On numerical stationary distribution of overdamped Langevin equation in harmonic system

De-Zhang Li(李德彰) and Xiao-Bao Yang(杨小宝)[†]

Department of Physics, South China University of Technology, Guangzhou 510640, China

(Received 12 December 2022; revised manuscript received 22 January 2023; accepted manuscript online 2 March 2023)

Efficient numerical algorithm for stochastic differential equation has been an important object in the research of statistical physics and mathematics for a long time. In this work we study the highly accurate numerical algorithm for the overdamped Langevin equation. In particular, our interest is in the behaviour of the numerical schemes for solving the overdamped Langevin equation in the harmonic system. Based on the large friction limit of the underdamped Langevin dynamic scheme, three algorithms for overdamped Langevin equation are obtained. We derive the explicit expression of the stationary distribution of each algorithm by analysing the discrete time trajectory for both one-dimensional case and multi-dimensional case. The accuracy of the stationary distribution of each algorithm is illustrated by comparing with the exact Boltzmann distribution. Our results demonstrate that the "BAOA-limit" algorithm generates an accurate distribution of the harmonic system in a canonical ensemble, within a stable range of time interval. The other algorithms do not produce the exact distribution of the harmonic system.

Keywords: numerical stationary distribution, overdamped Langevin equation, exact solution, harmonic system

PACS: 05.20.–y, 05.10.Gg, 02.30.Jr **DOI:** 10.1088/1674-1056/acc05c

1. Introduction

As the first published stochastic differential equation, Langevin equation[1,2] plays an important role in studying statistical mechanics and mathematical physics. It offers a heuristic mathematical description of the Brownian motion,[3–5] and becomes a widely used tool in the fields of natural science, mathematics, and social science.[6] Specifically, the Boltzmann distribution in a canonical ensemble can be obtained from the stationary state distribution of Langevin equation.[7] Hence, Langevin equation is an efficient approach to solving the problem of canonical sampling in statistical mechanics.

In this work, we focus on a certain type of Langevin equation, called overdamped Langevin equation. It can be seen as the overdamped limit of the usual underdamped Langevin equation.[8] For the system with potential energy U and mass matrix $\boldsymbol{M}$, the overdamped Langevin equation takes the following form:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\frac{1}{\gamma}\boldsymbol{M}^{-1}\nabla U(\boldsymbol{x}) + \sqrt{\frac{2}{\beta\gamma}}\boldsymbol{M}^{-1/2}\boldsymbol{\eta}(t). \tag{1}$$

Here γ is the friction coefficient, $\beta = 1/k_{\mathrm{B}}T$ with the Boltzmann constant k_{B} and the temperature T, and $\boldsymbol{\eta}(t)$ is the white noise random vector associated with the Wiener process satisfying

$$\langle \eta_i(t)\rangle = 0, \quad \left\langle \eta_i(t)\eta_j(t')\right\rangle = \delta_{ij}\delta(t-t') \tag{2}$$

or equivalently

$$\langle \boldsymbol{\eta}(t)\rangle = 0, \quad \left\langle \boldsymbol{\eta}(t)\boldsymbol{\eta}^{\mathrm{T}}(t')\right\rangle = \delta(t-t')\,\mathbf{1}. \tag{3}$$

Here, $\mathbf{1}$ denotes the identity matrix hereinafter. Considering the probability density function of the system $\rho(\boldsymbol{x},t)$, associated with Eq. (1), the time-dependent probability density $\rho(\boldsymbol{x},t)$ of the system evolves according to the well-known Fokker–Planck equation[6,7,9–12]

$$\frac{\partial}{\partial t}\rho(\boldsymbol{x},t) = \frac{\partial}{\partial \boldsymbol{x}}\cdot\left[\frac{1}{\gamma}\boldsymbol{M}^{-1}\frac{\partial U}{\partial \boldsymbol{x}}\rho(\boldsymbol{x},t) + \frac{1}{\beta\gamma}\boldsymbol{M}^{-1}\frac{\partial \rho(\boldsymbol{x},t)}{\partial \boldsymbol{x}}\right]. \tag{4}$$

The relevant time evolution operator is then

$$\mathcal{L}(\sim) = \frac{\partial}{\partial \boldsymbol{x}}\cdot\left[\frac{1}{\gamma}\boldsymbol{M}^{-1}\frac{\partial U}{\partial \boldsymbol{x}}(\sim) + \frac{1}{\beta\gamma}\boldsymbol{M}^{-1}\frac{\partial(\sim)}{\partial \boldsymbol{x}}\right], \tag{5}$$

where $\rho(\boldsymbol{x},t)$ evolves from an initial distribution (*e.g.*, $\rho(\boldsymbol{x},0) = \delta(\boldsymbol{x}-\boldsymbol{x}_0)$) to a stationary state $\rho(\boldsymbol{x},\infty)$, and satisfies $\mathcal{L}\rho(\boldsymbol{x},\infty) = 0$. The stationary distribution is usually called the invariant distribution in the mathematical literature.[13,14] We denote the stationary distribution as $\rho_{\mathrm{st}}(\boldsymbol{x})$ in this work. In our case, $\rho_{\mathrm{st}}(\boldsymbol{x})$ is simply the Boltzmann distribution

$$\rho_{\mathrm{st}}(\boldsymbol{x}) = \frac{1}{N}\exp\left[-\beta U(\boldsymbol{x})\right], \tag{6}$$

where N is the normalization constant. Then one can see that the term $\sqrt{2/\beta\gamma}\boldsymbol{M}^{-1/2}$ associated with $\boldsymbol{\eta}(t)$ in Eq. (1) fulfils the fluctuation–dissipation relation, which ensures that the overdamped Langevin equation generates the Boltzmann distribution in a stationary state. Therefore, overdamped

[†]Corresponding author. E-mail: scxbyang@scut.edu.cn

Langevin equation can be an effective tool to sample the equilibrium canonical ensemble.

