Haussdorf School on MCMC: Recent developments and new connections, September 21–25, 2020

NUMERICAL INTEGRATORS FOR THE HAMILTONIAN MONTE CARLO METHOD, LECTURE III

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

14: IMPROVING ON LEAPFROG?

• The computational cost of HMC lies mainly in carrying out the numerical integrations to find the proposals (more precisely, in evaluating $-\nabla V(q)$ within the numerical integrator).

• Hence the efficiency of the numerical integrator is of crucial importance and it is natural to consider the following problem:

Is leapfrog the best integrator to use within HMC?

• As we will see leapfrog is very good but often not the best.

THE MATERIAL IS ALSO RELEVANT FOR UNADJUSTED HMC AND EVEN FOR MOLECULAR DYNAMICS

• Important: satisfactory HMC integrations may be inaccurate.

• With an energy error $\Delta(q^{(n)}, p^{(n)}) = 0.5$ the proposal will be accepted with probability $\exp(-1) > 60\%$.

• If the value of H on the trajectory is ≈ 1, $\Delta(q^{(n)}, p^{(n)}) = 0.5$ means a relative error several order of magnitude larger than those one aims at in typical ODE simulations.

• However note that typically, as the dimensionality increases, the value of H and, by implication, the value of Δ for given h increases. For instance, in the model situation examined before of many replicas of the same system, Δ grows linearly with the number of replicas. As this number grows one needs to reduce h on accuracy grounds, as we saw.

- Leapfrog is very good for HMC...
- ...because it has an optimality property.

Assume that for the standard univariate Gaussian target,

$\star$ you wish to advance the numerical solution over a time interval of length h,
$\star$ you wish to do so in a stable way (the matrix that effects the advancement has bounded powers),
$\star$ you are prepared to use s evaluations of ∇V and you are not prepared to solve implicit equations,

then:

If you take s consecutive steps of length h/s of leapfrog, then h needs to be $< 2s$ (so that $h/s < 2 = \eta$).

If you use an integrator other than leapfrog, then the restriction on h will be more severe.

• In other words, leapfrog provides sensible (perhaps not too accurate) integrations for large step-sizes for which other integrators just give nonsense.

• In HMC we may live with those inaccurate stable integrations (unless the high dimensionality kicks in).

• But high-dimensional problems are precisely those where computational cost is a more pressing concern.

• Is then leapfrog the best?

• I interpret what's the 'best' integrator in the following way:

What is the (volume-preserving and time-reversible) integrator that delivers the highest number of accepted proposals for a given computational budget? [I will discuss later why acceptance rate is a good metric.]

⋆ The answer depends on the target distribution.

⋆ The answer cannot be found analytically because it hinges on the global error of the different integrators, that as we know are difficult to pin down, particularly so when they are large.

• Therefore we follow an approach that is common in numerical analysis: use a model problem to select suitable integrators and then use numerical experiments to see how the selected integrators behave in general problems.

15: A CRITERION FOR SELECTING HMC INTEGRATORS

(Blanes, Casas & SS 2014)

• Use the Gaussian model problem to discriminate between integrators.

• Recall that, in the standard univariate case, the expected acceptance rate for given h is a function of the expected energy error and this has a bound

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

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

(χ_h is an integrator-dependent value) and the expected acceptance rate is a (monotonically decreasing) function of $\mathbb{E}(\Delta)$.

• This suggest minimizing $\rho(h)$ over all integrators with a given computational effort.

• In the multivariate Gaussian case, if $\omega_j \Delta t < \eta$,

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

• Numerical experiments reveal that, over a large range of Gaussian targets, leapfrog works well in HMC if $\forall j,\ \omega_j \Delta t < 1$ (which is half the maximum allowed for stability). When Δ is chosen in this way

$$\mathbb{E}(\Delta) \leq d \max_{0<h<1} \rho(h).$$

• Therefore if s is the number of gradient evaluations per time-step, Blanes et al. suggest choosing the integrator that minimizes:

$$\boxed{\max_{0<h<s} \rho(h)}$$

• Integrators whose stability interval is shorter than $(0, s)$ are non-starters.

16: CONSTRUCTING HMC INTEGRATORS

(Blanes, Casas & SS 2014)

• Leapfrog

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

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:

$$\begin{aligned} &(A) \qquad (d/dt)q = M^{-1}p, \quad (d/dt)p = 0, \\ &(B) \qquad (d/dt)q = 0, \quad (d/dt)p = -\nabla V(q). \end{aligned}$$

• Consider two-parameter family of splitting integrators (three kicks, two drifts):

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

• Volume-preserving and time-reversible (palindromic).

