SIAM J. NUMER. ANAL.
Vol. 43, No. 3, pp. 1139–1154

NUMERICAL INTEGRATION OF STOCHASTIC DIFFERENTIAL EQUATIONS WITH NONGLOBALLY LIPSCHITZ COEFFICIENTS*

G. N. MILSTEIN†‡ AND M. V. TRETYAKOV‡

Abstract. We propose a new concept which allows us to apply any numerical method of weak approximation to a very broad class of stochastic differential equations (SDEs) with nonglobally Lipschitz coefficients. Following this concept, we discard the approximate trajectories which leave a sufficiently large sphere. We prove that accuracy of any method of weak order p is estimated by $\varepsilon + O(h^p)$, where ε can be made arbitrarily small with increasing radius of the sphere. The results obtained are supported by numerical experiments.

Key words. SDEs with nonglobally Lipschitz coefficients, numerical integration of SDEs in the weak sense

AMS subject classifications. 60H35, 65C30, 60H10

DOI. 10.1137/040612026

1. Introduction.

Stochastic differential equations (SDEs) with nonglobally Lipschitz coefficients possessing unique solutions make up a very important class in applications. For instance, Langevin-type equations and gradient systems with noise belong to this class [10, 9, 1, 5, 11]. At the same time, most numerical methods for SDEs are derived under the global Lipschitz condition [3, 7]. If this condition is violated, the behavior of many standard numerical methods in the whole space can lead to incorrect conclusions (see, for instance, [9, 1, 5, 11, 4]). This situation is very alarming since we are forced to refuse many effective methods and/or to resort to some comparatively complicated and inefficient numerical procedures. In [6] (see also Example 3.3 here), applying an explicit quasi-symplectic method of weak approximation to a Langevin equation with nonglobally Lipschitz coefficients for calculating an ergodic limit, the authors found an explosive behavior of some approximate trajectories. The explosions are observed outside of a comparatively large sphere after a relatively large time and very rarely. Clearly, the exploding approximate trajectories badly reproduce the actual behavior of the considered system. We have also found that if these rare trajectories are discarded, then the explicit quasi-symplectic method gives much better results than the implicit Euler method, which does not have any exploding trajectories. From the heuristic point of view, this is rather natural. Roughly speaking, the value of an ergodic limit depends, on the whole, on the behavior of trajectories in a bounded (though large) domain on a finite (though large) time interval. Consequently, any method that is effective for systems with globally Lipschitz coefficients has to work well for systems with nonglobally Lipschitz coefficients as well if one rejects a small number of "bad" trajectories.

In this paper, we propose a new concept which allows us to apply any method of weak approximation to a very broad class of SDEs with nonglobally Lipschitz coef-

*Received by the editors July 21, 2004; accepted for publication (in revised form) January 13, 2005; published electronically September 23, 2005. This research was supported by EPSRC Research Grant GR/R90017/01.

http://www.siam.org/journals/sinum/43-3/61202.html

†Department of Mathematics, Ural State University, Lenin Str. 51, 620083 Ekaterinburg, Russia (Grigori.Milstein@usu.ru).

‡Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK (M.Tretiakov@le.ac.uk).

ficients. Roughly speaking, we require for SDEs from this class just to have regular solutions on a time interval $[t_0, T]$ and to have sufficiently smooth coefficients; i.e., the assumptions made (which are given in terms of Lyapunov functions in section 2) usually hold for SDEs of applicable interest. (We note that convergence of the explicit Euler method is proved under more restrictive assumptions in [1].) Following the concept proposed here, we discard the approximate trajectories which leave a sufficiently large sphere $S_R := \{x : |x| < R\}$. The theoretical justification of the concept is given in section 2. We prove that accuracy of any method of weak order p is estimated by $\varepsilon + O(h^p)$, where ε can be made arbitrarily small with growing R, and $|O(h^p)| \le Kh^p$ can be made arbitrarily small with decreasing h (of course, K depends on R). Thus, we obtain that the violation of the global Lipschitz condition is not fatal for applying any method of numerical integration. Since the Monte Carlo technique is used for simulation of a mean $Ef(X(T))$, the error estimation $\varepsilon + O(h^p)$ should be increased by the Monte Carlo error. It turns out that in practice the error given by ε is much smaller than the joint numerical integration and Monte Carlo error. Furthermore, in principle, due to the concept of rejecting "bad" trajectories, we can choose a suitable method for solving a system of SDEs with nonglobally Lipschitz coefficients, taking into account all the known methods of numerical integration [3, 7]. The application of the concept is discussed in section 3, where some numerical experiments are presented.

An implication of the concept proposed here for the calculation of ergodic limits will be considered in a separate publication.

2. Integration via paths in a bounded domain. Consider the system of Ito SDEs

$$dX = a(t, X)dt + \sum_{l=1}^{q} \sigma_l(t, X)dw_l(t), \qquad X(t_0) = x, \tag{2.1}$$

where X, a, σ_l are d-dimensional column-vectors and $w_l(t)$, $l = 1, \dots, q$, are independent standard Wiener processes.

We suppose the coefficients of (2.1) to be sufficiently smooth functions in $[t_0, T] \times \mathbf{R}^d$, and any solution $X(t; t_0, x)$ of (2.1) to be regular on $[t_0, T]$. We recall that a process is called regular if it is defined for all $t_0 \le t \le T$. Denote by $\mathbf{C}^2$ the class of functions defined on $[t_0, T] \times \mathbf{R}^d$ and twice continuously differentiable with respect to x and once with respect to t. A sufficient condition of regularity (see [2]) consists of the existence of a Lyapunov function $V \in \mathbf{C}^2$, $V \ge 0$, which satisfies the inequality

$$LV(t, x) \le c_0 V(t, x) + c_1, \qquad (t, x) \in [t_0, T] \times \mathbf{R}^d, \tag{2.2}$$

and

$$V_R := \min_{t_0 \le t \le T,\ |x| \ge R} V(t, x), \qquad \lim_{R \to \infty} V_R = \infty, \tag{2.3}$$

where c_0 and c_1 are some constants and L is the following generating operator:

$$LV(t, x) = \frac{\partial V}{\partial t}(t, x) + \sum_{i=1}^{d} a^i(t, x) \frac{\partial V}{\partial x^i}(t, x) + \frac{1}{2} \sum_{i,j=1}^{d} a^{ij}(t, x) \frac{\partial^2 V}{\partial x^i \partial x^j}(t, x), \tag{2.4}$$

$$a^{ij} := \sum_{l=1}^{q} \sigma_l^i \sigma_l^j.$$

Moreover, if (2.2) and (2.3) are fulfilled and if an initial distribution for x (x can be random) is such that $EV(t_0, x)$ exists, then $EV(t, X(t; t_0, x))$ exists for all $t_0 \le t \le T$. For instance, if V has an m-polynomial growth at infinity, then there exist the moments of order m for X. We note that it is not required that c_0 be negative. For definiteness, we consider $c_0 > 0$ and $c_1 > 0$.

