COMM. MATH. SCI.
Vol. 1, No. 2, pp. 385–391

FAST COMMUNICATIONS

NUMERICAL TECHNIQUES FOR MULTI-SCALE DYNAMICAL SYSTEMS WITH STOCHASTIC EFFECTS*

ERIC VANDEN-EIJNDEN†

Abstract. Numerical schemes are presented for dynamical systems with multiple time-scales. Two classes of methods are discussed, depending on the time interval on which the evolution of the slow variables in the system is sought. On rather short time intervals, the slow variables satisfy ordinary differential equations. On longer time intervals, however, fluctuations become important, and stochastic differential equations are obtained. In both cases, the numerical methods compute the evolution of the slow variables without having to derive explicitly the effective equations beforehand; rather, the coefficients entering these equations are obtained on the fly using short simulations of appropriate auxiliary systems.

1. Introduction

Systems evolving on widely separated time-scales represent a challenge for numerical simulations. A generic example is

$$\begin{cases} \dot{X}_t^\varepsilon = f(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon), & X_{t=0}^\varepsilon = x \\ \dot{Y}_t^\varepsilon = \dfrac{1}{\varepsilon} g(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon), & Y_{t=0}^\varepsilon = y, \end{cases} \tag{1.1}$$

where f and g are of order one, ε is a small parameter measuring the separation of time-scale in the system, and we have assumed that the phase-space can be decomposed into slow degrees of freedom, x, and fast ones, y. Systems of this type arise, e.g., in molecular dynamics, material sciences, atmosphere-ocean sciences, etc. Standard computational schemes fail due to the wide separation between the $O(\varepsilon)$ time-scale one must compute with, and the $O(1)$, or even $O(\varepsilon^{-1})$, time-scales one is typically interested in analyzing the solutions.

On analytical grounds, the following is known about (1.1) (see, e.g., [1, 7] and references therein). If the dynamics for Y_t^ε alone at $X_t^\varepsilon = x$ fixed has an invariant measure, $\mu_x^\varepsilon(dy)$, and in addition

$$\lim_{\varepsilon \to 0} \int f(x, y, \varepsilon) \mu_x^\varepsilon(dy) = B(x), \tag{1.2}$$

then in the limit as $\varepsilon \to 0$, X_t^ε converges to the solution of the following equation:

$$\dot{X}_t = B(X_t), \qquad X_{t=0} = x. \tag{1.3}$$

This effective equation is valid on bounded time-intervals, $t \in [0, T)$ with $T \leq \infty$; however, this equation becomes invalid if the evolution of X_t^ε is sought on very large time intervals, of the order of ε^{-1}. In particular, it is always inappropriate if

$$\int f(x, y, \varepsilon) \mu_x^\varepsilon(dy) = O(\varepsilon), \tag{1.4}$$

* Received: January 6, 2003; accepted: January 7, 2003.
† Courant Institute, New York University, New York, NY 10012, and Institute for Adanced Study, Princeton, NJ 08540.

since in these situations the evolution of X_t arises only on the ε^{-1}-time-scale. Over such long time intervals, fluctuations become important, and the effective equation for X_t^ε in the limit of small ε on the rescaled time $s = t\varepsilon$ is a stochastic differential equation,

$$dX_s = b(X_s)ds + \sigma(X_s)dW_s, \qquad X_{s=0} = x, \tag{1.5}$$

whose coefficients b and σ are expressed in terms of limits of expectations similar to (1.2) – see below.

(1.3) or (1.5) can be used directly to simulate the evolution of X_t^ε, provided one is able to compute analytically the coefficients $B(x)$, $b(x)$, and $\sigma(x)$ entering these equations. It might be the case, however, that this cannot be done because the dynamics of the fast variables Y_t^ε are too complicated. One way to proceed then is to make some suitable approximations on (1.1) to simplify the dynamics of Y_t^ε so as to make possible the calculation of $B(x)$, $b(x)$, and $\sigma(x)$ (see, e.g., [6] for application of this philosophy to problems arising in atmosphere-ocean science). Another way, which is what this paper is about, is to estimate $B(x)$, $b(x)$, and $\sigma(x)$ numerically, via simulation of the equation for Y_t^ε (or some auxiliary variables, suitably defined) at $X_t^\varepsilon = x$ fixed on the rescaled time-scale $\tau = t/\varepsilon$. For small ε, such numerical schemes are advantageous since Y_t^ε reaches equilibrium on a interval of time which is smaller than the time-step used to evolve X_t; therefore, the simulation of Y_t^ε is only necessary on a small portion of the time interval one wishes to evolve X_t. Note that a similar philosophy is applied in [3, 4, 5] for the deterministic situations on short time intervals; to the best of our knowledge, no schemes of this type have been proposed for the stochastic situations on longer time intervals.