• The computational cost is 3 gradient evaluations per time-step.

• Easy implementation: just as leapfrog. Sequence of kicks/drifts.

• When $a = 1/3$ $b = 1/3$, integrator is concatenation of three steps (each of step-length $h/3$) of leapfrog. Then $\eta = 6$.

• Possible to choose parameters so as to achieve order $r = 4$. This integrator has been considered repeteadly in the literature but it is very poor, specially in the HMC context, because its stability interval is very short.

• In fact the stability interval is very short unless $a + b - 6ab = 0$ (Campos & SS 2017). So we impose this relation and are left with b as a free parameter.

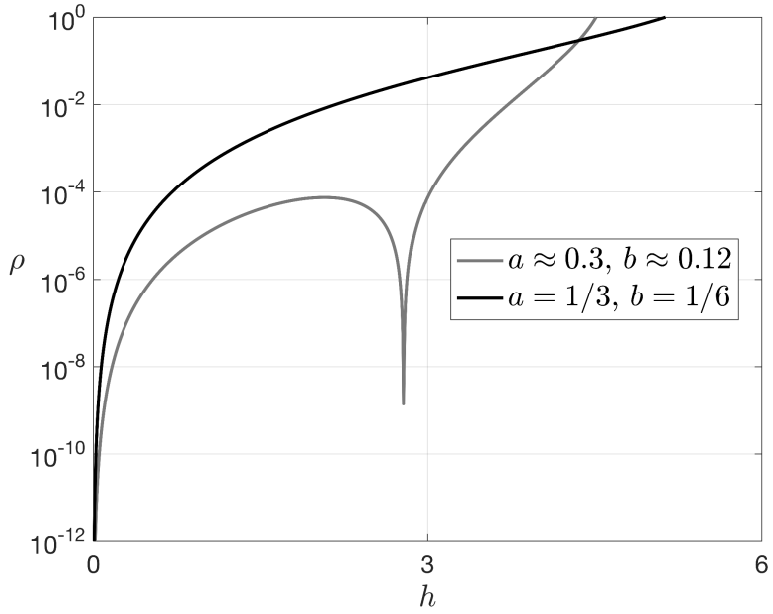
• Predescu et al 2012 suggest $b = 0.39100857459657 5$ for molecular dynamics simulations.

For quadratic potential this value of b guarantees that after a single time-step the energy error is pointwise $\mathcal{O}(h^5)$ as $h \downarrow 0$. [Hence, for Gaussian targets and over an integration leg, the pointwise energy error is $\mathcal{O}(h^4)$ with expectation $\mathcal{O}(h^8)$.] It has $\eta \approx 4.584$.

• As discussed above Blanes, Casas & SS 2014 choose b to minimize $\max_{0<h<3} \rho(h)$, which leads to $b = 0.38111989033452$, with $\eta \approx 4.662$.

• This value is not too different from Predescu's.

• Both Predescu et al and Blanes et al use energy error for harmonic oscillator to determine b. Predescu et al look at integrator in the limit $h \downarrow$, Blanes et al over a full interval $0 < h < 3$ (motivated by the fact that HMC integrations do not require very small h).



ρ as a function of h for leapfrog and Blanes. Leapfrog has vertical asymptote at $h = 6.00$ (not shown) and Blanes at $h = 4.66$. For $h < 3$ Blanes is a clear improvement on leapfrog.

