

AN ANALYSIS OF NOISE PROPAGATION IN THE MULTISCALE SIMULATION OF COARSE FOKKER-PLANCK EQUATIONS

YVES FREDERIX¹, GIOVANNI SAMAELY¹ AND DIRK ROOSE¹

Abstract. We consider multiscale systems for which only a fine-scale model describing the evolution of individuals (atoms, molecules, bacteria, agents) is given, while we are interested in the evolution of the population *density* on coarse space and time scales. Typically, this evolution is described by a coarse Fokker-Planck equation. In this paper, we consider a numerical procedure to compute the solution of this Fokker-Planck equation directly on the coarse level, based on the estimation of the unknown parameters (drift and diffusion) using only appropriately chosen realizations of the fine-scale, individual-based system. As these parameters might be space- and time-dependent, the estimation is performed in every spatial discretization point and at every time step. If the fine-scale model is stochastic, the estimation procedure introduces noise on the coarse level. We investigate stability conditions for this procedure in the presence of this noise and present an analysis of the propagation of the estimation error in the numerical solution of the coarse Fokker-Planck equation.

Mathematics Subject Classification. 65L20, 35Q84, 60H30, 35R60.

Received October 24, 2009. Revised May 3, 2010 and July 26, 2010.

Published online November 30, 2010.

1. INTRODUCTION

While multiscale problems naturally arise in many application areas (biochemistry, material science, atmospheric simulations...), their solution poses a challenge to computational science. Often only the slow, macroscopic dynamics of the system is of interest and the exact dynamics of the fine-scale fluctuations is less important. Here, we deal with microscopic (fine-scale) individual-based multiscale models, while we are only interested in the evolution of the population density on a macroscopic, coarse level. As the analytical derivation of a closed coarse model from the fine-scale description might not be feasible, alternative approaches have to be explored.

A classical solution strategy is to first assume a coarse-scale model, which in our context is typically a convection-diffusion equation. This equation is then discretized on a coarse mesh, on which the unknown convection (drift) and diffusion coefficients are precomputed in every grid point from simulations of the microscopic model. The solution can then be computed using a standard solver. However, when these coefficients are also time-dependent, this is no longer possible. In such a case, drift and diffusion not only have to be estimated in every grid point, but also *at every time step*. Due to stochasticity in the fine-scale model, the estimation error

Keywords and phrases. Multiscale computing, stochastic systems, Fokker-Planck equation, uncertainty propagation.

¹ Department of Computer Science, K.U. Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium. Yves.Frederix@cs.kuleuven.be; Giovanni.Samaely@cs.kuleuven.be; Dirk.Roose@cs.kuleuven.be

appears on the coarse level in the form of numerical noise. The aim of this paper is to investigate the effect of this noise on the stability and accuracy of the coarse-scale simulation.

For convenience of the analysis, we consider processes for which the coarse dynamics $X(t) \in \mathbb{R}^N$ satisfies an effective stochastic differential equation (SDE) of the form

$$dX = \gamma(X)dt + \sigma(X)dW_t, \quad (1.1)$$

with γ the drift coefficient, σ the diffusion coefficient and W_t a d -dimensional Wiener process.

The goal is to compute the evolution of the *probability density* $\rho(X, t)$, which is described by the corresponding Fokker-Planck equation [21],

$$\rho(X, t)_t = -\nabla_X \cdot (\gamma(X)\rho(X, t)) + \frac{1}{2}\nabla_X \cdot [\nabla_X \cdot (\sigma(X)\sigma(X)^T \rho(X, t))]. \quad (1.2)$$

A particular class of models that fits in this framework are singularly perturbed systems of stochastic differential equations [24],

$$\begin{aligned} dx^\epsilon &= a(x^\epsilon, y^\epsilon)dt + b(x^\epsilon, y^\epsilon)dU_t, \\ dy^\epsilon &= \frac{1}{\epsilon}f(x^\epsilon, y^\epsilon)dt + \frac{1}{\sqrt{\epsilon}}g(x^\epsilon, y^\epsilon)dV_t, \end{aligned} \quad (1.3)$$

with $x^\epsilon \in \mathbb{R}^N$ and $y^\epsilon \in \mathbb{R}^M$ the slow and fast variables respectively and the time scale separation given by a small parameter ϵ . U_t and V_t are mutually independent d_1 - and d_2 -dimensional Wiener processes. Systems of the form (1.3) have been studied extensively within the framework of averaging for SDEs, see [12,18] and references therein. If the y -dynamics is ergodic for fixed values of $x^\epsilon = X$ and generates a unique invariant measure $\mu^X(dy^\epsilon)$, it is well known that for $\epsilon \rightarrow 0$, the slow dynamics $x^\epsilon(t)$ converges weakly to $X(t)$ when $\gamma(X)$ and $\sigma(X)\sigma(X)^T$ in (1.1) are defined as the averages of $a(X, y^\epsilon)$ and $b(X, y^\epsilon)b(X, y^\epsilon)^T$ with respect to the measure $\mu^X(dy^\epsilon)$; see, *e.g.*, [23]. Although averaging theory provides expressions to compute γ and σ , it might not always be possible to obtain an analytical solution. They can, however, be estimated, which has been the topic of extensive research, see, *e.g.*, [1,19,20] for some recent results with references to earlier literature. In [4,7,26] the simulation of a coarse SDE (1.1) is accelerated by estimating the coarse drift and diffusion coefficients from local simulations of the fine-scale model at each coarse time step. There, the authors propose to generate a time series of the fast variables by simulating the y -dynamics for fixed value of x^ϵ and to compute the unknown coefficients by averaging over this time series. Here, we do not allow changes to the fine-scale simulator and use estimators based on an ensemble average of short time series of the full fine-scale model.

We consider the numerical behavior of a multiscale procedure in which γ and σ are estimated in each time step in a finite difference discretization of (1.2). Hence, the algorithm can also be applied to systems for which the coarse Fokker-Planck equation, possibly in multiple space dimensions, contains space- and time-dependent coefficients,

$$\rho(X, t)_t = -\nabla_X \cdot (\gamma(X, t)\rho(X, t)) + \frac{1}{2}\nabla_X \cdot [\nabla_X \cdot (\sigma(X, t)\sigma(X, t)^T \rho(X, t))]. \quad (1.4)$$

The analysis, however, is carried out for a one-dimensional version of the model problem (1.1)–(1.2), which simplifies the analysis while retaining the essential properties of the propagation of the estimation error. This work forms a natural extension of the work carried out in the equation-free [15,16] and heterogeneous multiscale frameworks [3,5], enabling a better understanding of the propagation of numerical errors due to repeated estimation of coarse-scale quantities based on fine-scale simulation.

This paper is organized as follows. In Section 2, we introduce estimators for the drift and diffusion coefficients and motivate the main assumption about their statistical properties, which will be used in the analysis. Section 3 then gives a detailed description of the multiscale procedure. In Sections 4 and 5, we present theoretical stability results and an analysis for the propagation of the estimation error in the coarse solution for a limiting, pure diffusive case of (1.4). The results are illustrated with numerical experiments. In Section 6, we provide numerical results for the more general case of (1.4) and point out similarities with the derived theoretical bounds. Finally, in Section 7, we present conclusions and discuss possible future research directions.

2. DRIFT AND DIFFUSION ESTIMATION

2.1. The estimators

This section introduces estimators to compute the coarse drift and diffusion coefficients. The performance of the estimators is illustrated with a numerical example that motivates the characterization of the estimation error that will be used later on in the analysis.

We assume the availability of a simulator for the fine-scale system (1.3) of which all variables $v^\epsilon = (x^\epsilon, y^\epsilon)$ can be initialized at will. We can then define a “time stepper” $\phi_{\Delta t}(v^\epsilon(t))$ that returns an (approximate) solution $v^\epsilon(t + \Delta t)$ for the system with initial condition $v^\epsilon(t)$. To relate the slow variables X to the fine-scale variables v^ϵ , we define the restriction operator \mathcal{M} so that $X = \mathcal{M}(x^\epsilon, y^\epsilon) = x^\epsilon$. The coefficients $\gamma(X)$ and $\sigma(X)\sigma(X)^\top$ are estimated from simulations with the fine-scale model, starting from an ensemble of R initial conditions consistent with $\mathcal{M}(v_r^\epsilon) = X$. For the estimates to converge, it must be ensured that the fast variables y_r^ϵ sample the invariant measure $\mu^X(dy^\epsilon)$ of y^ϵ for fixed $x^\epsilon = X$; see, e.g., [23,26].

Here, we are interested in accurate estimation of drift and diffusion coefficients that may depend on space, without changing the fine-scale simulator. However, as only their value around a given spatial discretization point is required and locally these can be assumed to be approximately constant, the estimators should preferably be based on short localized simulations. For the systems of the form (1.3) considered in this work, we will use estimators based on ensemble averaging that fulfill this condition,

$$\begin{aligned} \bar{\gamma}(X) &:= \frac{1}{R\Delta t} \sum_{r=1}^R (\mathcal{M}(\phi_{\Delta t}(X, y_r^\epsilon)) - X) \approx \gamma(X), \\ \bar{\sigma}(X)\bar{\sigma}(X)^\top &:= \frac{1}{R\Delta t} \sum_{r=1}^R (\mathcal{M}(\phi_{\Delta t}(X, y_r^\epsilon)) - X) (\mathcal{M}(\phi_{\Delta t}(X, y_r^\epsilon)) - X)^\top \approx \sigma(X)\sigma(X)^\top. \end{aligned} \tag{2.1}$$

Provided that the initial conditions y_r^ϵ represent a good sampling of the invariant measure $\mu^X(dy^\epsilon)$, the estimates converge in the limit $R \rightarrow \infty$, $\epsilon \rightarrow 0$ and $\Delta t \rightarrow 0$. We assume that the measure $\mu^X(dy^\epsilon)$ can be computed analytically (or numerically) so that we can indeed choose appropriate y_r^ϵ by sampling from the corresponding distribution. If the fine-scale system has the form (1.3), it is then optimal to use Δt equal to the time step used for the integration of the fine-scale system.