Let $S_R := \{x : |x| < R\}$ be an open sphere in $\mathbf{R}^d$, $Q_R = [t_0, T) \times S_R$ be a cylinder in $\mathbf{R}^{d+1}$, $\Gamma_R = \bar{Q}_R \backslash Q_R$, where $\bar{Q}_R$ is the closure of Q_R. The set Γ_R is a part of the boundary of cylinder Q_R consisting of the upper base and the lateral surface. Let τ_R be the first-passage time of the process $(t, X(t; t_0, x))$, $t_0 \le t \le T$, to Γ_R. Clearly, $t_0 \le \tau_R \le T$. We introduce the following events:

$$\begin{aligned} \Omega_R &:= \{\omega : |X(s; t_0, x)| < R,\ t_0 \le s < T\} = \{\omega : \tau_R = T\}, \\ \Lambda_R &:= \{\omega : \exists\ s \in [t_0, T) \ \text{ such that } \ |X(s; t_0, x)| \ge R\} = \{\omega : \tau_R < T\}. \end{aligned} \tag{2.5}$$

Let us obtain an upper bound for the probability

$$p_R := P(\tau_R < T) = P(\Lambda_R), \tag{2.6}$$

assuming (2.2) and (2.3) (see [2]). Introduce the nonnegative function

$$U(t, x) = (c_1 + 1)(T - t) + \exp(c_0(t_0 - t))V(t, x), \tag{2.7}$$

where $V(t, x)$ is a function satisfying (2.2)–(2.3). We get

$$LU(t, x) = -(c_1 + 1) - c_0 \exp(c_0(t_0 - t))V(t, x) + \exp(c_0(t_0 - t))LV(t, x) \le -1. \tag{2.8}$$

Due to the Ito formula, we have

$$\begin{aligned} dU(t, X(t; t_0, x)) &= LU(t, X(t; t_0, x))dt \\ &+ \exp(c_0(t_0 - t)) \sum_{i=1}^{d} \frac{\partial V}{\partial x^i}(t, X(t; t_0, x)) \sum_{l=1}^{q} \sigma_l^i(X(t; t_0, x))dw_l(t). \end{aligned} \tag{2.9}$$

Hence

$$U(\tau_R, X(\tau_R; t_0, x)) - U(t_0, x) = \int_0^{\tau_R} LU dt + \int_0^{\tau_R} \exp(c_0(t_0 - t)) \sum_{i=1}^{d} \frac{\partial V}{\partial x^i} \sum_{l=1}^{q} \sigma_l^i \ dw_l(t). \tag{2.10}$$

The expectation of the second integral on the right-hand side of (2.10) is equal to zero according to the martingale property. Therefore, due to (2.8), we get

$$EU(\tau_R, X(\tau_R; t_0, x)) \le U(t_0, x) = (c_1 + 1)(T - t_0) + V(t_0, x). \tag{2.11}$$

By Chebyshev's inequality, we obtain from (2.11)

$$\begin{aligned} p_R \exp(c_0(t_0 - T)) \min_{t_0 \le t \le T,\ |x| \ge R} V(t, x) &\le p_R \min_{t_0 \le t \le T,\ |x| \ge R} U(t, x) \\ &\le (c_1 + 1)(T - t_0) + V(t_0, x), \end{aligned} \tag{2.12}$$

whence

$$p_R \le \exp(c_0(T-t_0))\frac{(c_1+1)(T-t_0)+V(t_0,x)}{\min_{t_0\le t\le T,\ |x|\ge R} V(t,x)}, \tag{2.13}$$

and therefore

$$\lim_{R\to\infty} p_R = 0.$$

PROPOSITION 2.1. *Let* (2.2) *and* (2.3) *be fulfilled. Let* $f(x)$ *be a function such that*

$$|f(x)| \le V(t,x), \qquad t_0 \le t \le T,\ x \in \mathbf{R}^d. \tag{2.14}$$

Then for any $x \in R^d$ *and* $\varepsilon > 0$ *there exists* $R(x,\varepsilon) > 0$ *such that for any* $R > R(x,\varepsilon)$

$$|Ef(X(T;t_0,x)) - E[f(X(T;t_0,x))\chi_{\Omega_R}(\omega)]| < \varepsilon. \tag{2.15}$$

Proof. Clearly,

$$\lim_{R\to\infty} f(X(T;t_0,x))\chi_{\Omega_R}(\omega) = f(X(T;t_0,x)),\ \text{a.s.}$$

Now the conclusion of this proposition follows from the existence of $EV(T,X(T;t_0,x))$ and the Lebesgue theorem on majorized convergence. □

The significance of this proposition consists of the capability to disregard the trajectories running off too far. We are about to show that when systems under consideration are numerically integrated, the approximating trajectories running off too far can also be discarded. Due to this possibility, we are able, in principle, to use any known method of numerical integration for calculating means. In this respect we shall rest on the developed theory of weak approximation for SDEs with globally Lipschitz coefficients [3, 7]. To this aim we introduce an auxiliary system with globally Lipschitz coefficients, which coincides with the original system in a sphere $S_{R'}$ somewhat wider than S_R: $S_{R'} \supset S_R$, where $R' = R + r$ and $r > 0$ is a constant.

Let the coefficients a^i, σ_l^i and the function f have continuous derivatives up to some order. The requirement on smoothness depends on a particular numerical method used; in general, the higher the order of the method, the more derivatives are needed. We construct coefficients $a_R^i, (\sigma_l^i)_R$ and function f_R so that in $[t_0,T]\times S_{R'}$ they coincide with a^i, σ_l^i and f, respectively, and, in addition, they are bounded in $[t_0,T]\times\mathbf{R}^d$ together with their derivatives up to the same order. This can be done in the following way. Introduce the function $\varphi(z)$ of one variable z:

$$\varphi(z) = \begin{cases} z, & -R' \le z \le R', \\ R' + \displaystyle\int_{R'}^{z} \frac{dz'}{1+(z'-R')^k}, & z > R', \\ -R' - \displaystyle\int_{z}^{-R'} \frac{dz'}{1+(-R'-z')^k}, & z < -R', \end{cases} \tag{2.16}$$

where $k \ge 2$ is a natural number. Clearly, $\varphi(z)$ is bounded on $\mathbf{R}^1$ together with its derivatives up to order k. Let $g(t,x^1,\dots,x^d)$ be a function with some continuous derivatives defined in $[t_0,T]\times\mathbf{R}^d$. It is easily seen that

$$g_R(t,x^1,\dots,x^d) := g(t,\varphi(x^1),\dots,\varphi(x^d))$$

satisfies the above-mentioned conditions. Moreover, there exists a constant $\rho > r$ (which does not depend on R) such that for any $x = (x^1, \dots, x^d) \in \mathbf{R}^d$ the point $(\varphi(x^1), \dots, \varphi(x^d)) \in S_{R+\rho}$. Therefore

$$\sup_{x\in\mathbf{R}^d} |f_R(x)| \le \max_{|x|\le R+\rho} |f(x)|\,. \tag{2.17}$$

Introduce the auxiliary system of SDEs

$$dX_R = a_R(X_R)dt + \sum_{l=1}^{q} (\sigma_l^i)_R(X_R)dw_l(t)\,. \tag{2.18}$$

We emphasize that this system is used in our theoretical proofs only; it is not used in simulation.

