NUMERICAL INTEGRATORS FOR THE HAMILTONIAN MONTE CARLO METHOD

J. M. Sanz-Serna
Universidad Carlos III de Madrid

Collaborators: S. Blanes (U. Politécnica Valencia), F. Casas (U. Jaume I), E. Akhmatskaya (BCAM), M. Fernández-Pendás (BCAM), T. Radivojevic (BCAM), N. Bou-Rabee (U. Rutgers, Camden)

I: THE PROBLEM

• Wish to sample from target pdf $\propto \exp(-V(q))$.

• Computational effort in HMC mostly spent when numerically integrating the Hamiltonian dynamics associated with the Hamiltonian function $H(q,p) = (1/2)p^2 + V(q)$, i.e. the differential system

$$(d/dt)q = p, \qquad (d/dt)p = -\nabla V(q).$$

• Which integrator shall we use?

[N. Bou-Rabee & JMSS, Geometric integrators and the Hamiltonian Monte Carlo method, Acta Numerica, 2018.]

• Verlet is the algorithm of choice. For velocity form, one time-step is

$$\begin{aligned} p_{i+1/2} &= p_i - \frac{h}{2}\nabla_q V(q_i), \quad \text{(kick)} \\ q_{i+1} &= q_i + h\, p_{i+1/2}, \quad \text{(drift)} \\ p_{i+1} &= p_{i+1/2} - \frac{h}{2}\nabla_q V(q_{i+1}). \quad \text{(kick)} \end{aligned}$$

• This is obviously a splitting integrator. Over one time step numerical solution is advanced by map

$$\psi_h = \varphi^B_{h/2} \circ \varphi^A_h \circ \varphi^B_{h/2},$$

where φ^A, φ^B are exact solution maps (flows) of the (Hamiltonian) split systems:

$$\begin{aligned} (A) \qquad & (d/dt)q = p, \quad (d/dt)p = 0, \\ (B) \qquad & (d/dt)q = 0, \quad (d/dt)p = -\nabla V(q). \end{aligned}$$

- ψ_h volume preserving as composition of volume-preserving flows.

- ψ_h reversible due to palindromic structure of composition.

- These properties allow for simple accept-reject rule. Proposal (q^*, p^*) accepted with probability

$$a^{(n)} = \min\left(1, \exp\left(H(q^{(n)}, p^{(n)}) - H(q^*, p^*)\right)\right).$$

($(q^{(n)}, p^{(n)})$ current state of Markov chain).

[For useful dynamics that do not preserve volume see Y. Fang, JMSS & RD Skeel, Compressible generalized HMC, J. Chem. Phys. 2014.]

- Points to remember when choosing an integrator:

1. Interested in *energy errors* $\Delta(q,p)$ after I time-steps:

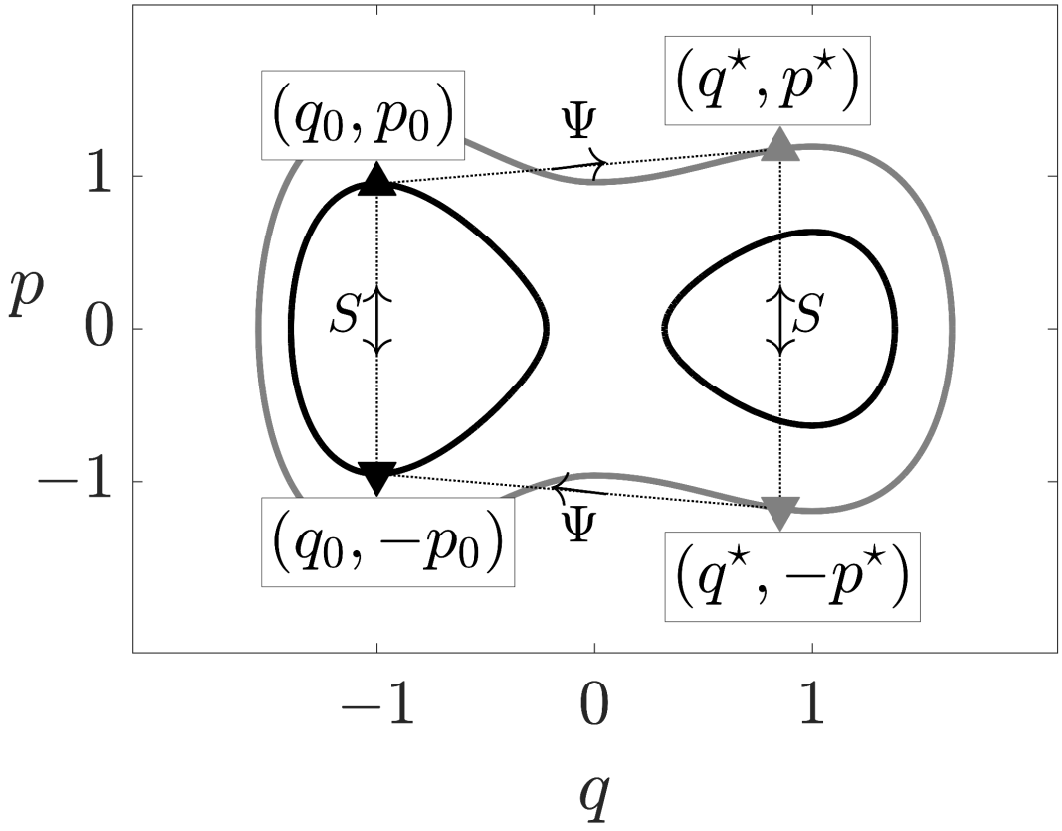
   $$\Delta = H(\Psi_{h,I}(q,p)) - H(\varphi_{Ih}^{H}(q,p)) = H(\Psi_{h,I}(q,p)) - H(q,p),$$

