SUPPORTING INFORMATION
Adaptive splitting integrators for enhancing sampling efficiency of modified Hamiltonian Monte Carlo methods in molecular simulation

Elena Akhmatskaya*[1,2], Mario Fernández-Pendás[1], Tijana Radivojević[1], and J.M. Sanz-Serna[3]

[1]BCAM - Basque Center for Applied Mathematics, Alameda de Mazarredo 14, E-48009 Bilbao, Spain

[2]Ikerbasque, Basque Foundation for Science, María Díaz de Haro 3, E-48013 Bilbao, Spain

[3]Departamento de Matemáticas, Universidad Carlos III de Madrid, Avenida de la Universidad 30, E-28911 Leganés (Madrid), Spain

Contents

1 Derivation of the function ρ used in MAIA and e-MAIA algorithms

In order to derive integrators with optimal conservation properties, we adopt a strategy similar to the one proposed in ref 1, namely to find the parameters of integrators that minimize the expected value of the energy error. In the present study, the energy error resulting from numerical integration is in terms of the modified Hamiltonian and the expected value is taken with respect to the modified density

$$\tilde{\pi}(\mathbf{q},\mathbf{p}) \propto \exp\left(-\beta \tilde{H}^{[k]}(\mathbf{q},\mathbf{p})\right).$$

As in the case when considering the error in the true Hamiltonian,[1] one may prove that the expected error in the modified Hamiltonian $\mathbb{E}[\Delta\tilde{H}^{[4]}]$ is also positive. Our objective is, therefore, to find a function $\rho(h,b)$ that upperbounds $\mathbb{E}[\Delta\tilde{H}^{[4]}]$, i.e.,

$$0 \le \mathbb{E}[\Delta\tilde{H}^{[4]}] \le \frac{1}{\beta}\rho(h,b),$$

where b is the parameter of the two-stage integrators family

$$\psi_{\Delta t} = \phi^B_{b\Delta t} \circ \phi^A_{\Delta t/2} \circ \phi^B_{(1-2b)\Delta t} \circ \phi^A_{\Delta t/2} \circ \phi^B_{b\Delta t}. \tag{S1}$$

*E-mail: akhmatskaya@bcamath.org

Such family is defined in terms of solution flows of the equations of motion[1]

$$\phi_t^A(\mathbf{q},\mathbf{p}) = (\mathbf{q}, \mathbf{p} - t\nabla U(\mathbf{q})) \tag{S2}$$

and

$$\phi_t^B(\mathbf{q},\mathbf{p}) = (\mathbf{q} + tM^{-1}\mathbf{p}, \mathbf{p}). \tag{S3}$$

We consider the one-dimensional harmonic oscillator with potential $U(q) = (k/2)q^2$ ($k > 0$ a constant) and mass M, whose equations of motion are

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{p}{M}, \quad \frac{\mathrm{d}p}{\mathrm{d}t} = -kq. \tag{S4}$$

Using a linear change of variables $\bar{q} = \sqrt{k}q, \bar{p} = p/\sqrt{M}$ and denoting the non-dimensional time step as $h = \omega\Delta t$, where $\omega = \sqrt{k/M}$, lead to the dynamics

$$\frac{\mathrm{d}\bar{q}}{\mathrm{d}t} = \omega\bar{p}, \quad \frac{\mathrm{d}\bar{p}}{\mathrm{d}t} = -\omega\bar{q}, \tag{S5}$$

with Hamiltonian

$$H(\bar{q},\bar{p}) = \frac{1}{2}\bar{p}^2 + \frac{1}{2}\bar{q}^2$$

and the fourth order modified Hamiltonian for integrators of the family eq S1

$$\tilde{H}^{[4]}(\bar{q},\bar{p}) = \frac{1}{2}\bar{p}^2 + \frac{1}{2}\bar{q}^2 + h^2\lambda\bar{p}^2 + h^2\mu\bar{q}^2. \tag{S6}$$

The numerical integration of the system eq S5 in the new variables $\bar{q}, \bar{p}$ is equivalent to the application of the above change of variables to the numerical solution q, p obtained by integration of the system eq S4.

