

**An analysis of noise propagation in the  
multiscale simulation of coarse  
Fokker–Planck equations**

*Yves Frederix    Giovanni Samaey    Dirk Roose*

*Report TW 549, October 2009 (Rev. November 2009)*



**Katholieke Universiteit Leuven**  
**Department of Computer Science**

Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

# An analysis of noise propagation in the multiscale simulation of coarse Fokker–Planck equations

*Yves Frederix    Giovanni Samaey    Dirk Roose*

*Report TW 549, October 2009 (Rev. November 2009)*

Department of Computer Science, K.U.Leuven

## **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 investigate 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 solution-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 and present an analysis of the propagation of the estimation error in the numerical solution of the coarse Fokker–Planck equation. The results show that for decreasing spatial discretization error, the total error grows rapidly due to the use of estimated coefficients. This effect can be avoided by increasing the quality of the estimates when the spatial discretization decreases. Although the procedure is illustrated for a specific class of multiscale stochastic systems, it is devised so that it can easily be generalized to other stochastic or particle models.

**Keywords :** multiscale, coarse Fokker–Planck equation, stochastic differential equation, singular perturbation

# An analysis of noise propagation in the multiscale simulation of coarse Fokker–Planck equations

Yves Frederix      Giovanni Samaey      Dirk Roose

November 24, 2009

## 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 investigate 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 solution-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 and present an analysis of the propagation of the estimation error in the numerical solution of the coarse Fokker–Planck equation. The results show that for decreasing spatial discretization error, the total error grows rapidly due to the use of estimated coefficients. This effect can be avoided by increasing the quality of the estimates when the spatial discretization decreases. Although the procedure is illustrated for a specific class of multiscale stochastic systems, it is devised so that it can easily be generalized to other stochastic or particle models.

## 1 Introduction

While multiscale problems naturally arise in many application areas (biochemistry, material science, atmospheric simulations, ...), their solution poses a challenge to computational science due to the presence of multiple scales in the system. Often, however, 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 individual-based multiscale models, while we are only interested in the evolution of the population density on a 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 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 from simulations of the microscopic model. The solution can then be computed using a standard PDE solver. However, when these coefficients are time- or solution-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 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

$$dX = \gamma(X)dt + \sigma(X)dW_t, \quad (1)$$

with  $\gamma$  the drift coefficient,  $\sigma$  the diffusion coefficient and  $W_t$  an  $N$ -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 [18],

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

A particular class of models that fits in this framework are singularly perturbed systems of stochastic differential equations [21],

$$\begin{aligned} dx^\epsilon &= a(x^\epsilon, y^\epsilon)dt + b(x^\epsilon, y^\epsilon)dU_t, \\ dy^\epsilon &= \frac{1}{\epsilon}f(x^\epsilon, y^\epsilon)dt + \frac{1}{\sqrt{\epsilon}}g(x^\epsilon, y^\epsilon)dV_t, \end{aligned} \quad (3)$$

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  $\epsilon$ .  $U_t$  and  $V_t$  are mutually independent  $N$ - and  $M$ -dimensional Wiener processes. Systems of the form (3) have been studied extensively within the framework of averaging for SDEs [15, 12]. 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  $\sigma(X)$  in Eq. (1) are defined as the averages of  $a(x^\epsilon, y^\epsilon)$  and  $b(x^\epsilon, y^\epsilon)$  with respect to the measure  $\mu^X(dy^\epsilon)$ ; see, e.g., [20]. Although averaging theory provides theoretical expressions to compute  $\gamma$  and  $\sigma$ , it might not always be possible to obtain analytical expressions. They can, however, be estimated, which has been the topic of extensive research, see, e.g., [17, 1, 16]. For the accelerated simulation of a coarse SDE (1), the coarse drift and diffusion coefficients are estimated from local simulations of the fine-scale model at each coarse time step. In [23, 5, 7], the authors propose to generate a time series of the fast variables by simulating the  $y$ -dynamics for fixed value of  $x^\epsilon$  and compute the unknown coefficients by averaging over this

time series. In [11], the coefficients are estimated from an ensemble averaging of short time series of the slow variables  $X$ . The latter approach has the advantage that the precise form of the underlying model is irrelevant as long as a coarse model of the form (1) exists locally around every point in time.

We investigate the numerical behavior of a general multiscale procedure that uses the above estimation techniques for  $\gamma$  and  $\sigma$  at each time step in a finite difference discretization of the coarse Fokker–Planck equation (2). Hence, the algorithm can be applied to a general class of nonlinear PDEs, with time- or solution dependent coefficients,

$$\rho(X, t)_t = -\nabla_X \cdot (\gamma(X; \rho(X, t)) \rho(X, t)) + \frac{1}{2} \nabla_X \cdot \left[ \nabla_X \cdot \left( \sigma(X; \rho(X, t))^2 \rho(X, t) \right) \right]. \quad (4)$$

Equations of this form occur, for instance, in the context of bacterial chemotaxis models, where  $\gamma$  depends on an external food gradient  $S$ , which in turn is influenced by the local particle density [6, 2]. The analysis, however, will be carried out on the model problem (1)–(2), which greatly simplifies the analysis, while retaining the essential properties of the propagation of the estimation error. The results are thus not limited to cases for which the coarse dynamics is of this form. This work forms a natural extension of the work carried out in the equation-free [13, 14] and heterogeneous multiscale framework [3, 4].

To quantify the uncertainty on the solution, it would be possible to view the procedure in the framework of stochastic PDEs (SPDEs) and use techniques such as a polynomial chaos expansion [25] or collocation methods [24]. A key assumption for these techniques to work in practice, however, is the possibility to expand the stochasticity in the PDE using a limited number of random variables. However, as the estimates are independent in all spatial and temporal discretization points, this corresponds on the coarse PDE level to a white noise process in space and in time. As the latter cannot be represented by an expansion with a finite number of random variables, SPDE uncertainty quantification techniques are not applicable to our case.

This paper is organized as follows. In Section 2, we introduce the estimators for the drift and diffusion coefficients and motivate the main assumption about their statistical properties, required for the analysis. Next, we give a detailed description of the multiscale procedure in Section 3. In sections 4 and 5, we present theoretical stability results and an analysis for the propagation of the estimation error in the coarse solution. The results are illustrated with numerical experiments. Finally, in Section 6 we present conclusions and discuss possible future research directions.

## 2 Drift and diffusion estimation

This section introduces the estimators to compute the coarse drift and diffusion coefficients. The performance of the estimators is then illustrated with a numerical example, which 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 (3), of which all variables  $v = (x^\epsilon, y^\epsilon)$  can be initialized at will. We can then define a “time stepper”  $\phi_{\Delta t}(v(t))$  that returns an (approximate) solution  $v(t + \Delta t)$  for the system with initial condition  $v(t)$ . It is also assumed that there exists a known mapping  $X = \mathcal{M}(x^\epsilon, y^\epsilon) = x^\epsilon$  from the fine-scale variables  $v$  to the slow variables  $X$ .

The coefficients  $\gamma(X)$  and  $\sigma^2(X)$  are estimated from simulations with the fine-scale model, starting from an ensemble of  $R$  initial conditions consistent with  $\mathcal{M}(v_r) = X$ . For the estimates to converge, it must be ensured that the fast variables  $y_r^\epsilon$  sample the invariant measure  $\mu^X(dy^\epsilon)$  of  $y^\epsilon$  for fixed  $x^\epsilon = X$ ; see, e.g., [20, 23, 11]. In [11], the invariant measure is sampled by initializing  $y_r$  randomly and, as it is assumed that  $y^\epsilon$  evolves quickly (over a time scale  $\eta\Delta t$ ) towards its invariant measure, only data from times  $t > t_0 + \eta\Delta t$  is used in the estimation. In [23], the sampling is performed by simulating the  $y$ -dynamics for fixed  $x^\epsilon = X$ .

