Rational Construction of Stochastic Numerical Methods for Molecular Sampling

Rational Construction of Stochastic Numerical Methods for Molecular Sampling

Benedict Leimkuhler and Charles Matthews *

March 24, 2012

Abstract

In this article, we focus on the sampling of the configurational Gibbs-Boltzmann distribution, that is, the calculation of averages of functions of the position coordinates of a molecular N-body system modelled at constant temperature. We show how a formal series expansion of the invariant measure of a Langevin dynamics numerical method can be obtained in a straightforward way using the Baker-Campbell-Hausdorff lemma. We then compare Langevin dynamics integrators in terms of their invariant distributions and demonstrate a superconvergence property (4th order accuracy where only 2nd order would be expected) of one method in the high friction limit; this method, moreover, can be reduced to a simple modification of the Euler-Maruyama method for Brownian dynamics involving a non-Markovian (coloured noise) random process. In the Brownian dynamics case, 2nd order accuracy of the invariant density is achieved. All methods considered are efficient for molecular applications (requiring one force evaluation per timestep) and of a simple form. In fully resolved (long run) molecular dynamics simulations, for our favoured method, we observe up to two orders of magnitude improvement in configurational sampling accuracy for given stepsize with no evident reduction in the size of the largest usable timestep compared to common alternative methods.

keywords: molecular dynamics; sampling; Langevin dynamics; Brownian dynamics; stochastic dynamics; diffusion equation.

*School of Mathematics and Maxwell Institute of Mathematical Sciences, James Clerk Maxwell Building, Kings Buildings, University of Edinburgh, Edinburgh, EH9 3JZ, UK

1 Introduction

Let $U : \mathbf{R}^N \to \mathbf{R}$ be the potential energy function of a classical model for a molecular system. A fundamental challenge is to sample the configurational Gibbs-Boltzmann (canonical) distribution with density

$$\bar{\rho}_\beta(x) = Z^{-1} \exp(-\beta U(x)), \tag{1}$$

where $\beta^{-1} = k_B T$ where k_B is Boltzmann's constant, T is temperature, and Z is a normalization constant so that $\bar{\rho}_\beta$ has unit integral over the entire configuration space. A wide variety of methods are available to calculate averages with respect to $\bar{\rho}_\beta$, among these some of the most popular are based on Brownian dynamics or Langevin dynamics (defined in the phase space of positions and momenta) [1, 2, 3, 4, 5]. (Note: in this article we focus exclusively on molecular dynamics techniques; molecular models can also be sampled using Monte-Carlo methods, and, more generally, using hybrid algorithms which combine molecular dynamics with a Metropolis-Hastings test in order to correct averages. For a recent review of such schemes see [6].) Recall that Brownian dynamics (overdamped Langevin dynamics) is a system of Itō-type stochastic differential equations of the form

$$\mathrm{d}x = -M^{-1}\nabla U(x)\mathrm{d}t + \sqrt{2\beta^{-1}}M^{-1/2}\mathrm{d}W, \quad x(0) = x_0, \tag{2}$$

where $\mathrm{d}W(t)$ is the infinitesimal increment of a vector of stochastic Wiener processes $W(t)$, and M is a positive (we assume here diagonal) mass matrix.[1] A simple and popular method for numerical solution of (2) is the Euler-Maruyama method

$$x_{n+1} = x_n - hM^{-1}\nabla U(x_n) + \sqrt{2k_B T h}M^{-1/2}R_n,$$

where R_n is a vector of random variables with standard normal distribution. This produces a sequence of points $x_0, x_1, x_2, \ldots$, which, following a certain relaxation period, are approximately distributed according to the canonical invariant distribution. Euler-Maruyama has the property that the time averages along discrete trajectories, in the limit of large time (under appropriate conditions on the potential $U(x)$ and assuming no effects from floating point rounding error), have error proportional to h. One of the observations of this article is that the simple modification

$$x_{n+1} = x_n - hM^{-1}\nabla U(x_n) + \sqrt{\frac{k_B T h}{2}}M^{-1/2}(R_n + R_{n+1}) \tag{3}$$

provides a second order approximation of stationary averages.

[1] In terms of sampling the Gibbs distribution, M is in fact arbitrary, but it may be useful to allow for coordinate scaling.

We arrive at this scheme by considering the large friction limit in a particular numerical method for Langevin dynamics. Recall that Langevin dynamics is a stochastic-dynamical system involving both positions and momenta p of the form

$$\mathrm{d}x = M^{-1}p\mathrm{d}t, \quad \mathrm{d}p = [-\nabla U(x) - \gamma p]\mathrm{d}t + \sigma M^{1/2}\mathrm{d}W, \tag{4}$$

where $W = W(t)$ is again a vector of N independent Wiener processes, and $\gamma > 0$ is a free parameter, the friction coefficient. The methods that are in fact the primary focus of this paper are splitting integrators that decompose the stochastic vector field of Langevin dynamics into simpler vector fields which can be solved exactly. The composition method that results cannot be directly related to a stochastic differential equation, so the analogy with backward error analysis for deterministic problems [16, 17] is incomplete, but nonetheless the invariant measure associate to the numerical method can be derived, using the Baker-Campbell-Hausdorff (BCH) expansion, as an asymptotic series in two parameters: the stepsize and the reciprocal of the friction coefficient. The superconvergence result alluded to above is then obtained in the high friction limit. The Langevin stepsize must be understood to be proportional to the square root of the stepsize that appears in Eq. 3, so in Langevin dynamics an effective 4th order approximation is obtained, but only for the marginal configurational invariant distribution, Eq. 1. Our approach also provides a simple method for comparative assessment of the invariant measure of a class of Langevin integrators.

In large scale simulations, it must be acknowledged that the statistical error (dependent on the number of samples used) is typically the dominant concern. Our approach focuses instead on the truncation error of the invariant distribution, thus the greatest benefit would be seen only when the statistical error is well controlled. Nonetheless we observe in our numerical experiments that the least biased scheme from the point of view of the bias introduced in configurational sampling is also as efficient as the alternatives and the most robust with respect to variation of the parameters (stepsize, friction coefficient), thus there is effectively no price for the improvement in accuracy. Moreover, it would be possible to complement the methods proposed here by procedures such as importance sampling [7] to further reduce the statistical error.

With regard to the approximation of canonical averages, methods have previously been constructed for Brownian dynamics with order > 1 and for Langevin dynamics with order > 2 [2, 4], but these require multiple evaluations of the force; for this reason they are not normally viewed as competitive alternatives for molecular sampling [8]. By contrast all of the methods described in this article use a single force evaluation at each timestep.

The approach used here may be compared to other recent works on stochastic numerical methods, and, in particular [9, 10, 11, 12, 13, 14, 15]. Our technique differs from these in (i)

the direct focus on the stationary configurational distribution, and (ii) the use of the BCH expansion. Other articles (see e.g. [5]) which address the invariant measure in Langevin-type stochastic differential equations do not use the backward error analysis (and do not find the superconvergent scheme).

The rest of this article is as follows. In Section 2 we review necessary background on stochastic differential equations for sampling from the Gibbs measure. Section 3 presents our expansion of the associated perturbed invariant measure and calculations involving the use of the Baker-Campbell-Hausdorff theorem. Section 4 describes the reduction of the methods in the case of overdamped Langevin Dynamics. Section 5 demonstrates the theory obtained using numerical experiments to verify the results.

2 Background

The Ornstein-Uhlenbeck (OU) stochastic differential equation is an Itō equation of the form

$$\mathrm{d}u = -\gamma u \mathrm{d}t + \sigma \mathrm{d}W,$$

where we assume that U is a random variable defined for each time t, γ and σ are positive parameters and $\mathrm{d}W$ represents the infinitesimal increment of a Wiener process. Langevin dynamics (Eq. 4) combines the Ornstein-Uhlenbeck stochastic vector field with conservative dynamics.