The aim of this work is to study the stationary distributions of the algorithms for numerically solving the overdamped Langevin equation. Various numerical algorithms for overdamped Langevin dynamics (also called Brownian dynamics) have been proposed in previous studies.[8,13–17] Systematic errors arise from the finite time interval Δt of the numerical algorithms. Our interest is in the exact solution of the stationary distribution with finite Δt, in the harmonic system. The rest of this paper is organized as follows. In Section 2, we introduce three algorithms based on the splitting method of the integrators for underdamped Langevin equation. In Section 3, the exact stationary distribution of each algorithm in one-dimensional harmonic system is derived. We employ the stochastic analysis of the discrete time trajectory in the configurational space. The extension of the results to the multi-dimensional harmonic system is done in Section 4, by making use of the normal mode coordinate transformation. The accuracy of the stationary distribution generated from each algorithm is illustrated. Conclusions are outlined in Section 5.

2. Numerical algorithms

First we recall the splitting method to construct the integrator of underdamped Langevin equation by Leimkuhler and Matthews,[13]

$$\begin{bmatrix} \frac{d\boldsymbol{x}}{dt} \\ \frac{d\boldsymbol{p}}{dt} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{M}^{-1}\boldsymbol{p} \\ 0 \end{bmatrix}}_{\mathrm{A}} + \underbrace{\begin{bmatrix} 0 \\ -\nabla U(\boldsymbol{x}) \end{bmatrix}}_{\mathrm{B}} + \underbrace{\begin{bmatrix} 0 \\ -\tilde{\gamma}\boldsymbol{p} + \sqrt{\frac{2\tilde{\gamma}}{\beta}}\boldsymbol{M}^{1/2}\boldsymbol{\eta}(t) \end{bmatrix}}_{\mathrm{O}}, \tag{7}$$

where $\tilde{\gamma}$ is the friction coefficient, $\boldsymbol{\eta}(t)$ is the random vector satisfying the conditions in Eq. (3), and "O" represents the Ornstein–Uhlenbeck[18,19] part. Each of the three parts of underdamped Langevin equation can be solved "exactly", and the integrator can be obtained from a certain composition of these three parts. As an example, BAOAB denotes the integrator of the order of composition $\mathrm{B}\left(\frac{1}{2}\delta t\right)\mathrm{A}\left(\frac{1}{2}\delta t\right)\mathrm{O}(\delta t)\mathrm{A}\left(\frac{1}{2}\delta t\right)\mathrm{B}\left(\frac{1}{2}\delta t\right)$, where we use δt as the time interval for the underdamped Langevin dynamics. It is straightforward to verify that $\mathrm{B}\left(\frac{1}{2}\delta t\right)\mathrm{A}(\delta t)\mathrm{B}\left(\frac{1}{2}\delta t\right)$ is the famous velocity–Verlet algorithm[20] for integrating the Hamiltonian equation. Alternative splitting methods which divide the underdamped Langevin equation into two parts have also been discussed.[21,22] Among the integrators constructed from the splitting methods, BAOAB scheme has been shown to produce the most accurate stationary distribution in the configurational space.[13,21–28] Recently, a scheme named leapflog "middle" was studied.[29,30] This scheme is equivalent to BAOA in the stationary state and leads to the same configurational distribution as that of BAOAB. The numerical algorithms of overdamped Langevin equation considered in this work can be obtained from the underdamped Langevin integrators constructed from Eq. (7).

The BAOA scheme of underdamped Langevin dynamics is expressed as

$$\begin{aligned}
\boldsymbol{p}_{n,1} &= \boldsymbol{p}_n - \delta t \nabla U(\boldsymbol{x}_n), \\
\boldsymbol{x}_{n,1} &= \boldsymbol{x}_n + \frac{\delta t}{2}\boldsymbol{M}^{-1}\boldsymbol{p}_{n,1}, \\
\boldsymbol{p}_{n+1} &= \mathrm{e}^{-\tilde{\gamma}\delta t}\boldsymbol{p}_{n,1} + \sqrt{\frac{1}{\beta}\left(1-\mathrm{e}^{-2\tilde{\gamma}\delta t}\right)}\boldsymbol{M}^{1/2}\boldsymbol{\mu}_{n+1}, \\
\boldsymbol{x}_{n+1} &= \boldsymbol{x}_{n,1} + \frac{\delta t}{2}\boldsymbol{M}^{-1}\boldsymbol{p}_{n+1},
\end{aligned} \tag{8}$$

with the order $\mathrm{B}(\delta t)\mathrm{A}\left(\frac{1}{2}\delta t\right)\mathrm{O}(\delta t)\mathrm{A}\left(\frac{1}{2}\delta t\right)$ in a time step. Here $\boldsymbol{\mu}_{n+1}$ is a standard normal random vector. The random vector in each time step is independent. Arranging Eq. (8) into a compact expression leads to

$$\begin{aligned}
\boldsymbol{x}_{n+1} =\ & \boldsymbol{x}_n + \frac{\delta t}{2}\boldsymbol{M}^{-1}\left(1+\mathrm{e}^{-\tilde{\gamma}\delta t}\right)\left[\boldsymbol{p}_n - \delta t\nabla U(\boldsymbol{x}_n)\right] \\
& + \frac{\delta t}{2}\sqrt{\frac{1}{\beta}\left(1-\mathrm{e}^{-2\tilde{\gamma}\delta t}\right)}\boldsymbol{M}^{-1/2}\boldsymbol{\mu}_{n+1}, \\
\boldsymbol{p}_{n+1} =\ & \mathrm{e}^{-\tilde{\gamma}\delta t}\left[\boldsymbol{p}_n - \delta t\nabla U(\boldsymbol{x}_n)\right] \\
& + \sqrt{\frac{1}{\beta}\left(1-\mathrm{e}^{-2\tilde{\gamma}\delta t}\right)}\boldsymbol{M}^{1/2}\boldsymbol{\mu}_{n+1}.
\end{aligned} \tag{9}$$