Here, we will use related estimators based on ensemble averaging, which are particularly suited for systems of the form (3),

$$\begin{aligned}\bar{\gamma}(X) &:= \frac{1}{R\Delta t} \sum_{r=1}^R (\mathcal{M}(\phi_{\Delta t}(X, y_r^\epsilon)) - X) \approx \gamma(X), \\ \bar{\sigma}(X)^2 &:= \frac{1}{R\Delta t} \sum_{r=1}^R (\mathcal{M}(\phi_{\Delta t}(X, y_r^\epsilon)) - X)^2 \approx \sigma(X)^2.\end{aligned}\tag{5}$$

Provided that the initial conditions  $y_r$  represent a good sampling of the invariant measure  $\mu^X(dy)$ , the estimates converge in the limit  $R \rightarrow \infty$ ,  $\epsilon \rightarrow 0$  and  $\Delta t \rightarrow 0$ . We will assume that the measure  $\mu^X(dy^\epsilon)$  can be computed analytically (or numerically) so that we can indeed choose appropriate  $y_r$  by sampling from the corresponding distribution. If the fine-scale system has the form (3), it is then optimal to use  $\Delta t = \delta t$ , namely the time step used for the integration of the detailed multiscale system.

Note that the choice of initial  $y_r$  is the stochastic equivalent of what in equation-free computing is called the *lifting problem* [13]. For the deterministic case, solutions have been proposed for some specific settings [10, 22], but in general, lifting is a difficult problem [8].

**Example 1.** We now illustrate the parameter estimation for a multiscale system of the form (3)

$$\begin{aligned}dx &= (x - x^3 + Ay) dt + By^2 dU_t, \\ dy &= \frac{-y}{\epsilon} dt + \sqrt{\frac{2}{\epsilon}} dV_t,\end{aligned}\tag{6}$$

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  $\delta t = \epsilon^2 = 0.01$ .

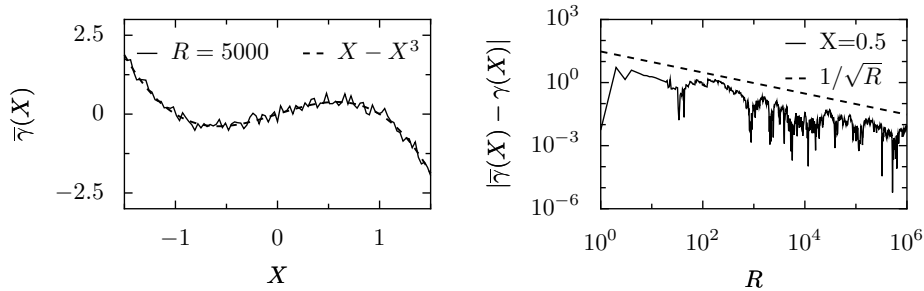


Figure 1: Left: Estimation of the drift for  $X \in [-1.5, 1.5]$ . The exact value (dashed) and the estimate for  $R = 5000$  (solid) are shown. Right: Error on the estimate  $\bar{\gamma}$  with respect to the exact value as a function of the number of replicas  $R$  for  $X_0 = 0.5$ .

The slow dynamics describes the movement of a particle in a bistable symmetrical double-well potential driven by the fast dynamics and  $U_t$ . 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 for small  $\epsilon$  by Eq. (1) with [20]

$$\gamma(X) = X - X^3 \quad \sigma(X) = \sqrt{3}B. \quad (7)$$

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

$$\rho(X, t)_t = -((X - X^3)\rho(X, t))_X + \frac{3B^2}{2}\rho(X, t)_{XX}. \quad (8)$$

We estimate drift  $\gamma(X)$  and diffusion  $\sigma(X)$  using Eq. (5) and compare to the exact values given by Eq. (7). Fig. 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 an appropriate statistical library. The results for the diffusion coefficient are similar (not shown).

Based on the above discussion, we now propose the main assumption that will serve as the basis for the analysis.

**Assumption 2.1.** *The estimated drift and diffusion are the sum of their exact value and a Gaussian distributed estimation error, i.e.,*

$$\begin{aligned} \bar{\gamma}(X) &= \gamma(X) + A_\gamma \cdot \xi_\gamma, \\ \bar{\sigma}(X)^2 &= \sigma(X)^2 + A_\sigma \cdot \xi_\sigma, \end{aligned} \quad (9)$$

with Gaussian variables  $\xi_\gamma, \xi_\sigma \sim \mathcal{N}(0, 1)$ .

The noise amplitudes  $A_\gamma$  and  $A_\sigma$  determine the quality of the estimate and are directly related to the number of realizations in the estimation procedure. Furthermore, the above assumption implies that, at the coarse level, we assume that the estimation errors in each spatial and temporal discretization point are mutually independent.

Note that the analysis does not rely on a specific estimation procedure as it is sufficient that Assumption 2.1 holds. This is the case for systems of the form (3) with estimators (5).

### 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 assuming the form of the equation governing this dynamics and estimating the unknown drift and diffusion coefficients in every spatial discretization point at each time step.

If the slow dynamics can be described by an effective equation Eq. (1), this evolution is given by the corresponding Fokker–Planck (FP) equation (2). For a more general fine-scale description, however, Eq. (1) might not be valid and the drift and diffusion coefficients in Eq. (2) might depend on the solution and the macroscopic equation would look like Eq. (4). After discretization on a coarse grid, (4) is solved using a “classical” PDE solver. Each time the solver requires the value of an unknown coefficient, it is estimated from local simulations with the fine-scale model using the estimators from the previous section. For the analysis, we will consider a finite difference discretization with 1st order upwinding for the convective part and 2nd order central differences for the diffusive part. A forward-Euler scheme is used for the time integration. While the procedure is general, we restrict ourselves to the one-dimensional case  $N = 1$ , so that Eq. (2) takes the form

$$\rho(X, t)_t = -(\gamma(X)\rho(X, t))_X + \frac{1}{2}(\sigma(X)^2\rho(X, t))_{XX}. \quad (10)$$

The discretized equation then becomes

$$\rho_i^n = \rho_i^{n-1} + a_i^n(\rho_n, \Delta X)\Delta t, \quad (11)$$

with

$$\begin{aligned} a_i^n(\rho_n, \Delta X) := & -\frac{1}{\Delta X} (\gamma_i^n \rho_i^n - \gamma_{i-1}^n \rho_{i-1}^n) \\ & + \frac{1}{2\Delta X^2} ((\sigma_{i-1}^n)^2 \rho_{i-1}^n - 2(\sigma_i^n)^2 \rho_i^n + (\sigma_{i+1}^n)^2 \rho_{i+1}^n), \end{aligned} \quad (12)$$

in which for ease of representation we assumed  $\gamma_i^n \geq 0$ . The solution is computed over the computational domain  $\Omega \subset \mathbb{R}$  with Dirichlet boundary conditions  $\rho(x, t) \equiv 0$  for  $x \in \delta\Omega$ .