Note that the choice of initial y_r^ϵ is the stochastic equivalent of what in equation-free computing is called the *lifting problem* [16]. For the deterministic case, solutions have been proposed for some specific settings [11,25], but in general, lifting is a difficult problem [9].

Example 2.1. We now illustrate the parameter estimation for a multiscale system of the form (1.3),

$$\begin{aligned} dx &= (x - x^3 + Ay) dt + By^2 dU_t, \\ dy &= \frac{-y}{\epsilon} dt + \sqrt{\frac{2}{\epsilon}} dV_t, \end{aligned} \tag{2.2}$$

with $x, y \in \mathbb{R}$ and U_t, V_t independent Wiener processes, in which we choose $A = 1/10$, $B = 1/2$ and $\epsilon = 0.1$. The system is integrated using an Euler-Maruyama scheme with time step $\delta t = \epsilon^2 = 0.01$.

The slow dynamics describes the movement of a particle in a bistable symmetrical double-well potential driven by the fast dynamics and by the noise term U_t . In this case, the fast dynamics is a fast Ornstein-Uhlenbeck (OU) process, which, for simplicity, is independent of the slow variable x . Also, as the OU process is ergodic with invariant measure equal to the standard normal distribution, the slow dynamics can be approximated [23] for small ϵ by (1.1) with

$$\gamma(X) = X - X^3 \qquad \sigma(X) = \sqrt{3}B. \tag{2.3}$$

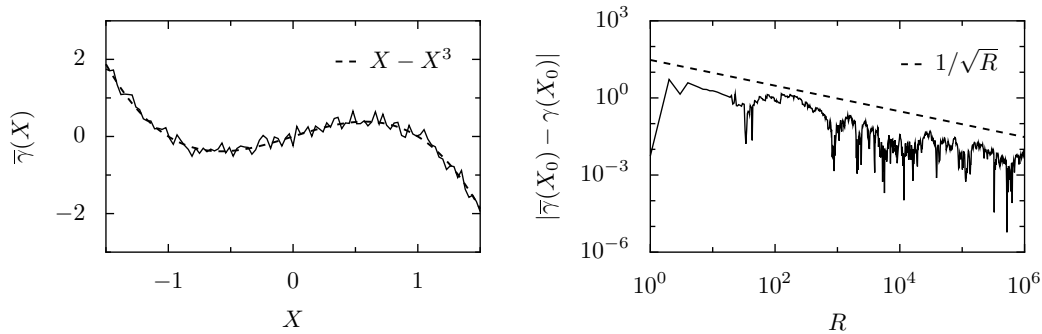


FIGURE 1. Left: Estimation of the drift for $X \in [-1.5, 1.5]$. The exact value (dashed) and the estimate for $R = 5000$ (solid) are shown. Right: Error on the estimate $\bar{\gamma}(X_0)$ with respect to the exact value as a function of the number of replicas R for $X_0 = 0.5$.

The evolution of the probability density $\rho(X, t)$ is given by the corresponding Fokker-Planck equation

$$\rho(X, t)_t = -((X - X^3)\rho(X, t))_X + \frac{3B^2}{2}\rho(X, t)_{XX}. \quad (2.4)$$

We estimate drift $\gamma(X)$ and diffusion $\sigma(X)$ using (2.1) and compare to the exact values given by (2.3). Figure 1 (left) shows the estimated drift $\bar{\gamma}(X)$ for $R = 5000$. The right figure illustrates the $1/\sqrt{R}$ convergence of the estimate for fixed X . Note that as the invariant measure $\mu^X(dy)$ is the standard normal distribution, initialization of the multiscale system amounts to generating high quality samples from this distribution, which can be accomplished easily using a standard statistical library. The results for the diffusion coefficient are similar (not shown).

2.2. Distribution of the estimates

To be able to investigate the error propagation, a formal link has to be established between the fine-scale system (from which the drift and diffusion are estimated) and the Fokker-Planck description on the coarser level. This can be done by considering the distribution of the estimated values. Unfortunately, finding an accurate quantification of this distribution is not an easy task for general fine-scale systems and estimators. For systems of the form (1.3), however, we can draw some conclusions for estimators of the form (2.1). Indeed, as these formulas are based on the computation of an average, it follows from the central limit theorem that, for the number of realizations $R \rightarrow \infty$, the distribution of the estimates can be approximated by a Gaussian. Furthermore, it was illustrated in the above numerical example that for increasing R , the estimators indeed converge to the correct value of drift and diffusion, so that in fact the estimation error is normally distributed with zero mean. From this observation, we now propose the following assumption on the distribution of the estimation error, which will serve as the basis for the analysis below. This way, the analysis remains valid also for more general fine-scale systems and estimators, as long as this assumption holds.

Assumption 2.2. *The estimated drift and diffusion are normally distributed around their exact values, i.e.,*

$$\begin{aligned} \bar{\gamma}(X) &= \gamma(X) + A_\gamma \cdot \xi_X^\gamma, \\ \bar{\sigma}(X)\bar{\sigma}(X)^T &= \sigma(X)\sigma(X)^T + A_\sigma \cdot \xi_X^\sigma, \end{aligned} \quad (2.5)$$

with $\xi_X^\gamma \in \mathbb{R}^N$, $\xi_X^\sigma \in \mathbb{R}^{N \times N}$ a vector and matrix with independent standard normal variables. It is also assumed that the quality of the estimates is identical for all X , i.e., A_γ and A_σ are constants (this can in principle be accomplished by adjusting the number of realizations R in the estimation procedure).

Note that, as the estimation involves independent localized procedures around every spatial discretization point X_i , the random variables ξ_i^γ and ξ_j^γ (resp. ξ_i^σ and ξ_j^σ) are necessarily independent for any two points X_i and X_j . Furthermore, as we will use independent simulations to estimate γ and σ , ξ_i^γ and ξ_i^σ are uncorrelated for all i . If, on the other hand, they are estimated from the same time series, correlations should be expected. In this case, a detailed quantification of these correlations is required before attempting the analysis of the propagation of the estimation errors. This, however, is beyond the scope of this paper.

By modeling the distribution of the estimation error, the fine-scale model and related estimators are effectively decoupled from the error on the coarse level. This has advantages, as now the propagation of this error on the coarse level can be analyzed independently of the exact nature of the fine-scale model. The only constraint is that Assumption 2.2 should hold for the used estimators. For systems (1.3) this is indeed the case. For more general fine-scale multiscale systems such as SDE homogenization problems or particle models, it might be hard to find *good* estimators, *i.e.*, unbiased estimators that asymptotically converge to the correct value with increasing computational effort, see, *e.g.*, [8,19].

3. SOLUTION OF THE COARSE FOKKER-PLANCK EQUATION

Our aim is to compute the evolution of the probability density of the slow variable of the fine-scale model directly on the coarse level. This can be accomplished by discretizing (1.4) and estimating the unknown drift and diffusion coefficients in every spatial discretization point at each time step using the estimators from the previous section. The total error of this scheme consists of a systematic error, which depends on the discretization parameters, and a statistical error, which is due to the estimation.

The goal of this paper is to analyze the statistical error. To this end, we consider (1.2), discretized using finite differences with first order upwinding for the convective part and second order central differences for the diffusive part. A forward-Euler scheme is used for the time integration. Clearly, higher-order discretization schemes would decrease the systematic error, but these schemes cannot be analyzed easily. Therefore, we present the analysis for a simple discretization and provide numerical results for some higher order schemes.

We also restrict ourselves to problems in one space dimension. A treatment of the N -dimensional case would involve additional complications to the analysis, without providing more insight in the problem. In N space dimensions, the computational work increases at the same rate as the number of spatial discretization points (due to the estimations at each time step). The Fokker-Planck equation (1.2) then takes the form

$$\rho(X, t)_t = -(\gamma(X)\rho(X, t))_X + \frac{1}{2}(\sigma(X)^2\rho(X, t))_{XX}. \tag{3.1}$$

The discretized equation becomes

$$\rho_j^{k+1} = \rho_j^k + \Delta t a_j^k(\rho_k; \Delta X), \tag{3.2}$$

with

$$\begin{aligned} a_j^k(\rho_k; \Delta X) := & -\frac{1}{\Delta X} (\gamma_j^k \rho_j^k - \gamma_{j-1}^k \rho_{j-1}^k) \\ & + \frac{1}{2\Delta X^2} ((\sigma_{j-1}^k)^2 \rho_{j-1}^k - 2(\sigma_j^k)^2 \rho_j^k + (\sigma_{j+1}^k)^2 \rho_{j+1}^k), \quad j = 0, \dots, m-1, \end{aligned} \tag{3.3}$$

with m the number of spatial discretization points and in which, for ease of representation, we assumed $\gamma_j^k \geq 0$. We note that if both sub- and superscript indices are present, we use the notation $a_j^k := a(x_j, t_k)$. The presence of only a subindex indicates vector notation, *i.e.*, $a_k := [a_0^k, \dots, a_{m-1}^k]$. We restrict the computational domain to $\Omega = [x_l, x_r] \subset \mathbb{R}$ with suitable (Dirichlet, no-flux or periodic) boundary conditions. When x_l and x_r are appropriately chosen, *i.e.*, Ω is sufficiently large to accommodate the important dynamics, this only introduces a negligible approximation error.

