Variational schemes in the Fokker - Planck equation

1996 J. Phys. A: Math. Gen. 29 1859

(http://iopscience.iop.org/0305-4470/29/9/003)

J. Phys. A: Math. Gen. **29** (1996) 1859–1872. Printed in the UK

Variational schemes in the Fokker–Planck equation

T Blum and A J McKane

Department of Theoretical Physics, University of Manchester, Manchester M13 9PL, UK

Received 30 June 1995, in final form 28 September 1995

Abstract. We investigate variational methods for finding approximate solutions to the Fokker–Planck equation, especially in cases lacking detailed balance. These schemes fall into two classes: those in which a Hermitian operator is constructed from the (non-Hermitian) Fokker–Planck operator, and those which are based on soluble approximations to this operator. The various approaches are tested on a simple quantum-mechanical problem, a toy Fokker–Planck equation and finally on the problem of a particle moving in a potential and subject to external non-white noise.

1. Introduction

Variational schemes, such as the Rayleigh–Ritz procedure [1] are well known methods for finding approximate solutions to the Schrödinger equation. There has been far less work done in applying analogous procedures to the Fokker–Planck equation, even though the equations share a similar structure. One of the reasons for this situation is that the differential operator in the Fokker–Planck equation is not, in general, self-adjoint, making the formalism more complicated: eigenvalues are not necessarily real, nor are the right and left eigenfunctions equal. Another reason is the existence of a zero eigenvalue, corresponding to the steady state of the system. Under certain conditions (the 'potential conditions') the determination of the steady-state probability distribution function reduces to quadratures. Since all eigenvalues have a non-negative real part [2], in this case the lowest eigenvalue and eigenfunction are known exactly, and approximation schemes only have to be developed for 'excited' states. The formalism for this has been developed and applied to a number of problems [2, 3]. However, for many systems of interest the potential conditions do not hold, and as a consequence the steady-state distribution cannot be determined in closed form. It is desirable to have variational schemes in this case as well. Some work has been carried out by Seybold [4], but no systematic discussion exists in the literature. The purpose of this paper is to investigate the usefulness of a number of variational schemes in the case where detailed balance does not hold. These are illustrated on simple two-dimensional systems, which are the simplest systems for which potential conditions need not hold.

The schemes we consider are of two types. In the first type a Hermitian operator is constructed from the (non-Hermitian) Fokker–Planck operator, and one then proceeds as in variational calculation in quantum mechanics. The second type can be called 'variationally improved perturbation theory'; they are based on a soluble approximation to the Fokker–Planck operator and the perturbation theory based there on. Both are viable options when attacking a problem with a finite number of degrees of freedom as will be illustrated by the examples provided at the end of this paper. The second type comes into its own when dealing with a *functional* Fokker–Planck equation (as the first type is generally plagued

0305-4470/96/091859+14$19.50

with divergence problems). In a separate work [5] we have applied variationally improved perturbation theory to the Fokker–Planck formulation of the Kardar–Parisi–Zhang equation.

The outline of this paper is as follows. In section two we review some general formalism for the Fokker–Planck equation and in addition discuss the potential conditions and show how they lead to an exact solution for the stationary probability distribution. Section three introduces the variational methods that are the heart of this work. In section four, we consider some examples of these schemes in action.

2. Formalism

In this section we first present some of the formalism associated with the Fokker–Planck equation and then treat the potential conditions, demonstrating how they lead to an exact solution for the steady-state probability distribution. Here, we will see that making the Fokker–Planck equation look as much as possible like quantum mechanics is very natural.

2.1. General formalism

The Fokker–Planck equation for a system with N degrees of freedom is [2, 6]

$$\partial_t\, p(\boldsymbol{x}, t) = L\, p(\boldsymbol{x}, t) \tag{2.1}$$

where L is a differential operator of the form

$$L = -\partial_i A_i(\boldsymbol{x}) + \partial_i \partial_j B_{ij}(\boldsymbol{x}) \tag{2.2}$$

where the summation convention is understood. Here $A_i(\boldsymbol{x})$ and $B_{ij}(\boldsymbol{x}) = B_{ji}(\boldsymbol{x})$ are the drift vector and the diffusion matrix respectively, ∂_i means $\partial/\partial x_i$ and $i, j = 1, 2, \ldots, N$. Looking for separable solutions of the form

$$p(\boldsymbol{x}, t) = P_n(\boldsymbol{x})\, \exp\{-\lambda_n t\} \tag{2.3}$$

leads to the eigenfunction equation

$$L\, P_n(\boldsymbol{x}) = -\lambda_n\, P_n(\boldsymbol{x}) \tag{2.4}$$

where we have assumed for convenience a discrete eigenvalue spectrum with $n = 0, 1, 2, \ldots$. Since L cannot, in general, be brought into Hermitian form, the eigenfunctions $Q_m(\boldsymbol{x})$ of the adjoint operator have to be found as well:

$$L^\dagger\, Q_m(\boldsymbol{x}) = -\lambda_m\, Q_m(\boldsymbol{x})\,. \tag{2.5}$$

The set of eigenvalues in (2.4) and (2.5) are equal; moreover, since the operator is real, the complex conjugate of an eigenvalue is also an eigenvalue. The set of functions $P_n(\boldsymbol{x})$ and $Q_m(\boldsymbol{x})$ are bi-orthogonal:

$$\int \mathrm{d}^N \boldsymbol{x}\, P_n(\boldsymbol{x})\, Q_m(\boldsymbol{x}) = \delta_{n,m}\,. \tag{2.6}$$

Equation (2.3) shows that the stationary state corresponds to $\lambda_0 = 0$, i.e. $P_0(\boldsymbol{x}) = p_{st}(\boldsymbol{x})$. From the form of L we see that $Q_0(\boldsymbol{x})$ is a constant, and then from (2.6) that $Q_0(\boldsymbol{x}) = 1$. Henceforth, we will frequently denote operations such as those found in equations (2.4), (2.5) and (2.6) in the following notation:

$$L\,|P_n\rangle = -\lambda_n\,|P_n\rangle \qquad \langle Q_m|\,L = -\lambda_m\,\langle Q_m| \qquad \langle Q_m|P_n\rangle = \delta_{n,m}\,. \tag{2.7}$$