For stochastic differential equations, the density evolves according to an evolution equation (the Fokker-Planck or forward Kolmogorov equation) of the form

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho,$$

where $\mathcal{L}^*$ is a second order differential operator. In the case of Langevin dynamics, the relevant Kolmogorov operator is defined by its action on a function ϕ of the variables of the system by

$$\mathcal{L}^*_{\mathrm{LD}} \phi = -M^{-1} p \cdot \nabla_x \phi + \nabla U(x) \cdot \nabla_p \phi + \gamma \nabla_p \cdot (p\phi) + \frac{\sigma^2}{2} \Delta_p \phi,$$

where Δ_p is the mass-weighted Laplacian in the momenta: $\Delta_p = \sum_i m_i \frac{\partial^2}{\partial p_i^2}$. By choosing $\sigma = \sqrt{2 k_B T \gamma}$ one easily checks that the Gibbs distribution with density $\rho_\beta = \tilde{Z}^{-1} \exp(-\beta H)$ is a steady state of the Kolmogorov equation, where $\tilde{Z}$ is a normalising constant and $H(x,p) = p^T M^{-1} p/2 + U(x)$ represents the Hamiltonian energy function. For later reference, we define averages with respect to the Gibbs distribution by

$$\mathbb{E}\left(\phi(x,p)\right) = \int \phi(x,p) \rho_\beta \, \mathrm{d}x \, \mathrm{d}p.$$

It is possible to demonstrate that the operator $\mathcal{L}_{\mathrm{LD}}^*$ is hypoelliptic by using Hörmander's criterion [18] based on iterated commutators, and this implies that the Gibbs measure is the unique steady state (up to normalization). Even stronger is the result of [19] which demonstrates the existence of a spectral gap for the operator $\mathcal{L}_{\mathrm{LD}}$, under certain conditions on U. Many of the challenges related to obtaining formal analytical results for stochastic differential equations relate to singularities of the potential and unboundedness of the solution domain. However, with periodic boundary conditions and strong repulsive potentials (e.g. Lennard-Jones potentials) we observe that configurations typical evolve in a bounded set and remain far from singular points (the radial distribution vanishes in a large interval around the origin). For the purpose of deriving practical methods we assume that (i) the positions are confined to a periodic simulation box $\Omega = L_x\mathbb{T} \times L_y\mathbb{T} \times L_z\mathbb{T}$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is the one-dimensional torus, and (ii) U is C^∞ on Ω. These assumptions, which are realistic in most molecular dynamics applications, allow us to use recent results in hypocoercivity [20, 21] to establish the regularity properties of $\mathcal{L}_{\mathrm{LD}}$. Specifically, we have as a consequence of these articles, that

- $\mathcal{L}_{\mathrm{LD}}^*$ has a compact resolvent [22],
- $\mathcal{L}_{\mathrm{LD}}^* f = 0$ has a unique solution in $C^\infty(\Omega)$, i.e., the Gibbs density ρ_β.

Note that, if, for some given g, there are two solutions of $\mathcal{L}_{\mathrm{LD}}^* f = g$ then, using the linearity of the operator these may differ only in a constant scaling of the Gibbs density.

2.1 Timestepping Methods.

In a splitting method for a deterministic system $\dot{z} = f(z)$, one divides the vector field f into exactly solvable parts, i.e. $f = f_1 + f_2$, which are treated sequentially within a timestep. An example of such a splitting method for the Hamiltonian system with energy $H(x,p) = \sum_i p_i^2/(2m_i) + U(x_1, x_2, \ldots, x_N)$ is the "symplectic Euler" method defined by $f_1 = \sum_i m_i^{-1} p_i \partial_{x_i}$, $f_2 = -\sum_i \frac{\partial U}{\partial x_i}\partial_{p_i}$. By dividing the vector field as $f = \frac{1}{2}f_1 + f_2 + \frac{1}{2}f_1$, solving each vector field in turn, we obtain the so-called position Verlet method, and by switching the roles of f_1 and f_2 we obtain the velocity Verlet method. Splitting methods like these are explicit and this feature is of particular importance in molecular dynamics, where the force calculation is the usual measure of per-timestep computational complexity.

In a similar way, Langevin dynamics may be treated by splitting [9, 12, 13]. For example, one may divide the Langevin system (Eq. 4) into three parts

$$\mathrm{d}\begin{bmatrix} x \\ p \end{bmatrix} = \underbrace{\begin{bmatrix} M^{-1}p \\ 0 \end{bmatrix}\mathrm{d}t}_{\mathrm{A}} + \underbrace{\begin{bmatrix} 0 \\ -\nabla U \end{bmatrix}\mathrm{d}t}_{\mathrm{B}} + \underbrace{\begin{bmatrix} 0 \\ -\gamma p \mathrm{d}t + \sigma M^{1/2}\mathrm{d}W \end{bmatrix}}_{\mathrm{O}}, \tag{5}$$

and each of the three parts may be solved 'exactly'. In the case of the OU part (labelled here simply as O) we mean by this that we realize the stochastic process by the equivalent formula

$$p(t) = e^{-\gamma t}p(0) + \sigma\sqrt{(1-e^{-2\gamma t})\beta^{-1}}M^{1/2}R(t), \tag{6}$$

where $R(t)$ is a vector of uncorrelated independent standard normal random processes (white noise). One method based on the splitting (Eq. 5) is defined by the composition

$$\psi^{\delta t}_{\mathrm{ABAO}} = \exp((\delta t/2)\mathcal{L}_{\mathrm{A}})\exp(\delta t\mathcal{L}_{\mathrm{B}})\exp((\delta t/2)\mathcal{L}_{\mathrm{A}})\exp(\delta t\mathcal{L}_{\mathrm{O}}),$$

where $\exp(\delta t\mathcal{L}_f)$ represents the phase space propagator associated to the (deterministic or stochastic) vector field f, and we use δt as the timestep in Langevin dynamics (later we use h for the timestep of an associated Brownian dynamics). The deterministic part is approximated by the position Verlet method. This is referred to as a Geometric Langevin Algorithm of order two (GLA-2) following [13]; an alternative is to use velocity Verlet for the Hamiltonian part ($\psi^{\delta t}_{\mathrm{BABO}}$). A simple generalization of GLA methods is obtained by interspersing integrators associated to parts of the Hamiltonian vector field with exact OU solves, thus we have

$$\psi^{\delta t}_{\mathrm{ABOBA}} = \exp((\delta t/2)\mathcal{L}_{\mathrm{A}})\exp((\delta t/2)\mathcal{L}_{\mathrm{B}})\exp(\delta t\mathcal{L}_{\mathrm{O}})\exp((\delta t/2)\mathcal{L}_{\mathrm{B}})\exp((\delta t/2)\mathcal{L}_{\mathrm{A}}),$$

and ψ_{BAOAB} defined in an analogous way.

An alternative integrator termed the Stochastic Position Verlet (SPV) method [9, 12], relies on the splitting

$$\mathrm{d}\begin{bmatrix} x \\ p \end{bmatrix} = \begin{bmatrix} M^{-1}p \\ 0 \end{bmatrix}\mathrm{d}t + \begin{bmatrix} 0 \\ -\nabla U\mathrm{d}t - \gamma p\mathrm{d}t + \sigma M^{1/2}\mathrm{d}w \end{bmatrix}.$$

SPV is not a (generalized) GLA-type method, although it, as each of the generalized GLA schemes, is *quasisymplectic* in the language of [10]. Likewise the commonly used method of Brunger, Brooks and Karplus (BBK) [1] is not of the (generalized) GLA family. Details of all methods examined are given in the Appendix.

To slightly simplify the presentation that follows, we make the change of variables $q \to M^{-1/2}q$, $p \to M^{+1/2}p$, with a corresponding adjustment of the potential; this is equivalent to assuming $M = I$.

3 Expansion of the Invariant Measure

We shall work here with formal series expansions, however we expect that our derivation could be rigorously founded using techniques found in [23] and [24]. Associated to any given

splitting-based method for Langevin dynamics, we define the operator $\hat{\mathcal{L}}^*$ that characterises the propagation of density by an expansion of the form:

$$\hat{\mathcal{L}}^* = \mathcal{L}^*_{\mathrm{LD}} + \delta t \mathcal{L}^*_1 + \delta t^2 \mathcal{L}^*_2 + O(\delta t^3), \tag{7}$$

For example, for the method labelled by the string ABAO, we have

$$\exp(\delta t \hat{\mathcal{L}}^*_{\mathrm{ABAO}}) = \exp((\delta t/2)\mathcal{L}^*_{\mathrm{A}}) \exp(\delta t \mathcal{L}^*_{\mathrm{B}}) \exp((\delta t/2)\mathcal{L}^*_{\mathrm{A}}) \exp(\delta t \mathcal{L}^*_{\mathrm{O}}).$$

The perturbation series may be found by successive applications of the BCH expansion [16] and linearity properties of the Kolmogorov operator. However, unlike in the deterministic case, the terms that appear in the series cannot be associated to modified vector fields or even SDEs [25].