In order to find the error in the modified Hamiltonian after L integration steps of the dynamics eq S5 with a time step h, i.e.,

$$\Delta\tilde{H}^{[4]} = \tilde{H}^{[4]}(\Psi_{h,L}(\bar{q},\bar{p})) - \tilde{H}^{[4]}(\bar{q},\bar{p}), \tag{S7}$$

we first find the numerical solution to the dynamics eq S5 for a single time step $(\bar{q}_{n+1}, \bar{p}_{n+1}) = \psi_h(\bar{q}_n, \bar{p}_n)$. In matrix form this is given by

$$\begin{bmatrix} \bar{q}_{n+1} \\ \bar{p}_{n+1} \end{bmatrix} = \tilde{M}_h \begin{bmatrix} \bar{q}_n \\ \bar{p}_n \end{bmatrix}, \quad \tilde{M}_h = \begin{bmatrix} A_h & B_h \\ C_h & A_h \end{bmatrix},$$

where the coefficients A_h, B_h, C_h depend on the integrator. After L integration steps the state of the system $(\bar{q}_L, \bar{p}_L) = \Psi_{h,L}(\bar{q},\bar{p})$ is given by

$$\begin{bmatrix} \bar{q}_L \\ \bar{p}_L \end{bmatrix} = \underbrace{\tilde{M}_h \ldots \tilde{M}_h}_{L \text{ times}} \begin{bmatrix} \bar{q} \\ \bar{p} \end{bmatrix} = \tilde{M}_h^L \begin{bmatrix} \bar{q} \\ \bar{p} \end{bmatrix}. \tag{S8}$$

For the two-stage family of integrators the matrix $\tilde{M}_h$ can be calculated as

$$\tilde{M}_h = B\left(b\right) \cdot A\left(\frac{1}{2}\right) \cdot B\left(1 - 2b\right) \cdot A\left(\frac{1}{2}\right) \cdot B\left(b\right),$$

where

$$A(a) = \begin{bmatrix} 1 & ah \\ 0 & 1 \end{bmatrix}, \; B(b) = \begin{bmatrix} 1 & 0 \\ -bh & 1 \end{bmatrix}$$

correspond to the flows ϕ_h^A and ϕ_h^B, respectively (eqs S2 and S3). The resulting entries of $\tilde{M}_h$ are

$$\begin{aligned} A_h &= \frac{h^4}{4}b(1-2b) - \frac{h^2}{2} + 1 \\ B_h &= -\frac{h^3}{4}(1-2b) + h \\ C_h &= -\frac{h^5}{4}b^2(1-2b) + h^3 b(1-b) - h. \end{aligned} \tag{S9}$$

It is well known that if time step h is such that $|A_h| < 1$ the integration is stable.[2,3] In that case one may define variables

$$\zeta_h := \arccos A_h, \qquad \chi_h := B_h / \sin \zeta_h,$$

for which the one-step and L-steps integration matrices $\tilde{M}_h$ and $\tilde{M}_h^L$, respectively, are

$$\tilde{M}_h = \begin{bmatrix} \cos(\zeta_h) & \chi_h \sin(\zeta_h) \\ -\chi_h^{-1}\sin(\zeta_h) & \cos(\zeta_h) \end{bmatrix}$$

and

$$\tilde{M}_h^L = \begin{bmatrix} \cos(L\zeta_h) & \chi_h \sin(L\zeta_h) \\ -\chi_h^{-1}\sin(L\zeta_h) & \cos(L\zeta_h) \end{bmatrix}. \tag{S10}$$

In order to calculate the expected value of the error eq S7 we follow ideas from the proof of Proposition 3 in ref 1 and denote

$$\begin{aligned} c &= \cos(L\zeta_h), \\ s &= \sin(L\zeta_h), \\ S_1 &= 1 + 2h^2\mu, \\ S_2 &= 1 + 2h^2\lambda. \end{aligned}$$

Substituting eqs S6, S10 and S8 into eq S7 leads to

$$\begin{aligned} 2\Delta\tilde{H}^{[4]} &= S_1\left(c\bar{q} + \chi_h s\bar{p}\right)^2 + S_2\left(-\frac{1}{\chi_h}s\bar{q} + c\bar{p}\right)^2 - S_1\bar{q}^2 - S_2\bar{p}^2 \\ &= s^2\left(\frac{1}{\chi_h^2}S_2 - S_1\right)\bar{q}^2 + s^2\left(\chi_h^2 S_1 - S_2\right)\bar{p}^2 + 2sc\left(S_1\chi_h - S_2\frac{1}{\chi_h}\right)\bar{q}\bar{p}. \end{aligned}$$

Since the expectations are taken with respect to the modified density $\tilde{\pi}$,

$$\mathbb{E}[\bar{q}^2] = \frac{1}{\beta S_1},\ \mathbb{E}[\bar{p}^2] = \frac{1}{\beta S_2},\ \mathbb{E}[\bar{q}\bar{p}] = 0,$$

it follows that

$$2\mathbb{E}[\Delta\tilde{H}^{[4]}] = \frac{1}{\beta}s^2\Big(\frac{1}{\chi_h^2}\frac{S_2}{S_1} + \chi_h^2\frac{S_1}{S_2} - 2\Big).$$