In choosing the overdamped limit $\tilde{\gamma} = +\infty$, the momentum turns into a random vector from the Maxwell distribution, *i.e.*, $\boldsymbol{p}_{n+1} = \sqrt{1/\beta}\boldsymbol{M}^{1/2}\boldsymbol{\mu}_{n+1}$. Then the updated configurational variable in a time step becomes

$$\begin{aligned}
\boldsymbol{x}_{n+1} =\ & \boldsymbol{x}_n - \frac{\delta t^2}{2}\boldsymbol{M}^{-1}\nabla U(\boldsymbol{x}_n) \\
& + \frac{\delta t}{2}\sqrt{\frac{1}{\beta}}\boldsymbol{M}^{-1/2}\left(\boldsymbol{\mu}_n + \boldsymbol{\mu}_{n+1}\right).
\end{aligned} \tag{10}$$

Note that the friction coefficient and time interval of overdamped Langevin equation are γ and Δt, respectively. If we take the time interval of underdamped Langevin dynamics as $\delta t = \sqrt{2\Delta t/\gamma}$, equation (10) converts into

$$\begin{aligned}
\boldsymbol{x}_{n+1} =\ & \boldsymbol{x}_n - \frac{\Delta t}{\gamma}\boldsymbol{M}^{-1}\nabla U(\boldsymbol{x}_n) \\
& + \sqrt{\frac{\Delta t}{2\beta\gamma}}\boldsymbol{M}^{-1/2}\left(\boldsymbol{\mu}_n + \boldsymbol{\mu}_{n+1}\right),
\end{aligned} \tag{11}$$

which is a numerical solution of overdamped Langevin equation in a time interval Δt. For convenience, we call this algorithm "BAOA-limit". Remark that BAOA-limit is non-Markovian due to the correlation between the stochastic part $\boldsymbol{\mu}_n + \boldsymbol{\mu}_{n+1}$ in a couple of time steps. This algorithm was first proposed in Ref. [13] by the overdamped limit of BAOAB, and rederived from the postprocessed integrators in Ref. [15]. It was further studied from the perspective of the high accuracy for the stationary distribution.[8,14,17,23] Our analysis here clearly shows the equivalence between BAOA and BAOAB in the configurational space.

Now we come to consider the Euler–Maruyama method (EM), which is the most common numerical algorithm for overdamped Langevin dynamics. The updated configurational variable in a time step reads

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \frac{\Delta t}{\gamma}\boldsymbol{M}^{-1}\nabla U(\boldsymbol{x}_n) + \sqrt{\frac{2\Delta t}{\beta\gamma}}\boldsymbol{M}^{-1/2}\boldsymbol{\mu}_n. \quad (12)$$

Using the analysis similar to that of BAOA-limit, one can verify that EM is actually OBAB-limit, also in choosing $\tilde{\gamma} = +\infty$ and $\delta t = \sqrt{2\Delta t/\gamma}$. In the same overdamped limit condition, OABA-limit algorithm has also been introduced in Ref. [16]. The OABA-limit takes the form of

$$\begin{aligned}\boldsymbol{x}_{n+1} &= \boldsymbol{x}_n - \frac{\Delta t}{\gamma}\boldsymbol{M}^{-1}\nabla U\left(\boldsymbol{x}_n + \sqrt{\frac{\Delta t}{2\beta\gamma}}\boldsymbol{M}^{-1/2}\boldsymbol{\mu}_n\right)\\ &\quad + \sqrt{\frac{2\Delta t}{\beta\gamma}}\boldsymbol{M}^{-1/2}\boldsymbol{\mu}_n. \end{aligned} \quad (13)$$

The EM and OABA-limit are Markovian.

In the rest of this paper, we compare the long-time behaviours of BAOA-limit [Eq. (11)], EM [Eq. (12)], and OABA-limit [Eq. (13)] in the harmonic system. The numerical stationary distributions of these three algorithms are explicitly derived, and the stable conditions for Δt are verified. Numerical tests in the quartic model, which is an anharmonic system, are shown in the Appendix A.

3. Numerical stationary distributions in one-dimensional harmonic system

Exact solutions for the probability distribution of stochastic numerical algorithm are rare. Harmonic system is one of the exactly solvable models in statistical physics. In the one-dimensional harmonic system $U(x) = \frac{1}{2}m\omega^2x^2$, where ω is the frequency, the force is simply linear $-\nabla U(x) = -m\omega^2 x$. The overdamped Langevin equation Eq. (1) for this system is now given by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\omega^2}{\gamma}x^2 + \sqrt{\frac{2}{\beta\gamma m}}\eta(t). \quad (14)$$

The Fokker–Planck equation Eq. (4) becomes

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{\partial}{\partial x}\left[\frac{\omega^2}{\gamma}x\cdot\rho(x,t) + \frac{1}{\beta\gamma m}\frac{\partial\rho(x,t)}{\partial x}\right]. \quad (15)$$