   as only these determine the acceptance probability.

2. High accuracy may not be required, unless *number of degrees of freedom is very high.* With an energy error $\Delta(q^{(n)},p^{(n)}) = 1$ the proposal q^* will be accepted with probability $\exp(-1) > 36\%$. And for $\Delta(q^{(n)},p^{(n)}) = 2$ the probability of acceptance is still larger than 13%.

3. The *sign* of the energy error matters: $\Delta(q^{(n)},p^{(n)}) < 0$ leads to acceptance of the proposal.

- Conservation of volume and reversibility have an impact on energy errors:



- Expected energy error at stationarity of chain:

$$\mathbb{E}(\Delta) = \int_{\mathbb{R}^{2d}} \Delta(q,p) \exp\big(-H(q,p)\big)\, dq\, dp;$$

from figure we see $\mathbb{E}(\Delta)$ may also be written

$$-\int_{\mathbb{R}^{2d}} \Delta(q,p) \exp\big(-H(\Psi(q,p))\big)\, dq\, dp,$$

or, averaging both expressions,

$$\tfrac{1}{2}\int_{\mathbb{R}^{2d}} \Delta(q,p)\Big[\exp\big(-H(q,p)\big) - \exp\big(-H(\Psi(q,p))\big)\Big]\, dq\, dp$$
$$= \tfrac{1}{2}\int_{\mathbb{R}^{2d}} \Delta(q,p)\Big[1 - \exp\big(-\Delta(q,p)\big)\Big] \exp\big(-H(q,p)\big)\, dq\, dp.$$

From here one may prove

$$0 \le \mathbb{E}(\Delta) \le \int_{\mathbb{R}^{2d}} \Delta(q,p)^2 \exp\big(-H(q,p)\big)\, dq\, dp.$$

Conclusions:

For a volume-preserving, reversible integrator, energy errors are, on average, much smaller than one would think.

Asymptotic properties in the limit $h \to 0$ (i.e. order, leading coefficients of local error expansion) of limited interest as in practice h will not be 'small'.

Hence: analyze integrators without assuming smallness of h.

But this is only possible for model problems . . .

II: A MODEL PROBLEM

- Harmonic oscillator:

$$H = \frac{1}{2}(p^2 + q^2), \qquad q, p \in \mathbb{R},$$

$$\frac{d}{dt}q = -p, \qquad \frac{d}{dt}p = q.$$

- From sampling point of view, target is the standard univariate Gaussian distribution. In matrix form, the solution flow is given by

$$\begin{bmatrix} q(t) \\ p(t) \end{bmatrix} = M_t \begin{bmatrix} q(0) \\ p(0) \end{bmatrix}, \qquad M_t = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}.$$

- Then, *assuming stability,* the one-step numerical matrix is:

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

and, over i steps:

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

numerical solution stays on an ellipse.