PROPOSITION 2.2. *Assume that $V(t,x)$ satisfies* (2.2), (2.3), *and*

$$\frac{\min_{t_0\le t\le T,\ |x|\ge R+\rho} V(t,x)}{\min_{t_0\le t\le T,\ |x|\ge R} V(t,x)} \le c, \tag{2.19}$$

where c is a constant which is independent of R. Let $f(x)$ be a function such that

$$\lim_{R\to\infty} \frac{\max_{|x|\le R} |f(x)|}{\min_{t_0\le t\le T,\ |x|\ge R} V(t,x)} = 0. \tag{2.20}$$

Then for any $x \in R^d$ and $\varepsilon > 0$ there exists $R(x,\varepsilon) > 0$ such that for any $R > R(x,\varepsilon)$

$$|Ef_R(X_R(T;t_0,x)) - Ef(X(T;t_0,x))| < \varepsilon. \tag{2.21}$$

Proof. Since the solutions $X(t;t_0,x)$ and $X_R(t;t_0,x)$ and also the functions $f(X(t;t_0,x))$ and $f_R(X_R(t;t_0,x))$ coincide on the interval $t \in [t_0,\tau_R]$, we have

$$E[f_R(X_R(T;t_0,x))\chi_{\Omega_R}(\omega)] = E[f(X(T;t_0,x))\chi_{\Omega_R}(\omega)]. \tag{2.22}$$

Hence

$$\begin{aligned} &|Ef_R(X_R(T;t_0,x)) - Ef(X(T;t_0,x))| \\ &\le |E[f_R(X_R(T;t_0,x))\chi_{\Lambda_R}(\omega)]| + |E[f(X(T;t_0,x))\chi_{\Lambda_R}(\omega)]|. \end{aligned} \tag{2.23}$$

Proposition 2.1 implies

$$\lim_{R\to\infty} |E[f(X(T;t_0,x))\chi_{\Lambda_R}(\omega)]| = 0. \tag{2.24}$$

Further, due to (2.17) and (2.13), we obtain

$$\begin{aligned} E|f_R(X_R(T;t_0,x))\chi_{\tau_R<T}(\omega)| &\le \sup_{x\in\mathbf{R}^d} |f_R(x)|p_R \le \max_{|x|\le R+\rho} |f(x)|p_R \\ &\le \frac{\max_{|x|\le R+\rho} |f(x)|}{\min_{t_0\le t\le T,\ |x|\ge R+\rho} V(t,x)} \times \frac{\min_{t_0\le t\le T,\ |x|\ge R+\rho} V(t,x)}{\min_{t_0\le t\le T,\ |x|\ge R} V(t,x)} \\ &\quad \times \exp(c_0(T-t_0))[(c_1+1)(T-t_0) + V(t_0,x)]. \end{aligned} \tag{2.25}$$

Now, by using the conditions (2.19) and (2.20), we complete the proof. ∎

Our next step is to show that approximating paths obtained by a numerical method applied to the system (2.1) belong to the bounded domain with a large probability and that averaging via these paths gives a good approximation for the mean $Ef(X(T;t_0,x))$.

Let us start with some necessary auxiliary knowledge of the Markov chains generated by numerical methods. Consider the system of SDEs in the sense of Ito:

$$dY = b(t,Y)dt + \sum_{l=1}^{q} \gamma_l(t,Y)dw_l(t). \tag{2.26}$$

We assume that the functions $b(t,y)$ and $\gamma_l(t,y)$, $(t,y) \in [t_0,T] \times \mathbf{R}^d$, have bounded derivatives with respect to t,y up to some order. In particular, the system (2.18) satisfies this assumption. In most cases a method (both mean-square and weak) can be defined by a one-step approximation of the form

$$\bar{Y}(t+h;t,y) = y + A(t,y,h;\xi), \qquad t_0 \le t < t+h \le T, \tag{2.27}$$

where ξ is a random vector having moments of a sufficiently high order and A is a vector function of dimension d.

Partition the interval $[t_0,T]$ into N equal parts with the step $h = (T-t_0)/N$: $t_0 < t_1 < \cdots < t_N = T$, $t_{k+1} - t_k = h$. According to (2.27), we construct the sequence

$$\bar{Y}_0 = Y(t_0) = y, \quad \bar{Y}_{k+1} = \bar{Y}_k + A(t_k,\bar{Y}_k,h;\xi_k), \qquad k = 0,\dots,N-1, \tag{2.28}$$

where ξ_0 is independent of $\bar{Y}_0$, while ξ_k for $k > 0$ are independent of $\bar{Y}_0,\dots,\bar{Y}_k$, $\xi_0,\dots,\xi_{k-1}$. The sequence $\bar{Y}_k$ is a Markov chain. Its transition probability function is defined by

$$P(t,y,s,D) = P(\bar{Y}(s;t,y) \in D), \qquad s \ge t,\ t,s = t_0, t_0+h,\dots,T, \tag{2.29}$$

where $\bar{Y}(s;t,y)$ is the process with discrete time starting at the moment t from y and defined by (2.28). The generating operator of the Markov chain is defined by

$$\begin{aligned} L_h U(t,y) &= \frac{1}{h}\int P(t,y,t+h,dz)[U(t+h,z) - U(t,y)] \\ &= \frac{1}{h}[EU(t+h,\bar{Y}(t+h;t,y)) - U(t,y)]. \end{aligned} \tag{2.30}$$

Denote by $\bar{\tau}_R$ the first exit time of the process $(t_k,\bar{Y}(t_k;t_0,y))$, $k = 0,\dots,N$, from $[t_0,T) \times S_R$. Due to (2.30), we have (see [8])

$$EU(\bar{\tau}_R,\bar{Y}(\bar{\tau}_R;t_0,y)) - U(t_0,y) = E\sum_{k=0}^{\kappa-1} L_h U(t_k,\bar{Y}(t_k;t_0,y))h, \tag{2.31}$$

where κ is defined by $t_\kappa = \bar{\tau}_R$.

In what follows we use the following assumption.