We can then use (2.5) and substitute the estimated values $\bar{\gamma}$ and $\bar{\sigma}$ in (3.2) to obtain

$$\rho_j^{k+1} = \rho_j^k + \Delta t L_j^k(\rho_k; \Delta X), \quad (3.4)$$

with the discrete right-hand side (RHS) operator L given by

$$\begin{aligned} L_j^k(\rho_k; \Delta X) = & -\frac{1}{\Delta X} ((\gamma_j^k + A_\gamma \xi_{\gamma,j}^k) \rho_j^k - (\gamma_{j-1}^k + A_\gamma \xi_{\gamma,j-1}^k) \rho_{j-1}^k) \\ & + \frac{1}{2\Delta X^2} (((\sigma_{j-1}^k)^2 + A_\sigma \xi_{\sigma,j-1}^k) \rho_{j-1}^k - 2((\sigma_j^k)^2 + A_\sigma \xi_{\sigma,j}^k) \rho_j^k + ((\sigma_{j+1}^k)^2 + A_\sigma \xi_{\sigma,j+1}^k) \rho_{j+1}^k). \end{aligned} \quad (3.5)$$

This expression can be split into a sum of a deterministic and a stochastic part,

$$L_j^k(\rho_k; \Delta X) = a_j^k(\rho_k; \Delta X) + b_j^k(\rho_k; \Delta X, \xi_\gamma^k, \xi_\sigma^k), \quad (3.6)$$

in which

$$\begin{aligned} b_j^k(\rho_k; \Delta X, \xi_\gamma^k, \xi_\sigma^k) := & -\frac{A_\gamma}{\Delta X} (\xi_{\gamma,j}^k \rho_j^k - \xi_{\gamma,j-1}^k \rho_{j-1}^k) \\ & + \frac{A_\sigma}{2\Delta X^2} (\xi_{\sigma,j-1}^k \rho_{j-1}^k - 2\xi_{\sigma,j}^k \rho_j^k + \xi_{\sigma,j+1}^k \rho_{j+1}^k), \end{aligned} \quad (3.7)$$

where the dependence of b_j^k on A_γ and A_σ is not made explicit.

For future reference, we introduce a short hand notation Φ for the above procedure,

$$\rho_{k+1} = \Phi(\rho_k) := \rho_k + \Delta t (a_k(\rho_k; \Delta X) + b_k(\rho_k; \Delta X, \xi_\gamma^k, \xi_\sigma^k)), \quad (3.8)$$

which can also be written as

$$\rho_{k+1} = \Phi^{k+1}(\rho_0). \quad (3.9)$$

In the next sections, we investigate the numerical behavior of Φ in the presence of estimation errors, *i.e.*, we quantify the error propagation and derive a suitable stability criterion.

Note that the chosen discretization might not be optimally suited to compute an accurate solution of the (deterministic) Fokker-Planck equation. Our goal, however, is solely to investigate the effect of the estimation errors on the computed result. Hence, instead of attempting the rigorous treatment of a general discretization scheme, we investigate in detail the error propagation for a simpler scheme. These results can then serve as an indicator for the expected behavior if higher-order schemes are used.

4. STABILITY ANALYSIS

In this section, we assume stability of the *deterministic* solver and only consider the effects due to the estimation errors. We show that for fixed Δt and ΔX the procedure Φ becomes unstable for A_γ and A_σ above certain critical values \hat{A}_γ and \hat{A}_σ of the noise level. We explore how (3.8) can become unstable and under which conditions on the estimation error a stable algorithm is obtained.

4.1. Derivation of the stability condition

We derive a theoretical stability criterion to find \hat{A}_σ for the case of pure diffusion, *i.e.*, $\gamma \equiv 0$. Afterwards, in Section 6, we show numerically how the results carry over to the case $\gamma \neq 0$.

To avoid technical issues in the analysis, we will use periodic boundary conditions in this section. This change does not influence the nature nor the behavior of the error propagation significantly and merely serves to facilitate the derivation below.

If we define the diffusion coefficient $D(X) := D = \sigma^2/2$ and set $\gamma \equiv 0$, the Fokker-Planck equation (3.1) becomes

$$\rho(X, t)_t = (D\rho(X, t))_{XX}. \tag{4.1}$$

Although we choose the exact diffusion D to be constant, it is nevertheless estimated in every spatial discretization point. Using Assumption 2.2 and suppressing the subscript σ , we can then write the estimated diffusion

$$\overline{D}_j^k = D + A \cdot \xi_j^k, \tag{4.2}$$

with $\xi_j \sim \mathcal{N}(0, 1)$, $j = 0, \dots, m - 1$ and m the number of spatial discretization points. Equation (3.8) then becomes

$$\rho_j^{k+1} = \rho_j^k + D\Delta t \frac{\rho_{j-1}^k - 2\rho_j^k + \rho_{j+1}^k}{\Delta X^2} + A\Delta t \frac{\rho_{j-1}^k \xi_{j-1}^k - 2\rho_j^k \xi_j^k + \rho_{j+1}^k \xi_{j+1}^k}{\Delta X^2}. \tag{4.3}$$

For convenience, we rewrite this expression in matrix form

$$\rho_{k+1} = (I + \mu DU) \cdot \rho_k + \mu AU \cdot \begin{bmatrix} \rho_0^k \cdot \xi_0^k \\ \vdots \\ \rho_{m-1}^k \cdot \xi_{m-1}^k \end{bmatrix}, \tag{4.4}$$

with $I \in \mathbb{R}^{m \times m}$ the unit matrix, $\mu = \Delta t/\Delta X^2$ and $U \in \mathbb{R}^{m \times m}$ the discretization matrix for the diffusion equation with central differences and periodic boundary conditions. We note for later use that the matrix U is both symmetric and circulant and has an orthonormal eigenspace with known eigenvalues λ_j and eigenvectors P_j [13],

$$\lambda_j = -2 + 2 \cos\left(\frac{2\pi j}{m}\right), \tag{4.5}$$

$$P_j = \frac{1}{\sqrt{m}} \left[1, e^{-2\pi i j/m}, \dots, e^{-2\pi i j(m-1)/m} \right]^T, \quad j = 0, \dots, m - 1, \tag{4.6}$$

with i the imaginary unit. Note that some of the λ_j are equal. By taking appropriate linear combinations of the corresponding eigenvectors, it is possible to find a set of orthonormal *real* eigenvectors.

To be able to analyze the stability, we first need to formulate a formal stability condition. Inspired by SDE literature, we choose a definition in the mean-square sense.

Definition 4.1 (mean-square stability). The procedure Φ is called mean-square stable for given Δt and ΔX if

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\|\rho_n\|_{\Delta X}^2 \right] \leq C,$$

with $\rho_n = \Phi^n(\rho_0)$, $\mathbb{E}[\cdot]$ the expected value operator and C a constant that depends on the used discretization scheme and the quality of the parameter estimation.

As customary in stability analysis, a modified discrete norm $\|\cdot\|_{\Delta X} = \|\cdot\|_2 \sqrt{\Delta X}$ is used.

For the analysis, we thus need to consider $\mathbb{E} \left[\|\rho_n\|_{\Delta X}^2 \right]$ and investigate the result for increasing n . Our strategy consists in first making an ansatz for the form of the solution,

$$\mathbb{E} \left[\|\rho_n\|_{\Delta X}^2 \right] = \rho_0^T Q_n \rho_0 \cdot \Delta X. \tag{4.7}$$

The stability condition can then be derived from the spectral properties of the matrix Q_n .

Before stating the main stability result, we establish the following lemma.

Lemma 4.2. *Let $B \in \mathbb{R}^{m \times m}$ be a circulant matrix and h a scalar, then all elements on the diagonal of $H_k = B(I + hB)^{2k} B$ are equal.*

Proof. It can easily be verified that H_k is a circulant matrix. Equality of the diagonal elements then follows trivially. \square

For D the diffusion coefficient, $\mu = \Delta t / \Delta X^2$ and $I \in \mathbb{R}^{m \times m}$ the unit matrix, we define the matrix M_k as it will appear in the analysis below,

$$M_k = U (I + \mu DU)^{2k} U. \tag{4.8}$$

As the discretization matrix U is circulant, it follows from Lemma 4.2 that all elements on the diagonal of M_k are equal.

Definition 4.3. Define v_k to be the value of the diagonal elements of M_k , with M_k given by (4.8), i.e.,

$$\text{diag}(M_k) = v_k \text{diag}(I). \tag{4.9}$$

We are now ready to state the main stability result.

Theorem 4.4. Let $\rho_n = \Phi^n(\rho_0)$ be computed via (4.4). If we have stability of the deterministic integrator ($\mu D < 0.5$), it holds that

$$\mathbb{E} \left[\|\rho_n\|_{\Delta X}^2 \right] = \rho_0^T Q_n \rho_0 \cdot \Delta X, \tag{4.10}$$

with

$$Q_n = (I + \mu DU)^{2n} + (\mu A)^2 \left(\sum_{k=0}^{n-1} v_k Q_{n-k-1} \right), \tag{4.11}$$

and v_k given by (4.9).

Proof. To derive the desired result, we follow a top-down approach and repeatedly apply (4.4) to rewrite $\mathbb{E} \left[\|\rho_n\|_{\Delta X}^2 \right]$ in terms of the initial density ρ_0 , assuming that Q_k is known for $k < n$. For convenience, we introduce a shorter notation for the stochastic factor in (4.4),

$$\rho_{\xi,n} := [\rho_0^n \cdot \xi_0^n, \dots, \rho_{m-1}^n \cdot \xi_{m-1}^n]^T. \tag{4.12}$$

We have

$$\mathbb{E} \left[\|\rho_n\|_{\Delta X}^2 \right] = \mathbb{E} \left[\rho_n^T \rho_n \right] \Delta X.$$

Application of (4.4), while taking into account that $\mathbb{E} [\rho_{\xi,n-1}] \equiv 0$, results in

$$\mathbb{E} \left[\|\rho_n\|_{\Delta X}^2 \right] = \underbrace{\mathbb{E} \left[\rho_{n-1}^T (I + \mu DU)^2 \rho_{n-1} \right] \Delta X}_{E_1} + \underbrace{(\mu A)^2 \mathbb{E} \left[\rho_{\xi,n-1}^T U^2 \rho_{\xi,n-1} \right] \Delta X}_{E_2}.$$

First consider E_2 . Because of the presence of $\mathbb{E} [\cdot]$ and the independence of the estimation errors, only the elements on the diagonal of U^2 (that contain terms in ξ_j^2 with $j = 0, \dots, m - 1$) have a contribution. Using Lemma 4.2 and equation (4.9), this value on the diagonal is equal to v_0 . The term E_2 can then be written as

$$E_2 = (\mu A)^2 v_0 \cdot \mathbb{E} \left[\rho_{\xi,n-1}^T \rho_{\xi,n-1} \right] \Delta X = (\mu A)^2 v_0 \cdot \rho_0^T Q_{n-1} \rho_0 \cdot \Delta X. \tag{4.13}$$