- θ_h governs phase errors (here irrelevant).

- χ_h governs shape of numerical orbits/energy errors. $\chi_h \equiv 1$ would be ideal (then numerical solution stay on circles, no energy error).

The expectation of the random variable $\Delta(q_0, p_0)$ is given by:

[S. Blanes, F. Casas, JMSS, SIAM J. Sci. Comput. 2014]

$$\mathbb{E}(\Delta) = \sin^2(I\theta_h)\, \rho(h),$$

where

$$\rho(h) = \frac{1}{2}\left(\chi_h^2 + \frac{1}{\chi_h^2} - 2\right) = \frac{1}{2}\left(\chi_h - \frac{1}{\chi_h}\right)^2 \geq 0.$$

Accordingly

$$0 \leq \mathbb{E}(\Delta) \leq \rho(h).$$

Illustration:

Velocity Verlet is stable for $0 < h < 2$, which is *optimal*.

For stable values of h:

$$\mathbb{E}(\Delta) \leq \frac{h^4}{32(1 - \frac{h^2}{4})}.$$

For $h \leq 1$ the expected energy error is $\leq 1/24$.

Halving h to $h \leq 1/2$, leads to an expected energy error $\leq 1/480$!

Extension: For a d-variate Gaussian target distribution (d coupled linear oscillators), assuming stability,

$$\mathbb{E}(\Delta) \leq \sum_{j=1}^{d} \rho(h\omega_j),$$

where ω_j are the angular frequencies of the oscillators (inverses of the standard deviations).

Note $h\omega_j$ is a *nondimensional* combination and stability requires $h\omega_j < 2$ for each j.

III: IMPROVING ON VERLET

• Split-step methods suggest themselves. Very easy implementation (sequence of drifts/kicks just as Verlet). They are symplectic, and, if palindromic, reversible.

• Many antecedents in literature: free parameters have been used to boost order and/or reduce error constants.

• Here we minimize

$$\|\rho\|_{(\bar{h})} = \max_{0<h<\bar{h}} \rho(h),$$

where $\bar{h}$ is suitable nondimensional maximum step-length ($\bar{h} \leq$ length of stability interval).

• If method uses r evaluations of ∇V per step, we choose $\bar{h} = r$, since Verlet works well with $h \approx 1$ for relevant numbers of degrees of freedom.

Two evaluations of ∇V per step:

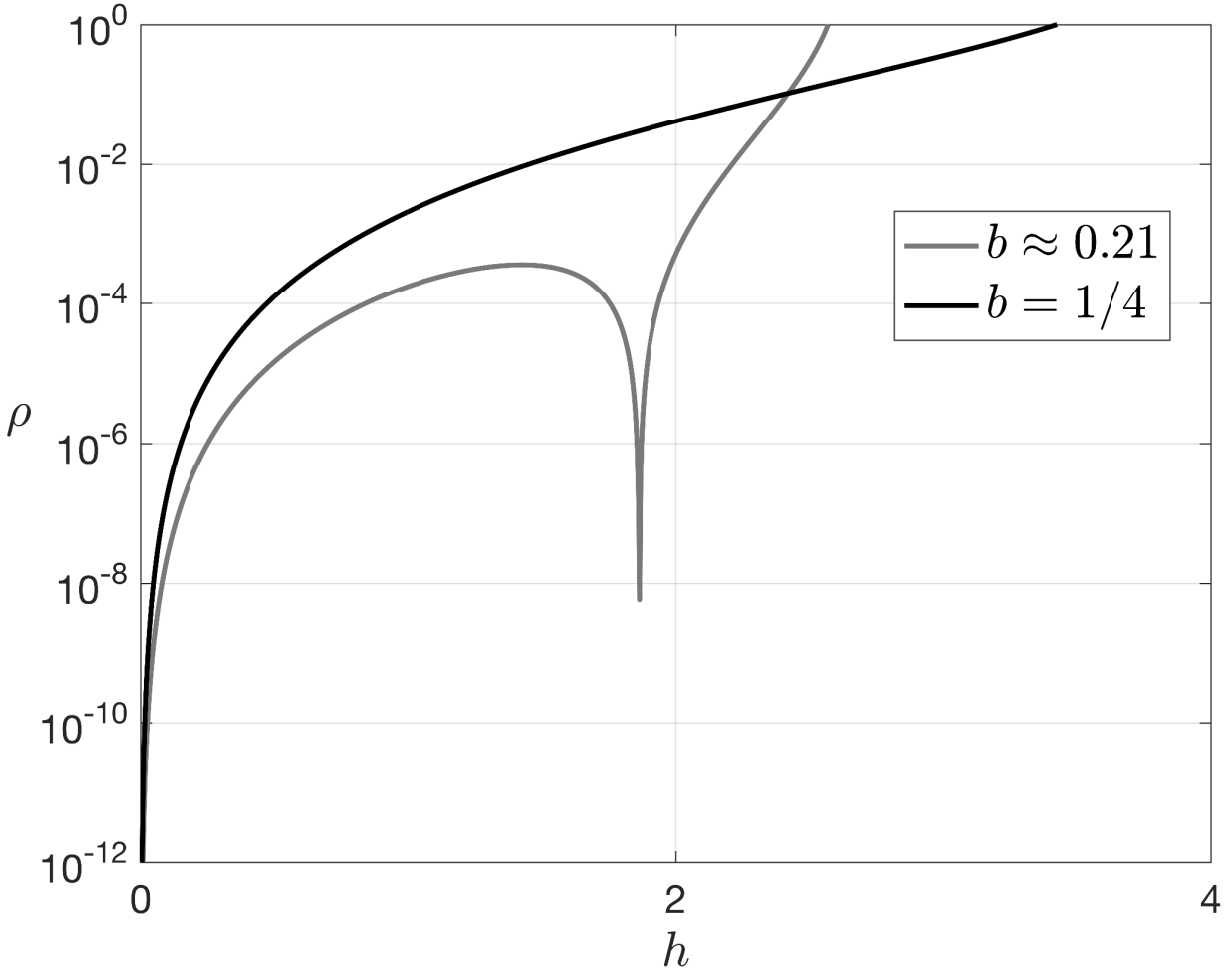
- One-parameter family of palindromic formulae (three kicks, two drifts):