We can use Assumption 2.1 and substitute the estimated values  $\bar{\gamma}$  and  $\bar{\sigma}$  in Eq. (11) to obtain

$$\rho_i^n = \rho_i^{n-1} + \Delta t L(\rho_n)_i, \quad (13)$$

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

$$\begin{aligned} L(\rho_n)_i = & -\frac{1}{\Delta X} \left( (\gamma_i^n + A_\gamma \xi_{\gamma,i}^n) \rho_i^n - (\gamma_{i-1}^n + A_\gamma \xi_{\gamma,i-1}^n) \rho_{i-1}^n \right) \\ & + \frac{1}{2\Delta X^2} \left( ((\sigma_{i-1}^n)^2 + A_\sigma \xi_{\sigma,i-1}^n) \rho_{i-1}^n - 2((\sigma_i^n)^2 + A_\sigma \xi_{\sigma,i}^n) \rho_i^n \right. \\ & \left. + ((\sigma_{i+1}^n)^2 + A_\sigma \xi_{\sigma,i+1}^n) \rho_{i+1}^n \right), \end{aligned} \quad (14)$$

This expression can be split into a sum of a deterministic and a stochastic part. First, we collect the deterministic terms to recover  $a_i^n(\rho_n, \Delta X)$ .

Next, after collection of the stochastic terms, we obtain

$$\begin{aligned} b_i^n(\rho_n, \Delta X, \xi_\gamma^n, \xi_\sigma^n) := & -\frac{A_\gamma}{\Delta X} (\xi_{\gamma,i}^n \rho_i^n - \xi_{\gamma,i-1}^n \rho_{i-1}^n) \\ & + \frac{A_\sigma}{2\Delta X^2} (\xi_{\sigma,i-1}^n \rho_{i-1}^n - 2\xi_{\sigma,i}^n \rho_i^n + \xi_{\sigma,i+1}^n \rho_{i+1}^n), \end{aligned} \quad (15)$$

where the dependence of  $b_i^n$  on  $A_\gamma$  and  $A_\sigma$  is not made explicit. The RHS operator can then be written as

$$L(\rho_n)_i = a_i^n(\rho_n, \Delta X) + b_i^n(\rho_n, \Delta X, \xi_\gamma^n, \xi_\sigma^n). \quad (16)$$

For future reference, we introduce a short hand notation  $\Phi$  for the above procedure. Written in vector form, this results in

$$\rho_n = \Phi(\rho_{n-1}) := \rho_{n-1} + \Delta t (a_n(\rho_n, \Delta X) + b_n(\rho_n, \Delta X, \xi_\gamma^n, \xi_\sigma^n)), \quad (17)$$

which can also be written as

$$\rho_n = \Phi^n(\rho_0). \quad (18)$$

The use of the estimated coefficients introduces an extra error in the coarse solution of Eq. (2). In the next sections, we investigate this error and analyze the numerical behavior of  $\Phi$ . This includes the quantification of the error propagation and the derivation of a stability criterion.

Note that the chosen discretization might not be optimally suited to compute an accurate solution of the Fokker–Planck equation. However, as our goal is solely to investigate the effect of the estimation errors on the computed result, this is not an issue.

## 4 Stability analysis

It can be numerically established that for fixed  $\Delta t$  and  $\Delta X$  the procedure  $\Phi$  becomes unstable for  $A_\gamma$  and  $A_\sigma$  above a certain critical value of the noise level

$\hat{A}_\gamma$  and  $\hat{A}_\sigma$  respectively. In this section, we examine how Eq. (17) can become unstable and under which conditions on the estimation error a stable algorithm is obtained.

#### 4.1 Derivation of the stability condition

We first derive a theoretical stability criterion to find  $\hat{A}_\gamma$  and  $\hat{A}_\sigma$  for the case of pure diffusion, i.e.,  $\gamma \equiv 0$ . Afterwards, we show numerically that the results carry over to the case of a general Fokker–Planck equation ( $\gamma \neq 0$ ) and that  $\hat{A}_\gamma$  and  $\hat{A}_\sigma$  exhibit identical asymptotic behavior.

Consider Eq. (10) with  $\gamma \equiv 0$  and periodic boundary conditions. Defining the diffusion term  $D(X) := \sigma(X)^2/2$ , we then have

$$\rho(X, t)_t = (D(X)\rho(X, t))_{XX}. \quad (19)$$

Although the exact diffusion  $D$  is chosen to be constant, it is estimated in every spatial discretization point. Using Assumption 2.1, we can then write the estimated diffusion

$$\bar{D}_i = D + A \cdot \xi_i, \quad (20)$$

with  $\xi_i \sim \mathcal{N}(0, 1)$ ,  $i = 0, \dots, m-1$  and  $m$  the number of spatial discretization points. Eq. (17) then becomes

$$\rho_i^{k+1} = \rho_i^k + D\Delta t \frac{\rho_{i-1}^k - 2\rho_i^k + \rho_{i+1}^k}{\Delta X^2} + A\Delta t \frac{\rho_{i-1}^k \xi_{i-1}^k - 2\rho_i^k \xi_i^k + \rho_{i+1}^k \xi_{i+1}^k}{\Delta X^2}. \quad (21)$$

For convenience, we rewrite this expression in matrix form

$$\rho_{k+1} = (I_m + \mu DU) \cdot \rho_k + \mu AU \cdot \begin{bmatrix} \rho_0^k \cdot \xi_0^k \\ \vdots \\ \rho_m^k \cdot \xi_{m-1}^k \end{bmatrix}, \quad (22)$$

with  $I_m \in \mathbb{R}^m$  the unit matrix,  $\mu = \Delta t/\Delta X^2$  and  $U$  the discretization matrix for the diffusion equation with central differences and periodic boundary conditions. We note for later use that  $U$  is a circulant matrix and hence has an orthonormal eigenspace with known eigenvalues  $\lambda_k$  and eigenvectors  $P_k$ ,

$$\lambda_k = -2 + 2 \cos\left(\frac{\pi k}{m}\right), \quad (23)$$

$$P_k = \frac{1}{\sqrt{m}} \left[ 1, e^{-2\pi i k/m}, \dots, e^{-2\pi i k(m-1)/m} \right]^T, \quad k = 0, \dots, m-1. \quad (24)$$

Note that some of the  $\lambda_k$  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  $\Phi$  is called mean-square stable for given  $\Delta t$  and  $\Delta X$  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.

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

**Remark 4.2.** The constant  $C$  in the above definition is typically non-zero due to the addition of estimation errors in every time step. The cumulative effect of this noise prevents that  $\|\rho_n\|_2 \rightarrow 0$  for  $n \rightarrow \infty$ .

For the analysis, we thus need to consider  $\mathbb{E} \left[ \|\rho_n\|_{\Delta x}^2 \right]$  and investigate the result for increasing time  $n$ . Our strategy consists in first making an ansatz for the form of the solution, followed by the derivation of stability conditions from its properties. In particular, we propose an expression of the form

$$\mathbb{E} \left[ \|\rho_n\|_{\Delta X}^2 \right] = \rho_0^T Q_n \rho_0 \cdot \Delta X. \quad (25)$$

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

*Proof.* It can easily be verified that  $H_j$  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  $I_m \in \mathbb{R}^{m \times m}$  the unit matrix, we define the matrix  $M_j$  as it will appear in the analysis below,

$$M_j = U (I_m + \mu D U)^{2j} U. \quad (26)$$

As the discretization matrix  $U$  is circulant, it follows from Lemma 4.3 that all elements on the diagonal of  $M_j$  are equal. We define  $v_j$  to be this value so that

$$\text{diag}(M_j) = v_j \text{diag}(I_m). \quad (27)$$

We are now ready to state the main stability result.