Note that, when iterated $n+1$ times, the method ABAO produces a sequence of the form

$$e^{\frac{\delta t}{2}\mathcal{L}_{\mathrm{A}}} e^{\frac{\delta t}{2}\mathcal{L}_{\mathrm{B}}} [\psi^{\delta t}_{\mathrm{BAOAB}}]^n e^{\frac{\delta t}{2}\mathcal{L}_{\mathrm{B}}} e^{\frac{\delta t}{2}\mathcal{L}_{\mathrm{A}}} e^{\delta t \mathcal{L}_{\mathrm{O}}}$$

thus, with a minor coordinate transformation, the dynamics sample the same invariant density as BAOAB. Similarly BABO and OBAB are essentially the same method as ABOBA. For this reason we concentrate in the remainder of this article on ABOBA and BAOAB. For these two methods, the symmetry implies that the odd order terms in (7) vanish identically using the Jacobi identity in the BCH expansions.

After deriving $\hat{\mathcal{L}}^*$ in this way, we seek the invariant distribution which satisfies $\hat{\mathcal{L}}^* \hat{\rho} = 0$. For the BAOAB and ABOBA methods, we make the ansatz that the invariant measure of the numerical method has the simple form

$$\hat{\rho} \propto \exp(-\beta[H + \delta t^2 f_2 + \delta t^4 f_4 + \ldots]). \tag{8}$$

We may rewrite this as

$$\hat{\rho} \propto \rho_\beta \left(1 - \beta \delta t^2 f_2(x,p) + O\left(\delta t^4\right)\right).$$

This means that the equation $\hat{\mathcal{L}}^* \hat{\rho} = 0$ becomes

$$\left(\mathcal{L}^*_{\mathrm{LD}} + \delta t^2 \mathcal{L}^*_2 + \ldots\right)\left(\rho_\beta - \delta t^2 \beta \rho_\beta f_2 + \ldots\right) = 0.$$

Equating second order terms in δt gives

$$\mathcal{L}^*_{\mathrm{LD}}\left(\rho_\beta f_2\right) = \beta^{-1} \mathcal{L}^*_2 \rho_\beta. \tag{9}$$

The equation $\mathcal{L}^*_{\mathrm{LD}} g = 0$ has unique solution $g = \rho_\beta$, up to a constant multiple. Hence the homogeneous solution to the above PDE is $f_2(x,p) = c$, for some constant c; we therefore

require a particular solution f_2 of (9). By the Fredholm Alternative, the equation has a solution provided the solution of

$$\mathcal{L}_{\mathrm{LD}} g = 0$$

satisfies

$$\int g \mathcal{L}_2^* \rho_\beta = 0.$$

As the only solutions of $\mathcal{L}_{\mathrm{LD}} g = 0$ are the constants, we require

$$\int \mathcal{L}_2^* \rho_\beta = 0.$$

3.1 Calculation of the inhomogeneity

For a symmetric splitting method such as the BAOAB method, recall that we can use the Baker-Campbell-Hausdorff [16] formula to find the overall one-step perturbation operator for the scheme. For linear operators X, Y and Z, we have that

$$e^{\frac{\delta t}{2}X} e^{\frac{\delta t}{2}Y} e^{\delta t Z} e^{\frac{\delta t}{2}Y} e^{\frac{\delta t}{2}X} = e^{\delta t S},$$

such that

$$\begin{aligned} S =& X + Y + Z + \frac{\delta t^2}{12} \left([Z,[Z,Y+X]] + [Y,[Y,X]] + [Z,[Y,X]] + [Y,[Z,X]] \right. \\ & \left. -\tfrac{1}{2}[Y,[Y,Z]] - \tfrac{1}{2}[X,[X,Z]] - \tfrac{1}{2}[X,[X,Y]] \right) + \mathcal{O}(\delta t^4). \end{aligned}$$

where $\delta t > 0$ and $[X,Y] = XY - YX$ is the commutator of Y and X.

In the case of the BAOAB method, we use

$$X = \mathcal{L}_{\mathrm{B}}^* = \nabla U(x) \cdot \nabla_p, \;\; Y = \mathcal{L}_{\mathrm{A}}^* = -p \cdot \nabla_x,$$

and

$$Z = \mathcal{L}_{\mathrm{O}}^* = \gamma \nabla_p \cdot (p \cdot) + \frac{\sigma^2}{2} \Delta_p,$$

to compute the perturbed operator for the method. A similar analysis can be conducted for the other generalised GLA-type methods considered in this paper.

To compute $\hat{\mathcal{L}}^*$ we simply plug in our choices into the BCH formula to obtain:

$$\begin{aligned} \hat{\mathcal{L}}^* &= \mathcal{L}_{\mathrm{LD}}^* + \delta t^2 \mathcal{L}_2^* + O\left(\delta t^4\right), \\ &= \mathcal{L}_{\mathrm{LD}}^* + \frac{\delta t^2}{12} \left([\mathcal{L}_{\mathrm{O}}^*, [\mathcal{L}_{\mathrm{O}}^*, \mathcal{L}_{\mathrm{Det}}^*]] + [\mathcal{L}_{\mathrm{A}}^*, [\mathcal{L}_{\mathrm{A}}^*, \mathcal{L}_{\mathrm{B}}^*]] \right. \\ &\quad + [\mathcal{L}_{\mathrm{O}}^*, [\mathcal{L}_{\mathrm{A}}^*, \mathcal{L}_{\mathrm{B}}^*]] + [\mathcal{L}_{\mathrm{A}}^*, [\mathcal{L}_{\mathrm{O}}^*, \mathcal{L}_{\mathrm{B}}^*]] - \tfrac{1}{2}[\mathcal{L}_{\mathrm{A}}^*, [\mathcal{L}_{\mathrm{A}}^*, \mathcal{L}_{\mathrm{O}}^*]] \\ &\quad \left. -\tfrac{1}{2}[\mathcal{L}_{\mathrm{B}}^*, [\mathcal{L}_{\mathrm{B}}^*, \mathcal{L}_{\mathrm{O}}^*]] - \tfrac{1}{2}[\mathcal{L}_{\mathrm{B}}^*, [\mathcal{L}_{\mathrm{B}}^*, \mathcal{L}_{\mathrm{A}}^*]] \right) + O(\delta t^4), \end{aligned}$$

where recall that

$$\mathcal{L}^*_{\mathrm{Det}} = \mathcal{L}^*_{\mathrm{A}} + \mathcal{L}^*_{\mathrm{B}},$$

and

$$\mathcal{L}^*_{\mathrm{LD}} = \mathcal{L}^*_{\mathrm{Det}} + \mathcal{L}^*_{\mathrm{O}}.$$

The calculation of the inhomogeneity in (8) then amounts to a straightforward computation of the commutator series applied to ρ_β. The commutators needed are:

$$\begin{aligned}
[\mathcal{L}^*_{\mathrm{A}}, [\mathcal{L}^*_{\mathrm{A}}, \mathcal{L}^*_{\mathrm{B}}]]\, \rho_\beta =& 2\beta p^T U''(x)\nabla U(x)\rho_\beta - \rho_\beta \beta\, p\cdot\nabla_x p^T U''(x)p,\\
[\mathcal{L}^*_{\mathrm{B}}, [\mathcal{L}^*_{\mathrm{B}}, \mathcal{L}^*_{\mathrm{A}}]]\, \rho_\beta =& -2\beta p^T U''(x)\nabla U(x)\rho_\beta,\\
[\mathcal{L}^*_{\mathrm{O}}, [\mathcal{L}^*_{\mathrm{A}}, \mathcal{L}^*_{\mathrm{B}}]]\, \rho_\beta =& 2\gamma\left(\Delta_x U(x) - \beta p^T U''(x)p\right)\rho_\beta,\\
[\mathcal{L}^*_{\mathrm{A}}, [\mathcal{L}^*_{\mathrm{O}}, \mathcal{L}^*_{\mathrm{B}}]]\, \rho_\beta =& \gamma\beta\left(|\nabla U(x)|^2 - p^T U''(x)p\right)\rho_\beta,\\
[\mathcal{L}^*_{\mathrm{A}}, [\mathcal{L}^*_{\mathrm{A}}, \mathcal{L}^*_{\mathrm{O}}]]\, \rho_\beta =& 2\gamma\left(\beta|\nabla U(x)|^2 - \Delta_x U(x)\right)\rho_\beta,\\
[\mathcal{L}^*_{\mathrm{B}}, [\mathcal{L}^*_{\mathrm{B}}, \mathcal{L}^*_{\mathrm{O}}]]\, \rho_\beta =& 0\\
[\mathcal{L}^*_{\mathrm{O}}, [\mathcal{L}^*_{\mathrm{O}}, \mathcal{L}^*_{\mathrm{Det}}]]\, \rho_\beta =& 0,
\end{aligned}$$

where we have abbreviated the Hessian $\nabla_x \nabla_x^T U(x) =: U''(x)$.