$$\psi_h = \varphi^B_{bh} \circ \varphi^A_{(1/2)h} \circ \varphi^B_{(1-2b)h} \circ \varphi^A_{(1/2)h} \circ \varphi^B_{bh}.$$

- When $b = 1/4$ method is concatenation of two-steps (of step-length $h/2$) of (position) Verlet method.

- All methods are second order accurate. Minimal error constant $b \approx 0.1932$. As b increases from 0.19 to 0.25 integrator may operate with larger values of h but is less accurate.

- Here a chosen to minimize $\max \rho(h)$, $0 < h < 2$. This leads to the optimal choice $b \approx 0.2118$.


10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
ρ
0
2
4
h
$b \approx 0.21$
$b = 1/4$

Three evaluations of ∇V per step:

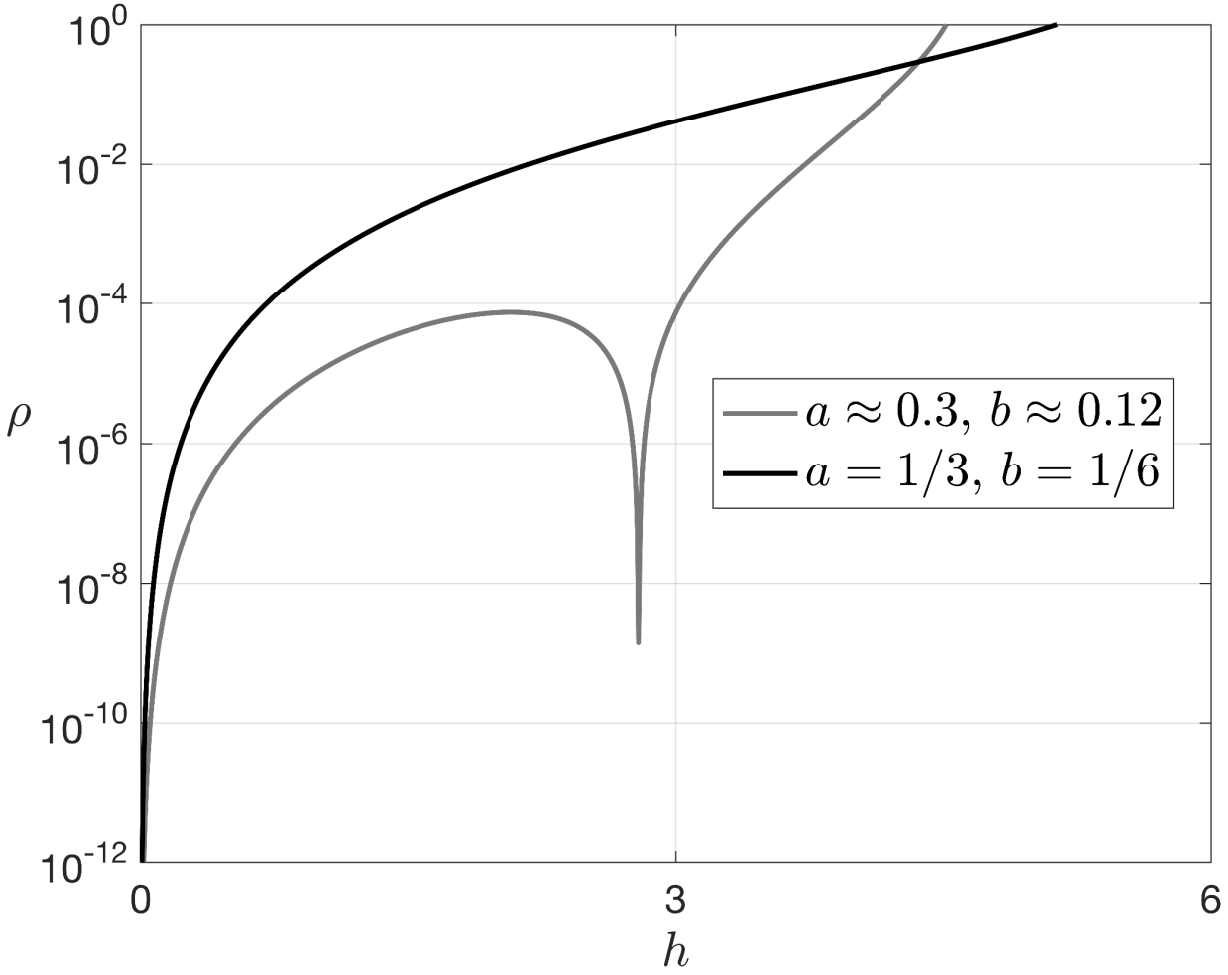
- Two-parameter family of palindromic formulae (three kicks, two drifts):

$$\varphi^B_{bh} \circ \varphi^A_{ah} \circ \varphi^B_{(1/2-b)h} \circ \varphi^A_{(1-2a)h} \circ \varphi^B_{(1/2-b)h} \circ \varphi^A_{ah} \circ \circ \varphi^B_{bh}.$$

- When $a = 1/3$ $b = 1/6$ method is concatenation of three-steps (of step-length $h/3$) of velocity Verlet method.

- Order *four* is possible and has been considered in this context.

- Here minimize $\max \rho(h)$, $0 < h < 3$. This is tricky: generically methods have stability intervals shorter than $0 < h < 3$. (The fourth-order method has stability interval $0 < h < 1.573$.)


10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
ρ
0
3
6
h
$a \approx 0.3,\ b \approx 0.12$
$a = 1/3,\ b = 1/6$

Numerical comparison of three-stage algorithms:

- Canonical distribution for an alkane molecule with 27 degrees of freedom.

(Not too many degrees of freedom and away from Gaussian model.)

- One force evaluation every 8 fs (good for Verlet).

- Average and standard deviation of acceptance rate:

Integrator	μ	σ
Verlet	77.7%	2.11%
Fourth order	0%	0%
Minimum ρ	96.7%	0.41%

AIA (adaptive integrator approach) [M Fernández-Pendás, E Akhmatskaya, JMSS, J. Comput. Phys. 2016]

Go back to two stage family. Above, free parameter b chosen once and for all to minimize ρ over $0 < h < 2$.

In AIA steplength chosen according to computational budget. Then estimate shortest interval $(0, h^\star)$ that contains all products $h\omega_j$ and minimize ρ over $0 < h < h^\star$.

Incorporated to molecular dynamics popular software GROMACS.

Sampling for a large biomolecule. If $\triangle t$ is large, AIA chooses Verlet. If user may operate smaller values of $\triangle t$, AIA automatically picks an integrator with a shorter stability interval and enhanced accuracy.