(A1) *The one-step approximation* (2.27) *is at least of order two in the weak sense, and the method defined by this approximation converges at least with order one.*

By the definition of weak approximation

$$|EU(t+h,\bar{Y}(t+h;t,y)) - EU(t+h,Y(t+h;t,y))| \le Kh^2, \tag{2.32}$$

where K is a positive constant, provided that y belongs to a compact set. At the same time (see [7])

$$EU(t+h,Y(t+h;t,y)) = U(t,y) + hLU(t,y) + O(h^2), \tag{2.33}$$

where $|O(h^2)| \le Kh^2$ and K is again independent of y belonging to a compact set. Using (2.32) and (2.33), we get

$$|EU(t+h,\bar{Y}(t+h;t,y)) - U(t,y) - hLU(t,y)| \le Kh^2,$$

and then we obtain from (2.30)

$$|L_hU(t,y) - LU(t,y)| \le Kh. \tag{2.34}$$

Let us proceed to the system (2.18). Apply a numerical method of weak order $p \ge 1$ to the systems (2.1) and (2.18). As a result, we obtain two Markov chains $\bar{X}_k$ and $(\bar{X}_R)_k$. For the Markov chain $(\bar{X}_R)_k$ (but not for $\bar{X}_k$) we have for $(t,x) \in [t_0,T] \times S_R$ (see [7])

$$|Ef(X_R(T;t_0,x)) - Ef(\bar{X}_R(T;t_0,x))| \le Kh^p. \tag{2.35}$$

Let L_R be the generating operator for (2.18), and $(L_R)_h$ for $\bar{X}_R$. According to (2.34), we get

$$|(L_R)_hU(t,x) - L_RU(t,x)| \le Kh, \qquad (t,x) \in [t_0,T] \times S_R. \tag{2.36}$$

If $(t,x) \in [t_0,T] \times S_R$, then

$$L_RU(t,x) = LU(t,x). \tag{2.37}$$

Due to (2.8), we obtain that $LU \le -1$. It follows from this inequality together with (2.36) and (2.37) that for all h small enough

$$(L_R)_hU(t,x) \le 0, \qquad (t,x) \in [t_0,T] \times S_R. \tag{2.38}$$

In future we need the following assumption.

(A2) *If* $(\bar{X}_R)_i \in S_R$, $i = 0,\dots,k$, *then* $\bar{X}_k = (\bar{X}_R)_k$, $k \le N$. (*Of course, the approximating trajectories are starting from the same point:* $\bar{X}_0 = (\bar{X}_R)_0 = x$.)

The assumption is evidently true, for instance, for the explicit Euler method. Moreover, for this method even $\bar{X}_{k+1} = (\bar{X}_R)_{k+1}$ if only $\bar{X}_k = (\bar{X}_R)_k \in S_R$, though $\bar{X}_{k+1}$ may not belong to S_R. The definition of φ (see (2.16)) ensures coincidence of the coefficients of (2.1) and (2.18) in the wider domain S_{R+r}. Due to this fact, (A2) is fulfilled for h small enough if ξ_k are bounded. This is the most typical case for weak methods, while, applying mean-square methods, we can use random variables such that $\xi_k h^{1/2}$ are small if h is small (see [7]). Thus, the condition (A2) is fulfilled for typical mean-square methods as well. Nevertheless, we should pay attention that a method of the type (2.27) with initial data from $[t_0,T] \times S_R$ may depend on behavior of a system's coefficients not only in $[t_0,T] \times S_{R+r}$ but, generally speaking, in $[t_0,T] \times \mathbf{R}^d$.

Let us take h ensuring (2.38). Denote by $\bar\tau_R$ the first exit time of the chain $(t_k, \bar X_R(t_k;t_0,x))$ from $[t_0,T)\times S_R$, i.e., $\bar X_R(t_k;t_0,x)\in S_R$, $k=0,1,\dots,\kappa-1$, and either $\bar X_R(t_\kappa;t_0,x)=\bar X_R(\bar\tau_R;t_0,x)\notin S_R$, where $\bar\tau_R=t_\kappa$, or $t_\kappa=T$. Applying (2.31) to $\bar X_R$ and using (2.38), we obtain

$$EU(\bar\tau_R,\bar X_R(\bar\tau_R;t_0,x))\le U(t_0,x)=(c_1+1)(T-t_0)+V(t_0,x). \tag{2.39}$$

Introduce the events

$$\begin{aligned}\tilde\Omega_R &:= \{\omega: |\bar X_R(t_k;t_0,x)|<R,\ k=0,\dots,N-1,\ \text{ and }\ |\bar X_R(T;t_0,x)|\le R\}\\ &= \{\omega:(\bar\tau_R=T)\setminus\big((\bar\tau_R=T)\cap\big(|\bar X_R(T;t_0,x)|>R\big)\big)\},\\ \tilde\Lambda_R &:= \{\omega:(\exists\ t_k,\ k=0,\dots,N-1,\ \text{ such that}\\ &\qquad |\bar X_R(t_k;t_0,x)|\ge R)\cup(|\bar X_R(T;t_0,x)|>R)\}\\ &= \{\omega:(\bar\tau_R<T)\cup\big((\bar\tau_R=T)\cap\big(|\bar X_R(T;t_0,x)|>R\big)\big)\}.\end{aligned} \tag{2.40}$$

The event $\tilde\Lambda_R$ consists of leaving S_R by $\bar X_R$ at one of the moments $t_0,\dots,t_{N-1}$ or leaving $\bar S_R$ at $t_N=T$. We note that in the continuous case (see (2.5)) the set $(\tau_R=T)\cap(|X_R(T;t_0,x)|>R)$ is empty.

Let

$$\bar p_R=P(\tilde\Lambda_R).$$

Analogously to (2.12) and (2.13), we apply Chebyshev's inequality and obtain from (2.39)

$$\bar p_R\le\exp(c_0(T-t_0))\frac{(c_1+1)(T-t_0)+V(t_0,x)}{\min_{t_0\le t\le T,\ |x|\ge R}V(t,x)}. \tag{2.41}$$

Further, analogously to (2.25), we obtain

$$E|f_R(\bar X_R(T;t_0,x))\chi_{\tilde\Lambda_R}(\omega)|\le\max_{|x|\le R+\rho}|f(x)|\bar p_R. \tag{2.42}$$

We see from the two last inequalities that the expectation $E|f_R(\bar X_R(T;t_0,x))\chi_{\tilde\Lambda_R}(\omega)|$ is as small as $E|f_R(X_R(T;t_0,x))\chi_{\tilde\Lambda_R}(\omega)|$ (cf. (2.25)) if only h ensures (2.38).