17: TESTING THE INTEGRATORS: GAUSSIAN MODELS

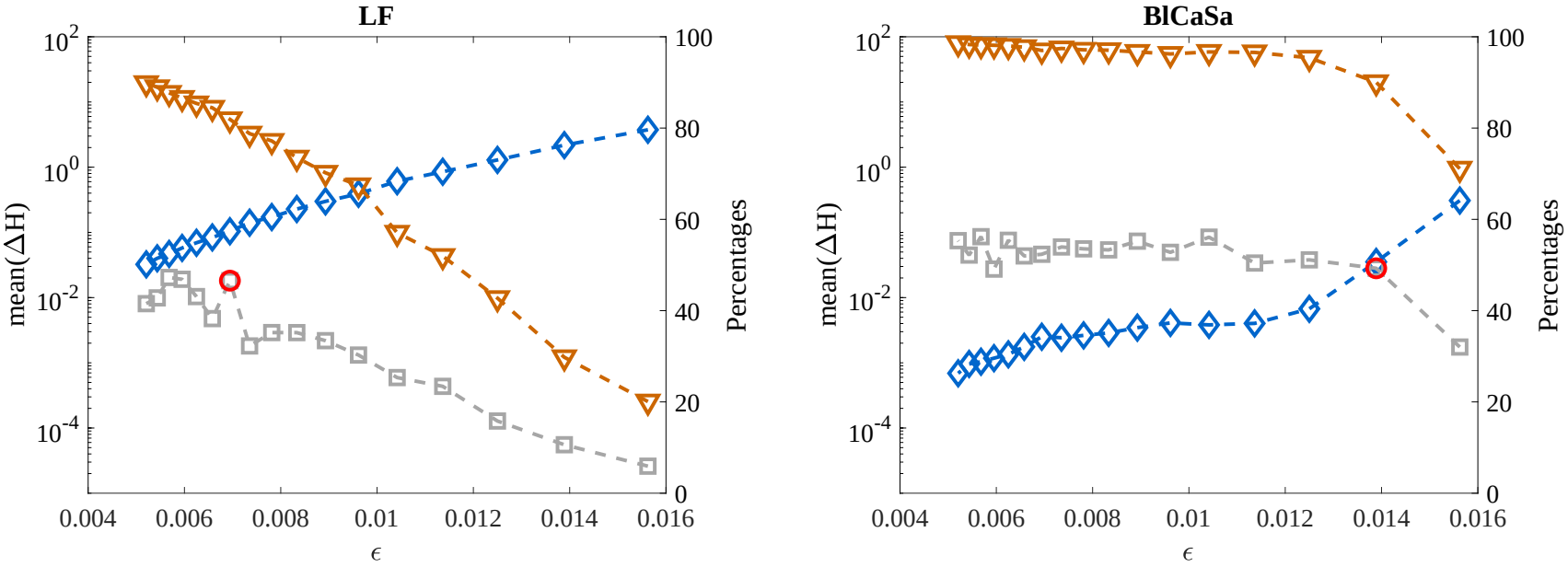
Calvo, Sanz-Alonso & SS 2019

• Use the model target density

$$\propto \exp\left(-\frac{1}{2}\sum_{j=1}^{d} j^2 q_j^2\right),$$

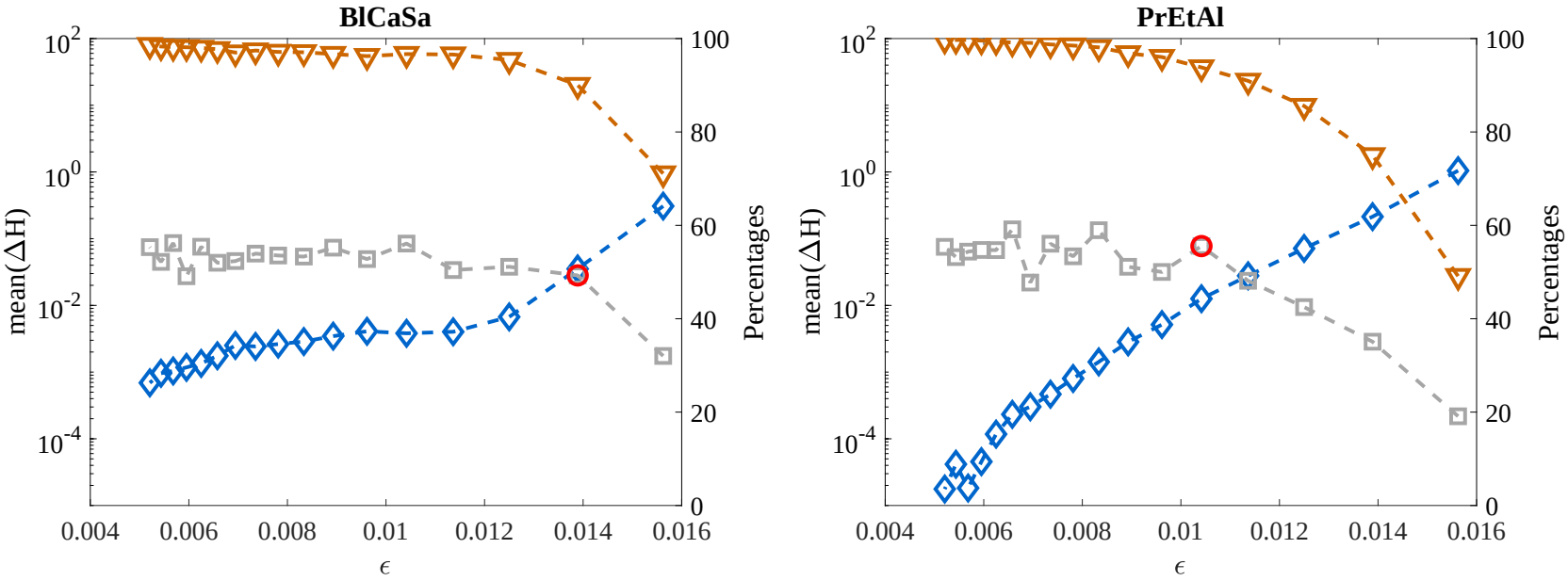
with $d = 256$ or $d = 1024$.