2. Short time intervals

Consider

$$\begin{cases} \dot{X}_t^\varepsilon = f(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon), & X_{t=0}^\varepsilon = x \\ \dot{Y}_t^\varepsilon = \dfrac{1}{\varepsilon} g(X_t^\varepsilon, Y_t^\varepsilon, t/\varepsilon, \varepsilon), & Y_{t=0}^\varepsilon = y, \end{cases} \tag{2.1}$$

where $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$, and f is a differentiable function. Assume that for each $x \in \mathbb{R}^n$,

$$\dot{Z}_\tau^\varepsilon = g(x, Z_\tau^\varepsilon, \tau, \varepsilon), \tag{2.2}$$

generates a stationary (ergodic) Markov process with invariant measure $\mu_x^\varepsilon(dy)$. (The statement below is rigorous in this case, but one has also in mind situations where the Markovianity, or even the randomness, of Z_τ is only approximated, for instance if (2.2) displays deterministic chaos with short memory effects.) Assume in addition that (1.2) holds. Then, for $t \in [0, T]$, with $0 \le T < \infty$, $\tau \in [0, T']$, with $0 \le T' < \infty$, and for any test function φ, one has

$$\left| \mathbf{E}\varphi(X_t^\varepsilon, Y_{t+\varepsilon\tau}^\varepsilon) - \mathbf{E}\varphi(\bar{X}_t^\varepsilon, Z_\tau^\varepsilon) \right| \to 0, \tag{2.3}$$

as $\varepsilon \to 0$. Here $\bar{X}_t^\varepsilon$ satisfies the ordinary differential equation:

$$\dot{\bar{X}}_t^\varepsilon = B(\bar{X}_t^\varepsilon, \varepsilon), \qquad \bar{X}_{t=0}^\varepsilon = x, \tag{2.4}$$

where

$$B(x, \varepsilon) = \int f(x, y, \varepsilon)\mu_x^\varepsilon(dy); \tag{2.5}$$

Z^ε_τ satisfies (2.2) with $x = \bar{X}^\varepsilon_t$ fixed and initial condition $Z^\varepsilon_{\tau=0}$ distributed according to $\mu^\varepsilon_{x=\bar{X}^\varepsilon_t}(dy)$. This result is a generalization of the one stated in the introduction – the common limit of X^ε_t and $\bar{X}^\varepsilon_t$ satisfies (1.3) – and it can be proven by analyzing the backward equation associated with (2.1); see [7] and references therein for details.

A class of numerical schemes can be devised in which one integrates (2.2) and (2.4) to approximate the solution of (2.1). These schemes involve three subroutines whose algorithms need to be decided beforehand and are organized as follows (we use here the terminology proposed in [3]):

1. A *macro-solver* for (2.4) will give us the desired evolution of the slow variables $\bar{X}^\varepsilon_t \approx X^\varepsilon_t$ on the $O(1)$-time-scale. The macro-solver can be explicit or implicit, with different order of accuracy, etc; it requires to estimate $B(\bar{X}^\varepsilon_t)$ at the given value of $\bar{X}^\varepsilon_t$. Each time this is necessary, one uses:

2. A *micro-solver* for (2.2), which can be different from the macro-solver, and will give us evolution of the fast variables $Z^\varepsilon_\tau \approx Y^\varepsilon_{t+\varepsilon\tau}$ on the $O(\varepsilon)$-time-scale; and:

3. An *estimator* to evaluate the expectation in (2.5) and obtain the desired estimate of $B(\bar{X}^\varepsilon_t)$. The estimator can use, e.g., different replica of (2.2) and evaluate (2.5) through average over this ensemble, or resort to ergodicity and replace the averages in (2.5) by time-averages.