2.2. The potential conditions

The study of the Fokker–Planck equation simplifies considerably if the drift vector $A_i(\boldsymbol{x})$ and the diffusion matrix $B_{ij}(\boldsymbol{x})$ satisfy the so-called potential conditions mentioned in section 1. There are various ways to introduce these conditions, but perhaps the simplest is first to note that the Fokker–Planck equation may be written in the form of a continuity equation, $\partial_t p(\boldsymbol{x}, t) + \partial_i J_i(\boldsymbol{x}, t) = 0$, where J_i is

$$J_i(\boldsymbol{x}) = \left[A_i(\boldsymbol{x}) - \partial_j B_{ij}(\boldsymbol{x})\right] p(\boldsymbol{x}, t) \tag{2.8}$$

and called the probability current. Stationarity implies $\nabla \cdot \boldsymbol{J} = 0$, but if it is also true that $\boldsymbol{J} = \boldsymbol{0}$, then

$$\partial_j \left[B_{ij}(\boldsymbol{x})\, p_{\rm st}(\boldsymbol{x})\right] = A_i(\boldsymbol{x})\, p_{\rm st}(\boldsymbol{x})\,. \tag{2.9}$$

(It turns out that an equivalent statement is $P_n(\boldsymbol{x}) = p_{\rm st}(\boldsymbol{x}) Q_n(\boldsymbol{x})$ for all n.) If in addition the diffusion matrix B has an inverse then

$$C_i(\boldsymbol{x}) \equiv \partial_i \ln p_{\rm st}(\boldsymbol{x}) = B^{-1}_{ij}(\boldsymbol{x}) \left[A_j(\boldsymbol{x}) - \partial_k B_{jk}(\boldsymbol{x})\right] \tag{2.10}$$

is a gradient. Necessary and sufficient conditions for this are

$$\frac{\partial C_i}{\partial x_j} = \frac{\partial C_j}{\partial x_i} \tag{2.11}$$

with $i, j = 1, \ldots, N$. Provided these conditions hold, the determination of the stationary probability distribution is reduced to quadratures:

$$p_{\rm st}(\boldsymbol{x}) = \exp\left\{\int^{\boldsymbol{x}} \mathrm{d}x_i'\, C_i(\boldsymbol{x}')\right\}. \tag{2.12}$$

Furthermore, when (2.9) holds, the operator

$$\mathcal{L} = (p_{\rm st}(\boldsymbol{x}))^{-1/2}\; L\; (p_{\rm st}(\boldsymbol{x}))^{1/2} \tag{2.13}$$

is Hermitian [2] with the real eigenfunctions

$$\psi_n(\boldsymbol{x}) = \frac{P_n(\boldsymbol{x})}{(p_{\rm st}(\boldsymbol{x}))^{1/2}} = (p_{\rm st}(\boldsymbol{x}))^{1/2}\; Q_n(\boldsymbol{x}) \tag{2.14}$$

and the eigenvalues are non-negative.

Of course, these simplifications occur only in exceptional cases; for most cases the operator L cannot be brought into Hermitian form as in (2.13). Nevertheless, it is still useful to define the operator $\mathcal{L}$, since it turns out that splitting this operator (as opposed to L) up into Hermitian and anti-Hermitian parts is natural in the sense that it is dictated by the underlying temporal symmetries of the system [2]. This decomposition can be achieved explicitly by writing $A_i = A_i^+ + A_i^-$, where A_i^+ is defined by $A_i^+ = p_{\rm st}^{-1}\partial_j[B_{ij}\, p_{\rm st}]$. Denoting the operator of the form (2.2), with A replaced by A^+, L by L^+ and defining $L^- = -\partial_i A_i^-$, the original Fokker–Planck operator may be written as $L = L^+ + L^-$ with the corresponding $\mathcal{L}$ operators defined by (2.13) having a similar decomposition: $\mathcal{L} = \mathcal{L}^+ + \mathcal{L}^-$. By construction $\mathcal{L}^+$ is Hermitian, since A^+ satisfies the conditions (2.9). It is also easy to check that $\mathcal{L}^-$ is anti-Hermitian using $L^- p_{\rm st}(\boldsymbol{x}) = 0$.

In general, the operators $\mathcal{L}$ and $\mathcal{L}^\dagger$ will have a different set of (complex) eigenfunctions: from (2.4), (2.5) and (2.13) the eigenfunctions of these operators are

$$\psi_n(\boldsymbol{x}) = \frac{P_n(\boldsymbol{x})}{(p_{\rm st}(\boldsymbol{x}))^{1/2}} \qquad \text{and} \qquad \chi_n(\boldsymbol{x}) = (p_{\rm st}(\boldsymbol{x}))^{1/2}\, Q_n(\boldsymbol{x})\,. \tag{2.15}$$

From equation (2.6) they satisfy the orthonormality condition

$$\langle \chi_m | \psi_n \rangle = \int \mathrm{d}^N \boldsymbol{x}\, \psi_n(\boldsymbol{x})\, \chi_m(\boldsymbol{x}) = \delta_{n,m}\,. \tag{2.16}$$

The eigenvalues, although complex, have a non-negative real part [2, 7] as expected on physical grounds. The eigenfunction of both $\mathcal{L}$ and $\mathcal{L}^\dagger$ corresponding to the zero eigenvalue is $(p_{\mathrm{st}}(\boldsymbol{x}))^{1/2}$.

3. Variational approaches

Variational approaches are quite useful for finding ground-state energies in quantum mechanics because: (i) an error of order ϵ in the variational wavefunction results in an error of order ϵ^2 in the variational energy; and (ii) the true ground-state energy is known to be lower than the variational energy. Since the lowest eigenvalue of a Fokker–Planck operator is identically zero, the focus becomes the eigenstates (the lowest being the stationary probability distribution) and/or excited eigenvalues (the first excited eigenvalue being the reciprocal of the longest relaxation time of the system [3]).

Some work on Fokker–Planck variational approaches has been carried out, mainly on one-component problems. In such cases the potential conditions automatically hold, the stationary probability distribution is known exactly, and the interest was naturally centred on higher eigenvalues. In these circumstances, progress is made by obtaining the Hermitian operator $\mathcal{L}$ and proceeding as in quantum mechanics. In calculations of the lowest non-zero eigenvalue, the nice features of a variational approach can be recovered simply by ensuring that the variational state is orthogonal to the stationary probability distribution. However, since most problems do not obey the potential conditions, variational schemes applicable to these more general situations need to be developed.