Hence we can see directly that

$$12\mathcal{L}^*_2\rho_\beta = 3\gamma\left(\Delta_x U(x) - \beta p^T U''(x)p\right)\rho_\beta + 3\beta p^T U''(x)\nabla U(x)\rho_\beta - \rho_\beta\, \beta p\cdot\nabla_x p^T U''(x)p,$$

giving

$$\mathcal{L}^*_2\rho_\beta \;=\; \rho_\beta\left[\frac{\gamma}{4}\left(\Delta_x U(x) - \beta p^T U''(x)p\right) + \frac{\beta}{4}p^T U''(x)\nabla U(x) - \frac{\beta}{12}\,p\cdot\nabla_x p^T U''(x)p\right]. \quad (10)$$

Observe that the condition of the Fredholm Alternative is satisfied since the average of the first term is equivalent to a canonical average which vanishes

$$\mathbb{E}\left(\Delta_x U(x) - \beta p^T U''(x)p\right) = 0.$$

The other terms in Eq. 10, being canonical averages of terms which are odd-order in p, necessarily also average to zero.

An analogous computation can be performed for the ABOBA method, giving a slightly (but crucially) different perturbation operator $\mathcal{L}^{*\,(\mathrm{ABOBA})}$, where

$$\mathcal{L}^{*\,(\mathrm{ABOBA})}_2\rho_\beta = -\rho_\beta\left[\frac{1}{4}\gamma\left(\Delta_x U(x) - \beta p^T U''(x)p\right) + \frac{1}{4}\beta p^T U''(x)\nabla U(x) - \frac{1}{6}\beta\, p\cdot\nabla_x p^T U''(x)p\right]$$

Here U'' is the Hessian matrix of U and $\Delta_x = \sum_{i=1}^N \frac{\partial^2}{\partial x_i^2}$ is the partial Laplacian in x. Again the Fredholm Alternative condition is easily seen to be satisfied.

3.2 Expansion in powers of γ^{-1}

Although we are not able to give the general analytical solution to the partial differential equation of Eq. 9, we can find a solution in an important limiting case: the high friction regime. To do this, we expand the invariant density of the numerical method further, viewing both δt and $\varepsilon = \gamma^{-1}$ as small parameters:

$$\hat{\rho} = \exp(-\beta[H + \delta t^2(f_{2,0} + f_{2,1}\varepsilon + O(\varepsilon^2)) + O(\delta t^4)]). \tag{11}$$

Dividing by γ, we may reduce Eq. 9 to

$$[\mathcal{L}_0^* + \varepsilon\mathcal{L}_1^*](f_{2,0} + \varepsilon f_{2,1} + O(\varepsilon^2)) = g_0 + \varepsilon g_1,$$

where

$$\mathcal{L}_0^* = \frac{1}{\beta}\Delta_p - p\cdot\nabla_p, \quad \mathcal{L}_1^* = \nabla U(q)\cdot\nabla_p - p\cdot\nabla_x,$$

and, for BAOAB,

$$\begin{aligned} g_0 &= \tfrac{1}{4}\left(\beta^{-1}\Delta_x U(x) - p^T U''(x)p\right), \\ g_1 &= \tfrac{1}{4}p^T U''(x)\nabla U(x) - \tfrac{1}{12}p\cdot\nabla_x p^T U''(x)p. \end{aligned}$$

Note that this is a singularly perturbed system as $\mathcal{L}_0$ is degenerate and it is only the combined operator that has the necessary regularity to define a unique solution. Nonetheless, as explained below it is possible to find the leading term $f_{2,0}$ by substituting a truncated expansion of fixed degree and solving the resulting equations.

Equating powers of ε, we find

$$\mathcal{L}_0^* f_{2,0} = g_0, \quad \mathcal{L}_1^* f_{2,0} + \mathcal{L}_0^* f_{2,1} = g_1,$$

and

$$\mathcal{L}_1^* f_{2,n-1} = -\mathcal{L}_0^* f_{2,n} \ \text{(for } n > 1).$$

Truncating at $n = 2$, for example, we find the following solution of these equations:

$$\begin{aligned} f_{2,0} \equiv f_{2,0}^{\mathrm{BAOAB}} &= \tfrac{1}{8}\left(p^T U''(x)p - \beta^{-1}\Delta U(x)\right), \\ f_{2,1} \equiv f_{2,1}^{\mathrm{BAOAB}} &= \tfrac{1}{24}\beta^{-1}p^T\nabla_x\Delta_x U(x) - \tfrac{1}{72}p^T\nabla_x p^T U''(x)p, \\ f_{2,2} \equiv f_{2,2}^{\mathrm{BAOAB}} &= \tfrac{1}{296}p^T\nabla_x p^T\nabla_x p^T U''(x)p - \tfrac{1}{48}\nabla U(x)\cdot\nabla_x p^T U''(x)p. \end{aligned}$$

For ABOBA, $f_{2,0}$ the solution would change to

$$f_{2,0}^{\mathrm{ABOBA}} = -\tfrac{1}{8}\left(p^T U''(x)p - 2\beta^{-1}\Delta U(x)\right).$$

3.3 Marginal distribution

We now turn out attention to the configurational marginal distribution obtained by integrating the density expansion (11) with respect to the momenta. Our interest is only in the leading term of this expansion, which defines the sampling behavior for large γ. Ignoring higher order terms in δt and ε, we would have the distribution

$$\hat{\rho} = \tilde{\rho} \times (1 + \mathcal{R}),$$

where, for BAOAB, $\tilde{\rho} = \rho_\beta \times \exp\left(-\beta\left[\frac{\delta t^2}{8}\left(p^T U''(x)p - \beta^{-1}\Delta U(x)\right)\right]\right)$, and $\mathcal{R} = O(\varepsilon\delta t^2) + O(\delta t^4)$.

Noting that the leading term in the exponent of $\tilde{\rho}$ is quadratic in momenta, we integrate out with respect to p to obtain

$$\begin{aligned}\int \tilde{\rho}\,\mathrm{d}^N p &= \int \exp\left(-\beta\left[\frac{1}{2}p^T(I + \frac{\delta t^2}{4}U'')p + U - \beta^{-1}\frac{\delta t^2}{8}\Delta U\right]\right)\mathrm{d}^N p \\ &= \sqrt{2\pi k_B T/\det\left(I + \frac{\delta t^2}{4}U''\right)}\exp\left(-\beta\left[U - \beta^{-1}\frac{\delta t^2}{8}\Delta U\right]\right)\end{aligned}$$

Using the identity $\det(M) = \exp(\mathrm{trace}(\log(M)))$, we find

$$\int \tilde{\rho}\,\mathrm{d}^N p \propto \exp\left(-\frac{1}{2}\mathrm{trace}\left(\log\left(I + \frac{\delta t^2}{4}U''\right)\right)\right) \times \exp\left(-\beta\left[U - \beta^{-1}\frac{\delta t^2}{8}\Delta U\right]\right).$$

We then Taylor expand the logarithm of $I + \frac{\delta t^2}{4}U''$ and take the trace to obtain a cancellation of the δt^2 terms, giving

$$\int \tilde{\rho}\,\mathrm{d}^N p \propto \exp(-\beta U + O(\delta t^4)).$$

The contribution to the configurational distribution error due to $\tilde{\rho}$ is $O(\delta t^4)$. This means that the overall error in the marginal distribution of $\hat{\rho}$ (which includes the neglected factor $1 + \mathcal{R}$) will be $O(\varepsilon\delta t^2) + O(\delta t^4)$.