Both equations (14) and (15) can be exactly solved. We treat $\eta(t)$ as a "function" of t, then the explicit expression of the solution for Eq. (14) can be obtained to be

$$\begin{aligned}x(t) &= \exp\left(-\frac{\omega^2 t}{\gamma}\right)x_0\\ &\quad + \int_0^t \exp\left(-\frac{\omega^2(t-s)}{\gamma}\right)\sqrt{\frac{2}{\beta\gamma m}}\eta(s)\,\mathrm{d}s \end{aligned} \quad (16)$$

with the initial condition x_0. The Wiener process ensures that

$$\int_0^t \exp\left(-\frac{\omega^2(t-s)}{\gamma}\right)\sqrt{\frac{2}{\beta\gamma m}}\eta(s)\,\mathrm{d}s$$

is a normal random variable. Therefore, the probability distribution of $x(t)$ is the Gaussian distribution with the mean $\mathrm{e}^{-\omega^2 t/\gamma}x_0$ and the variance

$$\begin{aligned}&\left\langle\left[\int_0^t \exp\left[-\frac{\omega^2(t-s)}{\gamma}\right]\sqrt{\frac{2}{\beta\gamma m}}\eta(s)\,\mathrm{d}s\right]^2\right\rangle\\ &= \frac{2}{\beta\gamma m}\int_0^t\int_0^t \exp\left[-\frac{\omega^2(t-s)}{\gamma}\right]\\ &\quad\times\exp\left[-\frac{\omega^2(t-s')}{\gamma}\right]\langle\eta(s)\eta(s')\rangle\,\mathrm{d}s\mathrm{d}s'\\ &= \frac{2}{\beta\gamma m}\int_0^t \exp\left[-\frac{2\omega^2(t-s)}{\gamma}\right]\mathrm{d}s\\ &= \frac{1}{\beta m\omega^2}\left(1-\exp\left[-\frac{2\omega^2 t}{\gamma}\right]\right). \end{aligned} \quad (17)$$

The time-dependent probability density is then

$$\begin{aligned}\rho(x,t) &= \sqrt{\frac{\beta m\omega^2}{2\pi\left(1-\mathrm{e}^{-(2\omega^2 t/\gamma)}\right)}}\\ &\quad\times\exp\left\{-\beta\frac{1}{2}m\omega^2\frac{1}{1-\mathrm{e}^{-(2\omega^2 t/\gamma)}}\right.\\ &\quad\left.\times\left[x-\exp\left(-\frac{\omega^2 t}{\gamma}\right)x_0\right]^2\right\}. \end{aligned} \quad (18)$$

One can easily examine that $\rho(x,t)$ in Eq. (18) is the solution of the Fokker–Planck equation Eq. (15) with the initial condition $\rho(x,0) = \delta(x-x_0)$. As expected, the long-time limit $t\to\infty$ of $\rho(x,t)$ leads to the Boltzmann distribution

$$\rho_{\mathrm{st}}(x) = \sqrt{\frac{\beta m\omega^2}{2\pi}}\exp\left[-\beta\frac{1}{2}m\omega^2x^2\right]. \quad (19)$$

Now we turn to the study of stationary distributions by using the numerical algorithms with a finite time interval Δt. The discrete time evolution of BAOA-limit algorithm in a time interval [Eq. (11)] is

$$x_{n+1} = \left(1-\frac{\omega^2\Delta t}{\gamma}\right)x_n + \sqrt{\frac{\Delta t}{2\beta\gamma m}}(\mu_n+\mu_{n+1}). \quad (20)$$

The discrete time trajectory begins from the initial condition x_0, then it is straightforward to show

$$x_n = \left(1-\frac{\omega^2\Delta t}{\gamma}\right)^n x_0 + \sqrt{\frac{\Delta t}{2\beta\gamma m}}$$

$$\times\sum_{j=0}^{n-1}\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{n-j-1}\left(\mu_j+\mu_{j+1}\right). \tag{21}$$

Rearranging Eq. (21) produces

$$\begin{aligned}x_n =& \left(1-\frac{\omega^2\Delta t}{\gamma}\right)^n x_0+\sqrt{\frac{\Delta t}{2\beta\gamma m}}\left\{\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{n-1}\mu_0\right.\\&+\sum_{j=1}^{n-1}\left[\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{n-j-1}\right.\\&\left.\left.+\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{n-j}\right]\mu_j+\mu_n\right\}.\end{aligned} \tag{22}$$

The configurational space point x_n at each step of the trajectory is a linear combination of standard normal random variables μ_j, thus x_n satisfies the Gaussian distribution. This probability distribution is denoted as $\rho_n(x)$. Since $\rho_n(x)$ is a Gaussian distribution, it is determined directly by the mean $\langle x_n\rangle$ and the variance $\left\langle (x_n-\langle x_n\rangle)^2\right\rangle$. From Eq. (22) it is easy to obtain the mean and the variance

$$\langle x_n\rangle=\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^n x_0, \tag{23}$$

$$\begin{aligned}\left\langle (x_n-\langle x_n\rangle)^2\right\rangle =& \frac{\Delta t}{2\beta\gamma m}\left\{\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{2n-2}+\sum_{j=1}^{n-1}\left[\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{n-j-1}+\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{n-j}\right]^2+1\right\}\\=&\frac{\Delta t}{2\beta\gamma m}\left[\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{2n-2}+\left(2-\frac{\omega^2\Delta t}{\gamma}\right)^2\sum_{j=1}^{n-1}\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{2n-2j-2}+1\right]\\=&\frac{1}{\beta m\omega^2}\left[1-\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{2n-1}\right].\end{aligned} \tag{24}$$

Taking the long-time limit $n\to\infty$, we expect $\rho_n(x)$ to approach to the stationary distribution of BAOA-limit. It is obvious that the stationary distribution exists when the stable condition $\left|1-(\omega^2\Delta t/\gamma)\right|<1$ is satisfied, *i.e.*,

$$\frac{\omega^2\Delta t}{\gamma}<2. \tag{25}$$

Within this stable regime of Δt, the long-time limit of the mean and the variance are simply

$$\langle x_n\rangle\overset{n\to\infty}{\to}0,\quad \left\langle (x_n-\langle x_n\rangle)^2\right\rangle\overset{n\to\infty}{\to}\frac{1}{\beta m\omega^2}. \tag{26}$$

The explicit expression for the stationary distribution is then

$$\rho_{\rm st}^{\rm BAOA-limit}(x)=\sqrt{\frac{\beta m\omega^2}{2\pi}}\exp\left[-\beta\frac{1}{2}m\omega^2x^2\right], \tag{27}$$

which is exactly the Boltzmann distribution Eq. (19) of the system.