3.1. Constructing a Hermitian operator: the $L^\dagger L$ approach

When the potential conditions do not hold, one can still mimic quantum mechanics by constructing a self-adjoint operator from the Fokker–Planck operator L. A natural choice is $L^\dagger L$ since it shares the right eigenstates with L and its eigenvalues Λ_n are related to those of L through $\Lambda_n = \lambda_n^* \lambda_n$. With a Hermitian operator in hand one simply proceeds as in quantum mechanics. In addition to the drawback that there is no longer any interest in finding the lowest eigenvalue (it is identically zero), another disadvantage of this scheme compared to Rayleigh–Ritz in quantum mechanics, is that the calculation requires twice as many operations—first operating with L and then with $L^\dagger$, rather than just with the Hamiltonian H.

It is natural to consider minimizing $\langle P_0^{\mathrm{tr}} | L^\dagger L | P_0^{\mathrm{tr}} \rangle / \langle P_0^{\mathrm{tr}} | P_0^{\mathrm{tr}} \rangle$ [4], where $P_0^{\mathrm{tr}}(\boldsymbol{x})$ is a trial distribution, as a means for finding a reasonable approximation to the steady-state distribution. This quantity is bounded from below by zero and equals zero only if the choice $P_0^{\mathrm{tr}}(\boldsymbol{x})$ is equal to the exact steady-state distribution $P_0(\boldsymbol{x})$.

Note that when examining the eigenvalue equation $LP_n = -\lambda_n P_n$ one has the freedom to transform the equation to $\tilde{L}\tilde{P}_n = -\lambda_n \tilde{P}_n$ where $\tilde{L} = R^{-1/2} L R^{1/2}$ and $\tilde{P}_n = R^{-1/2} P_n$ provided R has no zeros. Hence, one alternative would be to minimize $\langle \tilde{P}_0^{\mathrm{tr}} | \tilde{L}^\dagger \tilde{L} | \tilde{P}_0^{\mathrm{tr}} \rangle / \langle \tilde{P}_0^{\mathrm{tr}} | \tilde{P}_0^{\mathrm{tr}} \rangle$. Calculational convenience sometimes dictates using $R = P_0^{\mathrm{tr}}$.

The first excited eigenvalue λ_1 (the reciprocal of the longest relaxation time of the system) is accessible to a straightforward variational approach provided $|P_1\rangle$ has a different

symmetry to $|P_0\rangle$. In such cases $\langle P_1^{\rm tr}|L^\dagger L|P_1^{\rm tr}\rangle/\langle P_1^{\rm tr}|P_1^{\rm tr}\rangle$ yields an upper bound on Λ_1 provided one can guarantee that $|P_1^{\rm tr}\rangle$ is orthogonal to $|P_0\rangle$.

It is possible to lower the bounds on Λ_n by improving $|P_n^{\rm tr}\rangle$ (making a better choice or one with more variational parameters). A systematic procedure, which is similar to Lançzos tridiagonalization [8], uses

$$|\mathcal{P}_{\rm tr}\rangle = |P_{\rm tr}\rangle + \alpha L|P_{\rm tr}\rangle + \beta L^2|P_{\rm tr}\rangle + \cdots \tag{3.1}$$

as the improved trial distribution, where $\alpha, \beta, \ldots$ are additional variational parameters. Finally we note that it is possible to get some sense of how good the variational estimate is, since in the Fokker–Planck problem the lowest eigenvalue should be zero. In addition, for the higher eigenvalues there exists a procedure which finds lower as well as upper bounds on the eigenvalues [3, 9].

3.2. Variation within perturbation

The other variational approaches we will discuss have their foundations in perturbation theory. If the Fokker–Planck operator at hand is in some sense close to one which is solvable, then a perturbation approach is a viable scheme of approximation. One can make a perturbative expansion variational by introducing parameters into the solvable part and eventually choosing them to minimize some measure of the 'nearness' of the two problems.

The perturbative approach begins by splitting the Fokker–Planck operator as follows:

$$L = L^{(0)} + \epsilon\, L^{(1)} \tag{3.2}$$

where $L^{(0)}$ is the operator of some completely solvable problem, i.e. all of its eigenvalues $\lambda_n^{(0)}$ as well as its right and left eigenfunctions, $P_n^{(0)}(\boldsymbol{x})$ and $Q_m^{(0)}(\boldsymbol{x})$ are known. The variational aspect of this approach requires $L^{(0)}$ to contain some as yet undetermined parameters. Note that ϵ need not be small and is used here primarily as a counting device. Expanding the eigenfunctions and eigenvalues of L as a series in ϵ yields

$$\left(L^{(0)} + \epsilon L^{(1)}\right)\left(|P_n^{(0)}\rangle + \epsilon|P_n^{(1)}\rangle + \cdots\right) = -\left(\lambda_n^{(0)} + \epsilon\lambda_n^{(1)} + \cdots\right)\left(|P_n^{(0)}\rangle + \epsilon|P_n^{(1)}\rangle + \cdots\right). \tag{3.3}$$

Manipulating the terms of order ϵ in the usual ways leads to

$$\lambda_n^{(1)} = -\epsilon\,\langle Q_n^{(0)}|L^{(1)}|P_n^{(0)}\rangle \tag{3.4}$$

$$|P_n^{(1)}\rangle = -\epsilon \sum_{m\neq n} \frac{|P_m^{(0)}\rangle\langle Q_m^{(0)}|L^{(1)}|P_n^{(0)}\rangle}{(\lambda_n^{(0)} - \lambda_m^{(0)})} \tag{3.5}$$

$$\langle Q_n^{(1)}| = -\epsilon \sum_{m\neq n} \frac{\langle Q_n^{(0)}|L^{(1)}|P_m^{(0)}\rangle\langle Q_m^{(0)}|}{(\lambda_n^{(0)} - \lambda_m^{(0)})}\,. \tag{3.6}$$

The terms of order ϵ^2 lead to

$$\lambda_n^{(2)} = \epsilon^2 \sum_{m\neq n} \frac{\langle Q_n^{(0)}|L^{(1)}|P_m^{(0)}\rangle\langle Q_m^{(0)}|L^{(1)}|P_n^{(0)}\rangle}{(\lambda_n^{(0)} - \lambda_m^{(0)})} \tag{3.7}$$

and so on.