Next, we turn to E_1 and again apply (4.4) to obtain

$$E_1 = \underbrace{\mathbb{E} \left[\rho_{n-2}^T (I + \mu DU)^4 \rho_{n-2} \right] \Delta X}_{E_3} + \underbrace{(\mu A)^2 \mathbb{E} \left[\rho_{\xi,n-2}^T U (I + \mu DU)^2 U \rho_{\xi,n-2} \right] \Delta X}_{E_4}. \tag{4.14}$$

For E_4 we again note that due to the presence of $\mathbb{E}[\cdot]$, only the elements on the diagonal of $U(I + \mu DU)^2 U$ have a contribution. From Lemma 4.2 all elements on this diagonal are equal and we recognize v_1 from (4.9). Substitution in E_4 results in

$$E_4 = (\mu A)^2 v_1 \cdot \mathbb{E}[\rho_{\xi,n-2}^T \rho_{\xi,n-2}] \Delta X = (\mu A)^2 v_1 \cdot \rho_0^T Q_{n-2} \rho_0 \cdot \Delta X. \tag{4.15}$$

The above process can be repeated for E_3 until the term in ρ_0 is reached, giving rise to extra terms containing v_k . Using this notation and taking into account that the term in ρ_0 is deterministic, we obtain the desired result

$$\mathbb{E}[\|\rho_n\|_{\Delta X}^2] = \rho_0^T Q_n \rho_0 \cdot \Delta X, \tag{4.16}$$

with

$$Q_n = (I + \mu DU)^{2n} + (\mu A)^2 \left(\sum_{k=0}^{n-1} v_k Q_{n-k-1} \right), \tag{4.17}$$

and v_k given by (4.9). □

While the values v_k could be computed in a straightforward manner by many multiplications of large matrices (m and n are typically large), a more efficient way is to use the spectral properties of U .

Theorem 4.5. *Let v_k be defined by (4.9) and λ_j be the eigenvalues of U . It then holds that*

$$v_k = \frac{1}{m} \sum_{j=0}^{m-1} \lambda_j (1 + \mu D \lambda_j)^{2k} \lambda_j. \tag{4.18}$$

Proof. Let $U = P \Lambda P^T$ be the eigenvalue decomposition of U , with eigenvalues given by (4.5). The columns of P are normalized so that $PP^T = P^T P = I$, with identity matrix $I \in \mathbb{R}^{m \times m}$ and m the number of spatial discretization points.

We know from the definition of M_k in (4.8) that all its diagonal elements v_k are equal. Substituting the decomposition of U in (4.8) results in

$$M_k = P \Lambda P^T (I + \mu D P \Lambda P^T)^{2k} P \Lambda P^T. \tag{4.19}$$

The j th diagonal element is equal to v_k and can, e.g., be extracted by multiplication on the left and right by the j th canonical basis vector $e_j = [0, \dots, 0, 1, 0, \dots, 0]^T$,

$$\begin{aligned} v_k &= e_j^T \cdot P \Lambda P^T (I + \mu D P \Lambda P^T)^{2k} P \Lambda P^T \cdot e_j \\ &= p_j^T \Lambda P^T (I + \mu D P \Lambda P^T)^{2k} P \Lambda p_j \\ &= p_j^T \Lambda (I + \mu D \Lambda)^{2k} \Lambda p_j, \end{aligned} \tag{4.20}$$

with p_j the j th row of P .

It is important to note at this point that the above expression is valid for all choices of j . As we know from (4.6) that the first elements of the eigenvectors are all equal to $1/\sqrt{m}$, a good choice is $p_j = [1, \dots, 1]^T / \sqrt{m}$. After substitution in (4.20) we obtain the desired result. □

We now discuss some properties of the matrices Q_n . From their definition in Theorem 4.4, it follows that the eigenvectors of Q_n are equal to those of U . Furthermore, as these eigenvectors form an orthonormal basis

for $\mathbb{R}^{m \times m}$, it is possible to treat all modes of the initial density independently. This way, the properties of the matrix Q_n can be studied *via* the following scalar expression,

$$q_n(\lambda_j) = (1 + \mu D \lambda_j)^{2n} + (\mu A)^2 \left(\sum_{k=0}^{n-1} v_k q_{n-k-1}(\lambda_j) \right), \quad j = 0, \dots, m-1, \quad (4.21)$$

with λ_j the eigenvalues of U as given by (4.5) and $q_0(\lambda_j) = 1, \forall \lambda_j$.

For increasing n , $q_n(\lambda_j)$ contains two contributions: assuming the deterministic algorithm is stable, the first always decreases due to the diffusive damping, while the second may increase exponentially due to the addition of estimation errors in every time step. Hence, we can find the overall stability condition on A , by computing the maximal noise level for each individual mode λ_j , *i.e.*, the value \hat{A}_j for which $\lim_{n \rightarrow \infty} q_n(\lambda_j)$ is a constant, and taking the minimum $\hat{A} = \min_j \hat{A}_j$. It will be proven, however, that all \hat{A}_j are equal and independent of λ_j .

For later use, we state an auxiliary lemma, the proof of which is given in Appendix A.

Lemma 4.6. *Let a_k be a strictly decreasing series with $a_0 \geq 0$ and $\lim_{k \rightarrow \infty} a_k = 0$. Furthermore, let b_k be a series*

for which $\lim_{k \rightarrow \infty} b_k = K \in \mathbb{R}$. If $\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_k$ is finite, it then holds that

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_k b_{n-k-1} = K \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_k. \quad (4.22)$$

We are now ready to formulate the following theorem.

Theorem 4.7 (stability). *The algorithm (4.4) is mean-square stable for given $\mu = \Delta t / \Delta X^2$ if*

$$A^2 \leq \frac{D}{\mu} \left(\frac{1}{m} \sum_{j=0}^{m-1} \frac{-\lambda_j}{2 + \mu D \lambda_j} \right)^{-1}, \quad (4.23)$$

with λ_j the eigenvalues of U and m the number of spatial discretization points.

Proof. To compute the maximal noise level \hat{A} for stability of Φ , we find the value of A for which $\lim_{n \rightarrow \infty} q_n(\lambda_j)$ is equal to a constant $K_j \neq 0$,

$$\lim_{n \rightarrow \infty} q_n(\lambda_j) = \lim_{n \rightarrow \infty} \left((1 + \mu D \lambda_j)^{2n} + (\mu A)^2 \left(\sum_{k=0}^{n-1} v_k q_{n-k-1}(\lambda_j) \right) \right) = K_j. \quad (4.24)$$

Furthermore, using that for a stable deterministic algorithm $\mu D < 0.5$, and hence, that $|1 + \mu D \lambda_j| < 1$ for all λ_j , we know from equation (4.18) that the series v_k is strictly decreasing in k with maximum at $v_0 = 6$. As a result, in the sum in equation (4.24), the terms $q_{n-k-1}(\lambda_j)$ for large k are multiplied by increasingly small v_k so that the contribution of the initial $q_n(\lambda_j)$ (*i.e.*, for small n) in the overall sum decreases. By Theorem 4.6, we have that in the limit $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} v_k q_{n-k-1}(\lambda_j) = K_j \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} v_k. \quad (4.25)$$

Again using the fact that $|1 + \mu D \lambda_j| < 1$, the first term in the right-hand side of (4.24) disappears for $n \rightarrow \infty$ so that after substitution of (4.25) we have

$$K_j = (A\mu)^2 K_j \sum_{k=0}^{\infty} v_k. \tag{4.26}$$

We can then divide out the factor K_j and use (4.18) to obtain

$$1 = (A\mu)^2 \sum_{k=0}^{\infty} \frac{1}{m} \sum_{j=0}^{m-1} \lambda_j^2 (1 + \mu D \lambda_j)^{2k}.$$

After changing the order of the summations, we recognize a geometric series in the sum over k , so that

$$\begin{aligned} 1 &= (A\mu)^2 \frac{1}{m} \sum_{j=0}^{m-1} \lambda_j^2 \lim_{n \rightarrow \infty} \frac{1 - (1 + \mu D \lambda_j)^{2n}}{1 - (1 + \mu D \lambda_j)^2}, \\ &= A^2 \mu \frac{1}{m} \sum_{j=0}^{m-1} \frac{\lambda_j}{(-2 - \mu D \lambda_j) D}, \end{aligned}$$

in which we again used that $|1 + \mu D \lambda_j| < 1$. Equation (4.23) is obtained by simple rearrangement of the terms and by noting that the computed A is an upper limit for the noise level for stability.

Although (4.26) depends on λ_j via K_j , the resulting stability condition (4.23) does not. In other words, we obtain the *same* \hat{A} for all λ_j , indicating that all modes become instable at the same time. \square

It is interesting to note that for increasing number of discretization points m , the bound of Theorem 4.7 can be simplified even further. From the analytical expression of the eigenvalues equation (4.5), we find that the sum in the denominator of (4.23) defines a Riemann sum. Taking the limit $m \rightarrow \infty$ results in

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{j=0}^{m-1} \frac{-\lambda_j}{2 + \mu D \lambda_j} &= \int_0^1 \frac{2 - 2 \cos(\pi x)}{2 + \mu D (-2 + 2 \cos(\pi x))} dx, \\ &= \frac{1 - \sqrt{1 - 2\mu D}}{\mu D \sqrt{1 - 2\mu D}}. \end{aligned}$$

After substitution in (4.23), we obtain the following asymptotic stability condition

$$A^2 \leq D^2 \frac{\sqrt{1 - 2\mu D}}{1 - \sqrt{1 - 2\mu D}}. \tag{4.27}$$

Remark 4.8. The above analysis for the case $\gamma \equiv 0$ builds heavily on the structure and associated spectral properties of the matrix U . For $\gamma \neq 0$, the involved discretization matrices change and no longer possess these properties. Numerical experiments, however, indicate that the asymptotic behavior of the critical values \hat{A}_γ and \hat{A}_σ for $\gamma \neq 0$ is similar; see also Section 4.2 below.