Similar analysis can be applied to EM and OABA-limit. The mean and the variance of $\rho_n(x)$ for EM are

$$\begin{aligned}\langle x_n\rangle =& \left(1-\frac{\omega^2\Delta t}{\gamma}\right)^n x_0,\ \left\langle (x_n-\langle x_n\rangle)^2\right\rangle\\=&\frac{1}{\beta m\omega^2}\frac{1}{1-(\omega^2\Delta t/2\gamma)}\left[1-\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{2n}\right].\end{aligned} \tag{28}$$

The stationary distribution in the long-time limit within the stable regime Eq. (25) is

$$\begin{aligned}\rho_{\rm st}^{\rm EM}(x) =& \sqrt{\frac{\beta m\omega^2(1-\omega^2\Delta t/2\gamma)}{2\pi}}\\&\times\exp\left[-\beta\frac{1}{2}m\omega^2x^2\left(1-\frac{\omega^2\Delta t}{2\gamma}\right)\right].\end{aligned} \tag{29}$$

The result for OABA-limit is

$$\begin{aligned}&\langle x_n\rangle=\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^n x_0,\\&\left\langle (x_n-\langle x_n\rangle)^2\right\rangle\\&=\frac{1}{\beta m\omega^2}\left(1-\frac{\omega^2\Delta t}{2\gamma}\right)\left[1-\left(1-\frac{\omega^2\Delta t}{\gamma}\right)^{2n}\right],\end{aligned} \tag{30}$$

$$\begin{aligned}&\rho_{\rm st}^{\rm OABA\text{-}limit}(x)\\&=\sqrt{\frac{\beta m\omega^2}{2\pi(1-(\omega^2\Delta t/2\gamma))}}\\&\times\exp\left[-\beta\frac{1}{2}m\omega^2x^2\frac{1}{1-(\omega^2\Delta t/2\gamma)}\right]\end{aligned} \tag{31}$$

within the stable regime Eq. (25). Obviously, for both EM [Eq. (29)] and OABA-limit [Eq. (31)], the infinitesimal time interval limit $\Delta t\to 0$ leads to the exact Boltzmann distribution. Comparison of the accuracy associated with Δt among these three algorithms is straightforward. The BAOA-limit algorithm generates the exact Boltzmann distribution $\rho_{\rm st}(x)$ even when Δt is finite, as long as the stable condition is satisfied. The EM and OABA-limit fail to do so. The numerical error of the stationary distribution due to the finite Δt is the first order for both EM and OABA-limit. The accuracy analysis can be summarized as

$$\begin{aligned}&\rho_{\rm st}^{\rm BAOA\text{-}limit}(x)=\rho_{\rm st}(x),\\&\rho_{\rm st}^{\rm EM}(x)=\rho_{\rm st}(x)\times[1+{\rm O}(\Delta t)],\\&\rho_{\rm st}^{\rm OABA\text{-}limit}(x)=\rho_{\rm st}(x)\times[1+{\rm O}(\Delta t)].\end{aligned} \tag{32}$$

An observable or a thermodynamic quantity in configurational space can be used for indicating the accuracy of equilibrium sampling. Here we take the potential energy for example. The average potential energy for the stationary distribution of BAOA-limit can be simply obtained to be

$$\langle U\rangle^{\text{BAOA-limit}}=\int\rho_{\text{st}}^{\text{BAOA-limit}}(x)\,\frac{1}{2}m\omega^2x^2\,\mathrm{d}x=\frac{1}{2\beta},\tag{33}$$

which is the exact result for the Boltzmann distribution. The other two algorithms lead to

$$\begin{aligned}\langle U\rangle^{\text{EM}}&=\int\rho_{\text{st}}^{\text{EM}}(x)\,\frac{1}{2}m\omega^2x^2\,\mathrm{d}x\\&=\frac{1}{2\beta}\frac{1}{1-(\omega^2\Delta t/2\gamma)}=\frac{1}{2\beta}\left[1+\mathrm{O}(\Delta t)\right],\end{aligned}\tag{34}$$

$$\begin{aligned}\langle U\rangle^{\text{OABA-limit}}&=\int\rho_{\text{st}}^{\text{OABA-limit}}(x)\,\frac{1}{2}m\omega^2x^2\,\mathrm{d}x\\&=\frac{1}{2\beta}\left(1-\frac{\omega^2\Delta t}{2\gamma}\right)=\frac{1}{2\beta}\left[1+\mathrm{O}(\Delta t)\right].\end{aligned}\tag{35}$$

Therefore, the accuracy of average potential energy is consistent with that of the stationary distribution as shown in Eq. (32).

We remark that in Ref. [22] a trajectory-based approach was used to derive the numerical stationary distribution in the phase space (x,p), of the underdamped Langevin integrator in the harmonic system. The discrete time evolution of configuration and momentum according to the underdamped Langevin equation, with finite friction coefficient and time interval, were considered there. In this study we focus on the overdamped limit of the discrete equations of motion as stated in Section 2. Therefore, the derivation here deals with the overdamped case of Langevin algorithm, by using a similar technique to the trajectory-based approach. The analysis of the discrete time overdamped Langevin trajectory in this paper can be seen as the extension of the trajectory-based approach, which is originally used for underdamped Langevin dynamics.