The overall efficiency and accuracy of the schemes will vary depending on which methods one uses in the three subroutines. However, since the philosophy remains the same no matter what methods one chooses, we will here illustrate the technique using simple schemes; forward Euler for both the macro- and the micro-solvers, and the simplest time-averaging as estimator. The macro-solver then reads

$$\hat{X}^\varepsilon_{n+1} = \hat{X}^\varepsilon_n + \hat{B}(\hat{X}^\varepsilon_n, \varepsilon)\Delta t, \qquad \hat{X}^\varepsilon_{n=0} = x, \tag{2.6}$$

where $\hat{X}^\varepsilon_n$ is the numerical approximation of $\bar{X}^\varepsilon_{n\Delta t}$, Δt is the macro-time-step, and $\hat{B}(x,t,\varepsilon)$ denotes the approximate value of $\hat{B}(x,t,\varepsilon)$ which needs to be estimated at each macro-time-step. This is done by considering (2.2) with $x = \hat{X}^\varepsilon_n$, which if we pick the forward Euler scheme with micro-time-step $\Delta\tau$ as micro-solver can be expressed as

$$\hat{Z}^\varepsilon_{n,m+1} = \hat{Z}^\varepsilon_{n,m} + g(\hat{Z}^\varepsilon_{n,m}, m\Delta\tau, \varepsilon)\Delta\tau. \tag{2.7}$$

Since one is primarily interested in the invariant measure associated with this process, the initial condition only affects the rate of convergence to equilibrium; one may for instance take $\hat{Z}^\varepsilon_{0,0} = 0$ initially, then take the value reached by $\hat{Z}^\varepsilon$ at the previous macro-time-step, i.e. $\hat{Z}^\varepsilon_{n,0} = \hat{Z}^\varepsilon_{n-1,M-1}$, where M is the total number of steps used in the micro-solver. Finally, evaluating the expectation in (2.5) via time-averaging amounts to using as estimator

$$\hat{B}(\hat{X}^\varepsilon_n, \varepsilon) = \frac{1}{M}\sum_{m=0}^{M-1} f(\hat{X}^\varepsilon_n, \hat{Z}^\varepsilon_{n,m}, \varepsilon). \tag{2.8}$$

(2.8) is the most elementary procedure to evaluate time-averages; it can be improved e.g. by using proper filters which put less weight on the first steps in the micro-solver, which may significantly speed up convergence with M [2]. Note that M must be large enough so that the average in (2.8) is accurate, but it is $O(1)$ in ε, since it depends on the correlation time τ_c of Z^ε_τ.

The efficiency of the method can be assessed as follows. Since in most applications, the size of the phase-space associated with the fast $y \in \mathbb{R}^n$ variables will be much bigger than the one of the slow $x \in \mathbb{R}^m$ variables, i.e. $n \gg m$, it is natural to compare the number of evaluations of the functions g if (i) ones solves (2.1) by forward Euler, or (ii) one uses (2.6)–(2.8) – in this case, one also needs to compute (2.8), which requires mM function evaluations; we assume that the associated cost is negligible. Since (2.1) must be solved using a micro-time-step $\delta t = \varepsilon \Delta\tau$, advancing by one macro-time-step Δt requires $R = \lfloor \Delta t/\varepsilon\Delta\tau \rfloor = O(\varepsilon^{-1})$ micro-time-steps; hence, the number of evaluations of g is mR. On the other hand, the number of evaluations of g simply is mM using the scheme based on (2.7). Therefore, as soon as $M < R = O(\varepsilon^{-1})$, it becomes advantageous to use the new scheme instead of integrating (2.1) directly.

A more delicate question is the accuracy of the method. It appears that the main difficulty is to estimate how well and how fast the scheme approximate, the invariant measure $\mu_x^\varepsilon(dz)$. We will leave a detailed analysis of this question to a future publication.