If ε is small (or δt is relatively large), the error will be dominated by the quartic term in δt and we will observe 4th order accuracy in configurational averages. For small δt the method is always eventually second order.

In the case of the ABOBA method, the remarkable cancellation of the second order errors does not occur and the method always exhibits 2nd order configuration distribution error.

4 The Limit Method

We now consider the limit $\gamma \to \infty$, where the exact solution of the vector OU process reduces to $p = \sqrt{k_B T}M^{1/2}R$, where R is a vector whose components have a standard normal distribution

(Gaussian white noise). Alternatively, we could consider the limit of the particle mass going to 0, although this requires a reformulation of Eq. 4 so that the friction is proportional to the velocity instead of the momentum [6]. Whichever limit is taken, we would expect the ultimate result to be the same. (Here we have reintroduced the masses in order to present the method, since they may be useful scaling parameters in simulation.) In the configurations it is straightforward to show that the BAOAB method therefore becomes

$$\begin{aligned} x_{n+1} &= x_n + \frac{\delta t^2}{2} M^{-1} F(x_n) + \frac{\delta t}{2} M^{-1}(p_n + p_{n+1}) \\ &= x_n + \frac{\delta t^2}{2} M^{-1} F(x_n) + \frac{\delta t}{2} \sqrt{k_B T} M^{-1/2}(R_n + R_{n+1}), \end{aligned}$$

where the R_n are vectors of i.i.d. standard normal random variables. Replacing $\delta t^2/2$ by h we arrive at Eq. 3. Since the Langevin scheme gives 4th order accurate configurational averages in this limit, we expect the method of Eq. 3 to be second order accurate in modified timestep h. Moreover, since we completely remove the second order term in the Langevin dynamics configurational density expansion, we expect to observe this behaviour across all values of h.

By contrast the ABOBA scheme gives a much more complicated limit method as $\gamma \to \infty$ which is not in one-step form.

In the Euler-Maruyama method, the random perturbations introduced at each step are independent. In the method of Eq. 2, the random perturbation is a scaling of $Z_n = (R_n + R_{n+1})/\sqrt{2}$; the components $Z_n^{(i)}$ of these are independent of each other and decay linearly with timestep:

$$\begin{aligned} \langle Z_n^{(i)}, Z_n^{(i)} \rangle &= \frac{1}{2}(\langle R_{n+1}^{(i)}, R_{n+1}^{(i)} \rangle + \langle R_n^{(i)}, R_n^{(i)} \rangle) = 1 \\ \langle Z_n^{(i)}, Z_{n-1}^{(i)} \rangle &= \langle R_n^{(i)}, R_n^{(i)} \rangle = 1/2 \\ \langle Z_n^{(i)}, Z_{n-k}^{(i)} \rangle &= 0, \quad k = 2, 3, \ldots \end{aligned}$$

Thus, in this new method, we use a colored noise which has characteristics that directly depend on the stepsize, although the noise decorrelates in just a couple of timesteps. This is therefore no longer a Markov process, however it can be reformulated as such if one considers the appropriate extended space (eg. $y_n = [x_n, R_n, R_{n+1}]$).

5 Numerical Experiments

We implemented the methods ABOBA, BAOAB, SPV and BBK and compared the accuracy of configurational sampling for different values of γ and a range of timesteps. A brief analysis shows that the use of the harmonic oscillator leads to special cancellations in the BCH series of the splitting schemes, making it a poor test subject. Hence, in order to compare the order of

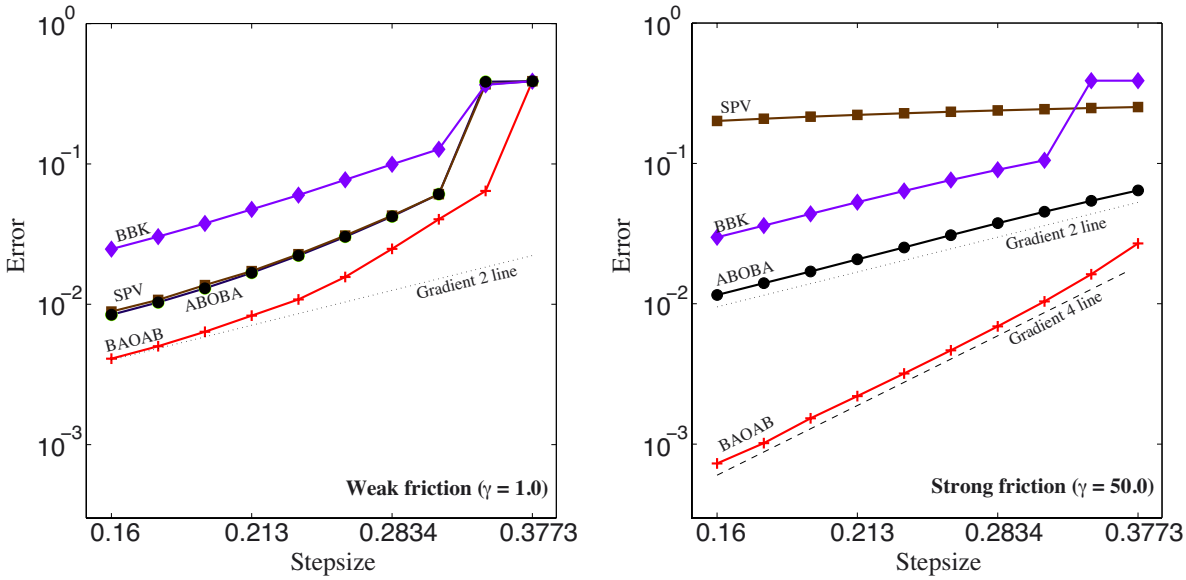


Figure 1: The graphs show the comparison of four different Langevin dynamics methods when applied using different stepsizes. The configurational distribution errors are plotted against the stepsize in a log-log scale. Here $k_BT = 1$. The simulation time was fixed for all runs at $t = 6.4 \times 10^7$, and five runs were averaged to further reduce sampling errors. At left, $\gamma = 1$, at right $\gamma = 50$. The graphs are entirely in keeping with the theory presented in the article.

accuracy of the different schemes, we first considered an oscillator model in 1D with potential $U(x) = x^4/4 + \sin(1 + 5x)$. This was accomplished by introducing M intervals ('bins') of equal length, and comparing the observed probability frequency with the exact frequency (obtained by integration of the probability density), using the Euclidean norm. If the observed frequency in bin i is $\hat{\alpha}_i$, and the exact expected frequency is α_i, then the error calculated is

$$\text{Error} = \sqrt{\sum_{\mathrm{i}=1}^{\mathrm{M}} (\alpha_{\mathrm{i}} - \hat{\alpha}_{\mathrm{i}})^2}. \tag{12}$$

In this one-dimensional example, we used 20 bins to cover the interval from -3.5 to 3.5. The configurational density error is plotted against stepsize in log-log scale. If Error $\propto \delta t^r$ then we expect this graph to be a line of slope r. Due to the relative simplicity of this model, we were able to perform highly resolved simulations to calculate accurate error estimates for the configurational distribution. The exact expected value that we compare experimental results against can be computed to arbitrary prescision, and we are able to run as many simulations as needed in order to drive the variance of results to a minimum. The variance in our results where the stepsize was less than 0.3 was consistently below 10^{-8}. Above this threshold some methods were found to be unstable. The results of our simulations are summarized in Figure 1.

As we can see, when γ is small, the methods perform somewhat similarly, at least in the qualitative sense, with all showing a 2nd order error in configurational sampling, and the ABOBA and SPV methods essentially identical. As γ is increased, the substantial difference between BAOAB and the other methods becomes apparent. In the limit of large γ the SPV method effectively annihilates the force which results in poor sampling. In the graph for $\gamma = 1$, we can see that for larger values of the stepsize, the graph steepens (indicating that the fourth order term is dominant); as the stepsize is decreased the method exhibits a 2nd order asymptotic decay. With $\gamma = 50$ the fourth order behavior is seen for for all indicated data points, although, again, this becomes second order for smaller values of δt. Note that the limit method Eq. 2 (with the substitution $h = \delta t^2/2$) gives an essentially identical behavior to the $\gamma = 50$ case. We also give a comparison of the actual computed configurational distributions, at different stepsizes using each method, for $\gamma = 20$ in Figure 2. .