Notice that the structure of the perturbation expansion is such that $\lambda_0 = 0$ and $\langle Q_0| = 1$ at every order, as these results are exact. Furthermore, it is worth remarking that the expansion for $|P_0\rangle$ does not necessarily remain everywhere positive order-by-order; thus there can arise difficulties in interpreting a truncated expansion of $|P_0\rangle$ as a probability distribution. Another feature of the perturbation expansion to note is that if the $\lambda_n^{(0)}$'s are

all real, then the perturbation expansion for the eigenvalues remains real (provided there are no degenerate $\lambda_n^{(0)}$'s which would necessitate degenerate perturbation theory); that is, the imaginary part of the eigenvalue is inaccessible to a perturbation theory that begins with purely real eigenvalues.

If L_0 has been chosen so as to satisfy the potential conditions, then it is convenient to consider the perturbative expansion of the operator transformed by $P_0^{(0)}$:

$$\left(\ell^{(0)}+\epsilon\ell^{(1)}\right)\left(|\psi_n^{(0)}\rangle+\epsilon|\psi_n^{(1)}\rangle+\cdots\right)=-\left(\lambda_n^{(0)}+\epsilon\lambda_n^{(1)}+\cdots\right)\left(|\psi_n^{(0)}\rangle+\epsilon|\psi_n^{(1)}\rangle+\cdots\right) \tag{3.8}$$

where

$$\begin{aligned}\ell^{(0)}&=\left(P_0^{(0)}(\boldsymbol{x})\right)^{-1/2}L^{(0)}\left(P_0^{(0)}(\boldsymbol{x})\right)^{1/2}\\ \ell^{(1)}&=\left(P_0^{(0)}(\boldsymbol{x})\right)^{-1/2}L^{(1)}\left(P_0^{(0)}(\boldsymbol{x})\right)^{1/2}\\ \psi_n^{(j)}(\boldsymbol{x})&=\frac{P_n^{(j)}(\boldsymbol{x})}{\left(P_0^{(0)}(\boldsymbol{x})\right)^{1/2}}\,.\end{aligned} \tag{3.9}$$

This is because the left and right eigenstates at zeroth order are identical, which is of considerable calculational convenience.

Of course, the more natural rotation is by the true stationary probability distribution, but that is unknown. In the perturbative approach it is possible to perform this transformation order-by-order as follows:

$$\begin{aligned}\mathcal{L}&=\left(P_0^{(0)}+\epsilon P_0^{(1)}+\cdots\right)^{-1/2}\left(L^{(0)}+\epsilon L^{(1)}\right)\left(P_0^{(0)}+\epsilon P_0^{(1)}+\cdots\right)^{1/2}\\ \mathcal{L}&=\mathcal{L}^{(0)}+\epsilon\mathcal{L}^{(1)}+\epsilon^2\mathcal{L}^{(2)}+\cdots\end{aligned} \tag{3.10}$$

where

$$\mathcal{L}^{(0)}=\left(P_0^{(0)}\right)^{-1/2}L^{(0)}\left(P_0^{(0)}\right)^{1/2} \tag{3.11}$$

$$\mathcal{L}^{(1)}=\left(P_0^{(0)}\right)^{-1/2}\left(L^{(1)}+\frac{1}{2}\left[L^{(0)},\frac{P_0^{(1)}}{P_0^{(0)}}\right]\right)\left(P_0^{(0)}\right)^{1/2} \tag{3.12}$$

and so on. Then one could proceed with

$$\begin{aligned}&\left(\mathcal{L}^{(0)}+\epsilon\mathcal{L}^{(1)}+\cdots\right)\left(|\Psi_n^{(0)}\rangle+\epsilon|\Psi_n^{(1)}\rangle+\cdots\right)\\ &\qquad=-\left(\lambda_n^{(0)}+\epsilon\lambda_n^{(1)}+\cdots\right)\left(|\Psi_n^{(0)}\rangle+\epsilon|\Psi_n^{(1)}\rangle+\cdots\right).\end{aligned} \tag{3.13}$$

This approach is rather cumbersome, especially when it is noted that finding $\mathcal{L}^{(1)}$ requires knowing $P_0^{(1)}$.

Since the lowest eigenvalue in the Fokker–Planck perturbation expansion is zero order-by-order, some other quantity must be found to vary. It is desirable, though not necessary, to vary a bounded expression, as the bound helps to ensure a sensible result. This motivation led us to consider varying $\langle P_0^{(1)}|P_0^{(1)}\rangle$ or alternatively $\langle\Psi_0^{(1)}|\Psi_0^{(1)}\rangle$ which is positive semi-definite by construction. We call this approach the 'minimal corrected wavefunction' criterion or MCW. The variation seeks to minimize $\langle P_0^{(1)}|P_0^{(1)}\rangle$—the philosophy being that the closer $|P_0^{(0)}\rangle$ is to the actual $p_{st}(\boldsymbol{x})$, the smaller its correction should be. The procedure might be continued by minimizing $\langle P_0^{(2)}|P_0^{(2)}\rangle$ to determine the parameters to be inserted in $|P_0^{(0)}\rangle+|P_0^{(1)}\rangle$, and so on. To calculate λ_1, $\langle P_1^{(n)}|P_1^{(n)}\rangle$ could be minimized to fix the parameters to be used in $\lambda_1^{(0)}+\lambda_1^{(1)}+\cdots+\lambda_1^{(n)}$.

Another method based on perturbation theory with undetermined parameters has been used by Edwards and co-workers on problems such as polymers with excluded volume

[10] and the Fokker–Planck formulation of the KPZ equation [11]. Stevenson has dubbed it the 'fastest apparent convergence' criterion or FAC [12]. First a number of terms in a perturbative expansion of the quantity of interest are calculated. FAC then assumes that the zeroth-order term (which depends on the input parameters) is exact; and so the rest of the expansion is set to zero. This last step determines the unknown parameters to be substituted into the zeroth-order term. Note that FAC is not a variational approach as the parameters are not determined by varying.

A scheme with the same starting point which is variational is the so-called 'principle of minimal sensitivity' or PMS [12]. After obtaining a truncated perturbation expansion, one varies it with respect to the undetermined parameter(s). Note that these parameters were introduced artificially and that the actual answer should not depend on them; however, any truncated expansion does depend on them. The PMS philosophy is then to search for the result that is 'least sensitive' to the parameters—and hence the variation. In a few select simple examples the PMS procedure has been proven to yield a convergent series of approximations even when the underlying perturbation expansion is asymptotic [13], but the general conditions for which it does so remains an open problem.