3. Long time intervals

Consider now situations where the evolution of X_t^ε in (2.1) is sought on longer time intervals of the order of $O(\varepsilon^{-1})$. Using the rescaled time $s = \varepsilon t$, (2.1) becomes

$$\begin{cases} \dot{X}_s^\varepsilon = \dfrac{1}{\varepsilon} f(X_s^\varepsilon, Y_s^\varepsilon, \varepsilon), & X_{s=0}^\varepsilon = x \\ \dot{Y}_s^\varepsilon = \dfrac{1}{\varepsilon^2} g(X_s^\varepsilon, Y_s^\varepsilon, s/\varepsilon^2, \varepsilon), & Y_{s=0}^\varepsilon = y. \end{cases} \tag{3.1}$$

We shall make two additional assumptions, which are unnecessary but simplify the discussion. First we assume that g can be decomposed as

$$g(x, y, \tau, \varepsilon) = g_1(y, \tau) + \varepsilon g_2(x, y, \tau, \varepsilon), \tag{3.2}$$

where the leading part, g_1, is differentiable in y and does not depend on x. Second, we assume that

$$\int_{\mathbb{R}^n} f(x, y, \varepsilon)\mu_x^\varepsilon(dy) = O(\varepsilon), \tag{3.3}$$

where $\mu_x^\varepsilon(dy)$ is the invariant measure of the process defined in (2.2); (3.3) guarantees that (3.1) is on the right time-scale on which the dynamics for X_s^ε arises – the procedure below can be generalized to situations where the integral in (3.3) is $O(1)$, in which case X_t^ε displays nontrivial dynamics on the $O(1)$-time-scale, as discussed in the previous section; these results will be presented elsewhere.

The system in (3.1) is more challenging to handle numerically since the coefficients entering the effective equations for X_s^ε are expressed in terms of expectations like (3.3), to be computed in the limit of small ε. This is nontrivial. Suppose that, relying on ergodicity, one tries to estimate (3.3) via time-averaging

$$\int_{\mathbb{R}^n} f(x, y, \varepsilon)\mu_x^\varepsilon(dy) \approx \frac{1}{T}\int_0^T f(x, Y_{\varepsilon^2\tau}^\varepsilon, \varepsilon)d\tau, \tag{3.4}$$

where $Y_{\varepsilon^2\tau}^\varepsilon$ is computed from (3.1) at $X_t^\varepsilon = x$ fixed. The error square one makes in such a computation is

$$\text{error}^2 = \text{var}(f)\frac{\tau_c}{T},$$

where τ_c is the correlation time of $Y^\varepsilon_{\varepsilon^2\tau}$; both τ_c and $\mathrm{var}(f)$ are typically $O(1)$ in ε, whereas from (3.3), the expectation of $f(x,y,\varepsilon)$ is $O(\varepsilon)$. Therefore, the averaging time T necessary to achieve accuracy diverges as ε^{-2}, i.e. expectations like (3.3) cannot be estimated directly as in (3.4) using a time interval to evolve Y^ε_t, which is smaller than the time step used to evolve X^ε_t. Similar difficulties arise also with other procedures than time-averaging – for instance, using different realizations.

The basic idea to compute with accuracy the expectation in (3.3) is to enlarge the system in (3.1). Let Y_s satisfy

$$\dot Y_s = \frac{1}{\varepsilon^2} g_1(Y_s, s/\varepsilon^2), \qquad Y_{s=0} = y.$$

(no superscript ε here, since this dependency is eliminated by rescaling time as $\tau = s/\varepsilon^2$), and define

$$\tilde Y^\varepsilon_s = \frac{1}{\varepsilon}\left(Y^\varepsilon_s - Y_s\right). \tag{3.5}$$

In terms of $(X^\varepsilon_s, Y_s, \tilde Y^\varepsilon_s)$, (3.1) becomes

$$\begin{cases} \dot X^\varepsilon_s = \dfrac{1}{\varepsilon} f(X^\varepsilon_s, Y_s, 0) + F(X^\varepsilon_t, Y_s, \tilde Y^\varepsilon_s, \varepsilon), & X^\varepsilon_{t=0} = x, \\ \dot Y_s = \dfrac{1}{\varepsilon^2} g_1(Y_s, s/\varepsilon^2), & Y_{s=0} = y, \\ \dot{\tilde Y}^\varepsilon_s = \dfrac{1}{\varepsilon^2}\left(G_1(Y_s, \tilde Y^\varepsilon_s, s/\varepsilon^2, \varepsilon) + g_2(X^\varepsilon_s, Y_s + \varepsilon \tilde Y^\varepsilon_s, s/\varepsilon^2, \varepsilon)\right), & \tilde Y^\varepsilon_{s=0} = 0. \end{cases} \tag{3.6}$$