4.2. Numerical example

To validate the above stability results for (4.1) with estimated diffusion coefficient D , we compare the theoretical bounds with a numerically determined estimate of the maximal noise level. The latter is computed using the fact that for A_γ and A_σ larger than \hat{A}_γ and \hat{A}_σ the algorithm is unstable, and hence, that the norm

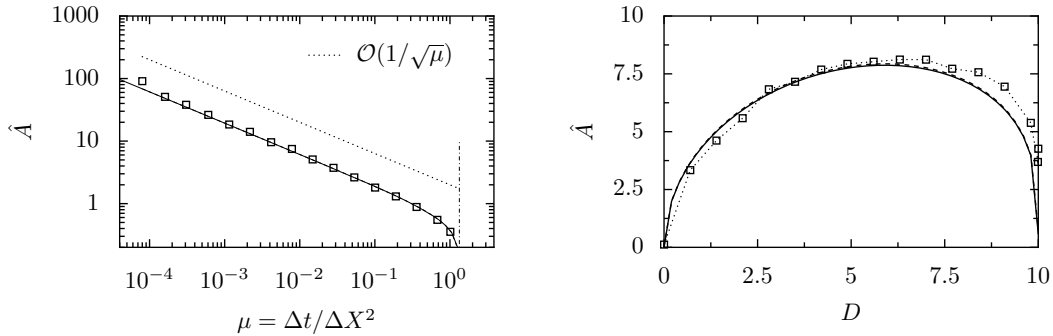


FIGURE 2. Left: Dependence on $\mu = \Delta t / \Delta X^2$ of the experimental (\square) and theoretical value of \hat{A} for $D = 3B^2/2$. The experimental value is computed from equation (4.28) with $t_e = 10$ and $C = 10$. The theoretical values are computed from (4.23) (solid) and (4.27) (dashed). Also shown are the $1/\sqrt{\mu}$ line (dotted) and the deterministic stability boundary $D\mu = 0.5$ (dash-dotted). Right: Dependence on D for fixed $\mu = 0.05$ of the experimental (\square) and theoretical values of \hat{A} . The theoretical values are computed from (4.23) (dashed) and (4.27) (solid).

of the ρ_n increases rapidly with n in such case. An approximation of \hat{A}_γ and \hat{A}_σ for fixed Δt , ΔX and t_n can thus be found by computing the values of the noise level for which

$$\mathbb{E} [\|\rho(t_n)\|_{\Delta X}] = C, \tag{4.28}$$

with C a sufficiently large constant. Note that \hat{A}_γ and \hat{A}_σ are typically overestimated, with the magnitude of the overestimation depending on the exact choice of C and t_n .

For efficient testing, we assume that the exact diffusion D is known and use (4.2) in each spatial discretization point and at every time step to emulate the estimation procedure. The computational domain $\Omega = [-30, 30]$ and the initial condition $\rho_0 = a \sin(2\pi(x + 30)/60)$, with a such that $\|\rho_0\|_{\Delta X} = 1$.

We now compare the experimentally determined \hat{A} with the theoretical bounds from (4.23) and (4.27). For the computation of the experimental value of \hat{A} , we apply the above-mentioned approach and solve (4.28) iteratively using a robust bisection method with $C = 10$. Both the dependence of \hat{A} on $\mu = \Delta t / \Delta X^2$ and on D are considered for a time horizon $t_e = 10$. In Figure 2 (left), \hat{A} is shown as a function of μ . The diffusion coefficient is chosen as in Example 2.1, *i.e.*, $D = 3B^2/2 = 0.375$. There is good correspondence between the experimental (\square) and theoretical values from (4.23) (solid) and (4.27) (dashed) for all μ . In fact, both theoretical values are indistinguishable from each other on the figure, indicating that the asymptotic approximation of (4.27) is very accurate. Furthermore, away from the deterministic stability condition $D\mu < 0.5$ (dash-dotted) we observe that $\hat{A} \propto 1/\sqrt{\mu}$. We found this behavior to be persistent as well for other time integrators such as, *e.g.*, a fourth order Runge-Kutta integrator. Figure 2 (right) shows the dependence of \hat{A} on D for $\mu = 0.05$ and $m = 380$. Again, we find good correspondence between the experimental (\square) and the theoretical results using (4.23) (dashed) and (4.27) (solid). The latter two are again almost indistinguishable. For small D , the added noise in every time step is only slightly damped, and the method quickly becomes unstable (\hat{A} is small). The same is true for larger D , as in that case, we are close to the deterministic stability condition.

4.3. Discussion

We have derived a stability condition for the procedure Φ (see Sect. 3) for the limiting case of (3.1) with $\gamma \equiv 0$. This was done by relating the expected norm of the solution at time t_n to the initial condition ρ_0 via a matrix Q_n . By studying the properties of this matrix and imposing that in the limit $n \rightarrow \infty$ its norm should be a constant, we obtained a closed expression for the maximum noise level (4.23) for the given discretization.

We also provided an approximate boundary if the number of spatial discretization points is large. In Section 6, we present numerical results for the more general case in which also $\gamma \neq 0$.

For simplicity, we also assumed a constant diffusion coefficient D . In the case that $D(X)$ depends on space, bounds for the maximal noise level can be obtained by using an effective constant diffusion coefficient $D = \min_j D(X_j)$ resp. $D = \max_j D(X_j)$ in Theorem 4.7.

5. ACCURACY ANALYSIS

We now investigate how the estimation error propagates in the numerical solution of the coarse Fokker-Planck equation and how it influences the order of convergence of the deterministic PDE solver. Both consistency and error propagation are considered, although for the theoretical analysis of the latter, we again limit ourselves to the case of a forward-Euler scheme with $\gamma \equiv 0$.

5.1. Consistency

We start by exploring consistency of Φ for (3.8) and consider the corresponding discrete RHS operator (3.6) in each spatial discretization point j and at time t_k . As we know that in the deterministic case the integrator is consistent, we only consider the stochastic term $b_j^k(\rho_k; \Delta X, \xi_\gamma^k, \xi_\sigma^k)$; see (3.7). As $\xi_{\gamma,j}^k$ and $\xi_{\sigma,j}^k$ are independent Gaussian random variables for all j , $b_j^k(\rho_k; \Delta X, \xi_\gamma^k, \xi_\sigma^k)$ is a Gaussian variable as well, so that, with slight abuse of notation, the discrete RHS operator for each spatial discretization point (3.6) can be rewritten as

$$L_j^k(\rho_k; \Delta X) = a_j^k(\rho_k; \Delta X) + b_j^k(\rho_k; \Delta X)\xi_j^k, \tag{5.1}$$

with $\xi_j^k \sim \mathcal{N}(0, 1)$, $a_j^k(\rho_k; \Delta X)$ as in (3.3) and $b_j^k(\rho_k; \Delta X)$ given by

$$b_j^k(\rho_k; \Delta X) = \left(\frac{\left((\rho_j^k)^2 + (\rho_{j-1}^k)^2 \right) A_\gamma^2}{\Delta X^2} + \frac{\left((\rho_{j-1}^k)^2 + 4(\rho_j^k)^2 + (\rho_{j+1}^k)^2 \right) A_\sigma^2}{4\Delta X^4} \right)^{1/2}. \tag{5.2}$$

Since the stochastic part disappears under the expected value operator, (5.1) shows that Φ is *expected* to be consistent for all t_k and for all choices of ΔX . For decreasing ΔX , however, the variance on the RHS goes to infinity, which hints at problems in the repeated evaluation of the RHS in the numerical solution of the equation. Indeed, it turns out that to avoid large errors and an unbounded increase of the variance, the accuracy of the estimates should increase with decreasing spatial discretization, so that A_γ and A_σ depend on ΔX . How this manifests itself on the coarse level is discussed in more detail in the error analysis below.

5.2. Error propagation

To investigate the propagation of the estimation error through the procedure Φ , we compare the computed solution with the deterministic solution at time t_e . As already mentioned, the analysis will be done for pure diffusion, *i.e.*, for (4.1).

Theorem 5.1 (error propagation). *Let ρ_n be the discretized density at $t = t_e$ computed via (4.4). If we have stability of both the deterministic integrator ($\mu D < 0.5$) and Φ (see Thm. 4.7), then the asymptotic behavior of the expected value of the squared error on ρ_n with respect to the exact, deterministic solution $\hat{\rho}_n$ is of the form*

$$\mathbb{E} \left[\|\hat{\rho}_n - \rho_n\|_{\Delta X}^2 \right] \lesssim \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta X^4) + \left(\frac{A}{D} \right)^2 M_Q \frac{1 - \sqrt{1 - 2D\Delta t/\Delta X^2}}{\sqrt{1 - 2D\Delta t/\Delta X^2}} \cdot \|\rho_0\|_{\Delta X}^2, \tag{5.3}$$

with M_Q a constant.

Proof. We rewrite the expected squared error as

$$\begin{aligned}\mathbb{E} \left[\|\hat{\rho}_n - \rho_n\|_{\Delta X}^2 \right] &= \mathbb{E} \left[(\hat{\rho}_n - \rho_n)^\top (\hat{\rho}_n - \rho_n) \right] \Delta X, \\ &= \mathbb{E} \left[\hat{\rho}_n^\top \hat{\rho}_n \right] \Delta X - 2\mathbb{E} \left[\hat{\rho}_n^\top \rho_n \right] \Delta X + \mathbb{E} \left[\rho_n^\top \rho_n \right] \Delta X.\end{aligned}$$

While the first term is completely deterministic, the second term contains ρ_n which is stochastic as it includes the accumulated effect of the estimation errors. After taking the expected value of this term, we recover the solution $\tilde{\rho}_n = \mathbb{E}[\rho_n]$ at time t_e , computed with the deterministic solver, which for now we assume to be of order p in Δt . We then have

$$\begin{aligned}\mathbb{E} \left[\|\hat{\rho}_n - \rho_n\|_{\Delta X}^2 \right] &= \hat{\rho}_n^\top \hat{\rho}_n \Delta X - 2\hat{\rho}_n^\top \tilde{\rho}_n \Delta X + \mathbb{E} \left[\rho_n^\top \rho_n \right] \Delta X, \\ &= \underbrace{(\hat{\rho}_n - \tilde{\rho}_n)^\top (\hat{\rho}_n - \tilde{\rho}_n) \Delta X}_{E_1} - \underbrace{\tilde{\rho}_n^\top \tilde{\rho}_n \Delta X + \mathbb{E} \left[\rho_n^\top \rho_n \right] \Delta X}_{E_2}.\end{aligned}\quad (5.4)$$

The first term E_1 represents the squared error for the deterministic solver. This error is bounded due to the stability of the algorithm and is of order $\mathcal{O}(\Delta t^{2p}) + \mathcal{O}(\Delta X^4)$.