The last expression can be simplified by defining

$$\tilde{\chi}_h^2 := \chi_h^2\frac{S_1}{S_2} = \chi_h^2 S,$$

to obtain

$$\mathbb{E}[\Delta\tilde{H}^{[4]}] = \frac{1}{\beta}s^2\rho(h,b),$$

where

$$\rho(h,b) = \frac{1}{2}\Big(\widetilde{\chi}_h - \frac{1}{\widetilde{\chi}_h}\Big)^2 = \frac{\Big(SB_h + C_h\Big)^2}{2S(1-A_h^2)} \tag{S11}$$

and

$$S = \frac{1+2h^2\mu}{1+2h^2\lambda}.$$

We note that the conditions for stable integration and positivity of $\rho(h,b)$ are that $|A_h| < 1$ and $S > 0$. For the two-stage integrators and the fourth order modified Hamiltonian this is equivalent to the following conditions

$$\begin{aligned} h &< \sqrt{12/(1-6b)} \quad \text{for } b < \frac{1}{6}, \\ h &> \sqrt{12/(1-6b)} \quad \text{for } b > \frac{1}{6}, \\ 0 < \quad h &< \min\{\sqrt{2/b}, \sqrt{1/(1-2b)}\}, \end{aligned}$$

which are always satisfied for $b \in (0, \frac{1}{2})$.

Finally, by substituting eq S9 into eq S11 we obtain the expression

$$\rho(h,b) = \frac{h^8\Big(b\big(12+4b(6b-5)+b(1+4b(3b-2))h^2\big)-2\Big)^2}{4\big(2-bh^2\big)\big(4+(2b-1)h^2\big)\big(2+b(2b-1)h^2\big)\big(12+(6b-1)h^2\big)\big(6+(1+6(b-1)b)h^2\big)}, \tag{S12}$$

which bounds the expected error in the modified Hamiltonian. This function is then used within an optimization routine to find the value b that provides the optimal conservation of the modified Hamiltonian for a specific system.