**Theorem 4.4.** *Let  $\rho_n = \Phi^n(\rho_0)$  be computed via Eq. (22). It then holds that*

$$\mathbb{E} \left[ \|\rho_n\|_{\Delta X}^2 \right] = \rho_0^T Q_n \rho_0 \cdot \Delta X, \quad (28)$$

with

$$Q_n = (I_m + \mu D U)^{2n} + (\mu A)^2 \left( \sum_{j=0}^{n-1} v_j Q_{n-j-1} \right), \quad (29)$$

and  $v_j$  given by Eq. (27).

*Proof.* To derive the desired result, we follow a top-down approach and repeatedly apply Eq. (22) to rewrite  $\mathbb{E} \left[ \|\rho_n\|_{\Delta X}^2 \right]$  in terms of the initial density  $\rho_0$ , assuming that  $Q_i$  is known for  $i < n$ . For convenience, we introduce a shorter notation for the stochastic factor in (22),

$$\rho_{\xi,k} := [\rho_0^k \cdot \xi_0^k, \dots, \rho_m^k \cdot \xi_m^k]^T. \quad (30)$$

We have

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

Application of Eq. (22), while taking into account that  $\mathbb{E} [\rho_{\xi,n-1}] \equiv 0$ , results in

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

First consider  $E_2$ . Because of the presence of  $\mathbb{E}[\cdot]$  and the independency of the estimation errors, only the elements on the diagonal of  $U^2$  have a contribution. Using Lemma 4.3 and Eq. (27), this value on the diagonal is equal to  $v_0$ . This term can be written as

$$E_2 = (\mu A)^2 v_0 \cdot \mathbb{E} [\rho_{\xi,n-1}^T \rho_{\xi,n-1}] \Delta X = (\mu A)^2 v_0 \cdot \rho_0^T Q_{n-1} \rho_0 \cdot \Delta X. \quad (31)$$

Next, we turn to  $E_1$  and again apply Eq. (22) to obtain

$$E_1 = \underbrace{\mathbb{E} \left[ \rho_{n-2}^T (I_m + \mu DU)^4 \rho_{n-2} \right] \Delta X}_{E_3} + \underbrace{(\mu A)^2 \mathbb{E} \left[ \rho_{\xi,n-2}^T U (I + \mu DU)^2 U \rho_{\xi,n-2} \right] \Delta X}_{E_4}. \quad (32)$$

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 4.3 all elements on this diagonal are equal and we recognize  $v_1$  from Eq. (27). Substitution in  $E_4$  results in

$$E_4 = (\mu A)^2 v_1 \cdot \mathbb{E} [\rho_{\xi,n-2}^T \rho_{\xi,n-2}] \Delta X = (\mu A)^2 v_1 \cdot \rho_0^T Q_{n-2} \rho_0 \cdot \Delta X. \quad (33)$$

The above process can be repeated for  $E_3$  until the term in  $\rho_0$  is reached, giving rise to extra terms containing  $v_j$ . Using this notation and taking into account that the term in  $\rho_0$  is deterministic, we obtain the desired result

$$\mathbb{E} \left[ \|\rho_n\|_{\Delta X}^2 \right] = \rho_0^T Q_n \rho_0 \cdot \Delta X, \quad (34)$$

with

$$Q_n = (I_m + \mu DU)^{2n} + (\mu A)^2 \left( \sum_{j=0}^{n-1} v_j Q_{n-j-1} \right), \quad (35)$$

and  $v_j$  given by Eq. (27).  $\square$

While the values  $v_j$  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  $U$ ; see Appendix A for the technical details.

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  $\mathbb{R}^m$ , 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,

$$q_n(\lambda_i) = (1 + \mu D \lambda_i)^{2n} + (\mu A)^2 \left( \sum_{j=0}^{n-1} v_j q_{n-j-1}(\lambda_i) \right), \quad i = 1, \dots, m, \quad (36)$$

with  $\lambda_i$  the eigenvalues of  $U$  as given by Eq. (23) and  $q_0(\lambda_i) = 1, \forall \lambda_i$ .

For increasing  $n$ ,  $q_n(\lambda_i)$  contains 2 terms. The 1st term always decreases due to the diffusive damping, while the 2nd term increases exponentially for an unstable algorithm due to the addition of estimation errors in every time step. Interestingly, for *all* modes  $\lambda_i$ ,  $q_n(\lambda_i)$  exhibits the same qualitative behavior, i.e., if the increasing component is present for *any* mode, it can be found for *all* modes  $\lambda_i$ . This fact enables us to formulate the following theorem.

**Theorem 4.5** (Stability). *The algorithm (22) is mean-square stable for given  $\mu = \Delta t / \Delta X^2$  and  $t_n = n \Delta t$  if for any eigenvalue  $\lambda_i$  of  $U$  it holds that  $\Delta q_n(\lambda_i) = q_n(\lambda_i) - q_{n-1}(\lambda_i) \leq 0$ . Otherwise it is unstable.*

Hence, to determine the critical noise level  $\hat{A}$  for given  $\mu$ ,  $D$  and  $n$ , it is sufficient to compute  $A$  so that

$$\Delta q_n(A, \lambda_i) = 0, \quad (37)$$

for any  $\lambda_i$ . In practice it is important to choose  $n$  sufficiently large. The reason for this will be illustrated in Section 4.2.

The above analysis was done for the case  $\gamma \equiv 0$ . Numerical experiments, however, indicate that the asymptotic behavior of  $\hat{A}_\gamma$  and  $\hat{A}_\sigma$  is identical when  $\gamma \neq 0$ ; see Section 4.2 below.

**Remark 4.6.** To validate the above stability results, it is necessary to have an numerical estimate of  $\hat{A}_\gamma$  and  $\hat{A}_\sigma$ . Based on Definition 4.1, we define the notion of practical stability, which can then be used to numerically compute these values.

**Definition 4.7** (Practical stability). *Algorithm  $\Phi_n$  is called practically stable for given  $\Delta t$ ,  $\Delta X$  and  $t_n$  if  $\|\rho(t_n)\|_{\Delta X} \leq C$ , with  $C$  a constant. Otherwise it is called unstable.*

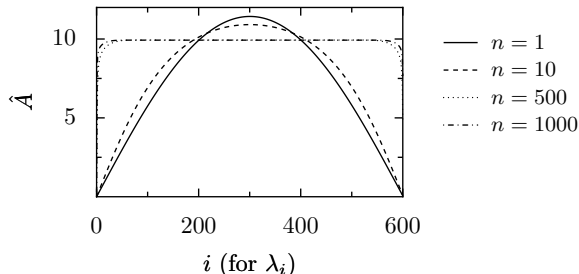


Figure 2: Maximal noise level  $\hat{A}$  for stability as a function of the eigenvalue  $\lambda_i$  (see Eq. (23)) for different values of  $n$ . Low frequency modes correspond to  $i$  near the end points of the  $i$ -axis, high frequency modes correspond to  $i$  near the middle.