• We report results for leapfrog, Predescu et al and Blanes et al, but several additional values of b were also considered—they are all less successful than leapfrog.

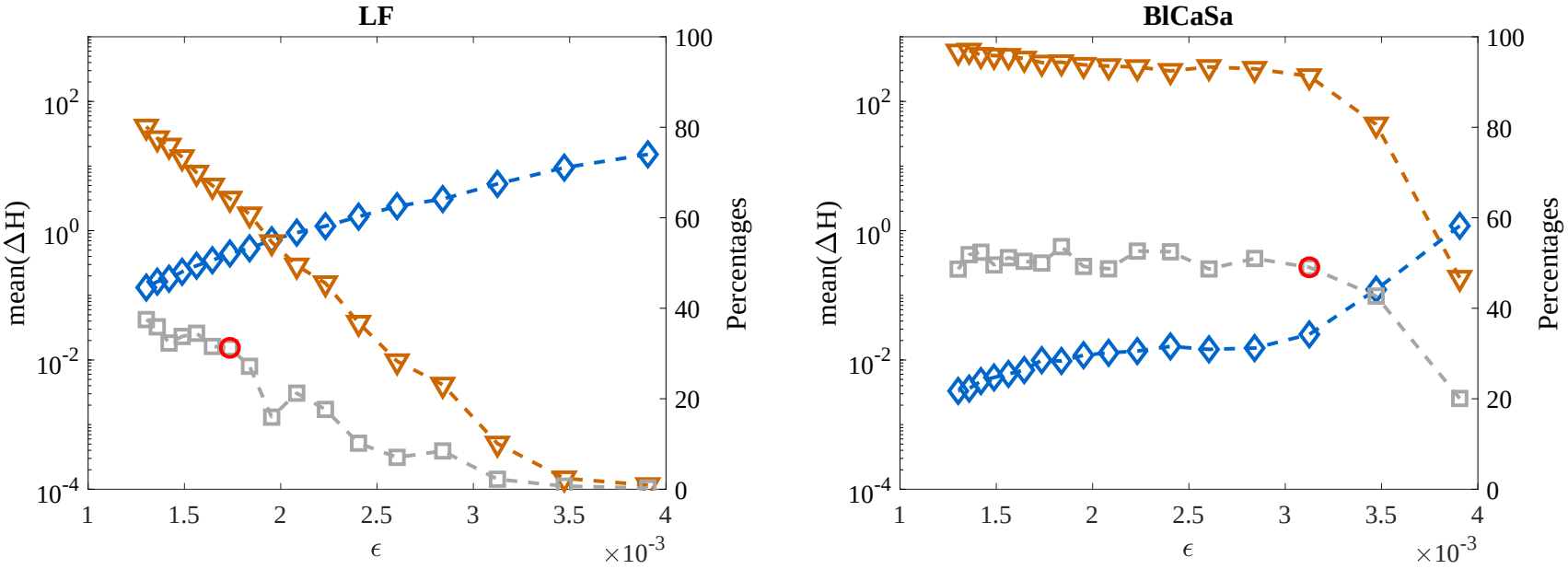


$d = 256$, leapfrog (left) against Blanes et al. Effective sample size of q_1 (squares), average energy error (diamonds) and acceptance rate (triangles) as functions of h.

Other metrics (such as mean square displacement) not reported, as they behave as the effective sample size for q_1. IN ALL EXPERIMENTS, sampling quality is found to be determined by the acceptance rate.