2 Flowchart of MAIA and e-MAIA algorithms

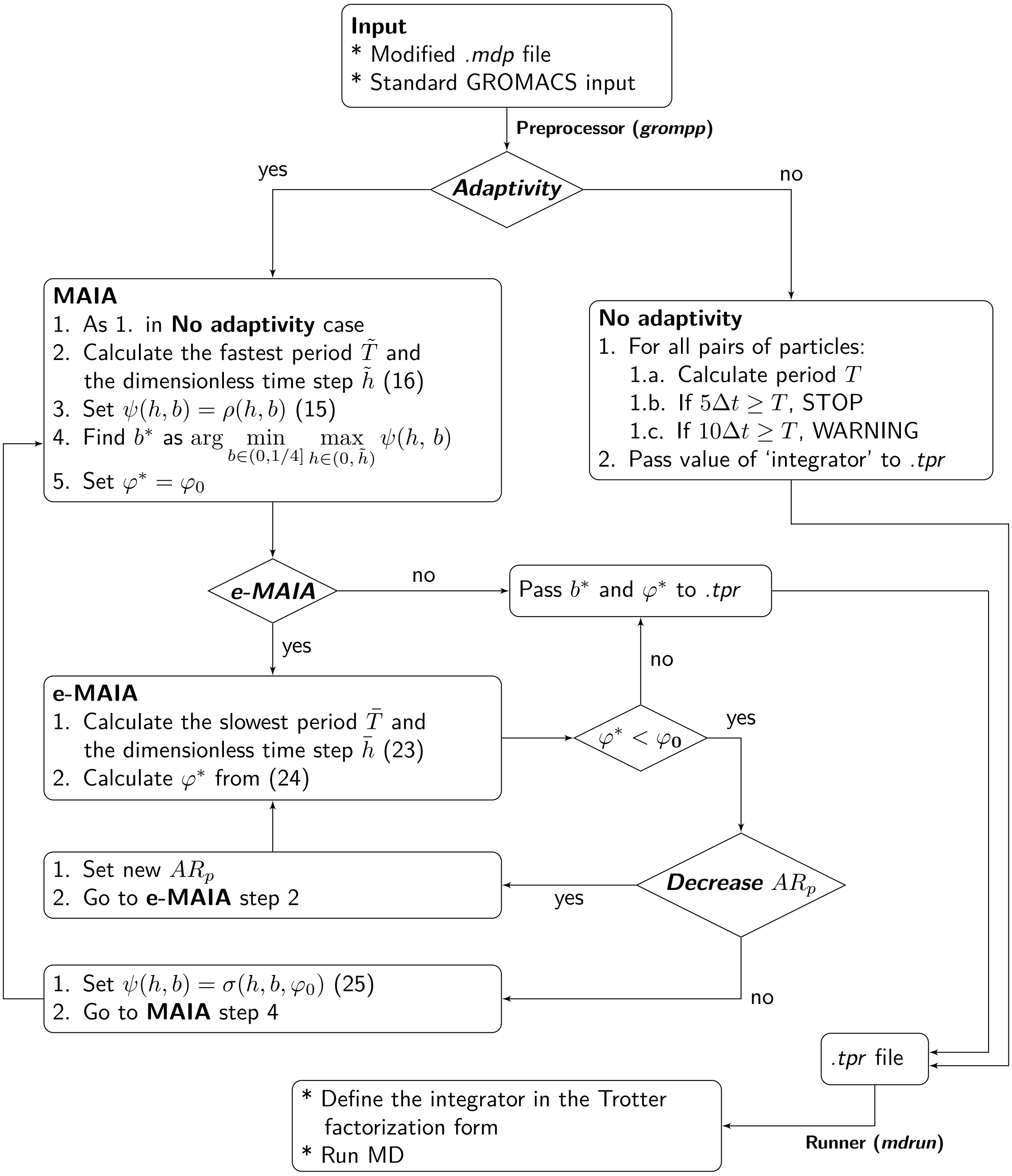

Figure S1: Flowchart of the Modified Adaptive Integration Approach (MAIA) and the extended MAIA (e-MAIA) as implemented in MultiHMC-GROMACS. The references (15)-(16) and (23)-(25) correspond to the equations in the paper.

3 Validation of the chosen simulation length for the villin benchmark

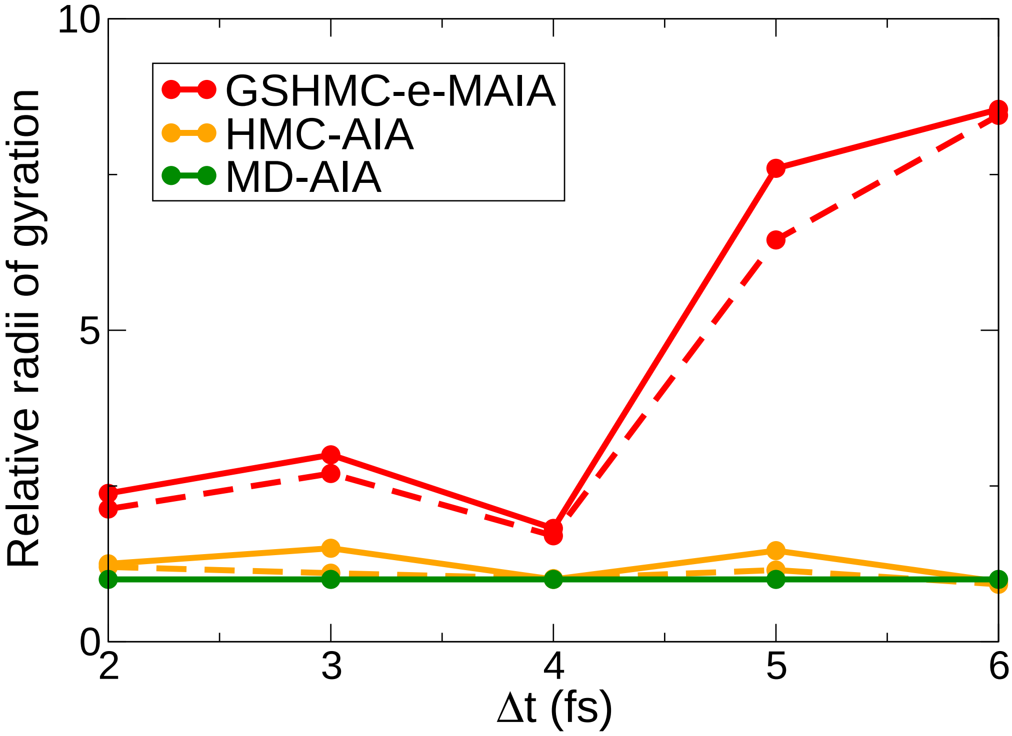

Figure S2: Evolution with time of the relative radii of gyration (RG) observed for each simulation method with respect to the RG found in MD simulations. The dashed lines represent the RG at half of the simulation time (2.5 ns) whereas the solid lines are used for the full simulations of 5 ns. The efficiency of GSHMC with e-MAIA, relative to MD, expressend in terms of radii of gyration, increases with simulation time. This suggests that simulations longer than those presented in this study are at least as favorable to the proposed algorithms as we claim; they may be even more favorable as the simulations become longer.

References

[1] Blanes, S.; Casas, F.; Sanz-Serna, J. M. Numerical Integrators for the Hybrid Monte Carlo Method. *SIAM J. Scient. Comp.* **2014**, *36*, A1556–A1580.

[2] Sanz-Serna, J. M.; Calvo, M. P. *Numerical Hamiltonian problems*, 1st ed.; Applied Mathematics and Mathematical Computation 7; Chapman & Hall: London, 1994.

[3] Hairer, E.; Lubich, C.; Wanner, G. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations; 2nd ed.*; Springer: Dordrecht, 2006.