theta\sqrt{q}}{1-q} \frac{1-\sqrt{q}}{\theta} \rightsquigarrow (e^{a\tau}\sigma_\infty - \sigma_\infty e^{-a^\dagger\tau})^{-1}\sigma_\infty(1 - e^{-a^\dagger\tau})$$

and

$$A(y) \rightsquigarrow -\sigma_\infty^{-1}y.$$

Therefore, in some altered way, (7) and (26) carry over.

This brings us on to another matter: our work concerns approximating the FP solution using the short- and long-term behaviour. Is it possible to find two OU models with the same short- and long-term behaviour, even if the medium-term behaviour is different? The answer is yes: indeed, all generators of the form

$$\mathbf{a} = (\mathbf{I} + \mathbf{u})\sigma_\infty^{-1}, \quad \mathbf{u} \in \mathfrak{A}$$

where \mathfrak{A} is the space of skew-symmetric matrices, give rise to the same behaviour in both limits. Formally, define two generators \mathbf{a} to be equivalent (notation \simeq and clearly an equivalence relation) if they give rise to the same asymptotic covariance matrix. It is easy to see that two equivalent generators have the same trace.

As an example: when $\mathbf{a} = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}$ we have

$$\sigma(t) = \begin{bmatrix} 1 - e^{-2t} + \frac{1}{2}a^2(1 - (1 + 2t + 2t^2)e^{-2t}) & -\frac{1}{2}a(1 - (1 + 2t)e^{-2t}) \\ -\frac{1}{2}a(1 - (1 + 2t)e^{-2t}) & 1 - e^{-2t} \end{bmatrix}$$

and

$$\sigma_\infty = \begin{bmatrix} 1 + a^2/2 & -a/2 \\ -a/2 & 1 \end{bmatrix}, \quad \sigma_\infty^{-1} = \frac{1}{1 + a^2/4} \begin{bmatrix} 1 & a/2 \\ a/2 & 1 + a^2/2 \end{bmatrix}. \quad (38)$$

So the equivalence class of \mathbf{a} under \simeq is

$$[\mathbf{a}]_\simeq = \frac{1}{1 + a^2/4} \begin{bmatrix} 1 - ab/2 & a/2 - b - a^2b/2 \\ b + a/2 & 1 + ab/2 + a^2/2 \end{bmatrix}, \quad b \in \mathbb{R}$$

which contains the following elements, as it must:

$$\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}; \quad \frac{1}{1 + a^2/4} \begin{bmatrix} 1 & a/2 \\ a/2 & 1 + a^2/2 \end{bmatrix}.$$

All elements of $[\mathbf{a}]_\simeq$ have trace 2.

In summary, every OU process (or generator) is equivalent under \simeq to a unique symmetric one; we obtain σ_∞ from the Lyapunov equation and then invert it to obtain the symmetric generator.

5. Conclusions and final remarks

We have described the solution to the Fokker–Planck equation with steady state in simple, intuitive terms and demonstrated the validity of our approach with numerical examples in a range of cases: The main results are (10) and its multidimensional analogue (30). The most striking, and potentially most useful, conclusion is that even a simple expansion without the intermediate correction terms—the ‘leading-order’ expansion—produces acceptable results for the great majority of cases, in what may be described as the ‘central zone’ where the process spends most of its time. Perhaps surprisingly it continues to work well even if the departure from the OU model is quite gross, such as for the double-well potentials that we consider. Given the explicit nature of the formulae we provide, they are very fast to compute by comparison with a PDE solver or Monte Carlo simulation, especially in higher dimensions. Hence we anticipate that our approach can be utilised to form the core of, say, a Kalman filter avoiding linearisation or high-dimensional computation.

We have also indicated in section 4 other developments that deal with extensions and important side-issues. These are: the far-field expansion, where the core result works less effectively; problems constrained to lie in the half-line or channel; and also the difficult case of when the force-field is non-conservative, i.e. not arising from a potential field.