## 4.2 Numerical example

**Pure diffusion.** We now present numerical validation of the stability results for (19) with estimated diffusion coefficient  $D$ . For efficient testing, we assume that the exact diffusion  $D$  is known and use Eq. (20) 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 first illustrate Theorem 4.5 by showing the maximal noise level  $\hat{A}$  allowed for stability for each eigenvalue  $\lambda_i$  of  $U$  and its dependence on the time horizon  $t_n = n\Delta t$ . The result for  $D = 1$ ,  $\mu = \Delta t/\Delta X^2 = 0.01$  and  $m = 601$  is given in Fig. 2. We see that, for small  $n$ , the maximal noise level can be larger for the modes that belong to  $\lambda_i$  near the middle of the  $i$ -axis (high frequencies); for modes that belong to  $\lambda_i$  near the endpoints of the axis (low frequencies),  $\hat{A}$  is smaller. However, we also observe that for increasing  $n$  this difference decreases and for large enough  $n$  the maximal noise level is nearly equal for all frequencies. Hence, if  $n$  is chosen sufficiently large,  $\hat{A}$  can be computed using only a single  $\lambda_i$ , e.g.,  $\lambda = -4$  for the mode with highest frequency.

Next we compare the experimentally determined  $\hat{A}$  with its predicted value. For the computation of the experimental value, we use Definition 4.7 with  $C = 3$  and determine  $\hat{A}$  iteratively via a robust bisection method. 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 Fig. 3 (left),  $\hat{A}$  is shown as a function of  $\mu$ . The diffusion coefficient is chosen as in Example 1, i.e.,  $D = 3B^2 = 0.375$ . There is good correspondence between the experimental ( $\square$ ) and theoretical ( $\triangle$ ) values for all  $\mu$ . Furthermore, away from the deterministic stability condition  $D\mu < 0.5$  (dash-dotted) we observe that  $\hat{A} \propto 1/\sqrt{\mu}$ . This behavior is persistent for other time integrators such as, e.g., a 4th order Runge–Kutta integrator (results not shown). Fig. 3 (right) shows the dependence of  $\hat{A}$  on  $D$  for  $\mu = 0.05$  and  $m = 380$ . Again, we find good correspondence between the experimental (solid) and theoretical (dashed) results. For small  $D$ , the added noise in every time step is only slightly damped,

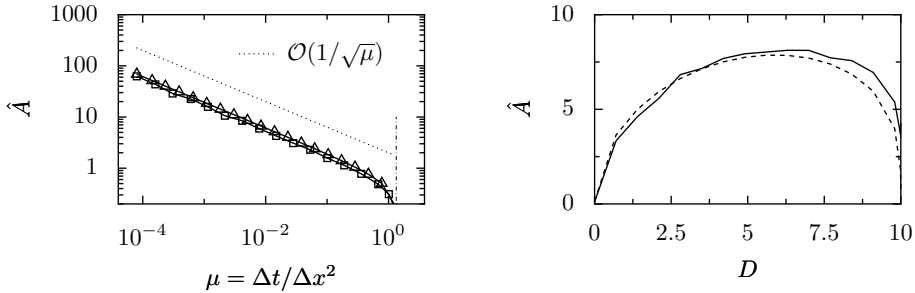


Figure 3: Left: Dependence of the experimental ( $\square$ ) and theoretical ( $\triangle$ ) value of  $\hat{A}$  on  $\mu = \Delta t / \Delta X^2$  for  $D = \sqrt{3}B$  and  $t_e = 10$ . The experimental value is determined via Th. 4.7 with  $C = 3$ . The theoretical value is the solution of Eq. (37) for  $n = 10/\Delta t$ . Also shown are the  $1/\sqrt{\mu}$  line (dotted) and the deterministic stability boundary  $D\mu = 0.5$  (dash-dotted). Right: Dependence on  $D$  of the experimental ( $\square$ ) and theoretical ( $\triangle$ ) value of  $\hat{A}$  for  $\mu = 0.05$  and  $t_e = 10$ .

and the method becomes unstable quickly ( $\hat{A}$  is small). The same is true for larger  $D$ , as in that case, we are close to the deterministic stability condition.

**Fokker–Planck equation.** To illustrate how the results from Section 4.1 generalize to the case of a general Fokker–Planck equation with  $\gamma \neq 0$ , we now return to Example 1 of Section 2. Recall that the Fokker–Planck equation for this example is given by Eq. (8). We compute the solution at  $t_n = n\Delta t$  over the computational domain  $\Omega = [-6, 6]$ . The error is defined as the difference between the computed solution and the exact deterministic solution. 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  $\|\rho_0\|_{\Delta X} = 1$ .

We now present numerical results for the dependence of  $\hat{A}$  on  $\mu = \Delta t / \Delta X^2$ . Its value is computed via Theorem 4.7 with  $C = 3$ ; see Fig. 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 both  $A_\sigma \neq 0$  and  $A_\gamma \neq 0$ . The lines almost coincide, which supports our earlier claim 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. 3).

### 4.3 Discussion

We have derived a stability condition for the algorithm  $\Phi$  (see Section 3). This was done by relating the expected norm of the solution at time  $t_n$  to the initial condition  $\rho_0$  via a matrix  $Q_n$ . By studying the properties of this matrix and imposing that its norm should not increase as a function of  $n$ , we obtained an equation (37), from which the maximum noise level could be computed. We note that our derivation is restricted to the case  $\gamma \equiv 0$ , whereas Section 3 deals

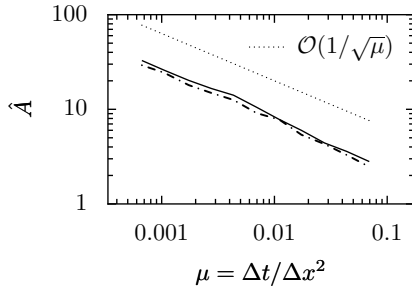


Figure 4: Dependence on  $\mu$  of the experimentally determined maximum noise level  $\hat{A}$  for  $D = \sqrt{3}B$  and  $t_e = 10$ . The value is computed via Th. 4.7 with  $C = 3$ . The result for perturbations on the diffusion coefficient only (solid) and on both diffusion and drift (dash-dotted) are plotted. Also shown is the  $1/\sqrt{\mu}$  line (dotted).

with a general FP equation, with both a convective and diffusive part. However, as shown in Section 4.2, the stability results extend to the general case in the sense that the asymptotic behavior of  $\hat{A}(\mu)$  in both cases is identical. This leads us to conclude that stability of  $\Phi$  is dominated by the error on the diffusion.

If for some reason the exact diffusion in the Fokker–Planck equation is known, and only the drift has to be estimated, the situation changes slightly. This situation could be analyzed in a way similar to the one given in Section 4.1, but as typically *both* drift and diffusion are unknown, we do not explicitly consider this case here. For simplicity we also assumed a constant diffusion coefficient  $D$ . In the case that  $D(X)$  depends on space, it is also possible to compute the maximal noise level  $\hat{A}^X$ . From a derivation along the line of the one given in the proof of Theorem 4.4, we obtain an upper bound for  $\hat{A}^X$  by solving Eq. (37) in which we use an effective *constant* diffusion  $D_e = \min_i D(X_i)$ .

## 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, it will be required to limit ourselves to the case of a forward-Euler scheme with  $\gamma \equiv 0$ . At the end of the section, however, we discuss how our conclusions generalize to other time integrators and to the general Fokker–Planck equation with  $\gamma \neq 0$ .