3.3. Comparison of approaches on an example from quantum mechanics

Let us test these approaches on a well known problem from quantum mechanics, the quartic oscillator: $H = -\frac{1}{2}\,\mathrm{d}^2/\mathrm{d}x^2 + \frac{1}{4}gx^4$ using the harmonic oscillator $H^{(0)} = -\frac{1}{2}\,\mathrm{d}^2/\mathrm{d}x^2 + \frac{1}{2}\omega^2 x^2$ as the basis for the perturbation expansion. By dimensional arguments, the eigenvalues of H can be seen to be proportional to $g^{1/3}$, and hereafter we scale this factor out. Let us focus our attention on the ground-state energy; the result is known to be $E_0^{\mathrm{direct}} = 0.420\,805\ldots$. Applying the MWC, FAC and PMS [13, 14] approaches to standard Rayleigh–Schrödinger perturbation theory (calculated to third order for the energy and second order for wavefunction) yield the results seen in table 1. We also include the Rayleigh–Ritz result and its first Lançzos-like correction (RRL).

Table 1. The results of four variational approaches applied to E_0 of the quartic oscillator.

Order	MWC	FAC	PMS	RRL
1	0.429 57	0.454 28	0.429 27	0.429 27
2	0.422 42	—	(0.421 43)	0.422 36
3	—	0.488 54	0.420 98	—

Let us make a few observations. PMS and RRL are identical at first order. At second order the FAC and PMS approaches have no physical solutions; in PMS one can search for inflection points when no extrema are found—at second order this yields 0.421 43. Notice that not only are the variational approaches better than FAC at first order but also that they were improved by going to second or third order, while FAC got worse. To the orders calculated here, PMS has led to the best results.

4. Some Fokker–Planck examples

In this section we will consider a few examples of the variational schemes applied to Fokker–Planck problems. As we wish to emphasize the techniques rather than any particular problem, we will first apply the techniques to a toy model constructed mainly for its

calculational simplicity. Then we will examine a more difficult problem seen recently in the physics literature—the coloured-noise problem.

4.1. A toy model

The toy model we consider is the Fokker–Planck equation corresponding to the following two coupled, non-linear Langevin equations:

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t} &= -\nu x_1 - g x_1 x_2^2 + \eta_1(t) \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -\nu x_2 + \eta_2(t)\end{aligned} \tag{4.1}$$

where the noise is Gaussian-correlated with zero mean and the following correlation:

$$\langle \eta_i(t)\eta_j(t')\rangle = 2D\delta_{ij}\delta(t-t')\,. \tag{4.2}$$

The associated Fokker–Planck operator is

$$L_{\mathrm{toy}} = D\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right) + \nu\left(2 + x_1\frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_2}\right) + g x_2^2\left(1 + x_1\frac{\partial}{\partial x_1}\right). \tag{4.3}$$

The first thing to note is that it does not satisfy the potential conditions. Since they are not satisfied, let us apply each of the variational methods suggested above to obtain the steady-state values of $\langle x_1^2\rangle$ and $\langle x_2^2\rangle$. For the approaches based on perturbation theory we will need an operator for which the eigenvalue problem is solvable. We choose as $L^{(0)}_{\mathrm{toy}}$ the Fokker–Planck operator for an Ornstein–Uhlenbeck process (the harmonic oscillator of Fokker–Planck problems):

$$L^{(0)}_{\mathrm{toy}} = \nu d_1\frac{\partial^2}{\partial x_1^2} + \nu d_2\frac{\partial^2}{\partial x_2^2} + \nu\left(2 + x_1\frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_2}\right). \tag{4.4}$$

The eigenstates of $L^{(0)}_{\mathrm{toy}}$ are

$$P^{(0)}_{n_1,n_2}(\boldsymbol{x}) = \frac{H_{n_1}\left(x_1/\sqrt{2d_1}\right)H_{n_2}\left(x_2/\sqrt{2d_2}\right)}{\left(2^{n_1+n_2+2}\pi^2 n_1!n_2!d_1d_2\right)^{1/2}}\exp\left\{-\frac{x_1^2}{2d_1} - \frac{x_2^2}{2d_2}\right\} \tag{4.5}$$

and its eigenvalues are $\lambda_{n1,n2} = -(n_1+n_2)\nu$ where $n_1, n_2 = 0, 1, 2, \ldots$. As we are interested in spatial quantities, we have selected parameters d_i which affect the spatial distribution $P^{(0)}_0(\boldsymbol{x})$ but leave the eigenvalues of $L^{(0)}_{\mathrm{toy}}$ unchanged.

The details of the variational calculations we performed on the toy model can be found in the appendix. The results are shown in table 2 along with the results from a simulation of the toy-model Langevin equation (4.1).

Table 2. Variational results for the toy model with $D = 0.5$, $v = 1.0$ and $g = 3.0$.

Method	$\langle x_1^2\rangle$	$\langle x_2^2\rangle$
Simulation	0.265 228	0.500 310
$L^\dagger L$	0.271 875	0.279 693
MCW	0.241 854	0.5
FAC	0.2	0.5
PMS	0.174 307	0.5

The simulation algorithm employed a second-order Runge–Kutta method to evolve the equations [15]. After a sufficiently long evolution, dependence on the initial conditions

is lost. We simulated the equation for 1 000 000 realizations and extracted the averages $\langle x_i^2(t_f)\rangle$ where $x_i(t_f)$ is the final position of the simulation. The results of a simulation with $D = 0.5$, $\nu = 1.0$ and $g = 3.0$ are $\langle x_1^2\rangle = 0.265\,228\,0$ and $\langle x_2^2\rangle = 0.500\,310\,4$.

Note that the values for d_1 and d_2 found from varying $\langle\psi_0^{(1)}|\psi_0^{(1)}\rangle$ were substituted into equation (A.9) at first order. The methods based on perturbation theory all agree that $\langle x_2^2\rangle = 0.5$ and are better than the $\langle\psi_0^{(0)}|\ell_{\rm toy}^\dagger\ell_{\rm toy}|\psi_0^{(0)}\rangle$ variation on this point; on the other hand, the latter variation produced the best result for $\langle x_1^2\rangle$. In fact, it is easy to see, by direct integration of the second equation (4.1), that $\langle x_2^2\rangle = D/\nu$, which is exactly 0.5 for the values of the parameters we chose. By beginning with this value at first order, the methods based on perturbation theory had a distinct advantage over the $L^\dagger L$ approach—although the values of the variational parameters d_1 and d_2 were very similar in the MCW and $L^\dagger L$ cases. One reason for the relatively poor result for $\langle x_2^2\rangle$ when not using the perturbative schemes, may be that the variation in this case is not carried out directly on the quantity of interest.