For the stochastic part E_2 , we again consider a forward-Euler scheme for the diffusion equation with estimated diffusion \bar{D} ; see (4.4). We obtain

$$\begin{aligned}E_2 &= -\rho_0^\top (I + \mu DU)^{2n} \rho_0 \cdot \Delta X + \rho_0^\top \left[(I + \mu DU)^{2n} + (\mu A)^2 \left(\sum_{k=0}^{n-1} v_k Q_{n-k-1} \right) \right] \rho_0 \cdot \Delta X, \\ &= (\mu A)^2 \cdot \rho_0^\top \left(\sum_{k=0}^{n-1} v_k Q_{n-k-1} \right) \rho_0 \cdot \Delta X\end{aligned}\quad (5.5)$$

with Q_{n-k-1} and v_k as in Theorem 4.4. If we use that the integration is mean-square stable (as in Thm. 4.7), it holds that $\|Q_k\|_2$ is bounded for all k and for all stable choices of Δt and ΔX . Denoting this maximum value by M_Q , we have

$$E_2 \leq (\mu A)^2 M_Q \left(\sum_{k=0}^{n-1} v_k \right) \cdot \|\rho_0\|_{\Delta X}^2.\quad (5.6)$$

Now consider only the sum over v_k . Using the alternative expression for v_k given by (4.18), this can be written as

$$\begin{aligned}\sum_{k=0}^{n-1} v_k &= \sum_{k=0}^{n-1} \frac{1}{m} \sum_{j=0}^{m-1} \lambda_j (1 + \mu D \lambda_j)^{2k} \lambda_j, \\ &= \frac{1}{m} \sum_{j=0}^{m-1} \lambda_j^2 \frac{1 - (1 + \mu D \lambda_j)^{2n}}{1 - (1 + \mu D \lambda_j)^2},\end{aligned}$$

with λ_j the eigenvalues of U . For n large, we obtain after substitution in (5.6):

$$E_2 \leq (\mu A)^2 M_Q \frac{1}{m} \sum_{j=0}^{m-1} \frac{\lambda_j^2}{1 - (1 + \mu D \lambda_j)^2} \cdot \|\rho_0\|_{\Delta X}^2.\quad (5.7)$$

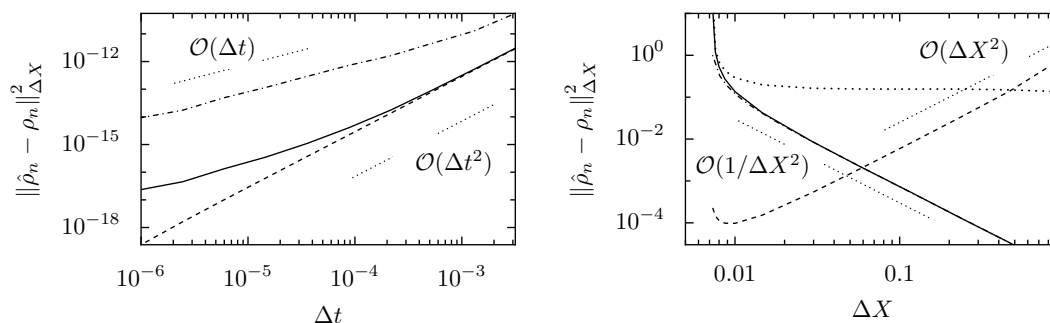


FIGURE 3. Left: Dependence of the error at $t_e = 10$ on the time step Δt for a forward-Euler integration of a pure diffusion equation with $D = 0.375$ and for different noise levels $A = 0$ (dashed), 5×10^{-7} (solid) and 10^{-5} (dash-dotted). Right: Dependence of the error at $t_e = 1$ on ΔX for a forward-Euler integration of a pure diffusion equation. Results are shown for constant $A = 0.2$ (solid), $A \propto \Delta X$ (thick dotted) and $A \propto \Delta X^2$ (dashed). The upper bound of (5.8) is shown in dash-dotted line.

As in Section 4.1, we can use the analytical expression for λ_j (see (4.5)) and approximate the Riemann sum in this equation by the analytical value of the corresponding integral. We then have

$$E_2 \gtrsim \left(\frac{A}{D}\right)^2 M_Q \frac{1 - \sqrt{1 - 2D\Delta t/\Delta X^2}}{\sqrt{1 - 2D\Delta t/\Delta X^2}} \cdot \|\rho_0\|_{\Delta X}^2. \tag{5.8}$$

Finally, substitution of this expression in (5.4) concludes the proof. □

Note that only to deal with the stochastic part of (5.4), we needed to assume a particular discretization of the Fokker-Planck equation. How the results generalize to other time integrators will be discussed and exemplified below.

5.3. Numerical example

To illustrate the error bound of equation (5.3), we consider the diffusion equation with estimated diffusion, see (4.1). We then compute the error for Φ as a function of Δt and ΔX with $D = 3B^2/2 = 0.375$. Figure 3 (left) shows the error as a function of Δt for fixed $\Delta X = 0.1$ and $t_e = 10$. The reference solution was computed from a deterministic simulation with $\Delta t = 10^{-8}$. As we want to assess the influence of the magnitude of the noise, the results are shown for different noise levels A . For (relatively) large $A = 10^{-5}$ (dash-dotted) the squared error behaves as $\mathcal{O}(\Delta t)$, while for $A = 0$ the convergence is $\mathcal{O}(\Delta t^2)$ as in the deterministic case. For a well chosen value of the noise level $A = 5 \times 10^{-7}$, both effects are observed: for large Δt there is deterministic convergence and for small Δt the slower convergence due to the stochastic estimation errors becomes apparent. We remark that for other time integrators such as, *e.g.*, a fourth order Runge-Kutta scheme, we found similar behavior, *i.e.*, for large noise level the error behaves like $\mathcal{O}(\Delta t)$, while for smaller values the deterministic $\mathcal{O}(\Delta t^8)$ convergence is recovered. This will be illustrated for the more general case in Section 6 below.

In Figure 3 (right), the dependence of the error on ΔX is shown for fixed $t_e = 1$. For constant $A = 0.2$ (solid), we observe that away from the deterministic stability boundary $D\mu < 0.5$, *i.e.*, for large ΔX , the error increases proportionally to $\mathcal{O}(1/\Delta X^2)$ with decreasing ΔX , as predicted by the bound in (5.3) (dash-dotted). Furthermore, we know from (5.3) that it is possible to obtain convergence for $\Delta X \rightarrow 0$ by increasing the quality of the estimates while decreasing ΔX . For $A \propto \Delta X$ (thick dotted), the error is more or less constant in ΔX . For $A \propto \Delta X^2$ (dashed), we observe convergence as $\mathcal{O}(\Delta X^2)$. For small ΔX , while approaching the deterministic stability boundary, there is divergence as $1/(1 - 2\mu D)$ for all cases.

5.4. Discussion

In this section, we have analyzed the propagation of the estimation error in the numerical solution of (3.1) with estimated coefficients. It was pointed out that, although there is consistency in every time step, the variance on the discretized RHS increases for $\Delta X \rightarrow 0$, which became clear in the error analysis. An expression for this error at fixed t_e was derived for a forward-Euler integrator in (5.3), which allowed us to predict the evolution of the error when $\Delta X \rightarrow 0$ and $\Delta t \rightarrow 0$.

The effect of increasing variance on the RHS for decreasing ΔX is clearly visible in the theoretical expression for the dependence of the error on ΔX for fixed Δt . To be able to refine the spatial discretization error with fixed Δt , without at the same time destroying the numerical result, the quality of the estimates must increase as ΔX decreases. For instance, to obtain the deterministic second order convergence, the estimation error must behave like $A \propto \Delta X^3$, without violating the stability condition on Φ . In terms of the multiscale estimation, this translates into an increase of the number of realizations $R \propto 1/\Delta X^6$. This is a serious limitation and is in a way inherent to the parameter estimation strategy. In Section 7, we discuss possible ways to improve on this point.

For fixed ΔX , equation (5.3) indicates that for $\Delta t \rightarrow 0$, the asymptotic error contains two components. The order of the first component is equal to the order of the deterministic integrator and is visible for small estimation error. The second component, originating from the estimation errors, has lower order. Although there always is convergence, the exact rate depends on the magnitude of the estimation error and on Δt .

The results for the error propagation were only rigorously established for a forward-Euler scheme and the case of pure diffusion ($\gamma \equiv 0$). In the next section, we present numerical results for the more general case in which also $\gamma \neq 0$.

6. NUMERICAL RESULTS FOR A GENERAL CASE

In the previous sections, we derived rigorous stability and error bounds for Φ for the case of pure diffusion. A general Fokker-Planck equation as described in Section 3, however, contains both a convective and diffusive part, in which case the spectral properties of the discretization matrices change, thus preventing the derivation of stability and error conditions using a similar strategy. In this section, we present numerical results for this more general case and illustrate their relation with the results for pure diffusion from Sections 4 and 5. It should be stressed that our experiments only illustrate what can be expected when Φ is used for the solution of a general Fokker-Planck equation, and that no theoretical analysis is provided.