On the more theoretical side there remains the open question of whether (7) converges in a neighbourhood of $\tau = 0$, or whether it is simply an asymptotic expansion for small τ that eventually blows up if too many terms are taken. The empirical evidence of [36] is that it is convergent when A is analytic, but this is far from clear. There is also the possibility, as indicated in [36], of using spectral methods to produce a higher-order expansion, rather than developing in a power series in $1 - e^{-2\theta\tau}$ or $1 - e^{-\theta\tau}$. Another avenue is attempting to combine the far-field expansion with the near-equilibrium expansion, essentially by allowing θ to be lower when y is a long way from equilibrium. If this is to be done then θ must be made symmetric in y and y_0 so as to preserve the reciprocity condition (8).

This paper is, therefore, not the last word on the subject, but provides opportunities for further work in an area that we consider still to be fertile.

Acknowledgments

RVC would like to thank the Leverhulme Trust for their support through a Research Fellowship.

ORCID iDs

R J Martin  <https://orcid.org/0000-0002-9272-1360>

R V Craster  <https://orcid.org/0000-0001-9799-9639>

A Pannier  <https://orcid.org/0000-0002-7328-7482>

M J Kearney  <https://orcid.org/0000-0002-9085-8638>

References

- [1] Alfarano S, Lux T and Wagner F 2005 Estimation of agent-based models: the case of an asymmetric herding model *Comput. Econ.* **26** 19–49
- [2] Boyd J P 2001 *Chebyshev and Fourier Spectral Methods* (New York: Dover)
- [3] Brics M, Kaupužs J and Mahnke R 2013 How to solve Fokker–Planck equation treating mixed eigenvalue spectrum? *Condens. Matter Phys.* **16** 1–13
- [4] Buch-larsen T, Nielsen J P, Guillén M and Bolancé C 2005 Kernel density estimation for heavy-tailed distributions using the champernowne transformation *J. Theor. Appl. Stat.* **39** 503–16
- [5] Burke T P 2016 Kernel density estimation techniques for Monte Carlo reactor analysis *PhD Dissertation* University of Michigan
- [6] Canuto C, Hussaini M Y, Quarteroni A and Zang T A 2006 *Spectral Methods: Fundamentals in Single Domains* (Berlin: Springer)
- [7] Carro A, Toral R and Miguel M S 2015 Markets, herding and response to external information *PLoS One* **10** e0133287
- [8] Caroli B, Caroli C and Roulet B 1979 Diffusion in a bistable potential: a systematic WKB treatment *J. Stat. Phys.* **21** 415
- [9] Cervera-Villanueva A *et al* 2002 Kalman filter tracking and vertexing in a silicon detector for neutrino physics *Nucl. Instrum. Methods Phys. Res. A* **486** 639–62
- [10] Chen N and Majda A J 2017 Beating the curse of dimension with accurate statistics for the Fokker–Planck equation in complex turbulent systems *Proc. Natl Acad. Sci.* **114** 12864–9
- [11] Coffey W T, Kalmykov Y P and Waldron J T 2003 *The Langevin Equation: with Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering* (Singapore: World Scientific)
- [12] Cox J C, Ingersoll J E and Ross S A 1985 A theory of the term structure of interest rates *Econometrica* **53** 385–407
- [13] Devroye L and Györfi L 1985 *Nonparametric Density Estimation: the L_1 View* (New York: Wiley)
- [14] Ditlevsen P D 1999 Observation of α -stable noise-induced millennial climate changes from an ice-core record *Geophys. Res. Lett.* **26** 1441–4
- [15] Doz C, Giannone D and Reichlin L 2011 A two-step estimator for large approximate dynamic factor models based on Kalman filtering *J. Econometr.* **164** 188–205
- [16] Dubkov A A, Hänggi P and Goychuk I 2009 Non-linear Brownian motion: the problem of obtaining the thermal Langevin equation for a non-Gaussian bath *J. Stat. Mech.* P01034
- [17] Dybiec B, Gudowska-Nowak E and Hänggi P 2007 Escape driven by α -stable white noise *Phys. Rev. E* **75** 021109
- [18] Einicke G A and White L B 1999 Robust extended Kalman filtering *IEEE Trans. Signal Process.* **47** 2596–9
- [19] Escande D F and Sattin F 2007 When can the Fokker–Planck equation describe anomalous or chaotic transport? *Phys. Rev. Lett.* **99** 185005
- [20] Flugge S 1971 *Practical Quantum Mechanics I* (Berlin: Springer)
- [21] Forman J L and Sørensen M 2008 The Pearson diffusions: a class of statistically tractable diffusion processes *Scand. J. Stat.* **35** 438–65
- [22] Friedrich R and Peinke J 1997 Description of a turbulent cascade by a Fokker–Planck equation *Phys. Rev. Lett.* **78** 863
- [23] Friedrich R, Peinke J and Renner C 2000 How to quantify deterministic and random influences on the statistics of the foreign exchange market *Phys. Rev. Lett.* **84** 5224
- [24] Ghasemi F, Sahimi M, Peinke J and Reza Rahimi Tabar M 2006 Analysis of non-stationary data for heart-rate fluctuations in terms of drift and diffusion coefficients *J. Biol. Phys.* **32** 117–28
- [25] Gihman I and Skorohod A 1972 *Stochastic Differential Equations* (Heidelberg: Springer)
- [26] Grewal M S and Andrews A P 1993 *Kalman Filtering Theory and Practice* (Englewood Cliffs, NJ: Prentice-Hall)
- [27] Harvey A C 1989 *Forecasting, Structural Time Series and the Kalman Filter* (Cambridge: Cambridge University Press)
- [28] Jones M C, Marron J S and Sheather S J 1996 A brief survey of bandwidth selection for density estimation *J. Am. Stat. Assoc.* **91** 401–7
- [29] Kirman A 1993 Ants, rationality and recruitment *Q. J. Econ.* **108** 137–56
- [30] Kelly F P 1979 *Reversibility and Stochastic Networks* (Chichester: Wiley)