Here

$$\begin{cases} F(x,z,u,\varepsilon) = \displaystyle\int_0^1 \left((u,\nabla_z) f(x, z+\varepsilon\theta u, \varepsilon\theta) + f_\varepsilon(x, z+\varepsilon\theta u, \varepsilon\theta)\right) d\theta, \\ G_1(z,u,\tau,\varepsilon) = \displaystyle\int_0^1 (u,\nabla_z) g_1(z+\varepsilon\theta u, \tau) d\theta, \end{cases} \tag{3.7}$$

where $f_\varepsilon(x,y,\varepsilon) = (\partial f/\partial\varepsilon)(x,y,\varepsilon)$. Furthermore, (3.4) can now be estimated from

$$\int_{\mathbb{R}^n} f(x,y,\varepsilon)\mu^\varepsilon_x(dy) \approx \frac{\varepsilon}{T}\int_0^T F(x, Y_{\varepsilon^2\tau}, \tilde Y^\varepsilon_{\varepsilon^2\tau}, \varepsilon) d\tau, \tag{3.8}$$

with equality in the limit as $T\to\infty$; the term involving $f(X^\varepsilon_s, Y_s, 0)$ is absent since, in order to satisfy (3.3), one must have $\int_{\mathbb{R}^n} f(x,y,0)\mu(dy) = 0$ (no ε here!), where $\mu(dy)$ is the invariant measure of Y_s. Because the right hand side in (3.8) is explicitly proportional to ε, the accuracy problem one had with (3.4) disappears.

We are now ready to state the main result which we will use to devise a numerical procedure to integrate (3.1) or, equivalently, (3.6). Consider

$$\begin{cases} \dot U_\tau = g_1(U_\tau, \tau), \\ \dot V^\varepsilon_\tau = G_1(U_\tau, V^\varepsilon_\tau, \tau, \varepsilon) + g_2(x, U_\tau + \varepsilon V^\varepsilon_\tau, \varepsilon), \end{cases} \tag{3.9}$$

and denote by $\mu^\varepsilon_x(du,dv)$ the invariant measure of the joint process $(U_\tau, V^\varepsilon_\tau)$ defined by (3.9), and by $\mu(du)$ the one of U_τ alone. Then, for $s\in[0,T]$, with $0\le T<\infty$, $\tau\in[0,T']$, with $0\le T'<\infty$, and any test function φ, one has

$$|\mathbf{E}\varphi(X^\varepsilon_s, Y^{1,\varepsilon}_{s+\varepsilon^2\tau}, Y^{2,\varepsilon}_{s+\varepsilon^2\tau}) - \mathbf{E}\varphi(\bar X^\varepsilon_s, U_\tau, V^\varepsilon_\tau)| \to 0, \tag{3.10}$$

as $\varepsilon \to 0$, where $\bar{X}^\varepsilon_s$ satisfies the stochastic differential equation:

$$d\bar{X}^\varepsilon_s = b(\bar{X}^\varepsilon_s, \varepsilon)ds + \sigma(\bar{X}^\varepsilon_s, \varepsilon)dW_s, \qquad \bar{X}^\varepsilon_{s=0} = x. \tag{3.11}$$

Here W_s is m-dimensional Wiener process; $b(x,\varepsilon) = b_1(x,\varepsilon) + b_2(x,\varepsilon)$, with

$$\begin{cases} b_1(x,\varepsilon) = \displaystyle\int_{\mathbb{R}^n \times \mathbb{R}^n} F(x,u,v,\varepsilon)\mu^\varepsilon_x(du,dv), \\ b_2(x,\varepsilon) = \displaystyle\int_0^\infty \int_{\mathbb{R}^n} \mathrm{div}_x f(x,u,\varepsilon)\mathbf{E}_u f(x,U_\tau,\varepsilon)\mu(du)d\tau, \end{cases} \tag{3.12}$$