4. Results in multi-dimensional harmonic system

We turn to the general k-dimensional harmonic system $U(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x}$, where $\boldsymbol{H}$ is the symmetric positive-definite Hessian matrix. The discrete time evolution in Δt of BAOA-limit is now

$$\begin{aligned}\boldsymbol{x}_{n+1}&=\boldsymbol{x}_n-\frac{\Delta t}{\gamma}\boldsymbol{M}^{-1}\boldsymbol{H}\boldsymbol{x}_n\\&\quad+\sqrt{\frac{\Delta t}{2\beta\gamma}}\boldsymbol{M}^{-1/2}(\boldsymbol{\mu}_n+\boldsymbol{\mu}_{n+1}).\end{aligned}\tag{36}$$

One can obtain the configurational space point $\boldsymbol{x}_n$, similar to the expressions in Eqs. (21) and (22) for one-dimensional case. Here we employ an alternative approach by using the normal mode coordinate transformation. The characteristic frequencies of the normal modes can be obtained from the eigenvalues of the matrix $\boldsymbol{M}^{-1/2}\boldsymbol{H}\boldsymbol{M}^{-1/2}$, *i.e.*,

$$\boldsymbol{M}^{-1/2}\boldsymbol{H}\boldsymbol{M}^{-1/2}\boldsymbol{T}=\boldsymbol{T}\left[\omega^2\right].\tag{37}$$

Here, $\boldsymbol{T}$ is an orthogonal matrix consisting of the eigenvectors of $\boldsymbol{M}^{-1/2}\boldsymbol{H}\boldsymbol{M}^{-1/2}$, and

$$\left[\omega^2\right]=\begin{bmatrix}\omega_1^2&&\\&\ddots&\\&&\omega_k^2\end{bmatrix}\tag{38}$$

is a diagonal matrix consisting of the eigenvalues with the characteristic frequency ω_i for each normal mode. Then the normal mode coordinate transformation is

$$\boldsymbol{q}=\boldsymbol{T}^{\mathrm{T}}\boldsymbol{M}^{1/2}\boldsymbol{x}\tag{39}$$

and the Hessian matrix $\boldsymbol{H}$ satisfies

$$\boldsymbol{H}=\boldsymbol{M}^{1/2}\boldsymbol{T}\left[\omega^2\right]\boldsymbol{T}^{\mathrm{T}}\boldsymbol{M}^{1/2}.\tag{40}$$

Rewriting the discrete time evolution Eq. (36) in the normal mode configurational space yields

$$\boldsymbol{q}_{n+1}=\boldsymbol{q}_n-\frac{\Delta t}{\gamma}\left[\omega^2\right]\boldsymbol{q}_n+\sqrt{\frac{\Delta t}{2\beta\gamma}}\boldsymbol{T}^{\mathrm{T}}(\boldsymbol{\mu}_n+\boldsymbol{\mu}_{n+1}).\tag{41}$$

Denoting $\tilde{\mu}_n=\boldsymbol{T}^{\mathrm{T}}\boldsymbol{\mu}_n$, we find that $\tilde{\mu}_n$ is still a standard normal random vector. It is easy to verify that the evolution of each degree of normal mode coordinate is independent in Eq. (41). For the degree i, the evolution reads

$$q_{i,n+1}=\left(1-\frac{\omega_i^2\Delta t}{\gamma}\right)q_{i,n}+\sqrt{\frac{\Delta t}{2\beta\gamma}}(\tilde{\mu}_{i,n}+\tilde{\mu}_{i,n+1}),\tag{42}$$

which takes the same form as that of Eq. (20) with a unit mass for the one-dimensional case. Then we can directly employ Eq. (27) to obtain the stationary distribution of q_i as follows:

$$\rho_{\text{st}}^{\text{BAOA-limit}}(q_i)=\sqrt{\frac{\beta\omega_i^2}{2\pi}}\exp\left[-\beta\frac{1}{2}\omega_i^2q_i^2\right]\tag{43}$$

with the stable regime

$$\frac{\omega_i^2\Delta t}{\gamma}<2.\tag{44}$$

Since the distribution of each degree of normal mode coordinate is independent, the stationary distribution in the normal mode configurational space is then

$$\begin{aligned}\rho_{\text{st}}^{\text{BAOA-limit}}(\boldsymbol{q})&=\left(\frac{\beta}{2\pi}\right)^{k/2}\left(\det\left[\omega^2\right]\right)^{1/2}\\&\quad\times\exp\left\{-\beta\frac{1}{2}\boldsymbol{q}^{\mathrm{T}}\left[\omega^2\right]\boldsymbol{q}\right\}.\end{aligned}\tag{45}$$

By the inverse coordinate transformation $\boldsymbol{x}=\boldsymbol{M}^{-1/2}\boldsymbol{T}q$, we obtain

$$\rho_{\text{st}}^{\text{BAOA-limit}}(\boldsymbol{x})$$

$$
\begin{aligned}
&= \left(\frac{\beta}{2\pi}\right)^{k/2} \left(\det\left[\boldsymbol{\omega}^2\right]\right)^{1/2} \left(\det \boldsymbol{M}\right)^{1/2} \\
&\quad \times \exp\left\{-\beta \frac{1}{2} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{M}^{1/2} \boldsymbol{T}\left[\boldsymbol{\omega}^2\right] \boldsymbol{T}^{\mathrm{T}} \boldsymbol{M}^{1/2} \boldsymbol{x}\right\} \\
&= \left(\frac{\beta}{2\pi}\right)^{k/2} \left(\det \boldsymbol{H}\right)^{1/2} \exp\left\{-\beta \frac{1}{2} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{x}\right\}. \qquad (46)
\end{aligned}
$$

The result demonstrates that BAOA-limit generates the exact Boltzmann distribution in the stationary state, when the stable condition of Δt Eq. (44) is satisfied for each characteristic frequency ω_i.