### 5.1 Consistency

We start by exploring consistency of  $\Phi$  for Eq. (17) and consider the corresponding discrete RHS operator (16) in each spatial discretization point  $i$  and at time

$t_n$ . As we know that in the deterministic case the integrator is consistent, we only consider the stochastic term  $b_i^n(\rho_n, \Delta X, \xi_\gamma^n, \xi_\sigma^n)$ ; see Eq. (15). As  $\xi_{\gamma,i}^n$  and  $\xi_{\sigma,i}^n$  are independent Gaussian random variables for all  $i$ ,  $b_i^n(\rho_n, \Delta X, \xi_\gamma^n, \xi_\sigma^n)$  is a Gaussian variable as well, so the discrete RHS operator for each spatial discretization point (16) can be rewritten as

$$L(\rho^n)_i = a_i^n(\rho_n, \Delta X) + b_i^n(\rho_n, \Delta X)\xi_i^n, \quad (38)$$

with  $\xi_i^n \sim \mathcal{N}(0, 1)$ ,  $a_i^n$  as in Eq. (12) and  $b_i^n(\rho_n, \Delta X)$  given by

$$b_i^n(\rho_n, \Delta X) = \left( \frac{\left( (\rho_{i+1}^n)^2 + (\rho_{i-1}^n)^2 \right) A_\gamma^2}{\Delta X^2} + \frac{\left( (\rho_{i-1}^n)^2 + 4(\rho_i^n)^2 + (\rho_{i+1}^n)^2 \right) A_\sigma^2}{4\Delta X^4} \right)^{1/2}. \quad (39)$$

Since the stochastic part disappears under the expected value operator, Eq. (38) shows that  $\Phi$  is *expected* to be consistent for all  $t_n$  and for all choices of  $\Delta X$ . For decreasing  $\Delta X$ , 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  $A_\gamma$  and  $A_\sigma$  depend on  $\Delta X$ . How this manifests itself on the coarse level will be discussed in more detail in the error analysis below.

## 5.2 Error propagation

We investigate the propagation of the estimation error through the procedure  $\Phi$ . To this end, we compare the computed solution with the deterministic solution at time  $t_e$ . As already mentioned, the analysis will be done for pure diffusion with  $\gamma \equiv 0$ .

**Theorem 5.1** (Error propagation). *Let  $\rho_n$  be the discretized density at  $t = t_e$  computed via Eq. (22). If we have stability of both the deterministic integrator ( $\mu D < 0.5$ ) and  $\Phi$  (see Theorem 4.5, then the asymptotic behavior of the expected value of the squared error on  $\rho_n$  with respect to the exact, deterministic solution  $\hat{\rho}_n$  is of the form*

$$\mathbb{E} \left[ \|\hat{\rho}_n - \rho_n\|_{\Delta X}^2 \right] \leq \mathcal{O}(\Delta t)^2 + \mathcal{O}(\Delta X^2)^2 + \frac{A^2 M_Q}{D} \frac{\Delta t}{\Delta X^2} \frac{2}{1 - 2\frac{\Delta t}{\Delta X^2} D} \cdot \|\rho_0\|_{\Delta X}^2, \quad (40)$$

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 2nd term contains  $\rho_n$  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  $\tilde{\rho}_n$  at time  $t_e$  computed with the deterministic solver, which for now we assume to be of order  $p$  in  $\Delta t$ . We then have

$$\begin{aligned}\mathbb{E} \left[ \|\hat{\rho}_n - \rho_n\|_{\Delta X}^2 \right] &= \hat{\rho}_n^T \hat{\rho}_n \Delta X - 2\hat{\rho}_n^T \tilde{\rho}_n \Delta X + \mathbb{E} [\rho_n^T \rho_n] \Delta X, \\ &= \underbrace{(\hat{\rho}_n - \tilde{\rho}_n)^T (\hat{\rho}_n - \tilde{\rho}_n) \Delta X}_{E_1} - \underbrace{\tilde{\rho}_n^T \tilde{\rho}_n \Delta X + \mathbb{E} [\rho_n^T \rho_n] \Delta X}_{E_2}.\end{aligned}\tag{41}$$

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  $\mathcal{O}(\Delta t^p)^2 + \mathcal{O}(\Delta X^2)^2$ .

For the stochastic part  $E_2$ , we again consider a forward-Euler scheme for the diffusion equation with estimated diffusion  $\bar{D}$ ; see Eq. (22). We obtain

$$\begin{aligned}E_2 &= -\rho_0^T (I + \mu DU)^{2n} \rho_0 \cdot \Delta X + \rho_0^T \left[ (I + \mu DU)^{2n} + (\mu A)^2 \left( \sum_{j=0}^{n-1} v_{n-j-1} Q_j \right) \right] \rho_0 \cdot \Delta X, \\ &= (\mu A)^2 \cdot \rho_0^T \left( \sum_{j=0}^{n-1} v_j Q_{n-j-1} \right) \rho_0 \cdot \Delta X\end{aligned}\tag{42}$$

with  $Q_{n-j-1}$  and  $v_j$  as in Theorem 4.4. For a stable integration,  $\|Q_j\|_2$  is bounded for all  $j$  and for all  $\Delta t$  and  $\Delta X$ . If we denote this maximum value by  $M_Q$ , we obtain

$$E_2 \leq (\mu A)^2 M_Q \left( \sum_{j=0}^{n-1} v_j \right) \cdot \|\rho_0\|_{\Delta X}^2.\tag{43}$$

Now consider only the sum over  $v_j$ . Using the alternative expression for  $v_j$  of Eq. (48), this can be written as

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

with  $\lambda_k$  the eigenvalues of  $U$ . After substitution in Eq. (43) and provided that  $n$  is large, we obtain

$$E_2 \leq (\mu A)^2 M_Q \frac{1}{m} \sum_{k=0}^{m-1} \lambda_k^2 \frac{1}{1 - (1 + \mu D \lambda_k)^2} \cdot \|\rho_0\|_{\Delta X}^2,\tag{44}$$

which can be bounded using the analytical expression for  $\lambda_k$  (see Eq. (23)),

$$\begin{aligned} E_2 &\leq (\mu A)^2 M_Q \frac{1}{m} \sum_{k=0}^{m-1} \frac{(-4)^2}{1 - (1 - 4\mu D)^2} \cdot \|\rho_0\|_{\Delta X}^2, \\ &\leq \frac{A^2 M_Q}{D} \frac{2\mu}{1 - 2\mu D} \cdot \|\rho_0\|_{\Delta X}^2. \end{aligned} \quad (45)$$

Finally, substitution of this expression in Eq. (41) concludes the proof.  $\square$

Note that only to deal with the stochastic part of Eq. (41), 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

**Pure diffusion.** To illustrate the error bound of Eq. (40), we consider the diffusion equation with estimated diffusion, see Eq. (19). We then compute the error for  $\Phi$  as a function of  $\Delta t$  and  $\Delta X$  with  $D = 3B^2 = 0.375$ . Fig. 5 (left) shows the error as a function of  $\Delta t$  for fixed  $\Delta X = 0.1$  and  $t_e = 10$ . The reference solution was determined 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  $\Delta t$  there is deterministic convergence and for small  $\Delta t$  the slower, stochastic convergence becomes apparent. We remark that for other time integrators such as, e.g., a 4th order Runge–Kutta scheme, we find similar behavior. For large noise level we find  $\mathcal{O}(\Delta t)$ , while for smaller values the deterministic  $\mathcal{O}(\Delta t^8)$  convergence is recovered.