where $\mathbf{E}_u$ denotes the expectation over the statistics of the process U_τ starting at $U_{\tau=0} = u$; $\sigma(x,\varepsilon)$ is a $\mathbb{R}^m \times \mathbb{R}^m$ matrix satisfying

$$\sigma(x,\varepsilon)\sigma^T(x,\varepsilon) = a(x,\varepsilon), \tag{3.13}$$

where

$$a(x,\varepsilon) = 2\int_0^\infty \int_{\mathbb{R}^n} f(x,u,\varepsilon) \otimes \mathbf{E}_u f(x,U_\tau,\varepsilon)\mu(du)d\tau. \tag{3.14}$$

Finally, in (3.10), $(U_\tau, V^\varepsilon_\tau)$ satisfies (3.9) at $x = \bar{X}^\varepsilon_s$ fixed, with initial condition $(U_{\tau=0}, V^\varepsilon_{\tau=0})$ distributed according to $\mu^\varepsilon_{x=\bar{X}^\varepsilon_s}(du,dv)$. We will omit the proof of this result; it involves straightforward generalizations of the techniques in [7].

As in the deterministic case, this result can be used to devise numerical schemes to approximate the solution of (3.1) by integrating (3.11). The schemes involve again three subroutines – macro-solver, micro-solver, and estimator – which must be decided beforehand, and we illustrate the technique using simple choices; forward Euler for both the macro- and the micro-solvers, and the simplest time-averaging for evaluating the expectations in (3.12) and (3.14). The macro-solver then reads

$$\hat{X}^\varepsilon_{n+1} = \hat{X}^\varepsilon_n + \hat{b}(\hat{X}^\varepsilon_n, \varepsilon)\Delta s + \hat{\sigma}(\hat{X}^\varepsilon_n, \varepsilon)\Delta W_n, \tag{3.15}$$

where $\hat{X}^\varepsilon_n$ is the numerical approximation of $\bar{X}^\varepsilon_{n\Delta s}$, Δs is the macro-time-step, and $\{\Delta W_n\}$ is a collection of i.i.d. Gaussian variables with zero mean and variance Δs. $\hat{b}(\hat{X}^\varepsilon_n, \varepsilon)$ and $\hat{\sigma}(\hat{X}^\varepsilon_n, \varepsilon)$ denotes the approximated values of $b(\hat{X}^\varepsilon_n, \varepsilon)$ and $\sigma(\hat{X}^\varepsilon_n, \varepsilon)$, which we estimated at each macro-time-step by considering (3.9) with $x = \hat{X}^\varepsilon_n$:

$$\begin{cases} \hat{U}_{n,m+1} = \hat{U}_{n,m} + g_1(\hat{U}_{n,m}, m\Delta\tau)\Delta\tau, \\ \hat{V}^\varepsilon_{n,m+1} = \hat{V}^\varepsilon_{n,m} + G_1(\hat{U}_{n,m}, \hat{V}^\varepsilon_{n,m}, m\Delta\tau, \varepsilon)\Delta\tau \\ \qquad\qquad + g_2(\hat{X}^\varepsilon_n, \hat{U}_{n,m} + \varepsilon\hat{V}^\varepsilon_{n,m}, m\Delta\tau, \varepsilon)\Delta\tau, \end{cases} \tag{3.16}$$

with suitable initial conditions (for instance taken similarly as in the deterministic case). Finally, the estimator to evaluate the expectations in (3.12) and (3.14) via time-averaging is

$$\begin{cases} \hat{b}_1(\hat{X}^\varepsilon_n, \varepsilon) = \dfrac{1}{M}\displaystyle\sum_{m=0}^{M-1} F(\hat{X}^\varepsilon_n, \hat{U}_{n,m}, \hat{V}^\varepsilon_{n,m}, \varepsilon), \\ \hat{b}_2(\hat{X}^\varepsilon_n, \varepsilon) = \dfrac{\Delta\tau}{M}\displaystyle\sum_{m=0}^{M-1}\sum_{m'=0}^{M-m-1} \mathrm{div}_x f(\hat{X}^\varepsilon_n, \hat{U}_{n,m}, \varepsilon) f(\hat{X}^\varepsilon_n, \hat{U}_{n,m+m'}, \varepsilon), \\ \hat{a}(\hat{X}^\varepsilon_n, \varepsilon) = \dfrac{2\Delta\tau}{M}\displaystyle\sum_{m=0}^{M-1}\sum_{m'=0}^{M-m-1} f(\hat{X}^\varepsilon_n, \hat{U}_{n,m}, \varepsilon) \otimes f(\hat{X}^\varepsilon_n, \hat{U}_{n,m+m'}, \varepsilon). \end{cases} \tag{3.17}$$