4.2. The coloured-noise example

Now we move on to consider one of the variational approaches for a more difficult situation—the coloured-noise problem [16–18]. Consider the following Langevin equation:

$$\dot{x} = -V'(x) + \xi(t) \tag{4.6}$$

which describes an overdamped particle subject to the force $f(x) = -V'(x)$ and an external noise $\xi(t)$. External noise is not intrinsically related to the system's evolution and is typically 'coloured' as opposed to 'white,' i.e. it is not delta-function correlated. As a consequence, many of the techniques and results familiar from the study of Markov processes are not applicable. For present purposes, we take $\xi(t)$ to be Gaussian distributed with zero mean and exponentially correlated:

$$\langle\xi(t)\xi(t')\rangle = \frac{D}{\tau}\exp\left\{-|t-t'|/\tau\right\}. \tag{4.7}$$

With this choice, the above one-dimensional non-Markov process (4.6) can be shown to be equivalent to the following two-dimensional Markov process [17]:

$$\dot{x} = -V'(x) + \xi(t) \qquad \dot{\xi} = -\frac{1}{\tau}\xi + \frac{1}{\tau}\eta(t) \tag{4.8}$$

where $\eta(t)$ is Gaussian distributed with zero mean and this time delta-function correlated:

$$\langle\eta(t)\eta(t')\rangle = 2D\delta(t-t'). \tag{4.9}$$

The corresponding Fokker–Planck equation is

$$\frac{\partial P(x,\xi;t)}{\partial t} = -\frac{\partial}{\partial x}\left[\left[-V'(x)+\xi\right]P\right] + \frac{1}{\tau}\frac{\partial}{\partial\xi}\left[\xi P + \frac{D}{\tau}\frac{\partial P}{\partial\xi}\right] \tag{4.10}$$

which upon the transformation to the velocity variable $\xi = \dot{x} + V'(x)$ becomes

$$\frac{\partial P(x,\dot{x};t)}{\partial t} = \left\{\frac{D}{\tau^2}\frac{\partial^2}{\partial\dot{x}^2} + \left[V''(x)+\frac{1}{\tau}\right]\left[1+\dot{x}\frac{\partial}{\partial\dot{x}}\right] - \dot{x}\frac{\partial}{\partial x} + \left[\frac{V'(x)}{\tau}\right]\frac{\partial}{\partial\dot{x}}\right\}P. \tag{4.11}$$

Again, it is readily confirmed that the potential conditions are not satisfied in this case. Nevertheless, there is a solvable case. When the operator L is quadratic (i.e. when the drift vector $A_i(\boldsymbol{x})$ is linear in and the diffusion matrix $B_{ij}(\boldsymbol{x})$ is independent of the x_i's), one can find a closed expression for the stationary distribution even if the potential conditions

are not met. For the coloured-noise problem, this situation occurs if $V(x) = \omega x^2/2$. The associated Fokker–Planck operator is

$$L_{\text{quad}} = \frac{D}{\tau^2}\frac{\partial^2}{\partial \dot{x}^2} + \left[\omega + \frac{1}{\tau}\right]\left[1 + \dot{x}\frac{\partial}{\partial \dot{x}}\right] - \dot{x}\frac{\partial}{\partial x} + \left[\frac{\omega x}{\tau}\right]\frac{\partial}{\partial \dot{x}} \tag{4.12}$$

and steady-state distribution is

$$P_0 \propto \exp\left\{-\frac{\omega(1+\omega\tau)}{2D}x^2 - \frac{\tau(1+\omega\tau)}{2D}\dot{x}^2\right\}. \tag{4.13}$$

This result can be useful for considering the behaviour near the minima for a more general potential $V(x)$.

Hereafter, we take $V(x)$ to be a bistable potential given by $V(x) = -x^2/2 + x^4/4$, with the following operator:

$$L_{\text{bistable}} = \frac{D}{\tau^2}\frac{\partial^2}{\partial \dot{x}^2} + \left[3x^2 - 1 + \frac{1}{\tau}\right]\left[1 + \dot{x}\frac{\partial}{\partial \dot{x}}\right] - \dot{x}\frac{\partial}{\partial x} + \left[\frac{x^3 - x}{\tau}\right]\frac{\partial}{\partial \dot{x}}. \tag{4.14}$$

For the region immediately surrounding the minima of the bistable $V(x)$ ($x = \pm 1$), the potential can be described by Taylor expansion truncated at the quadratic order, suggesting that

$$\lim_{x\to 1} P_0 \approx \exp\left\{-\frac{(1+2\tau)}{D}(x-1)^2 - \frac{\tau(1+2\tau)}{2D}\dot{x}^2\right\} \tag{4.15}$$

and similarly around $x = -1$.

The perturbation approaches require a soluble starting point. However, we know of no convenient solvable problem with a bistable potential. We could use L_{quad} but we expect that a few orders of quite cumbersome perturbation theory might be required before reasonable agreement is reached. Consequently, we will restrict our attention to the $L^\dagger L$ approach. All we need is some suitable trial steady-state distribution. We choose a variational stationary distribution of the form

$$P(x, \dot{x}) \sim \exp\left\{-f(x) - g(x)\dot{x}^2\right\}. \tag{4.16}$$

This is Gaussian in $\dot{x}$, so that the integrations over $\dot{x}$ are readily performed analytically. Note that stability requires that $g(x) > 0$ for all x and that symmetry suggests that $f(x)$ and $g(x)$ are even in x. Calculating the expectation of $L^\dagger L$ with respect to P yields

$$\begin{aligned}\frac{\langle P|L^\dagger L|P\rangle}{\langle P|P\rangle} = \int \mathrm{d}x \Bigg\{ &\frac{3\left(g'(x)\right)^2}{32\,(g(x))^{7/2}} + \frac{3}{4\,(g(x))^{1/2}}\left[\frac{2Dg(x)}{\tau^2} - 3x^2 + 1 - \frac{1}{\tau}\right]^2 \\ &+ \frac{1}{4\,(g(x))^{3/2}}\left[f'(x) + \frac{3}{4}\frac{g'(x)}{g(x)} - \frac{2(x^3 - x)g(x)}{\tau}\right]^2\Bigg\}\exp\{-2f(x)\} \\ &\times\left[\int \mathrm{d}x\,(g(x))^{-1/2}\exp\{-2f(x)\}\right]^{-1}.\end{aligned} \tag{4.17}$$