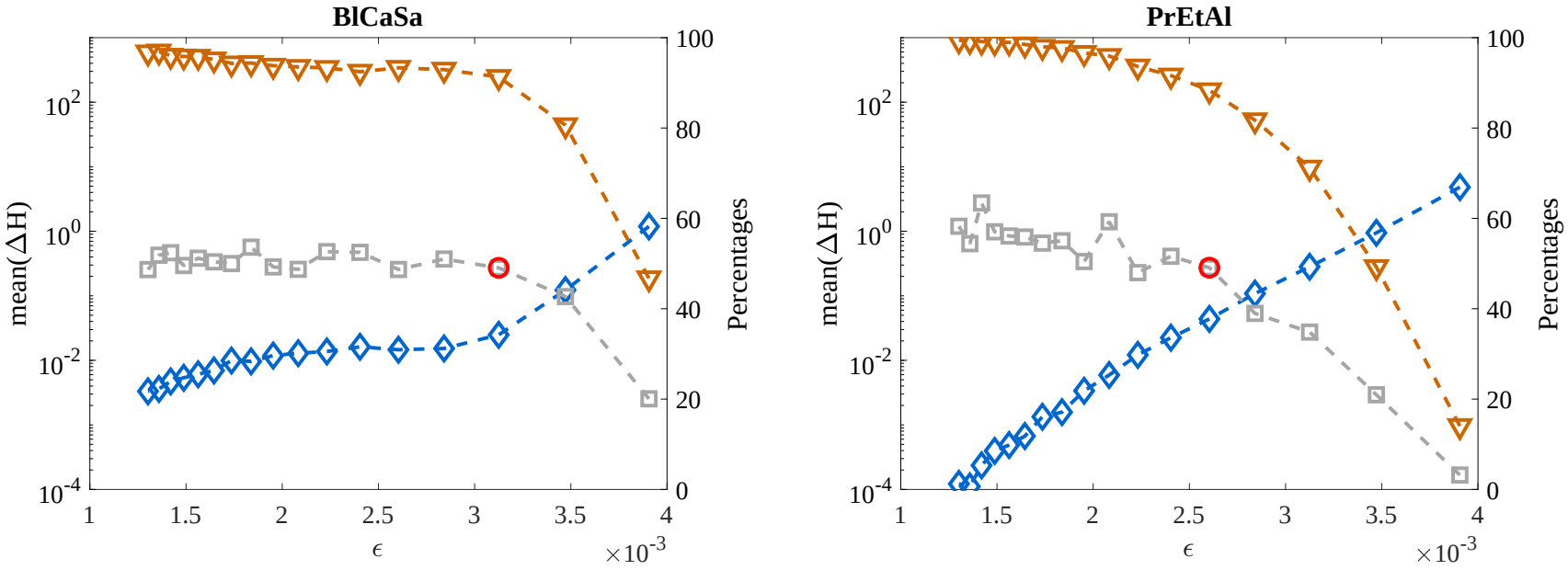


$d = 256$, Blanes et al. (left) against Predescu et al. Effective sample size of q_1 (squares), average energy error (diamonds) and acceptance rate (triangles) as functions of h. For h small, Predescu has extremely small energy errors, but those values of h have to be discarded on efficiency grounds (it would be better to increase h and have cheaper proposals).



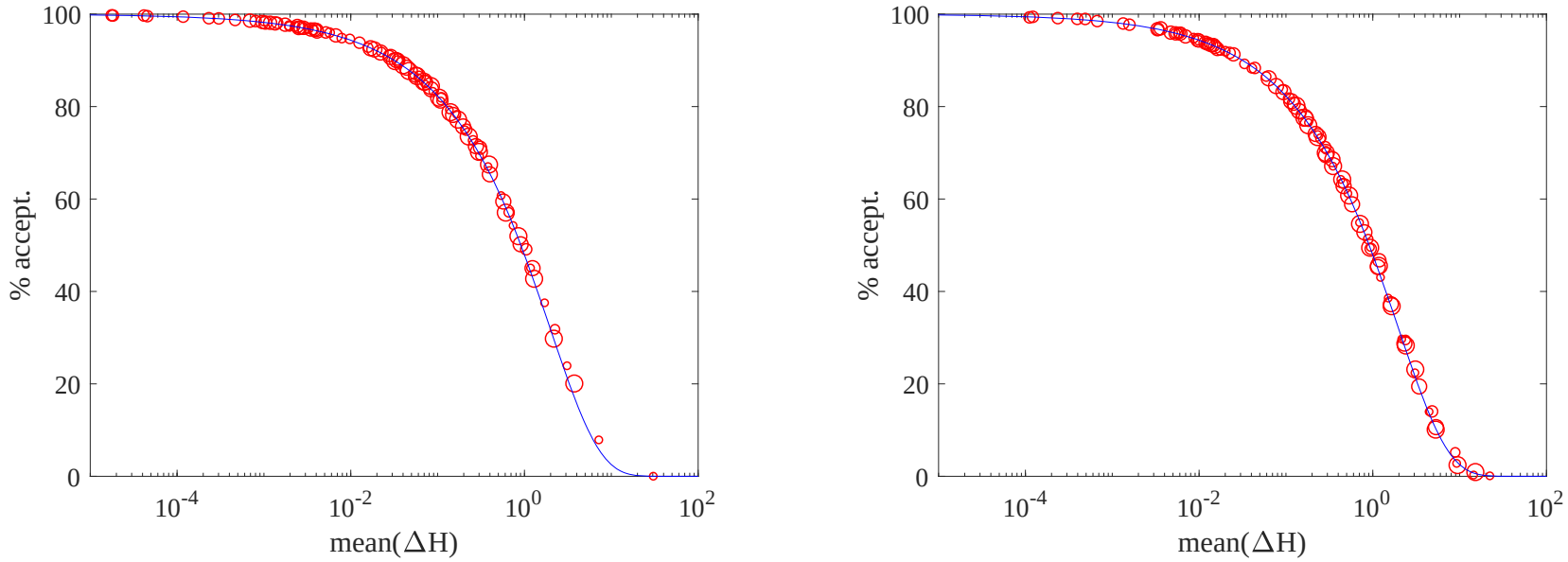
$d = 1024$, leapfrog (left) against Blanes et al.