For our experiments, we return to Example 2.1 of Section 2 and compute the solution at $t_n = n\Delta t$ over the computational domain $\Omega = [-6, 6]$ with homogeneous Dirichlet boundary conditions. The initial condition is always chosen to be $\rho_0(x) = a \cos(\pi x + 1)$ for $x \in [-1, 1]$ and zero otherwise, with a such that $\|\rho_0\|_{\Delta X} = 1$. Recall that the Fokker-Planck equation is given by (2.4).

Stability. We first present numerical results for the dependence of \hat{A} on $\mu = \Delta t/\Delta X^2$. Its value is the solution of (4.28) with $C = 10$; see Figure 4. The solid line represents $\hat{A}(\mu)$ with perturbation only on the diffusion coefficient, *i.e.*, $A_\sigma \neq 0$ and $A_\gamma = 0$. The dash-dotted line shows the result if the quality of the estimates is equal for drift and diffusion, *i.e.*, $A_\sigma = A_\gamma \neq 0$. The lines almost coincide, indicating that for stability the diffusive part dominates. Furthermore, we find the same asymptotic behavior $1/\sqrt{\mu}$ as we found in the case with $\gamma \equiv 0$ (compare Fig. 2). In other words, the stability result of Section 4 seems to extend to the case $\gamma \neq 0$ in the sense that the asymptotic behavior of $\hat{A}(\mu)$ in both cases is similar if the noise on the estimated drift and diffusion is of comparable magnitude. This leads us to conclude that typically the stability of Φ is dominated by the error on the diffusion.

If, however, the noise level on the drift is much larger than this of the diffusion, the effects due to convection start playing a larger role. Figure 4 illustrated this effect for $A_\gamma = 100A_\sigma$ (dotted). We observe that $\hat{A}(\mu)$ now behaves like $1/\mu^{1/4}$. For intermediate values such as $A_\gamma = 10A_\sigma$ (dashed), there is an interplay between both effects, which can clearly be seen in the figure. The observed behavior is in fact typical for large errors on the convection term and can fully be explained from an analysis of (3.4) with $\sigma \equiv 0$. In order not to overload

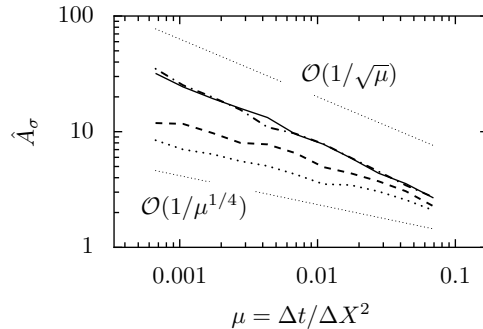


FIGURE 4. Dependence of the experimentally determined maximum noise level \hat{A} on μ for a Fokker-Planck equation with $D = 3B^2/2$ and $t_e = 10$. This value is computed from (4.28) with $C = 10$. The results for noise on the diffusion coefficient only (solid) and noise on both diffusion and drift with $A_\gamma = A_\sigma$ (dash-dotted) are plotted. Also shown are results for $A_\gamma = 10A_\sigma$ (dashed) and $A_\gamma = 100A_\sigma$ (dotted).

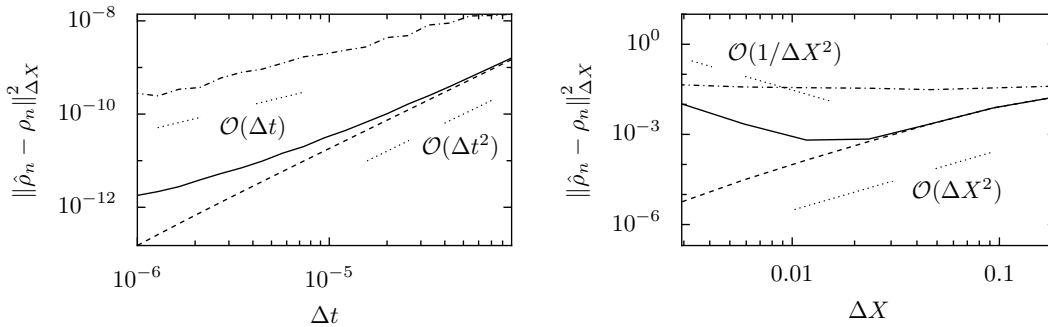


FIGURE 5. Left: Dependence of the error at $t_e = 10$ on the time step Δt for a forward-Euler integration of a Fokker-Planck equation with $D = 0.375$ and for different noise levels $A_\gamma = A_\sigma$. Used values are $A_\sigma = 0$ (dashed), 5×10^{-7} (solid) and 10^{-5} (dash-dotted). Right: Dependence of the error at $t_e = 1$ on ΔX for a forward-Euler integration of a Fokker-Planck equation with $D = 0.375$ and for different noise levels $A_\gamma = A_\sigma$. Results are shown for constant $A_\sigma = 0.15$ (solid), $A_\sigma \propto \Delta X$ (dash-dotted) and $A_\sigma \propto \Delta X^2$ (dashed).

this paper, we do not consider this case explicitly and only show results for the case in which the estimation errors for drift and diffusion are comparable, so that diffusion is expected to dominate. The full analysis and numerical experiments to deal with convective effects will be presented elsewhere.

Accuracy. We also illustrate the error propagation for the Fokker-Planck equation (2.4) from Example 2.1 with $\gamma \neq 0$, and apply Φ with $\Omega = [-6, 6]$ and Dirichlet boundary conditions. The reference solution is computed from a deterministic simulation with $\Delta t = 10^{-7}$. In Figure 5 (left), we show the dependence of the error on Δt for forward-Euler with $\Delta X = 0.1$ and $t_e = 10$. To illustrate the effects of using a higher-order time integration scheme, Figure 6 shows the results for the same experiment with a fourth order Runge-Kutta integrator. It is found that indeed for small noise levels $A_\gamma = A_\sigma$, the order of the deterministic integrator is recovered. For larger noise levels, we retain the $\mathcal{O}(\Delta t)$ convergence as expected. We also illustrate the dependence of the error as a function of the spatial discretization ΔX with constant $t_e = 1$; see Figure 5 (right). For constant $A_\sigma = A_\gamma = 0.15$ (solid), we initially see first order convergence (this is expected, as a first order upwind scheme is used). For smaller ΔX , as in the case of the diffusion equation, divergence proportional to $1/\Delta X^2$ is observed. For $A_\sigma \propto \Delta X$ (dash-dotted) a constant error is obtained. If $A_\sigma \propto \Delta X^2$ (dashed), we obtain the deterministic

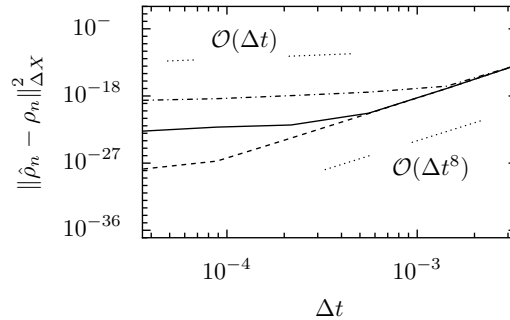


FIGURE 6. Dependence of the error at $t_e = 10$ on the time step Δt for a fourth order Runge-Kutta integration of a Fokker-Planck equation with $D = 0.375$ and for different noise levels $A_\gamma = A_\sigma$. Used values are $A_\sigma = 0$ (dashed), 10^{-8} (solid) and 10^{-10} (dash-dotted).

convergence rate $\mathcal{O}(\Delta X^2)$. Therefore, if the quality of the drift and diffusion estimate is comparable (*i.e.*, $A_\gamma \gg A_\sigma$), the numerical results suggest that the asymptotic error propagation and convergence behavior that were derived theoretically for the case of pure diffusion, are recovered for a Fokker-Planck equation (3.1) with $\gamma \neq 0$.

7. CONCLUSIONS

In this paper, we analyzed a procedure to compute the evolution of the probability density function of the slow components of a multiscale system directly on the coarse level. After assuming the form of the coarse PDE describing the dynamics, the unknown drift and diffusion coefficient therein are estimated from simulations of the fine-scale model using estimators based on ensemble averaging. The PDE is then discretized with finite differences and solved *via* standard solvers. As the unknown coefficients could be time-dependent, the estimation is not only carried out for every spatial discretization point, but also *at every time step*. We illustrated the procedure and its analysis for a class of multiscale stochastic systems for which the equation for the probability density, the coarse Fokker-Planck equation, is known and for which good estimators (unbiased, converging and based on localized short simulations with the fine-scale model) exist.

The procedure was analyzed to quantify its behavior in the presence of estimation errors. To make abstraction of the specific choice of the estimation procedure, we decoupled the analysis on the coarse level from the estimation procedure on the fine-scale level by modeling the distribution of the estimated coefficients. For the class of multiscale systems studied in this paper, good estimators are available, and it was argued that the estimation error in this case follows a Gaussian distribution. The latter was then proposed as the basic assumption for further analysis.

We considered stability of the procedure for a finite difference discretization of the coarse Fokker-Planck equation using a forward-Euler scheme in time. A condition was derived for the case of pure diffusion. The addition of a drift term did not change the numerically observed asymptotic behavior if the quality of the estimates was comparable for drift and diffusion. This indicated that stability for the Fokker-Planck case is typically diffusion dominated. If, however, there is an imbalance, and the noise on the drift estimate is much larger than this on the diffusion, we illustrated numerically that the convective effects cannot be ignored and should be taken into account. A full analysis of the latter case will be published elsewhere.

Furthermore, the use of higher-order discretization schemes complicates the derivation of an exact analytical stability condition. We found, however, numerical evidence that the qualitative behavior of the numerical solution is similar to this of the analyzed case, indicating that the mechanism that leads to instability of the overall procedure is identical in the case of other discretizations.

We also analyzed the propagation of the estimation error. It was shown that the spatial discretization error cannot be reduced without at the same time increasing the accuracy of the estimated coefficients. It was also found that for decreasing time step the scheme always converges. The exact order of convergence, however, depends on both the quality of the estimates and the time step itself. For large estimation error and for small Δt , the stochastic effects result in slow convergence. For small estimation error, the order of the deterministic integrator is recovered. Although the analysis was only performed for the diffusion equation and a forward-Euler scheme, we numerically observed similar behavior for the Fokker-Planck equation (3.2) and for other higher-order time integrators.