Our goal is to minimize (4.17). Note that the expression in the curly brackets above is the sum of three squares. The third term can be made zero by choosing

$$f(x) = -\frac{3}{4}\ln\left[g(x)\right] + \frac{2}{\tau}\int^x (y^3 - y)g(y)\,\mathrm{d}y \tag{4.18}$$

which we will do. On the other hand, the first two terms compete; the first of them would be zero if $g(x) =$ constant, while the second would be zero if $g(x) = [\tau(1+2\tau) + 3\tau^2(x^2 - 1)]/2D$. As a compromise let us try

$$g(x) = \frac{\tau(1+2\tau)}{2D} + \frac{b\tau^2(x^2-1)}{2D} \tag{4.19}$$

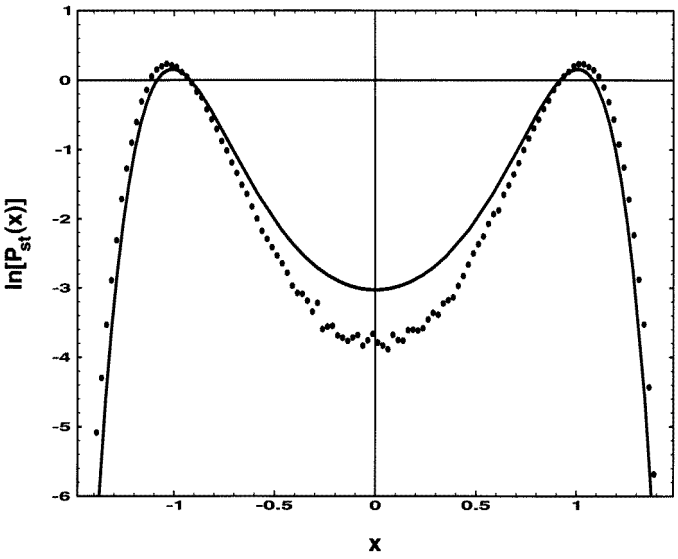

Figure 1. The value of $\ln[P_s t(x)]$ plotted for $D = 0.1$ and $\tau = 0.6$. The dots are from a numerical simulation of the Langevin equation and the full curve is the result of the variational calculation.

with b determined by minimizing (4.17). Note that these choices are consistent with what one expects from the Taylor expansion about the minima of the potential (4.15) and that they also recover the appropriate $\tau \to 0$ limit.

For $D = 0.1$ and $\tau = 0.6$, we find $\langle P|L^{\dagger}L|P\rangle = 0.320\,15$ when $b = 0$ (i.e. when P is given by (4.15); whereas varying b yielded a minimum of $\langle P|L^{\dagger}L|P\rangle = 0.037\,012\,1$, an order of magnitude smaller, at $b = 2.63$. The corresponding marginal stationary distribution

$$P_{mar}(x) = \int \mathrm{d}\dot{x}\; P(x,\dot{x}) \sim\; [g(x)]^{-1/2} \exp\{-f(x)\} \tag{4.20}$$

is shown in figure 1.

The results are reasonably good for a calculation that has one variational parameter and more importantly is restricted to a Gaussian form in $\dot{x}$. It has captured some features of the real marginal distribution including the shift (of order D) in the maxima away from $x = \pm 1$ (though this is slightly underestimated) and the rapid fall off for $x > 1.1$. However, the probability density is overestimated around $x = 0$. This shortcoming of the variational approach is understandable since this region has exponentially small contribution to $\langle P\mathcal{L}^{\dagger}L|P\rangle$. These results can be compared to those obtained by a systematic small-D expansion using path-integral techniques [18]. The path-integral approach is clearly superior for small D, but we expect it to become less reliable as D increases. In contrast, the variational approach can, in principle, be applied for any D. Of course, the variational calculation can be improved upon, for instance, by including terms proportional to $\dot{x}^3$ and $\dot{x}^4$ in the trial distribution. But this would merely complicate our illustration of the methods discussed earlier.

5. Conclusions

In this paper we have discussed a number of variational procedures which may be applied to the Fokker–Planck equation. They broadly fall into two classes: those in which a Hermitian operator was constructed from the generically non-Hermitian Fokker–Planck operator and those which were constructed with reference to an approximation to the full problem, which was exactly solvable. Both approaches have advantages and disadvantages. For example, in the former approach, where we used the Hermitian operator $L^\dagger L$, twice as many operations have to be carried out—the result of first operating with L and then with $L^\dagger$. On the other hand, this may be the only method available; in the coloured noise example, for instance, we were unable to construct a Fokker–Planck operator which was both solvable and which contained enough of the essential physics to make it a useful approximation. Whereas in a separate work on the KPZ equation [5], we found that the advantage clearly lay with the improved perturbation methods when considering functional Fokker–Planck problems, the variants in the second approach were tested on a simple quantum-mechanical example as well on a toy Fokker–Planck problem. One of these schemes (MCW) shares with $L^\dagger L$ the advantage of having a lower bound built in. From the few examples we have investigated, it is clear that variational procedures are capable of giving good quantitative, as well as qualitative, results for Fokker–Planck equations. However, more work is required to extend the range of the problems studied, as well as the order to which the calculations are carried out, before definitive statements about the relative usefulness of the various possible approaches can be made.

Acknowledgments

We wish to thank N Goldenfeld and M Moore for useful discussions and the Isaac Newton Institute for Mathematical Sciences, where this work was begun, for its hospitality. TB acknowledges the support of the Engineering and Physical Sciences Research Council under grant no GR/H40150.