Effective sample size of q_1 (squares), average energy error (diamonds) and acceptance rate (triangles) as functions of h.

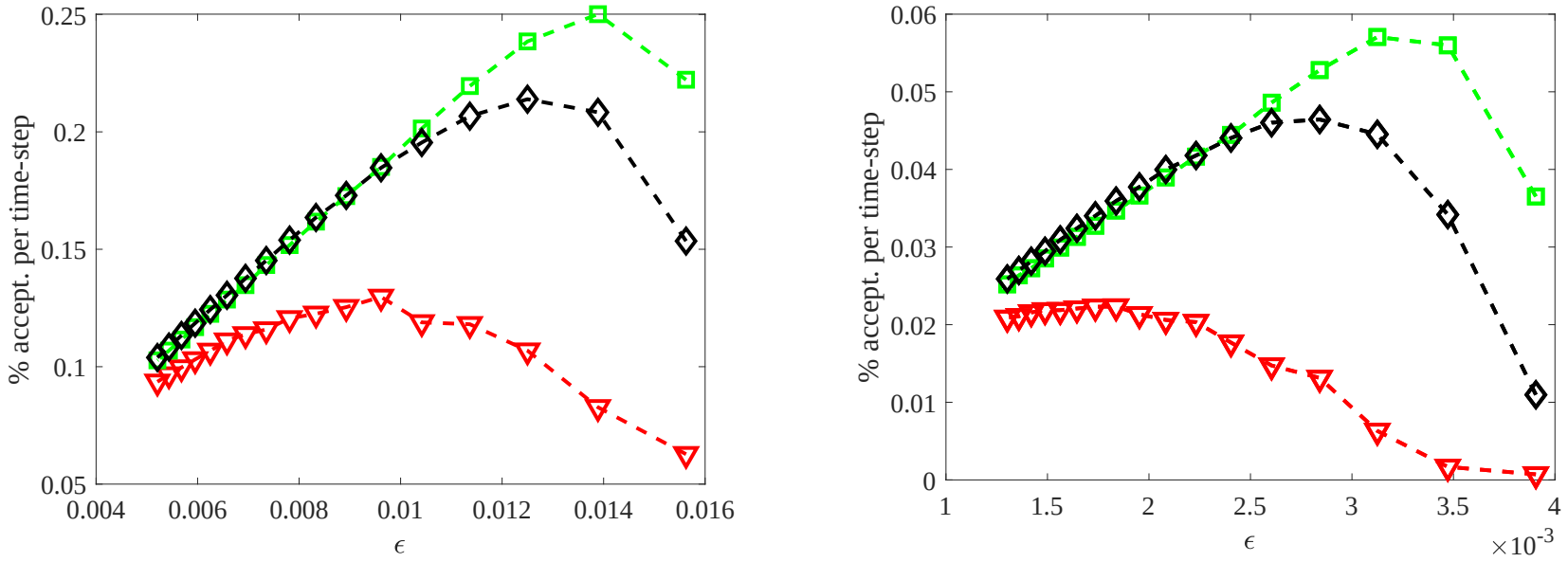


$d = 1024$, Blanes et al. (left) against Predescu et al.

Effective sample size of q_1 (squares), average energy error (diamonds) and acceptance rate (triangles) as functions of h.