Increasing the accuracy of the estimated coefficients to reduce the spatial discretization error can be expensive due to the nature of the estimation procedure (in every spatial discretization point and in every time step). To address this issue, it might be reasonable to assume that the coefficients do not change too much in space and time, and information from neighboring points may thus be interpolated. The interpolated values can then directly be given to the solver, or alternatively, might be used to increase the convergence speed of the estimation procedure. Clearly, the computational cost would decrease, but the above strategies involve a compromise between additional errors due to the interpolation and errors due to the use of estimated coefficients. Another strategy is to use identical random numbers to estimate the coefficients in neighboring points, as is done, *e.g.*, in [17] for the micro/macro simulation of dilute polymer solutions. As this introduces correlations between the estimates, the analysis of the resulting scheme would be more complicated, see, *e.g.*, [14] for an analysis in the context of polymeric fluids. Incorporation of these ideas into our setting will be explored in future research.

The procedure as presented in this paper and its analysis were motivated directly by observations made for a specific family of multiscale problems. However, in principle, it can be applied to other fine-scale models described by, *e.g.*, particle models or black box simulators. Possible dependence of drift and diffusion on time is dealt with by reestimating them in every time step. If they are space-dependent, the coefficients are estimated in every spatial discretization point as well. Although the analysis in this work is no longer exact for such general cases, it does provide a frame of reference and indicates what effects (due to estimation errors) can be expected in the solution of a coarse Fokker-Planck equation if the coefficients are estimated from simulations.

To conclude, we briefly comment on the requirements for the successful application of the procedure to more general problems. First, it is known that standard estimators do not perform well for a general multiscale problem (slow or no convergence) [19], so that it might be difficult to find good (localized) estimators for the considered fine-scale model. Second, to apply these estimators, every simulation should be assigned a “correct” internal state, *i.e.*, the unspecified fast variables of the system should be initialized consistently with the given coarse initial state. This might be achieved *via* a constrained runs algorithm [11], but the solution to this *lifting problem* is not straightforward. We hope to extend our approach to even more general problems for which the Fokker-Planck equation is non-linear and the coarse drift and diffusion depend on the density. Equations of this form occur, for instance, in the context of bacterial chemotaxis models, where γ depends on an external food gradient S , which in turn is influenced by the local particle density [2,6,22]. If good estimators are available, the density-dependence could also in this case be addressed by reestimating them in each time step. However, the additional difficulty lies in the initialization of the fine-scale system since locally around each discretization point the density of the reinitialized system should be consistent with the prescribed value, see, *e.g.*, [10].

A. APPENDIX A

Lemma A.1. *Let a_k be a strictly decreasing series with $a_0 \geq 0$ and $\lim_{k \rightarrow \infty} a_k = 0$. Furthermore, let b_k be a*

series for which $\lim_{k \rightarrow \infty} b_k = K \in \mathbb{R}$. If $\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_k$ is finite, it then holds that

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_k b_{n-k-1} = K \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_k. \quad (\text{A.1})$$

Proof. For convenience, we start by changing the indexation in the summation,

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_k b_{n-k-1} = \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_{n-k-1} b_k. \quad (\text{A.2})$$

From the fact that $\lim_{n \rightarrow \infty} b_n = K$, we know that $\forall \epsilon, \exists N : n > N \Rightarrow |b_n - K| < \epsilon$. The sum in (A.2) can then be split into two parts,

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_{n-k-1} b_k = \lim_{n \rightarrow \infty} \sum_{k=0}^N a_{n-k-1} b_k + \lim_{n \rightarrow \infty} \sum_{k=N+1}^{n-1} a_{n-k-1} b_k.$$

The first term is equal to zero as $\lim_{n \rightarrow \infty} a_n = 0$ for all finite N . For the second term, we have that

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=N+1}^{n-1} a_{n-k-1} b_k &= \lim_{n \rightarrow \infty} \left(\sum_{k=N+1}^{n-1} a_{n-k-1} b_k - K \sum_{k=N+1}^{n-1} a_{n-k-1} + K \sum_{k=N+1}^{n-1} a_{n-k-1} \right), \\ &\leq |b_k - K| \lim_{n \rightarrow \infty} \sum_{k=N+1}^{n-1} a_{n-k-1} + K \lim_{n \rightarrow \infty} \sum_{k=N+1}^{n-1} a_{n-k-1}, \\ &\leq \epsilon \lim_{n \rightarrow \infty} \sum_{k=N+1}^{n-1} a_{n-k-1} + K \lim_{n \rightarrow \infty} \sum_{k=N+1}^{n-1} a_{n-k-1}. \end{aligned}$$

The first term in this equation can be made arbitrarily small as $\lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_n$ is assumed finite and ϵ can be chosen freely. Furthermore, as $\lim_{n \rightarrow \infty} a_n = 0$, also $\lim_{n \rightarrow \infty} \sum_{k=0}^N a_{n-k-1} = 0$, so that we find that

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_{n-k-1} b_k &= \lim_{n \rightarrow \infty} K \sum_{k=N+1}^{n-1} a_{n-k-1} + K \lim_{n \rightarrow \infty} \sum_{k=0}^N a_{n-k-1}, \\ &= K \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} a_k. \quad \square \end{aligned}$$

Acknowledgements. The authors are grateful to D. Givon, Y. Kevrekidis and D. Xiu for useful discussions that led to this work. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. GS is a Postdoctoral Fellow of the Research Foundation – Flanders (FWO).

REFERENCES

- [1] Y. Ait-Sahalia, Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70** (2002) 223–262.
- [2] M. Alber, N. Chen, T. Glimm and P.M. Lushnikov, Multiscale dynamics of biological cells with chemotactic interactions: From a discrete stochastic model to a continuous description. *Phys. Rev. E* **73** (2006) 051901.
- [3] W. E and B. Engquist, The heterogeneous multi-scale methods. *Commun. Math. Sci.* **1** (2003) 87–132.
- [4] W. E, D. Liu and E. Vanden-Eijnden, Analysis of multiscale methods for stochastic differential equations. *Commun. Pure Appl. Math.* **58** (2005) 1544–1585.
- [5] W. E, B. Engquist, X. Li, W. Ren and E. Vanden-Eijnden, Heterogeneous multiscale methods: A review. *Commun. Comput. Phys.* **2** (2007) 367–450.
- [6] R. Erban and H.G. Othmer, From signal transduction to spatial pattern formation in *E. coli*: A paradigm for multiscale modeling in biology. *SIAM Multiscale Model. Simul.* **3** (2005) 362–394.

- [7] I. Fatkullin and E. Vanden-Eijnden, A computational strategy for multiscale systems with applications to Lorenz 96 model. *J. Comput. Phys.* **200** (2004) 605–638.
- [8] Y. Frederix and D. Roose, A drift-filtered approach to diffusion estimation for multiscale processes, in *Coping with complexity: model reduction and data analysis, Lecture Notes in Computational Science and Engineering* **75**, Springer-Verlag (2010).
- [9] Y. Frederix, G. Samaey, C. Vandekerckhove, T. Li, E. Nies and D. Roose, Lifting in equation-free methods for molecular dynamics simulations of dense fluids. *Discrete Continuous Dyn. Syst. Ser. B* **11** (2009) 855–874.
- [10] C. Gear, *Projective integration methods for distributions*. Technical report, NEC Research Institute (2001).
- [11] C.W. Gear, T.J. Kaper, I.G. Kevrekidis and A. Zagaris, Projecting to a slow manifold: Singularly perturbed systems and legacy codes. *SIAM J. Appl. Dyn. Syst.* **4** (2005) 711–732.
- [12] D. Givon, R. Kupferman and A. Stuart, Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **17** (2004) R55–R127.
- [13] R.M. Gray, Toeplitz and circulant matrices: A review. *Found. Trends Commun. Inf. Theory* **2** (2005) 155–239.
- [14] B. Jourdain, C.L. Bris and T. Lelièvre, On a variance reduction technique for micro-macro simulations of polymeric fluids. *J. Non-Newton. Fluid Mech.* **122** (2004) 91–106.
- [15] I.G. Kevrekidis and G. Samaey, Equation-free multiscale computation: Algorithms and applications. *Ann. Rev. Phys. Chem.* **60** (2009) 321–344.
- [16] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg and C. Theodoropoulos, Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.* **1** (2003) 715–762.
- [17] H.C. Öttinger, B.H.A.A. van den Brule and M.A. Hulsen, Brownian configuration fields and variance reduced CONNFESSIT. *J. Non-Newton. Fluid Mech.* **70** (1997) 255–261.
- [18] G. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization, Texts in Applied Mathematics* **53**. Springer, New York (2007).
- [19] G.A. Pavliotis and A.M. Stuart, Parameter estimation for multiscale diffusions. *J. Stat. Phys.* **127** (2007) 741–781.
- [20] Y. Pokern, A.M. Stuart and E. Vanden-Eijnden, Remarks on drift estimation for diffusion processes. *SIAM Multiscale Model. Simul.* **8** (2009) 69–95.
- [21] H. Risken, *The Fokker-Planck Equation: Methods of Solutions and Applications*. Springer Series in Synergetics, Second Edition, Springer (1989).
- [22] M. Rousset and G. Samaey, Simulating individual-based models of bacterial chemotaxis with asymptotic variance reduction. INRIA, inria-00425065, available at <http://hal.inria.fr/inria-00425065/fr/> (2009).
- [23] A. Skorokhod, *Asymptotic methods in the theory of stochastic differential equations, Translations of mathematical monographs* **78**. AMS, Providence (1999).
- [24] N. Van Kampen, Elimination of fast variables. *Phys. Rep.* **124** (1985) 69–160.
- [25] P. Van Leemput, W. Vanroose and D. Roose, Mesoscale analysis of the equation-free constrained runs initialization scheme. *SIAM Multiscale Model. Simul.* **6** (2007) 1234–1255.
- [26] E. Vanden-Eijnden, Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.* **1** (2003) 385–391.