The stationary distributions of EM and OABA-limit can be derived by using a similar technique

$$
\begin{aligned}
&\rho_{\mathrm{st}}^{\mathrm{EM}}(\boldsymbol{x}) \\
&= \left(\frac{\beta}{2\pi}\right)^{k/2} \left(\det\left[\left(\boldsymbol{1} - \boldsymbol{H}\boldsymbol{M}^{-1}\frac{\Delta t}{2\gamma}\right)\boldsymbol{H}\right]\right)^{1/2} \\
&\quad \times \exp\left\{-\beta \frac{1}{2} \boldsymbol{x}^{\mathrm{T}} \left(\boldsymbol{1} - \boldsymbol{H}\boldsymbol{M}^{-1}\frac{\Delta t}{2\gamma}\right) \boldsymbol{H}\boldsymbol{x}\right\}, \qquad (47) \\
&\rho_{\mathrm{st}}^{\mathrm{OABA\text{-}limit}}(\boldsymbol{x}) \\
&= \left(\frac{\beta}{2\pi}\right)^{k/2} \left(\det\left[\left(\boldsymbol{1} - \boldsymbol{H}\boldsymbol{M}^{-1}\frac{\Delta t}{2\gamma}\right)^{-1}\boldsymbol{H}\right]\right)^{1/2} \\
&\quad \times \exp\left\{-\beta \frac{1}{2} \boldsymbol{x}^{\mathrm{T}} \left(\boldsymbol{1} - \boldsymbol{H}\boldsymbol{M}^{-1}\frac{\Delta t}{2\gamma}\right)^{-1} \boldsymbol{H}\boldsymbol{x}\right\}. \qquad (48)
\end{aligned}
$$

For both algorithms, the stable condition is shown in Eq. (44) for each characteristic frequency ω_i, and the infinitesimal time interval limit $\Delta t \to 0$ leads to the exact Boltzmann distribution. It is trivial to find that the results proposed here for all the algorithms are consistent with those of the one-dimensional case when k reduces to 1. The advantage in the long-time behaviour of BAOA-limit algorithm is shown clearly. For both EM and OABA-limit, numerical error in the stationary distribution is the first order of Δt in comparison with the exact Boltzmann distribution. Therefore, the accuracy of these algorithms is consistent with that of the one-dimensional case as shown in Eq. (32).

The average potential energy for each algorithm is also derived conveniently from the one-dimensional case shown in Section 2. For BAOA-limit, we have the exact result for the Boltzmann distribution

$$
\begin{aligned}
\langle U\rangle^{\mathrm{BAOA\text{-}limit}} &= \int \rho_{\mathrm{st}}^{\mathrm{BAOA\text{-}limit}}(\boldsymbol{x}) \frac{1}{2} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{H}\boldsymbol{x}\, \mathrm{d}\boldsymbol{x} \\
&= \int \rho_{\mathrm{st}}^{\mathrm{BAOA\text{-}limit}}(\boldsymbol{q}) \frac{1}{2} \boldsymbol{q}^{\mathrm{T}} \left[\boldsymbol{\omega}^2\right] \boldsymbol{q}\, \mathrm{d}\boldsymbol{q} \\
&= \sum_{i=1}^{k} \int \rho_{\mathrm{st}}^{\mathrm{BAOA\text{-}limit}}(q_i) \frac{1}{2} \omega_i^2 q_i^2\, \mathrm{d}q_i \\
&= \frac{k}{2\beta}, \qquad (49)
\end{aligned}
$$

of which the last step can be easily verified by Eqs. (43) and (33). Similar derivations for the other two algorithms can be conducted by using Eqs. (34) and (35). The results are as follows:

$$
\begin{aligned}
\langle U\rangle^{\mathrm{EM}} &= \frac{1}{2\beta} \sum_{i=1}^{k} \frac{1}{1 - (\omega_i^2 \Delta t / 2\gamma)} \\
&= \frac{k}{2\beta}\left[1 + \mathrm{O}(\Delta t)\right], \qquad (50) \\
\langle U\rangle^{\mathrm{OABA\text{-}limit}} &= \frac{1}{2\beta} \sum_{i=1}^{k} \left(1 - \frac{\omega_i^2 \Delta t}{2\gamma}\right) \\
&= \frac{k}{2\beta}\left[1 + \mathrm{O}(\Delta t)\right]. \qquad (51)
\end{aligned}
$$

Once again, it can be seen that the accuracy of average potential energy is consistent with Eq. (32).

5. Conclusions

The high accuracy of BAOA-limit (or BAOAB-limit) algorithm in the equilibrium configurational sampling has been confirmed in previous studies. In this work, we present the evidence for this statement in an exactly solvable system. Within the stable regime of Δt, the numerical stationary distribution generated by BAOA-limit algorithm of the harmonic system is exactly the Boltzmann distribution, both for one-dimensional case and for multi-dimensional case. The EM (OBAB-limit) and OABA-limit, the other two algorithms considered here for comparison, both lead to the stationary distribution with the first order error of Δt. To derive the explicit expression of the numerical stationary distribution, we conduct the stochastic analysis of the discrete time trajectory. This method may be extended to other numerical solutions of stochastic differential equations.

Besides sampling the equilibrium canonical ensemble, the overdamped Langevin equation can be employed in the nonstationary process. For example, a conditional overdamped Langevin equation for generating transition paths from an initial state to a given final state is proposed.[31–36] Nonstationary distribution of overdamped Langevin equation deserves further study. The performance of BAOA-limit algorithm in sampling the transition path ensemble is of interest in future work.

Appendix A: Numerical tests in quartic potential model

In this appendix we present the numerical results for the quartic potential model $U(x) = \frac{1}{4}x^4$, for a comparison among three overdamped Langevin algorithms in an anharmonic system. For this model, we are not able to derive the exact solutions of the stationary distributions of three algorithms, therefore numerical simulations are necessary. The simulations are performed to calculate the average potential energy. The exact value of the average potential energy for the Boltzmann distribution of this model is simply

$$\begin{aligned}\langle U\rangle &= \frac{1}{Z}\int \exp\left\{-\beta\frac{1}{4}x^4\right\}\frac{1}{4}x^4\,\mathrm{d}x \\ &= -\frac{1}{\beta}\frac{1}{Z}\int \frac{1}{4}\frac{\partial\,\mathrm{e}^{-\beta(1/4)x^4}}{\partial x}x\,\mathrm{d}x \\ &= \frac{1}{4\beta}\frac{1}{Z}\int \mathrm{e}^{-\beta(1/4)x^4}\,\mathrm{d}x \\ &= \frac{1}{4\beta}. \end{aligned} \tag{A1}$$