To examine the performance of the limit method in more detail, we next considered small molecular clusters consisting of seven atoms (motion restricted to the plane), with both Morse (φ_M) and Lennard-Jones (φ_{LJ}) potentials, given by

$$\begin{aligned} \varphi_{\mathrm{M}}(r) &= (1 - \exp(-a(r - r_m)))^2, \\ \varphi_{\mathrm{LJ}}(r) &= \varepsilon \left(\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^{6} \right), \end{aligned}$$

where $a = 2, \varepsilon = 1$ and $r_m = 1$. The Morse potential gives a smoother dynamics compared to Lennard-Jones (the Morse forces were on average three times smaller than those for Lennard-Jones) and allows a more satisfying determination of the error scaling behavior with stepsize. To quantify the error in configurational sampling, we calculated the radial density $G(r)$ by binning the instantaneous interatomic distances at each step into 20 compartments and compared to a calculated reference value (Figure 3). Though not exact, the errors in the reference will be negligible compared with configurational distribution errors at higher stepsizes. For the Morse cluster the reference stepsize we used was $h_{\mathrm{ref}} = 0.001$, whereas for the Lennard-Jones cluster, we used $h_{\mathrm{ref}} = 0.00025$, both with the same integration time that was used in their respective test runs. In both cases the cluster was initialized with 6 particles placed on a regular hexagon and with the remaining particle in the centre. Initial velocities were randomly drawn from the canonical distribution.

Our results, presented in Figure 4, are entirely consistent with our analysis and show the second order dependence of the configurational sampling error on h (equivalent to fourth order in δt), as compared to the Euler-Maruyama's first order behavior. The results for the Morse cluster demonstrate a good agreement with our theoretical results. However the Lennard-Jones experiments, on the right of Figure 4, are not as well converged, owing to the difficulty in

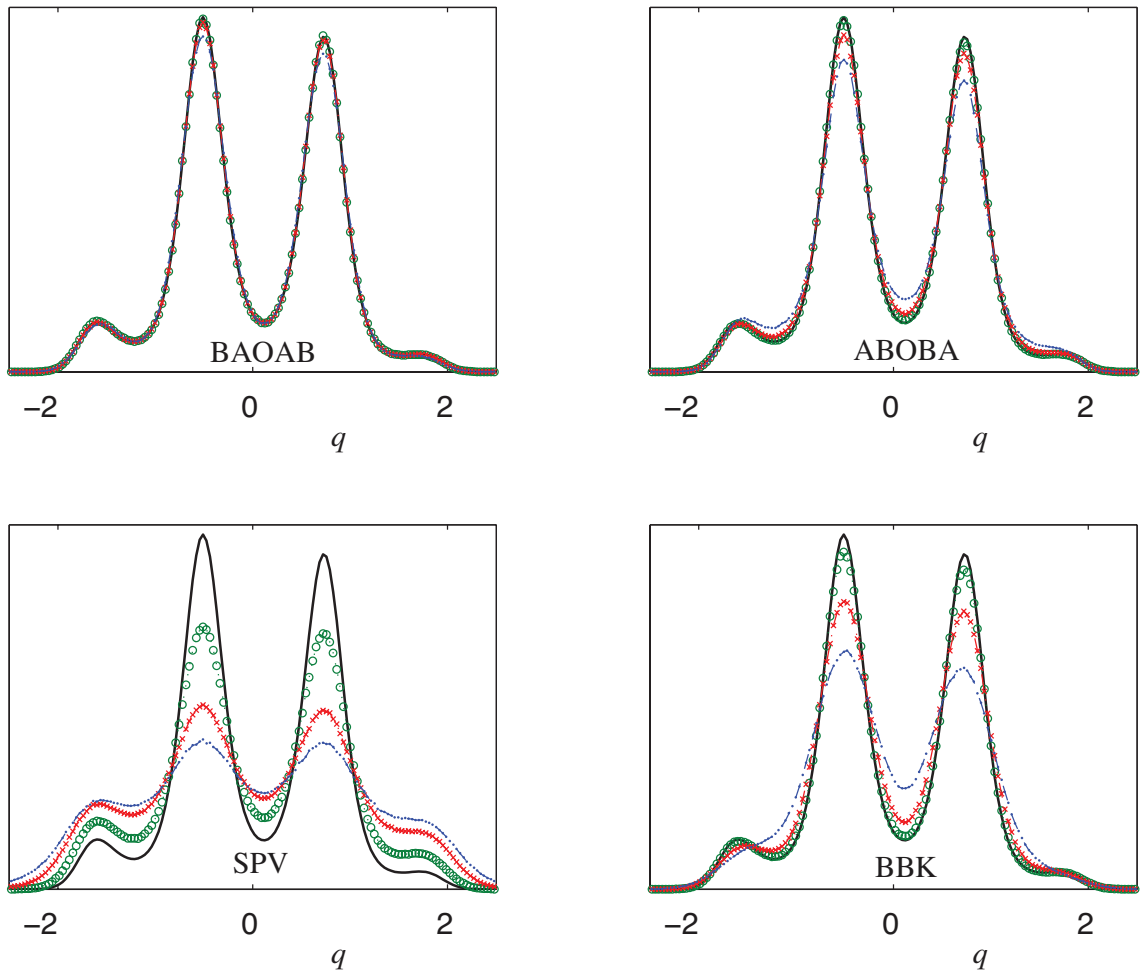


Figure 2: Computed distributions of the 1D model problem are compared for $\gamma = 20$. The three curves for each of the four methods show the results for three different stepsizes: $\delta t = 0.1$ (circles), $\delta t = 0.2$ (crosses), $\delta t = 0.3$ (dashed), compared to the dark, solid curve representing the exact distribution. The graph shows that the generalized GLA methods are superior to SPV and BBK in the moderate γ regime.

sampling this system accurately. (We are currently applying the method to more complicated systems with steep potentials, and with extensive use of computing power.) Nonetheless, in both cases, we observe a significant reduction in the error compared to Euler-Maruyama.

It might be a suggested that the improved accuracy seen in the high γ regime could potentially come at the price of a slower convergence to equilibrium due to a reduced rate of transition between metastable states, hence overall sampling of the configurational distribution might be impaired in favour of local sampling. To address this point, we have plotted in Figure 5 the error computed in our Lennard-Jones simulation as a function of the number of force evaluations (vertical) against the friction value γ used for the simulation (horizontal). Gridpoints in the plot are coloured according to the configurational error, computed using Eq. 12. The results indicate that the convergence rate for the BAOAB method is not diminished

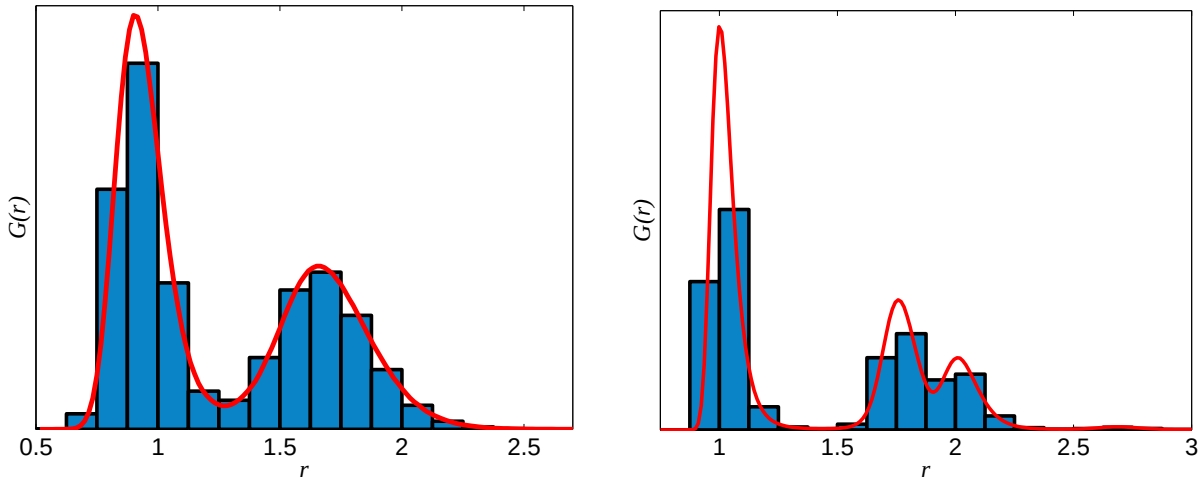


Figure 3: The diagrams illustrate the distributions of interatomic distances, $G(r)$, for Morse (left) and Lennard-Jones (right) clusters. They also show the choice of bins used in calculating the numerical distributions.

for large friction coefficient, so it does not appear that we are sacrificing sampling accuracy using this scheme. Note that the performance of ABOBA is also robust in the limit of large γ, but the achievable accuracy is reduced as it is only second order, consistent with what we have presented in this article.