In Fig. 5 (right), the dependence of the error on  $\Delta X$  is shown for fixed  $t_e = 10$ . For constant  $A = 0.2$  (solid) we observe that away from the deterministic stability boundary  $D\mu < 0.5$ , i.e., for large  $\Delta X$ , the error increases proportionally to  $\mathcal{O}(1/\Delta X^2)$  with decreasing  $\Delta X$ , as predicted by the bound in Eq. (40) (dash-dotted). Furthermore, we know from Eq. (40) that it is possible to obtain convergence for  $\Delta X \rightarrow 0$  by increasing the quality of the estimates while decreasing  $\Delta X$ . For  $A \propto \Delta X$  (thick dotted), the error is more or less constant in  $\Delta X$ . For  $A \propto \Delta X^2$  (dashed), we observe convergence as  $\mathcal{O}(\Delta X^2)$ . For small  $\Delta X$ , while approaching the deterministic stability boundary, there is divergence as  $1/(1 - 2\mu D)$  for all cases.

**Fokker–Planck equation.** We now illustrate the error propagation for the general case  $\gamma \neq 0$ . Again we consider the Fokker–Planck equation (8) from Example 1 and apply  $\Phi$ . In Fig. 6 (left), we show the dependence of the error on  $\Delta t$  for forward-Euler with  $\Delta X = 0.1$  and  $t_e = 10$ . Fig. 7 shows the results for the same experiment with a 4th order Runge–Kutta integrator. It is found that

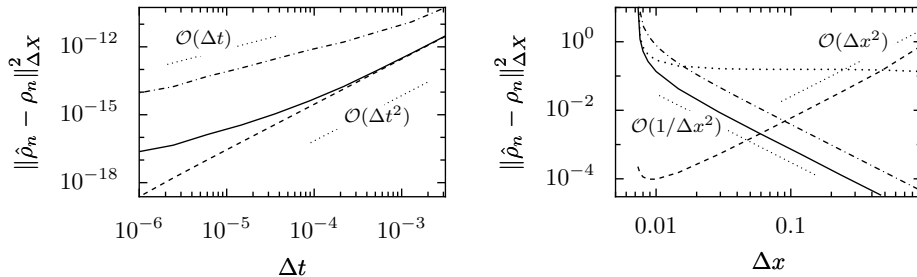


Figure 5: Left: Dependence of the error at  $t_e = 10$  on the time step  $\Delta t$  for a forward-Euler integration of a pure diffusion equation with  $D = 0.375$  and for different noise levels  $A = 0$  (dashed),  $5 \times 10^{-7}$  (solid) and  $10^{-5}$  (dash-dotted). Shown in dotted line are the  $\mathcal{O}(\Delta t)$  and  $\mathcal{O}(\Delta t^2)$  lines. Right: Dependence of the error at  $t_e = 1$  on  $\Delta X$  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 Eq. (45) is shown in dash-dotted line. Shown in dotted lines are the  $\mathcal{O}(1/\Delta X^2)$  and  $\mathcal{O}(\Delta X^2)$  lines.

indeed for small noise level, 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  $\Delta X$  with constant  $t_e = 1$ ; see Fig. 6(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  $\Delta X$ , 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 convergence rate  $\mathcal{O}(\Delta X^2)$ . Therefore, we conclude that the error propagation and convergence behavior that was derived theoretically for the case of pure diffusion, is recovered for the case of a general Fokker–Planck equation.

## 5.4 Discussion

In this section, we analyzed the propagation of the estimation error in the numerical solution of Eq. (10) 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 Eq. (40), 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  $\Delta X$  is clearly visible in the theoretical expression for the dependence of the error on  $\Delta X$  for fixed  $\Delta t$ . To be able to refine the spatial discretization error with fixed  $\Delta t$ , without at the same time destroying the numerical result, the quality of the estimates must increase as  $\Delta X$  decreases. For instance, to obtain the deterministic 2nd

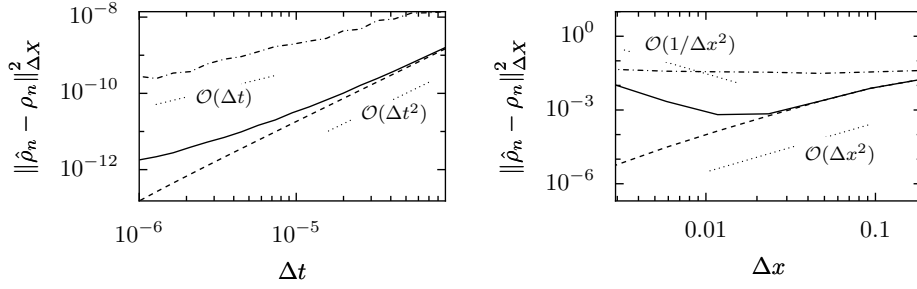


Figure 6: Left: Dependence of the error at  $t_e = 10$  on the time step  $\Delta t$  for a forward-Euler integration of a Fokker–Planck equation with  $D = 0.375$  and for different noise levels  $A = A_\gamma = A_\sigma$ . Used values are  $A = 0$  (dashed),  $5 \times 10^{-7}$  (solid) and  $10^{-5}$  (dash-dotted). Also shown in dotted line are the  $\mathcal{O}(\Delta t)$  and  $\mathcal{O}(\Delta t^2)$  lines. Right: Dependence of the error at  $t_e = 1$  on  $\Delta x$  for a forward-Euler integration of a pure diffusion equation. Results are shown for constant  $A = 0.15$  (solid),  $A \propto \Delta x$  (dash-dotted) and  $A \propto \Delta x^2$  (dashed). Also shown are the  $\mathcal{O}(1/\Delta x^2)$  and  $\mathcal{O}(\Delta x^2)$  lines.

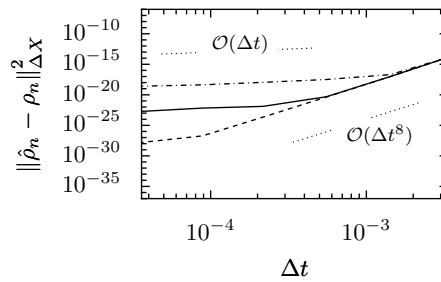


Figure 7: Dependence of the error at  $t_e = 10$  on the time step  $\Delta t$  for a 4th order Runge–Kutta integration of a Fokker–Planck equation with  $D = 0.375$  and for different noise levels  $A = A_\gamma = A_\sigma$ . Used values are  $A = 0$  (dashed),  $10^{-8}$  (dash-dotted) and  $10^{-10}$  (solid). Also shown in dotted line are the  $\mathcal{O}(\Delta t)$  and  $\mathcal{O}(\Delta t^8)$  lines.

order convergence, the estimation error must behave like  $A \propto \Delta X^3$ , without violating the stability condition on  $\Phi$ . 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 6, we discuss possible ways to improve on this point.

For fixed  $\Delta X$ , Eq. (40) 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  $\Delta t$ .

The results for the error propagation were only rigorously established for a forward-Euler scheme and the case of pure diffusion ( $\gamma \equiv 0$ ). However, the results can be generalized. First, similar asymptotic convergence behavior can be observed for other time integration schemes. For a general  $p$ th-order scheme, this is in fact reflected in the first term of Eq. (40) which then becomes  $\mathcal{O}(\Delta t^p)^2$ . Secondly, it was shown in Section 5.3 that the convergence behavior for the solution of a general Fokker–Planck equation (with  $\gamma \neq 0$ ) is identical to the case of pure diffusion.

## 6 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 a general form for 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 solution- or 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 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 assumed that the estimation error followed a Gaussian distribution. This assumption, however, is only valid if good estimators are available.

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 rigorously derived for the case of pure diffusion. The addition of a drift term, as in the general Fokker–Planck equation, did not change the asymptotic behavior, indicating that stability for the Fokker–Planck case is diffusion dominated.