We set $\beta = 1$ and $\gamma = 100$ in the simulations and employ the time intervals $\Delta t = 1,\ldots,10$. For each time interval, 20 independent samples are generated. The average and statistical error bar of any physical quantity are calculated over these 20 samples. The performances of three algorithms in calculating the average potential energy are shown in Fig. A1. Comparison of the numerical results with the exact value clearly demonstrates that the BAOA-limit algorithm generates the most accurate sampling in the quartic potential model. Thus, the BAOA-limit also exhibits high accuracy in the stationary distribution of the anharmonic system, while EM and OABA-limit both lead to a significant error of Δt. The numerical results here are consistent with the exact solutions for the stationary distributions of three algorithms in harmonic system.

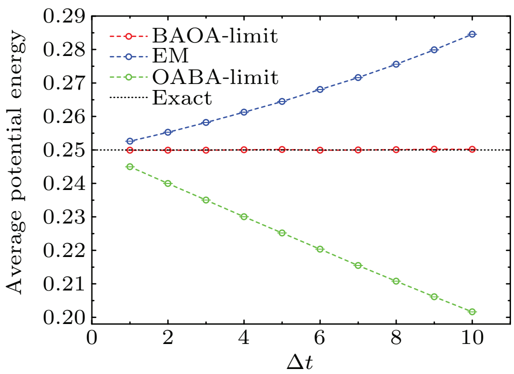

Fig. A1. Results of average potential energy of quartic potential model using different time intervals for three overdamped Langevin algorithms. Statistical error bars are included.

Availability of supporting data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

De-Zhang Li thanks Dr. Yun-Feng Xiong for the discussion about the content of this paper through the private communication.

Project supported by the Basic and Applied Basic Research Foundation of Guangdong Province, China (Grant No. 2021A1515010328), the Key-Area Research and Development Program of Guangdong Province, China (Grant No. 2020B010183001), and the National Natural Science Foundation of China (Grant No. 12074126).

References

[1] Langevin P 1908 *C. R. Acad. Sci.* (Paris) **146** 530
[2] Lemons D S and Gythiel A 1997 *Am. J. Phys.* **65** 1079
[3] Einstein A 1905 *Ann. Phys.* **322** 549
[4] Einstein A 1906 *Ann. Phys.* **324** 371
[5] Smoluchowski M v 1906 *Ann. Phys.* **326** 756
[6] Kampen N G v 2009 *Stochastic Processes in Physics and Chemistry* 3rd edn. (Amsterdam: Elsevier)
[7] Zwanzig R 2001 *Nonequilibrium statistical mechanics* (New York: Oxford University Press)
[8] Leimkuhler B and Sachs M 2022 *J. Sci. Comput.* **44** A364
[9] Fokker A D 1914 *Ann. Phys.* **348** 810
[10] Planck V 1917 *Akad. Wiss.* **24** 324
[11] Risken H 1989 *The Fokker–Planck Equation: Methods of Solution and Applications* (Berlin: Springer-Verlag)
[12] Pavliotis G A 2014 *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations* (New York: Springer)
[13] Leimkuhler B and Matthews C 2013 *Appl. Math. Res. Express* **2013** 34
[14] Leimkuhler B, Matthews C and Tretyakov M V 2014 *Proc. R. Soc. A* **470** 20140120
[15] Vilmart G 2015 *SIAM J. Sci. Comput.* **37** A201
[16] Fathi M and Stoltz G 2017 *Numer. Math.* **136** 545
[17] Shang X and Kröger M 2020 *SIAM Rev.* **62** 901
[18] Uhlenbeck G E and Ornstein L S 1930 *Phys. Rev.* **36** 823
[19] Wang M C and Uhlenbeck G E 1945 *Rev. Mod. Phys.* **17** 323
[20] Verlet L 1967 *Phys. Rev.* **159** 98
[21] Leimkuhler B and Matthews C 2013 *J. Chem. Phys.* **138** 174102
[22] Li D, Han X, Chai Y, Wang C, Zhang Z, Chen Z, Liu J and Shao J 2017 *J. Chem. Phys.* **147** 184104
[23] Leimkuhler B, Matthews C and Stoltz G 2016s *IMA J. Numer. Anal.* **36** 13
[24] Grønbech-Jensen N and Farago O 2013 *Mol. Phys.* **111** 983
[25] Liu J, Li D and Liu X 2016 *J. Chem. Phys.* **145** 024103
[26] Zhang Z, Liu X, Chen Z, Zheng H, Yan K and Liu J 2017 *J. Chem. Phys.* **147** 034109
[27] Li D, Chen Z, Zhang Z and Liu J 2017 *Chin. J. Chem. Phys.* **30** 735
[28] Grønbech-Jensen N 2020 *Mol. Phys.* **118** e1662506
[29] Zhang Z, Yan K, Liu X and Liu J 2018 *Chin. Sci. Bull.* **63** 3467
[30] Zhang Z, Liu X, Yan K, Tuckerman M E and Liu J 2019 *J. Phys. Chem. A* **123** 6056
[31] Orland H 2011 *J. Chem. Phys.* **134** 174114
[32] Majumdar S N and Orland H 2015 *J. Stat. Mech.* **2015** P06039
[33] Delarue M, Koehl P and Orland H 2017 *J. Chem. Phys.* **147** 152703
[34] Elber R, Makarov D E and Orland H 2020 *Molecular Kinetics in Condensed Phases: Theory, Simulation, and Analysis* (Hoboken, NJ: Wiley)
[35] Koehl P and Orland H 2022 *J. Chem. Phys.* **157** 054105
[36] Li D, Zeng J, Huang W, Yao Y and Yang X 2023 *Phys. Scr.* **98** 025218