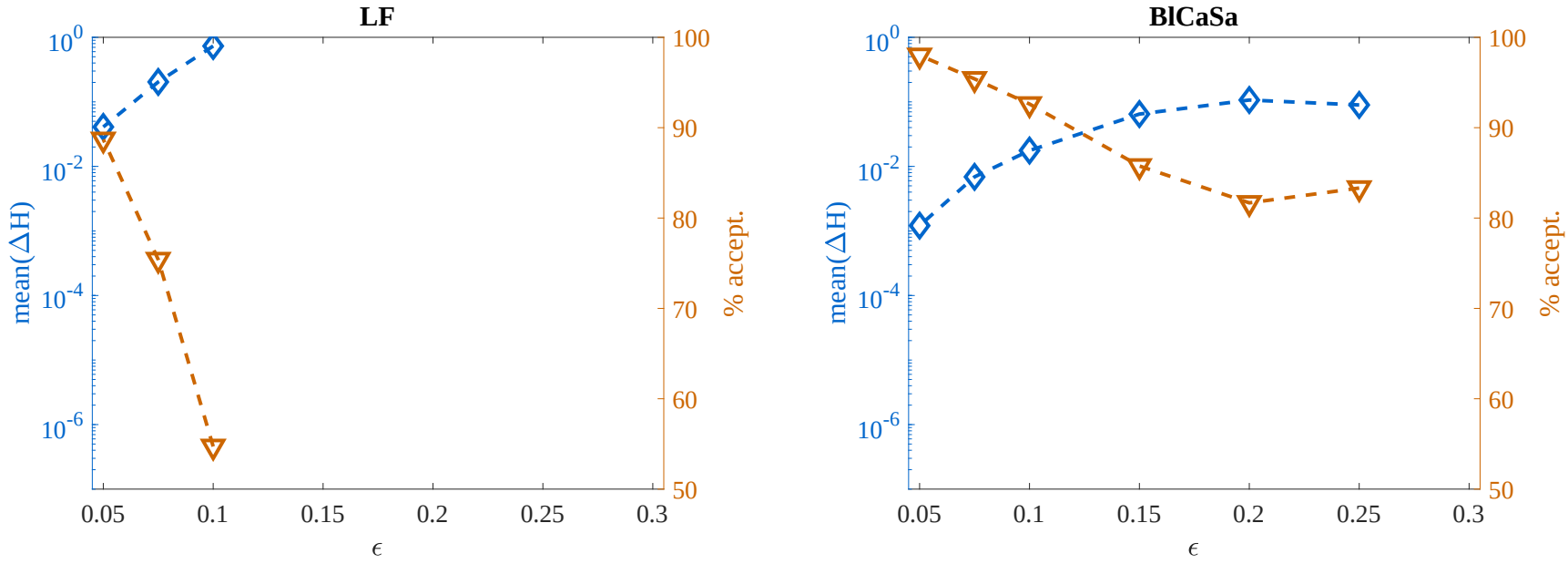
CLT in Calvo et al 2019 at work, $d = 256$ (left), $d = 1024$ (right). Acceptance percentage vs. mean energy error. Points come from six integrators and different step-lengths. The blue line corresponds to the curve in the CLT result $\mathbb{E}(a) = 2\Phi(-\sqrt{\mu/2})$. The acceptance rate is determined by the average energy error.



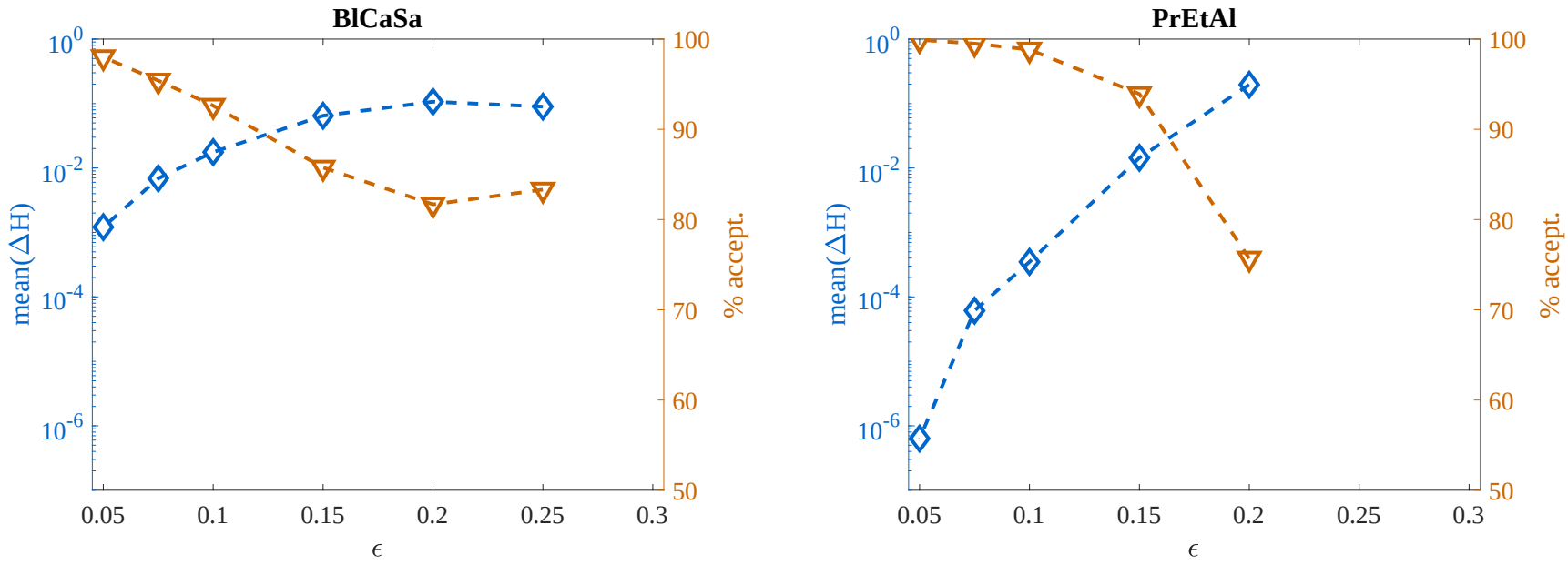
Efficiency compared: Acceptance percentage per unit computational effort for different step-lengths h. $d = 256$ (left), $d = 1024$ (right). Leapfrog (triangles), Blanes (squares), Predescu (diamonds). On the right, we see that, for a given computational budget, Blanes may deliver three times as many accepted proposals as leapfrog.