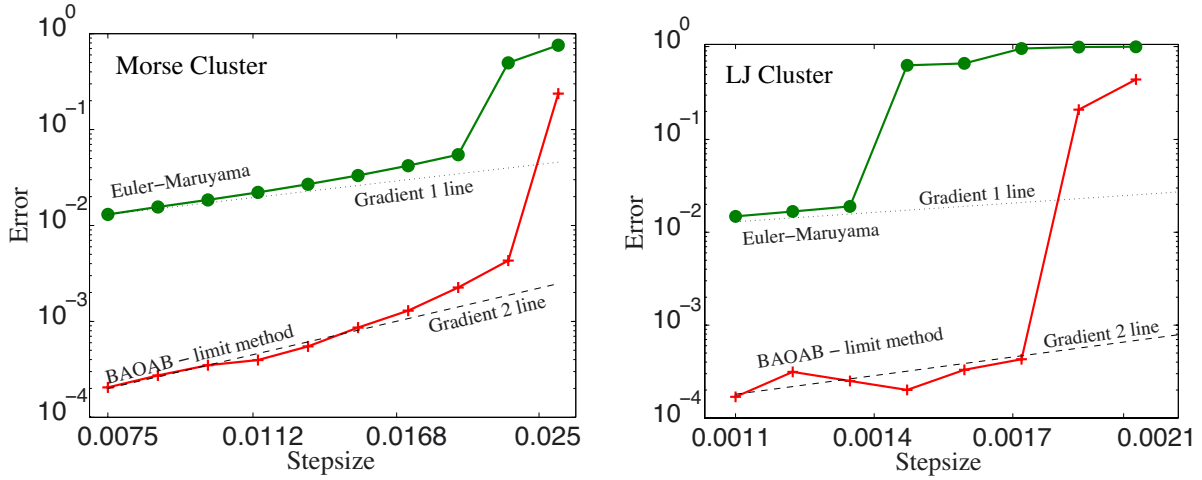
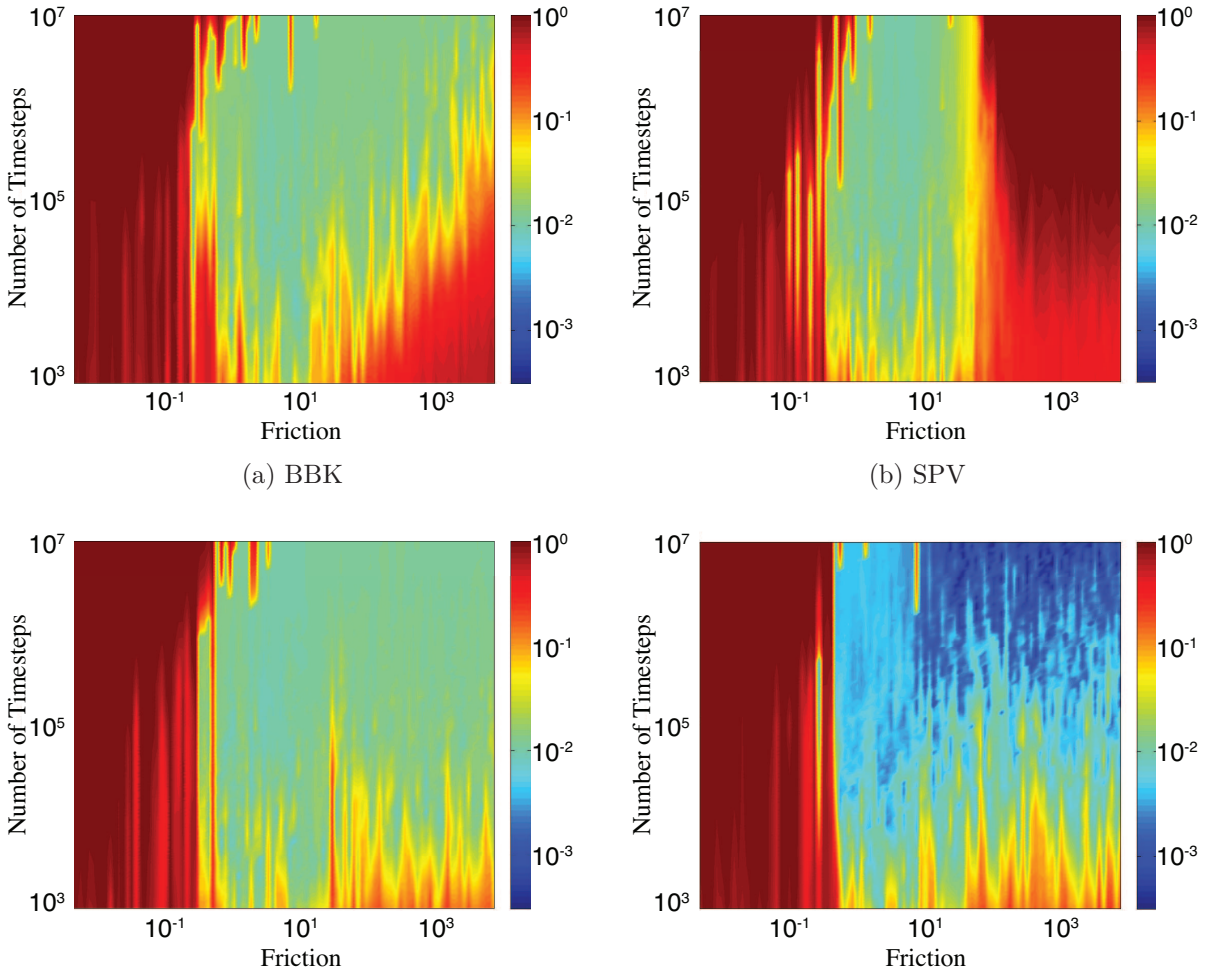


Figure 4: The radial distribution errors are plotted in log-log scale against stepsize, demonstrating the first order decay of the error in the case of Euler-Maruyama and the second-order behavior of the BAOAB limit method (in modified timestep h). Left: Morse potential; Right: Lennard-Jones. For Morse we used a temperature of $k_BT = 0.1$, a fixed time interval of $t = 10^5$, with stepsizes ranging from 0.0075 to 0.0264; each simulation was repeated 5 times. For Lennard-Jones the temperature was $k_BT = 0.15$, $t = 2.6 \times 10^5$ and stepsizes ranged from .0011 to 0.0021. Around two orders of magnitude of improvement are observed in the accurate regime, but, perhaps even more important, the BAOAB limit method is usable at substantially larger stepsizes than Euler-Maruyama in the Lennard-Jones case.



(a) BBK

(b) SPV

(c) ABOBA

(d) BAOAB

Figure 5: The diagrams illustrate the performance of the different algorithms (labelled) by showing the error in the computed radial distribution functions as a function of both γ and the number of timesteps (samples) taken, in the case of the 7 atom Lennard-Jones model. The graphs address the potential concern that the larger values of γ needed to give the superconvergence property may reduce the rate of convergence to equilibrium (it does not, in the case of BAOAB). The same stepsize of $\delta t = 0.044$ was used for all these simulations.

6 Conclusions

Our results confirm the theoretical results of Sections 3 and 4, in particular showing the higher order configurational sampling of the BAOAB method. The order in its Langevin formulation is effectively four in the large γ limit and large values of γ do not impair its stability due to the use of exact Ornstein-Uhlenbeck solves. Not only is the order of the method high, the constant multiplying the leading term must be, in the cases looked at here, of modest magnitude, since the errors are relatively small also at the large stepsize stability threshold. Since, in the context of molecular dynamics, BAOAB is a 'cheap' and easy to implement scheme using only a single force vector per timestep, we stress that there is no price to pay for its improved accuracy.