- [31] Klimontovich Y L 1994 Nonlinear Brownian motion *Phys.—Usp.* **37** 737–67
- [32] Lamberton D and Lapeyre B 2012 *Introduction au Calcul Stochastique Appliqué à la Finance* (Paris: Ellipses)
- [33] Lehmann E L and Casella G 1998 *Theory of Point Estimation* (Berlin: Springer)
- [34] Lefever R and Horsthemke R 1974 *Noise Induced Transitions* (Heidelberg: Springer)
- [35] Liu S C 1969 Solutions of Fokker–Planck equation with applications in nonlinear random vibration *Bell Syst. Tech. J.* **48** 2031–51
- [36] Martin R J, Kearney M J and Craster R V 2015 Infinite product expansion of the Fokker–Planck equation with steady-state solution *Proc. R. Soc. A* **471** 2179
- [37] Maybeck P S 1979 *Stochastic Models, Estimation, and Control* (New York: Academic)
- [38] Mehler F G 1866 Über die entwicklung einer funktion von beliebig vielen variablen nach Laplaceschen funktionen hoherer ordnung *J. Reine Angew. Math.* **566** 161–76
- [39] Moeller J K and Madsen H 2010 *From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform (IMM-Technical Report-2010-16)* (Kgs. Lyngby: Technical University of Denmark)
- [40] Risken H 1989 *The Fokker–Planck Equation: Method of Solution and Applications* (New York: Springer)
- [41] Ruess J, Miliias-Argeitis A, Summers S and Lygeros J 2011 Moment estimation for chemically reacting systems by extended Kalman filtering *J. Chem. Phys.* **135** 165102
- [42] Scott D W 1992 *Multivariate Density Estimation: Theory, Practice, and Visualization* (New York: Wiley)
- [43] Scherliess L, Schunk R W, Sojka J J and Thompson D C 2004 Development of a physics-based reduced state Kalman filter for the ionosphere *Radio Sci.* **39** RS1S04
- [44] Silverman B W 1986 *Density Estimation for Statistics and Data Analysis* (London: Chapman and Hall)
- [45] Stresing R, Peinke J, Seoud R E and Vassilicos J C 2010 Defining a new class of turbulent flows *Phys. Rev. Lett.* **104** 194501
- [46] Taranov K 2015 Rigid body dynamics and Kalman filtering *Master’s Thesis* Lappeenranta U. of Technology
- [47] Touchette H, Van der Straeten E and Just W 2010 Brownian motion with dry friction: Fokker–Planck approach *J. Phys. A: Math. Theor.* **43** 445002
- [48] Trefethen L N 2000 *Spectral Methods in Matlab* (Philadelphia, PA: SIAM)
- [49] Wong E 1964 The construction of a class of stationary Markov processes *16th Symp. in Applied Mathematics—Stochastic Processes in Mathematical Physics and Engineering (Providence, RI)* pp 264–76