THEOREM 2.3. *Consider any method satisfying* (A1) *which is weakly convergent with order p for systems with sufficiently smooth and bounded derivatives up to some order. Let the conditions of Propositions* 2.1 *and* 2.2 *and the assumption* (A2) *be fulfilled. Then for any $x\in R^d$ and $\varepsilon>0$ there exists $R(x,\varepsilon)>0$ such that for all $R\ge R(x,\varepsilon)$ and sufficiently small h*

$$|Ef(X(T;t_0,x))-E[f(\bar X(T;t_0,x))\chi_{\tilde\Omega_R}(\omega)]|\le Kh^p+\varepsilon, \tag{2.43}$$

where $K>0$ depends on x and R.

Proof. It has been proved (see Proposition 2.2) that for any $\varepsilon>0$

$$|Ef(X(T;t_0,x))-Ef_R(X_R(T;t_0,x))|\le\frac{\varepsilon}{2} \tag{2.44}$$

if R is sufficiently large.

Since the coefficients of system (2.18) and the function f_R can be taken so that they have bounded derivatives up to a sufficiently high order, the mentioned method gives for sufficiently small h (see [7])

$$|Ef_R(X_R(T;t_0,x))-Ef_R(\bar X_R(T;t_0,x))|\le Kh^p. \tag{2.45}$$

Let us choose $R(x,\varepsilon)$ so that for $R \geq R(x,\varepsilon)$ and sufficiently small h both inequality (2.44) and inequality

$$E|f_R(\bar{X}_R(T;t_0,x))\chi_{\tilde{\Lambda}_R}(\omega)| \leq \frac{\varepsilon}{2} \tag{2.46}$$

(see (2.41) and (2.42)) are fulfilled.

Since the assumption (A2) holds, $f_R(x) = f(x)$ for $x \in S_R$, and $\bar{X}_R = \bar{X}$ for $\omega \in \tilde{\Omega}_R$, we get

$$\begin{aligned} Ef_R(\bar{X}_R(T;t_0,x)) &= E[f_R(\bar{X}_R(T;t_0,x))\chi_{\tilde{\Omega}_R}(\omega)] + E[f_R(\bar{X}_R(T;t_0,x))\chi_{\tilde{\Lambda}_R}(\omega)] \\ &= E[f(\bar{X}(T;t_0,x))\chi_{\tilde{\Omega}_R}(\omega)] + E[f_R(\bar{X}_R(T;t_0,x))\chi_{\tilde{\Lambda}_R}(\omega)]. \end{aligned} \tag{2.47}$$

Inequality (2.43) follows from (2.44)–(2.47). Theorem 2.3 is proved. □

Remark 2.1. If a method for a particular stochastic system converges, then K in (2.43) is bounded for all R (and ε). However, as was discussed in the Introduction, a method applied to SDEs with nonglobally Lipschitz coefficients can be divergent. It is obvious that in this case K goes to infinity as $R \to \infty$ ($\varepsilon \to 0$). In practice (see, e.g., our experiments and also a comment on the choice of R in section 3), for a not too big R (and, consequently, not large K) the ε is negligibly small since the divergence is usually due to rare exploding approximate trajectories which have to be discarded. This concept of rejecting exploding trajectories is very practical; it allows us, in particular, to guarantee the accuracy of numerical results obtained even by "divergent" methods. We emphasize that the value of K depends on the choice of a numerical method, as is usual in the global Lipschitz case. Thanks to the above concept, we can exploit the whole arsenal of methods [3, 7] and choose an appropriate scheme depending on the system we are solving.

Remark 2.2. It is possible to prove that the proposed concept is also applicable in the case of the Talay–Tubaro extrapolation [12, 7]; i.e., for a sufficiently large R and all sufficiently small h the error can be expanded in powers of h:

$$\begin{aligned} &Ef(X(T;t_0,x)) - E[f(\bar{X}(T;t_0,x))\chi_{\tilde{\Omega}_R}(\omega)] \\ &\quad = \rho(R,h) + C_0h^p + \cdots + C_nh^{p+n} + O(h^{p+n+1}), \end{aligned} \tag{2.48}$$

where the constants $C_0, \ldots, C_n$ are independent of h and $\rho(R,h) \to 0$ as $R \to \infty$ uniformly with respect to h.

Due to Remark 2.1, ρ is negligibly small for a fixed R in comparison with the term $O(h^{p+n+1})$ (for realistic, not too small h, of course). Therefore it can be supposed that $|\rho| \leq Ch^{p+n+1}$, where C is a positive constant. Then we can use (2.48) in practice to estimate the global error as well as to improve the accuracy of the method [12, 7]. For example, simulating $u = Ef(X(T;t_0,x))$ twice by a first-order scheme (i.e., $p = 1$) with two different time steps $h_1 = h$, $h_2 = \alpha h$, $\alpha > 0$, $\alpha \neq 1$, we obtain $\bar{u}^{h_1} = E[f(\bar{X}^{h_1}(T;t_0,x))\chi_{\tilde{\Omega}_R}(\omega)]$ and $\bar{u}^{h_2} = E[f(\bar{X}^{h_2}(T;t_0,x))\chi_{\tilde{\Omega}_R}(\omega)]$, respectively. We can expand (see (2.48) with $p = 1$, $n = 0$):

$$u = \bar{u}^{h_1} + C_0h_1 + \delta_1, \quad u = \bar{u}^{h_2} + C_0h_2 + \delta_2,$$

where $|\delta_i| \leq Ch^2$, $i = 1, 2$. Hence C_0 can be estimated as $C_0 \simeq -\frac{\bar{u}^{h_2} - \bar{u}^{h_1}}{h_2 - h_1}$, and we get the improved value

$$\bar{u}_{imp} = \bar{u}^{h_1}\frac{h_2}{h_2 - h_1} - \bar{u}^{h_2}\frac{h_1}{h_2 - h_1}, \qquad u = \bar{u}_{imp} + \delta,$$

where $|\delta| \le Ch^2$.

Example 2.1. Consider the following system:

$$dP = -\nabla F(Q)dt - \nu P dt + \sum_{l=1}^{n} \sigma_l dw_l(t), \tag{2.49}$$
$$dQ = Pdt,$$

where ν is a positive constant and σ_l, $l = 1, \dots, n$, are n-dimensional constant linearly independent vectors. The authors of [5] prove exponential ergodicity of (2.49) assuming that $F \in \mathbf{C}^\infty(\mathbf{R}^n, \mathbf{R})$, $F(q) \ge 0$ for all $q \in \mathbf{R}^n$ and that there exist an $\alpha > 0$ and $0 < \beta < 1$ such that

$$\frac{1}{2}(\nabla F(q), q) \ge \beta F(q) + \nu^2 \frac{\beta(2-\beta)}{8(1-\beta)}|q|^2 - \alpha. \tag{2.50}$$

The Lyapunov function

$$V(x) = V(p,q) = \frac{1}{2}|p|^2 + F(q) + \frac{\nu}{2}(p,q) + \frac{\nu^2}{4}|q|^2 + 1 \tag{2.51}$$
$$\ge 1 + \frac{1}{8}|p|^2 + \frac{\nu^2}{12}|q|^2$$

is used to prove that for any $m \ge 1$ there exist positive $c_m,\ d_m$ such that

$$L[V(x)]^m \le -c_m[V(x)]^m + d_m. \tag{2.52}$$

The exponential ergodicity means that for any function f with a polynomial growth the following inequality holds:

$$\left| Ef(X(t;0,x)) - \int f(z)d\mu(z) \right| \le Ce^{-\lambda t}, \tag{2.53}$$

where $C > 0$ and $\lambda > 0$ are some constants. In (2.51)–(2.53), $x := (p,q),\ X := (P,Q)$, and μ is an invariant measure for the Markov process defined by (2.49).