In molecular modelling, there are other errors that play important roles, most importantly errors in the force fields (or, more fundamentally, the errors due to not modelling quantum mechanics properly) and sampling errors. Obviously these errors may dominate the overall method error and limit the relative benefit to be gained by using one integrator as compared to another, but it is also clear that both of the other types of errors are constantly being reduced through the design of better models and the use of more powerful computers. More important, one can ask the question: how can a practitioner know which part of the error in a given complicated simulation is due to sampling error and which part due to the truncation errors addressed here? In our experiments with molecular models, even where there was substantial sampling error still present, we nonetheless found the accuracy to be noticeably higher for the BAOAB method; it is likely that this improvement in sampling accuracy would be of direct benefit in many real world simulations. Finally we point out that the BAOAB scheme and its limit method (Eq. 2) was, in each case studied, stable at a larger stepsize than the alternatives, meaning that longer time intervals are made accessible. This was particular dramatic in the case of the Lennard-Jones system.

Acknowledgements. The first author acknowledges the support of a JTO Fellowship from the Institute for Computational Engineering and Sciences at the University of Texas. The second author was supported by the Centre for Numerical Algorithms and Intelligent Software (funded by EPSRC grant EP/G036136/1 and the Scottish Funding Council). This work was first presented at the EPSRC Symposium Workshop on Multiscale Systems: Theory and Applications held at Warwick University in December 2011; several of those present made helpful suggestions regarding the presentation of the results. We further acknowledge the comments received from A. Abdulle, G. Stoltz, and M. Tretyakov which have improved the article.

References

[1] Brünger, A, Brooks III, C, & Karplus, M. (1982) Stochastic boundary- conditions for molecular-dynamics simulations of ST2 water. *Chem. Phys. Letters* **105**: 495–500.

[2] Talay, D. (1995) Simulation and numerical analysis of stochastic differential systems: a review. in *Probabilistic Methods in Applied Physics*, P. Kree and W. Wedig (Eds.), vol. 451 of Lecture Notes in Physics, pp. 63–106. Springer-Verlag.

[3] Mishra, B. and Schlick, S. (1996) The notion of error in Langevin dynamics. I. Linear analysis *J. Chem. Phys.* **105**:299–318.

[4] Milstein, G. and Tretyakov, M. (2004) *Stochastic numerics for mathematical physics*, Springer.

[5] Burrage, K. and Lythe, G. (2009) Accurate stationary densities with partitioned numerical methods for stochastic differential equations. *SIAM J. Numer. Anal.* **47**:1601–1618.

[6] Lelièvre, T., Rousset, M., & Stoltz, G. (2010) *Free energy computations. A mathematical perspective.* Imperial College Press.

[7] Liu, J. (2001) *Monte carlo strategies in scientific computing.* Springer Series in Statistics.

[8] Larini, L., Mannella R., & Leporini, D. (2007) Langevin stabilization of molecular-dynamics simulations of polymers by means of quasisymplectic algorithms. *J. Chem. Phys.* **126**:104101.

[9] Skeel, R and Izaguirre, J. (2002) An impulse integrator for Langevin dynamics *Mol. Phys.* **100**:3885–3891.

[10] Milstein, G. and Tretyakov, M. (2003) Quasi-symplectic methods for Langevin-type equations *IMA J. Num. Anal.* **23**: 593–626.

[11] Shardlow, T., (2006) Modified equations for stochastic differential equations. *BIT* **46**, 111–125

[12] Melchionna, S. (2007) Design of quasi-symplectic propagators for Langevin dynamics. *J. Chem. Phys.* **127**:044108.

[13] Bou-Rabee, N. and Owhadi, H. (2010) Long-run accuracy of variational integrators in the stochastic context. *SIAM J. Num. Anal.* **48**: 278–297.

[14] Davidchack, R.L. (2010) Discretization errors in molecular dynamics simulations with deterministic and stochastic thermostats. *J. Comput. Phys.* **229**: 9323 – 9346.

[15] Zygalakis, K. (2011) On the existence and applications of modified equations for stochastic differential equations. *SIAM J. Sci. Comput.* **33**:102–130.

[16] Hairer, E, Lubich, C, & Wanner, G. (2006) *Geometric numerical integration.* Springer (New York), Second edition.

[17] Leimkuhler, B. and Reich, S. (2005) *Simulating Hamiltonian dynamics*, Cambridge University Press.

[18] Hörmander, L. (1967) Hypoelliptic second order differential equations. *Acta Math.* **119**: 147–171.

[19] Mattingly, J.C. Stuart, A.M. & Higham, D.J. (2002) Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appls.* **101**:185–232.

[20] Villani, C. (2009) Hypocoercivity. *Mem. Amer. Math. Soc.* **202**

[21] Pavliotis, G. and Hairer, M. (2008) From ballistic to diffusive behavior in periodic potentials. *J. Stat. Phys.* **131** 175–202

[22] Joubaud, R. and Stoltz, G. (2012) Nonequilibrium shear viscosity computations with Langevin dynamics. *SIAM MMS* **10**: 191–216

[23] Talay, D. (2002) Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields* **8**(2) 163–198.

[24] Mattingly, J.C. Stuart, A.M. & Tretyakov, M. (2010) Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Num. Anal.* **48**: 552–577.

[25] Debussche, A. and Faou, E. (2011) Weak backward error analysis for SDEs. *SIAM J. Num. Anal.* (to appear).

Appendix: Langevin Dynamics Integrators

The Langevin dynamics methods used for the numerical experiments in this paper are given here. R_i is an $N-$vector of i.i.d. Normal random numbers, with mean 0 and variance 1. The diagonal mass matrix is denoted M, and we assume a timestep δt is provided. Given the parameter γ, we define useful constants

$$c_1 = e^{-\gamma \delta t}, \quad c_2 = \gamma^{-1}(1 - c_1),$$

and

$$c_3 = \sqrt{k_B T (1 - c_1^2)}.$$

'BAOAB' method

$$p_{n+1/2} = p_n - \delta t\, \nabla U(x_n)/2;$$
$$x_{n+1/2} = x_n + \delta t\, M^{-1} p_{n+1/2}/2;$$
$$\hat{p}_{n+1/2} = c_1 p_{n+1/2} + c_3 M^{1/2} R_{n+1};$$
$$x_{n+1} = x_{n+1/2} + \delta t\, M^{-1} \hat{p}_{n+1/2}/2;$$
$$p_{n+1} = \hat{p}_{n+1/2} - \delta t\, \nabla U(x_{n+1})/2;$$

'ABOBA' method

$$x_{n+1/2} = x_n + \delta t\, M^{-1} p_n/2;$$
$$p_{n+1/2} = p_n - \delta t\, \nabla U(x_{n+1/2})/2;$$
$$\hat{p}_{n+1/2} = c_1 p_{n+1/2} + c_3 M^{1/2} R_{n+1};$$
$$p_{n+1} = \hat{p}_{n+1/2} - \delta t\, \nabla U(x_{n+1/2})/2;$$
$$x_{n+1} = x_{n+1/2} + \delta t\, M^{-1} p_{n+1}/2;$$

Stochastic Position Verlet (SPV)

$$x_{n+1/2} = x_n + \delta t\, M^{-1} p_n/2;$$
$$p_{n+1} = c_1 p_n - c_2 \nabla U(x_{n+1/2}) + c_3 M^{1/2} R_{n+1};$$
$$x_{n+1} = x_{n+1/2} + \delta t\, M^{-1} p_{n+1}/2;$$

The Method of Brunger-Brooks-Karplus (1982) (BBK)

$$p_{n+1/2} = (1 - \delta t\gamma/2) p_n - \delta t\, \nabla U(x_n)/2 + \sqrt{\delta t k_B T \gamma} M^{1/2} R_n/2;$$
$$x_{n+1} = x_n + \delta t\, M^{-1} p_{n+1/2};$$
$$p_{n+1} = [p_{n+1/2} - \delta t\, \nabla U(x_{n+1})/2 + \sqrt{\delta t k_B T \gamma} M^{1/2} R_{n+1}/2]/(1 + \delta t\gamma/2);$$

Euler-Maruyama

$$x_{n+1} = x_n - \delta t M^{-1} \nabla U(x_n) + \sqrt{2 k_B T \delta t} M^{-1/2} R_n;$$

'BAOAB' - Limit Method

$$x_{n+1} = x_n - \delta t M^{-1} \nabla U(x_n) + \sqrt{\frac{k_B T \delta t}{2}} M^{-1/2} (R_n + R_{n+1});$$