(3.17) can again be improved, e.g. by using proper filters as in the deterministic case, or by using fast Fourier transform techniques, which require $M \log M$ operations instead of M^2 for the last two averages in (3.17). Once (3.17) has been evaluated, the last step is to compute the square root (i.e. Cholesky's decomposition) of $\hat{a}(\hat{X}_n^\varepsilon, \varepsilon)$ to get $\hat{\sigma}(\hat{X}_n^\varepsilon, \varepsilon)$.

The efficiency of the method can be analyzed as in the deterministic case, by comparing the number of evaluations of the functions g and h in (3.1), and of g, G, and h in (3.9) – to that should be added the cost of computing (3.17), which requires $mM + mM^2 + \frac{1}{2}m^2M^2$ function evaluations, and taking the square root of $a(x,\varepsilon)$, which requires $\frac{1}{6}n^3$ operations; we assume that the cost associated with both is negligible. (3.1) must be solved using a micro-time-step $\delta s = \varepsilon^2 \Delta\tau$, and $R = \lfloor \Delta s/\varepsilon^2 \Delta\tau \rfloor = O(\varepsilon^{-2})$ micro-time-steps are required to advance by one macro-time-step Δs; hence, the number of evaluations of g and h is $2mR$. On the other hand, the number of evaluations of g, G, and h simply is $3mM$ using the scheme based on (3.9). Therefore, as soon as $M < \frac{3}{2}R = O(\varepsilon^{-2})$, it becomes advantageous to use the new scheme instead of integrating (3.1) directly. (Note also that this is the worst case scenario, and it will usually be possible to improve it greatly by taking advantage of the special symmetries of the problem under consideration – for instance, if $\mathrm{div}_x f(x,y,\varepsilon) = 0$, then $b_2(x,\varepsilon) = 0$, or if $f(x,y,\varepsilon) = f(y,\varepsilon)$, then $b_2(x,\varepsilon) = 0$ and $a(x,\varepsilon) = a(\varepsilon)$ must be computed only once, etc.)

Acknowledgments. The research reported here originated from discussions with Weinan E at a meeting at Oberwolfach in 1999; I thank Weinan for sharing many thoughts about these approaches since that time, and encouraging me to eventually write the present paper. I also thank Andy Majda for many interesting discussions. This work was supported in part by the NSF via grant DMS02-09959, and by the AMIAS (Association of Members of the Institute for Advanced Study).

REFERENCES

[1] A. Bensoussan, J.L. Lions, and G.C. Papanicolaou, *Asymptotic Analysis for Periodic Structures*. North-Holland, Amsterdam, 1978.

[2] W. E, *Analysis of the Heterogeneous Multiscale Method for ordinary differential equations*. submitted to Comm. Math. Sci., 2003.

[3] W. E and B. Engquist, *The heterogeneous multi-scale methods*. Comm. Math. Sci. 1(1): 88–134, 2003.

[4] C.W. Gear and I.G. Kevrekidis, *Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum*. submitted to SIAM J. Sci. Comp.

[5] V.I. Lebedev and S.I. Finogenov, *Explicit methods of second order for the solution of stiff systems of ordinary differential equations*. Zh. Vychisl. Mat. Mat Fiziki, 16:895–910, 1976.

[6] A.J. Majda, I. Timofeyev, and E. Vanden-Eijnden, *A Mathematical Framework for Stochastic Climate Models*. Comm. Pure Appl. Math., 54:891–974, 2001.

[7] G.C. Papanicolaou, *Introduction to the asymptotics Analysis of Stochastic Equations, in:Modern Modeling of Continuum Phenomena*. Lectures in Applied Mathematics, R. C. DiPrima Ed., vol. 16, 1976.