Resting on (2.52), it is not difficult to verify all the assumptions of Theorem 2.3 that concern the system under consideration. Due to (2.53), application of this theorem to calculation of $Ef(X(t;0,x))$ gives an approximate value of the ergodic limit. In [6] (see also Example 3.3 below) a numerical example connected with calculation of an ergodic limit is given.

3. Numerical experiments. Theorem 2.3 has the following practical implication for evaluating expectations of functionals of solutions to SDEs. We pick up a numerical method suitable for a stochastic system under consideration. We choose $R > 0$ such that the solution of the stochastic system equipped with some initial data leaves the sphere S_R of the radius R during a fixed time interval with a relatively small probability. In a lot of cases interesting from the applicable point of view (e.g., Langevin-type equations and gradient systems with noise) it is usually not difficult to guess this value of R by physical reasoning. Anyway, we can test the choice of R in practice as explained below. For the chosen R, we select a time step h for the numerical method, which ensures an accuracy appropriate for our purposes. As usual, the choice of time step h_* is appropriate if, by further decrease of the time step, we obtain a result which is close enough to the one obtained with h_*. The expectation of

a functional which we are aiming to find is evaluated according to the Monte Carlo technique by running M independent realizations of the numerical solution to the considered system. According to the concept proposed in this paper, the value of the functional corresponding to sample trajectories that left the sphere S_R is set to be zero when counted to the expectation. Finally, we say that the choice of R is appropriate if its increase does not essentially affect the result. We also note that there is Monte Carlo error in this procedure, which is controlled in the standard way by choosing an appropriate M. In practice, the procedure can be modified by assigning a certain value (not zero as we do here) for the trajectories which leave the sphere S_R. This value can be chosen/adjusted in response to experimental results or by physical reasoning.

As we will see in the numerical experiments presented below, the numerical integration and Monte Carlo errors affect accuracy of simulation much more than error due to canceling "bad" trajectories (ε in (2.43)), which is usually negligibly small.

Example 3.1. Consider the stochastic differential equation

$$dX = -X^3 dt + \sigma dw(t), \quad X(0) = X_0. \tag{3.1}$$

It is demonstrated in [5] (see also [9, 11]) that the explicit Euler method for (3.1),

$$X_{k+1} = X_k - X_k^3 h + \sigma \Delta_k w, \qquad \Delta_k w := w(t_{k+1}) - w(t_k), \tag{3.2}$$

can explode.

For test purposes, we evaluate the functional

$$F = \frac{1}{2} E X^2(T) + E \int_0^T X^4(t) dt. \tag{3.3}$$

It can be shown that

$$F = \frac{1}{2} \sigma^2 T.$$

To simulate this functional, we introduce the additional equation

$$dZ = X^4(t) dt, \qquad Z(0) = 0. \tag{3.4}$$

Then

$$F = E\left(\frac{1}{2} X^2(T) + Z(T)\right). \tag{3.5}$$

The solution of (3.4) is approximated as

$$Z_{k+1} = Z_k + X_k^4 h. \tag{3.6}$$

By taking $V(x, z) = x^6 + z^2$, it is not difficult to check that the conditions of Propositions 2.1 and 2.2 are satisfied for the system (3.1), (3.4). Also, the condition (A2) holds for the explicit Euler method (3.2). Then Theorem 2.3 is applicable here; i.e., we can evaluate F from (3.3) by using approximate trajectories $(\bar{X}(t), \bar{Z}(t))$, $0 \le t \le T$, which belong to the ball $\{(x, z) : x^2 + z^2 < R^2\}$. In fact, in the case of functionals like that in (3.3) it is enough to control the paths $\bar{X}(t)$ only; i.e., the following estimate takes place (cf. (2.43)):

$$\left| E\left(\frac{1}{2} X^2(T) + Z(T)\right) - E\left(\frac{1}{2} \bar{X}^2(T) + \bar{Z}(T)\right) \chi_{\bar{\Omega}_R}(\omega) \right| \le Kh + \varepsilon, \tag{3.7}$$

where $\tilde{\Omega}_R$ is defined by $\bar{\tau}_R$ being the first exit time of $(t, \bar{X}(t))$ from the rectangle $[0,T) \times (-R,R)$. This result is valid thanks to the fact that the right-hand sides of (3.1), (3.4) do not depend on Z. The proof is almost a word-by-word repetition of the proof of Theorem 2.3.

We also consider the weak Euler method:

$$X_{k+1} = X_k - X_k^3 h + \sigma \xi_k \sqrt{h}, \tag{3.8}$$

where ξ_k are i.i.d. (independently and identically distributed) random variables with the law $P(\xi = \pm 1) = 1/2$.

Let us choose a time step $h > 0$ for (3.8) such that

$$|X_0| \le \frac{1}{\sqrt{h}} \quad \text{and} \quad h < \frac{1}{\sigma}\left(1 - \frac{2}{3\sqrt{3}}\right).$$

Then one can directly show that $|X_1| \le \frac{1}{\sqrt{h}}$ and therefore $|X_k| \le \frac{1}{\sqrt{h}}$ for all k. Thus, trajectories of (3.8) do not explode, provided that the above conditions on the step h hold. The authors do not exclude a possibility that methods using bounded random variables (like the weak Euler method (3.8)) can weakly converge in some nonglobally Lipschitz cases. In general this question concerning convergence is rather complicated (see, e.g., the third example below and also [6]) and requires further investigation. We should stress that convergence of methods does not undermine the concept proposed in this paper. Indeed, suppose that a weak method converges but for a not very small time step it may have exploding trajectories; then results obtained with this time step should be disregarded unless this concept is applied. This is well illustrated in our examples. Further, the concept is universal. It allows us to use any numerical method in the nonglobally Lipschitz case straightaway for a very broad class of SDEs, without any additional analysis at all.