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  $\Delta t$ , the stochastic effects result in slow convergence. For small estimation error the order of the deterministic integrator is recovered. Although the analysis was performed for the diffusion equation and a forward-Euler integrator, this behavior turned out to be persistent for the Fokker–Planck equation and for other 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 estimation coefficients. This idea will be explored in future research.

The procedure as presented in this paper is kept general so that in principle it can be applied for fine-scale models described by particle models or black box simulators. For a particle-based bacterial chemotaxis model [6] with (indirect) particle interaction, the evolution of the density is given by a FP-like equation, with drift and diffusion now depending on the local density (and hence on time). The main challenge to be able to apply our procedure, is twofold. First, to estimate the coefficients the particles should be initialized locally in small boxes around the discretization points with the correct density [9]. Additionally, every particle should be assigned a “correct” internal state, i.e., consistent with the local density *and* its dynamics. It might be possible to achieve this via a constrained runs algorithm [10], but the solution to this *lifting problem* is all but straightforward. Secondly, finding good estimators for the unknown coefficients might be more involved. Their convergence might be slow due to large variance on the numerical results and many realizations might be required to obtain a sufficiently accurate estimate. For an efficient algorithm, specialized variance reduction techniques will prove unavoidable [19].

## Acknowledgments

The authors are grateful to D. Givon 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).

## Appendix

### A Properties and computation of $v_i$

Eq. (27) introduces the terms  $v_i$  that appear in the theoretical stability result in Theorem 4.5. We stated that in practice  $n$  should be large, and as a result, it is not feasible to carry out the  $2i + 1$  multiplications of possibly large matrices to compute their value. Here, we show an alternative (and cheaper) way to practically compute  $v_i$ , based on a spectral decomposition of the discretization matrix  $U$ .

Let  $U = P\Lambda P^T$  be the eigenvalue decomposition of  $U$ , with the eigenvalues and eigenvectors given by Eq. (23)-(24). We normalize the columns of  $P$  so that  $PP^T = P^T P = I_m$ , with identity matrix  $I_m \in \mathbb{R}^{m \times m}$  and  $m$  the number of spatial discretization points.

In Eq. (26), we defined the matrix  $M_i$ , for which we know that all diagonal elements  $v_i$  are equal. It is their value  $v_i$  that we are looking for. Substituting the decomposition of  $U$  in Eq. (26) results in

$$M_i = P\Lambda P^T (I_m + \mu DP\Lambda P^T)^{2i} P\Lambda P^T \quad (46)$$

The first diagonal element is equal to  $v_i$  and can, e.g., be extracted by multiplication on the left and right by  $e_1 = [1, 0, \dots, 0]^T$ ,

$$\begin{aligned} v_i &= e_1^T \cdot P\Lambda P^T (I_m + \mu DP\Lambda P^T)^{2i} P\Lambda P^T \cdot e_1 \\ &= \bar{P}_j^T \Lambda P^T (I_m + \mu DP\Lambda P^T)^{2i} P\Lambda \bar{P}_j \\ &= \bar{P}_j^T \Lambda (I_m + \mu D\Lambda)^{2i} \Lambda \bar{P}_j, \end{aligned} \quad (47)$$

with  $\bar{P}_j$  the  $j$ th row of  $P$ . It is important to remark at this point that this expression is valid *for all* choices of  $j$ . As we know from Eq. (24) that the first elements of the eigenvectors are all equal to  $1/\sqrt{m}$ , a good choice is  $\bar{P}_j = [1, \dots, 1]^T / \sqrt{m}$ . After substitution in Eq. (47) we obtain the following formula for  $v_i$ :

$$v_i = \frac{1}{m} \sum_{k=0}^{m-1} \lambda_k (1 + \mu D\lambda_k)^{2i} \lambda_k \quad (48)$$

Note that as the factor  $(1 + \mu D\lambda_k)^{2i}$  deals with diffusion, the series  $v_i$  is strictly decreasing with maximum at  $V_0 = 6$ .

## References

- [1] Y. Ait-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.

- [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. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 73(5):051901, 2006.
- [3] W. E and B. Engquist. The heterogeneous multi-scale methods. *Comm. Math. Sci.*, 1(1):87–132, 2003.
- [4] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden. Heterogeneous multiscale methods: A review. *Communications in Computational Physics*, 2(3):367–450, 2007.
- [5] W. E, D. Liu, and E. Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Communications on Pure and Applied Mathematics*, 58(11):1544–1585, 2005.
- [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 Modeling and Simulation*, 3(2):362–394, 2005.
- [7] I. Fatkullin and E. Vanden-Eijnden. A computational strategy for multiscale systems with applications to Lorenz 96 model. *J. Comput. Phys.*, 200(2):605–638, 2004.
- [8] 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 and Continuous Dynamical Systems Series B*, 11(4):855–874, June 2009. Preprint available at <http://www.cs.kuleuven.be/publicaties/rapporten/tw/TW525.pdf>.
- [9] C. Gear. Projective integration methods for distributions. Technical report, NEC Research Institute, nov 2001.
- [10] C. W. Gear, T. J. Kaper, I. G. Kevrekidis, and A. Zagaris. Projecting to a slow manifold: Singularly perturbed systems and legacy codes. *SIAM Journal of Applied Dynamical Systems*, 4(3):711–732, 2005.
- [11] D. Givon and I. G. Kevrekidis. Coarse-grained projective schemes for certain singularly perturbed stochastic problems. Submitted, 2008.
- [12] D. Givon, I. G. Kevrekidis, and R. Kupferman. Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. *Comm. Math. Sci.*, 4(4):707–729, 2006.
- [13] 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. *Comm. Math. Sci.*, 1(4):715–762, 2003.

- [14] I. G. Kevrekidis and G. Samaey. Equation-free multiscale computation: Algorithms and applications. *Annual Review of Physical Chemistry*, 60(1):321–344, 2009. PMID: 19335220.
- [15] G. Pavliotis and A. Stuart. *Multiscale Methods: Averaging and Homogenization*, volume 53 of *Texts in Applied Mathematics*. Springer New York, 2007.
- [16] G. A. Pavliotis and A. M. Stuart. Parameter estimation for multiscale diffusions. *Journal of Statistical Physics*, 127(4):741–781, 2007.
- [17] Y. Pokern, A. M. Stuart, and E. Vanden-Eijnden. Nonparametric drift estimation for diffusion processes. Accepted in *SIAM Multiscale Modelling and Simulation*, 2009.
- [18] H. Risken. *The Fokker-Planck Equation: Methods of Solutions and Applications*. Springer Series in Synergetics. Springer, 2 edition, 1989.
- [19] M. Rousset and G. Samaey. Individual-based models for bacterial chemotaxis and variance reduced simulations. Submitted, 2009.
- [20] A. Skorokhod. *Asymptotic methods in the theory of stochastic differential equations*, volume 78 of *Translations of mathematical monographs*. AMS, Providence, RI, 1999.
- [21] N. van Kampen. Elimination of fast variables. *Physics Reports*, 124(2):69–160, 1985.
- [22] P. Van Leemput, W. Vanroose, and D. Roose. Mesoscale analysis of the equation-free constrained runs initialization scheme. *SIAM Multiscale Modeling and Simulation*, 6(4):1234–1255, 2007.
- [23] E. Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.*, 1(2):385–391, 2003.
- [24] D. Xiu and J. S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.
- [25] D. Xiu and G. E. Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.