Appendix

In this appendix, some of the calculational details for the toy model are provided. The perturbative approaches split the operator L_{toy}, equation (4.3), as follows $L_{\text{toy}} = {L_{\text{toy}}}^{(0)} + {L_{\text{toy}}}^{(1)}$, where ${L_{\text{toy}}}^{(0)}$, equation (4.4), is the Fokker–Planck operator for the Ornstein–Uhlenbeck process. The associated $L^{(1)}_{\text{toy}}$ is

$$L^{(1)}_{\text{toy}} = \sum_{i=1,2} (D - \nu d_i) \frac{\partial^2}{\partial x_i^2} + g x_2^2 \left(1 + x_1 \frac{\partial}{\partial x_1} \right) . \tag{A.1}$$

It is convenient to work with the transformed operator $\left(P_0^{(0)}(\boldsymbol{x})\right)^{-1/2} L \left(P_0^{(0)}(\boldsymbol{x})\right)^{1/2}$; expressed in terms of annihilation and creation operators, it is

$$\ell^{(0)}_{\text{toy}} = -\nu(a_1^\dagger a_1 + a_2^\dagger a_2) \tag{A.2}$$

$$\ell^{(1)}_{\text{toy}} = \sum_{i=1,2} \left(\frac{D}{d_i} - \nu \right) a_i^\dagger a_i^\dagger - g d_2 (a_2 + a_2^\dagger)^2 (a_1^\dagger a_1 + a_1^\dagger a_1^\dagger) \tag{A.3}$$

where $x_i = d_i^{1/2}(a_i + a_i^\dagger)$ and $\partial/\partial x_i = (a_i - a_i^\dagger)/2d_i^{1/2}$.

From here one can readily calculate the following:

$$\langle\psi_0^{(0)}|\ell_{\rm toy}^{\dagger}\ell_{\rm toy}|\psi_0^{(0)}\rangle = 2\left[\frac{D}{d_1}-\nu-gd_2\right]^2+2\left[\frac{D}{d_2}-\nu\right]^2+4g^2d_2^2 \tag{A.4}$$

and similarly

$$\langle\psi_0^{(1)}|\psi_0^{(1)}\rangle = \frac{1}{2\nu^2}\left[\frac{D}{d_1}-\nu-gd_2\right]^2+\frac{1}{2\nu^2}\left[\frac{D}{d_2}-\nu\right]^2+\frac{g^2d_2^2}{4\nu^2}\,. \tag{A.5}$$

Minimization of $\langle\psi_0^{(0)}|\ell_{\rm toy}^{\dagger}\ell_{\rm toy}|\psi_0^{(0)}\rangle$ yields

$$d_1=\frac{D}{\nu+gd_2} \qquad \text{and} \qquad d_2^4+\frac{D\nu}{2g^2}d_2-\frac{D^2}{2g^2}=0 \tag{A.6}$$

while minimization of $\langle\psi_0^{(1)}|\psi_0^{(1)}\rangle$ results in

$$d_1=\frac{D}{\nu+gd_2} \qquad \text{and} \qquad d_2^4+\frac{2D\nu}{g^2}d_2-\frac{2D^2}{g^2}=0\,. \tag{A.7}$$

To include the FAC and PMS approaches, we calculate $\langle x_i^2\rangle$:

$$\langle x_i^2\rangle=\int \mathrm{d}x_1\,\mathrm{d}x_2\,x_i^2\big(P_0^{(0)}(\boldsymbol{x})+\epsilon P_0^{(1)}(\boldsymbol{x})+\epsilon^2P_0^{(2)}(\boldsymbol{x})+\cdots\big) \tag{A.8}$$

for $i=1,2$. To order ϵ^2 we find

$$\begin{aligned}\langle x_1^2\rangle &= d_1+\epsilon\left[\frac{d_1}{\nu}\left(\frac{D}{d_1}-\nu-gd_2\right)\right]+\epsilon^2\left[\frac{d_1}{\nu}\left(\frac{2g^2d_2^2}{\nu}+2gd_2-\frac{gd_2D}{\nu d_1}-\frac{gD}{\nu}\right)\right]\\ \langle x_2^2\rangle &= d_2+\epsilon\left[\frac{d_2}{\nu}\left(\frac{D}{d_2}-\nu\right)\right]+\epsilon^2[0]\,.\end{aligned} \tag{A.9}$$

FAC gives $\langle x_1^2\rangle = D\nu/(\nu^2+gD)$ and $\langle x_2^2\rangle = D/\nu$ at $\mathrm{O}(\epsilon)$, and has no physical solution at $\mathrm{O}(\epsilon^2)$. PMS, on the other hand, has no physical solution at $\mathrm{O}(\epsilon)$ and variation of $\langle x_1^2\rangle$ at $\mathrm{O}(\epsilon^2)$ with respect to d_1 and d_2 yields

$$\frac{2g^2d_2^2}{\nu^2}+\frac{gd_2}{\nu}-\frac{gD}{\nu^2}=0 \qquad \frac{4g^2d_1d_2}{\nu^2}+\frac{gd_1}{\nu}-\frac{gD}{\nu^2}=0\,. \tag{A.10}$$

References

[1] Schiff L I 1970 *Quantum Mechanics* 3rd edn (New York: McGraw-Hill)
[2] Risken H 1989 *The Fokker–Planck equation* 2nd edn (Berlin: Springer)
[3] Brand H, Schenzle A and Schröder G 1982 *Phys. Rev.* A **25** 2324
Agarwal G S, Dattagupta S and Murthy K P N 1984 *J. Phys. C: Solid State Phys.* **17** 6869
[4] Seybold K 1978 *Dissertation* Ulm
[5] Blum T and McKane A J 1995 *Phys. Rev.* E **52** 4741
[6] Gardiner C W 1990 *Handbook of Stochastic Methods* 2nd edn (Berlin: Springer)
[7] Graham R 1980 *Z. Phys.* B **40** 149
[8] Dagotto E and Moreo A 1985 *Phys. Rev.* D **31** 865
Duncan A and Roskies R 1985 *Phys. Rev.* D **31** 364
[9] Weinstein D H 1934 *Proc. Natl Acad. Sci., USA* **20** 529
[10] Edwards S F and Singh P 1979 *J. Chem. Soc. Faraday Trans.* **75** 1001
[11] Schwartz M and Edwards S F 1992 *Europhys. Lett*. **20** 301
[12] Stevenson P M 1981 *Phys. Rev.* D **23** 2916
[13] Buckley I R C, Duncan A and Jones H F 1993 *Phys. Rev.* D **47** 2554
Duncan A and Jones H F 1993 *Phys. Rev.* D **47** 2560
[14] Neveu A 1990 *Nucl. Phys. (Proc. Suppl.)* B **18** 242

[15] Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1986 *Numerical Recipes* (Cambridge: Cambridge University Press)
[16] Dykman M I and Lindenberg K 1994 *Contemporary Problems in Statistical Physics* ed G Weiss (Philadelphia: SIAM) p 41
[17] McKane A J, Luckock H C and Bray A J 1990 *Phys. Rev.* A **41** 644
[18] Rattray K M and McKane A J 1991 *J. Phys. A: Math. Gen.* **24** 4375