In our experiments we simulate F from (3.5) as follows:

$$\bar{F} = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{2}\left[\bar{X}^{(m)}(T)\right]^2 + \bar{Z}^{(m)}(T)\right)\chi_{\tilde{\Omega}_R}(\omega) + \rho_{mc}, \tag{3.9}$$

where M is the number of independent realizations $\bar{X}^{(m)}(T)$, $\bar{Z}^{(m)}(T)$ of $\bar{X}(T)$, $\bar{Z}(T)$ that are found due to a numerical method of our choice. The Monte Carlo error ρ_{mc} has zero bias, and its variance equals

$$Var(\rho_{mc}) = \frac{Var\ \left(\left(\frac{1}{2}\bar{X}^2(T) + \bar{Z}(T)\right)\chi_{\tilde{\Omega}_R}(\omega)\right)}{M}; \tag{3.10}$$

i.e., the simulated

$$\hat{F} := \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{2}\left[\bar{X}^{(m)}(T)\right]^2 + \bar{Z}^{(m)}(T)\right)\chi_{\tilde{\Omega}_R}(\omega)$$

belongs to the confidence interval

$$\hat{F} \in (E\bar{F} - c\sqrt{Var(\rho_{mc})}, E\bar{F} + c\sqrt{Var(\rho_{mc})}) \tag{3.11}$$

with the fiducial probability, for example, 0.997 for $c = 3$ and 0.95 for $c = 2$. For definiteness, we set $c = 2$ here.

TABLE 1

Simulation of (3.1), (3.4) *by the Euler methods* (3.2), (3.6) *and* (3.8), (3.6) *with various time steps* h. *See the other parameters in the text. The exact value* $F = 5$. *The "±" reflects the Monte Carlo error only; it does not reflect the error of the methods.*

h	(3.2), (3.6)		(3.8), (3.6)
	$\bar{F}$	Trajectories left $(-R, R)$	$\bar{F}$
0.25	6.640 ± 0.010	0.03%	5.962 ± 0.006
0.1	5.409 ± 0.007	0%	5.371 ± 0.007
0.02	5.069 ± 0.007	0%	5.073 ± 0.007
0.01	5.032 ± 0.007	0%	5.037 ± 0.007

In Table 1 some results of our numerical experiments are presented. We take $\sigma = 1$, $X_0 = 0$, $M = 400000$, $T = 10$, and $R = 50$. The "±" reflects the Monte Carlo error only; it gives the confidence interval with $c = 2$ (see (3.11)). If our concept were not applied in the case of the Euler method (3.2), (3.6) with $h = 0.25$, then there would be an overflow in computer calculations. We also note in passing that both Euler methods produce quite similar results. Of course, the Euler method (3.2) is computationally more expensive than (3.8) due to the need to simulate Gaussian random variables instead of very simple random variables for (3.8).

Example 3.2. As the second test model, we choose the equation

$$dX = -X \exp(X^2)dt + dw(t), \quad X(0) = X_0, \tag{3.12}$$

and evaluate $EX^2(t)$. See also some experiments and discussion concerning (3.12) in [5].

By taking $V(x) = x^4$, it is easy to check that the conditions of Propositions 2.1 and 2.2 are satisfied for this equation. Therefore Theorem 2.3 is applicable again for methods satisfying the condition (A2). Further, (3.12) is exponentially ergodic, and the second moment evaluated with respect to the invariant measure is equal to 0.2539 up to 4 decimal points (d.p.).

We simulate (3.12) by the explicit Euler method

$$X_{k+1} = X_k - X_k \exp(X_k^2)\, h + \Delta_k w \tag{3.13}$$

on the time interval $[0, 100]$ with $X_0 = 0$. In our numerical experiments we choose $R = 5$ (we note that the force $-x \exp(x^2)$ produces a very sharp "barrier"); the number of independent realizations $M = 400000$, and as the result the Monte Carlo error $2\sqrt{Var(\rho_{mc})} \le 0.001$. Figure 1 gives the averaged trajectories of $EX^2(t)$ for various time steps h. For $h = 0.2$, there are 356390 trajectories (i.e., 89%) that leave the interval $(-5, 5)$. Obviously, the obtained result cannot be considered reliable, and it is not presented in Figure 1. For $h = 0.1$, we have 36676 trajectories (i.e., 9%) that leave the interval $(-5, 5)$; for $h = 0.05$, 130 trajectories (i.e., 0.03%); and for $h = 0.02$, there are no trajectories out of 400000 that leave the interval $(-5, 5)$. An increase of R has almost no effect on the results. This indicates that our choice of R is appropriate. We see from Figure 1 that for $h = 0.05$ and $h = 0.02$ we obtain a quite good approximation of the ergodic limit. At the same time, we note that if our concept were not applied in this experiment, then the average trajectories for $h = 0.1$ and $h = 0.05$ would blow up. In the case of $h = 0.02$ we have not observed exploding trajectories, but this does not mean that if we continued the experiment further, we would not observe any exploding trajectories that lead to blow-up of the average trajectories. Further, we should note that the weak Euler method (cf. (3.8)) applied

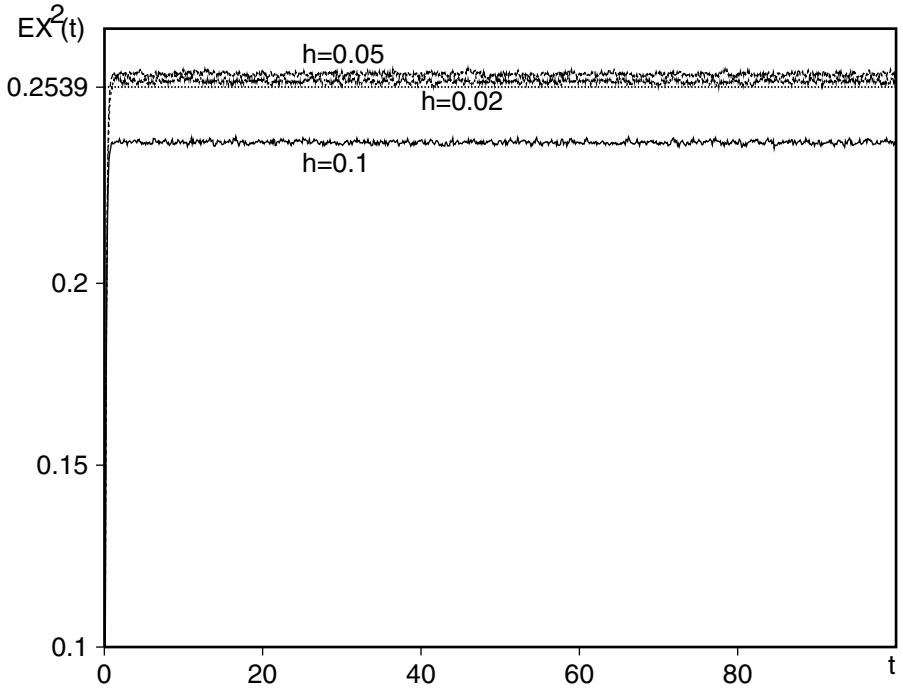

FIG. 1. *Result of simulations of* (3.12) *by the Euler method* (3.13) *with various time steps* h. *We take* $R = 5$*; see the other parameters in the text.*

to (3.12) does not explode even for $h = 0.2$. This can be explained by arguments similar to those used in the first example.