18: TESTING THE INTEGRATORS: LOG-GAUSSIAN COX PROBLEM (DIMENSION 4096)

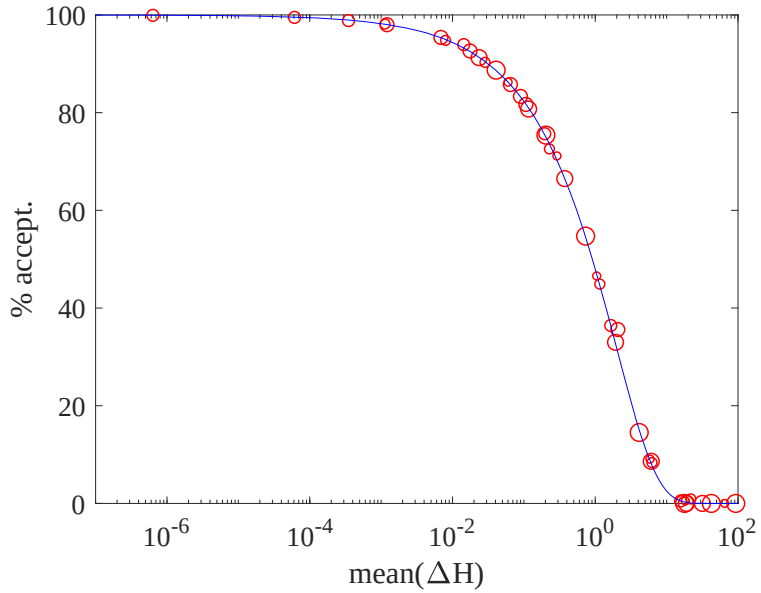
Calvo, Sanz-Alonso & SS 2019



Acceptance percentage (triangles) and the mean of ΔH (diamonds) as functions of the step-length. Leapfrog (left) vs. Blanes.



Acceptance percentage (triangles) and the mean of ΔH (diamonds) as functions of the step-length. Blanes (left) vs. Predescu.



Acceptance percentage vs. energy error. Points come from six integrators and different step-lengths. The continuous line corresponds to the case where ΔH is $\mathcal{N}(\mu, 2\mu)$. The CLT proved for Gaussian targets by Calvo, Sanz-Alonso & SS 2019 appears to hold for non-Gaussian targets.

19: THE AIA (ADAPTIVE INTEGRATOR APPROACH)

Fernández-Pendás, Akhmatskaya & SS 2016

• Blanes et al suggested to choose the integrator that minimizes ρ over the interval $(0, s)$. The upper limit of this interval was chosen because it was experimentally found that leapfrog operates well if $\omega\Delta t < s$.

• In AIA the user fixes a computational budget. This determines the value of Δt to use. And then the code chooses the parameters in the family of splitting integrators by minimizing ρ over the shortest interval that contains all products $\omega_j\Delta t$. (The ω_j are estimated by the code.)

• Hence the integrator changes with the target and with the available computational effort.

• Incorporated in the popular molecular dynamics package GROMACS and also useful out of the HMC context.

Sampling 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 smaller energy error.