Example 3.3. Consider the oscillator with cubic restoring force and additive noise

$$\begin{aligned} dQ &= Pdt, \\ dP &= \left(Q - Q^3\right) dt - \nu P dt + \sigma dw(t), \end{aligned} \tag{3.14}$$

where $w(t)$ is a standard Wiener process and ν and σ are positive constants. This system is exponentially ergodic, and the second moment EQ^2 evaluated with respect to the invariant measure is equal to 2.435 up to 3 d.p.

The system (3.14) belongs to the class of Langevin equations, for which quasi-symplectic methods are the most effective [6] (see also [7]). We apply an explicit quasi-symplectic method of weak order one to (3.14):

$$\begin{aligned} P_{k+1} &= (1 - \nu h)\left(P_k + h\left(Q_k - Q_k^3\right) + h^{1/2}\sigma\xi_k\right), \\ Q_{k+1} &= Q_k + h\left(P_k + h\left(Q_k - Q_k^3\right)\right), \end{aligned} \tag{3.15}$$

where ξ_k are i.i.d. random variables with the law $P(\xi = \pm 1) = 1/2$.

The results of simulating $EQ^2(t)$ by this method are presented in Figure 2. We take the parameters of (3.14) as follows: $Q(0) = P(0) = 0$, $\nu = 0.05$, and $\sigma = 1$. For realization of the proposed concept, we choose $R = 50$. The number of independent realizations M used to produce the picture is equal to 400000, which ensures the Monte Carlo error $2\sqrt{Var(\rho_{mc})} \le 0.008$. For $h = 0.2$, there are 28 trajectories (i.e., 0.007%) that leave the ball of radius 50. We see that, applying the proposed concept

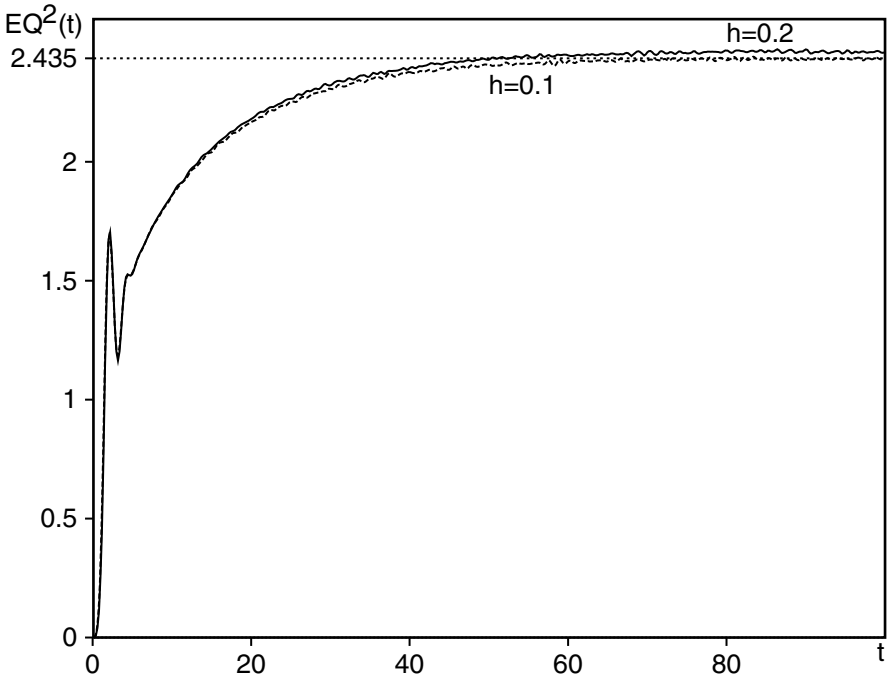

FIG. 2. *Result of simulations of* (3.14) *by the quasi-symplectic method* (3.15) *with various time steps* h. *We take* $R = 50$*; see the other parameters in the text.*

(i.e., not taking into account these 28 "bad" trajectories), we obtain a quite accurate approximation of the ergodic limit. If one did not exploit the concept here, then the results obtained with $h = 0.2$ could not be used, since the exploding trajectories lead to numerical overflow in computing the average. For $h = 0.1$, there are no trajectories out of 400000 that leave the ball of radius 50, but this does not exclude the possibility of having exploding trajectories in another series of Monte Carlo runs. Such uncertainty made it uncomfortable to use the results of such experiments before the concept developed in this paper. This concept gives us a rigorous basis for making use of any numerical method to solve nonlinear SDEs and for interpreting experiments in which occurrence of exploding trajectories is not excluded. Some other experiments with the model (3.14) are available in [6] and [5]. We also note that a further development of the concept of rejecting exploding trajectories specifically for calculation of ergodic limits will be addressed in a separate publication.

REFERENCES

[1] D. J. HIGHAM, X. MAO, AND A. M. STUART, *Strong convergence of Euler-type methods for nonlinear stochastic differential equations*, SIAM J. Numer. Anal., 40 (2002), pp. 1041–1063.

[2] R. Z. HASMINSKII, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Groningen, The Netherlands, 1980.

[3] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.

[4] H. LAMBA, J. C. MATTINGLY, AND A. M. STUART, *An Adaptive Euler–Maruyama Scheme for SDEs: Convergence and Stability*, working paper, Warwick University (England), 2005.

[5] J. C. MATTINGLY, A. M. STUART, AND D. J. HIGHAM, *Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise*, Stochastic Process. Appl., 101 (2002), pp. 185–232.

[6] G. N. Milstein and M. V. Tretyakov, *Quasi-symplectic methods for Langevin-type equations*, IMA J. Numer. Anal., 23 (2003), pp. 593–626.
[7] G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer, New York, 2004.
[8] M. B. Nevelson and R. Z. Hasminskii, *Stochastic Approximation and Recursive Estimation*, Transl. Math. Monogr. 47, American Mathematical Society, Providence, RI, 1973.
[9] G. O. Roberts and R. L. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernoulli, 2 (1996), pp. 341–363.
[10] C. Soize, *The Fokker–Planck Equation for Stochastic Dynamical Systems and Its Explicit Steady State Solutions*, World Scientific, London, 1994.
[11] D. Talay, *Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme*, Markov Process. Related Fields, 8 (2002), pp. 163–198.
[12] D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Anal. Appl., 8 (1990), pp. 483–509.