HMC and All That

A D Kennedy

Higgs Centre
School of Physics & Astronomy
University of Edinburgh

Field Theoretic Computer Simulations for
Particle Physics and Condensed Matter
Boston University
8–10 May 2014

What does HMC stand for?

A Comparison of two Markov Chain Monte Carlo samplers

Video by Tamara Broderick (Berkeley) and David Duvenaud (Cambridge)
https://www.youtube.com/watch?v=Vv3f0QNWvWQ

Hans Christian Andersen, ***Molecular dynamics simulations at constant pressure and/or temperature***, **J. Chem. Phys.**, **72**, **2384 (1980)**.

P. J. Rossky, J. D. Doll and H. L. Friedman, ***Brownian dynamics as smart Monte Carlo simulation***, **J. Chem. Phys.**, **69**, **4628**, **(1978)**.

- HMC is a Markov Chain Monte Carlo (MCMC) algorithm.
- It is widely used in Lattice Field Theory, Machine Learning, and several other fields.
- It was originally called Hybrid Monte Carlo.
  - Based on the Hybrid algorithm of Duane and Kogut (a hybrid of Molecular Dynamics and a momentum heatbath).
  - Made "exact" by Metropolis Monte Carlo accept/reject step.
- In the statistics and machine learning fields it is called Hamiltonian Monte Carlo, which is probably a more meaningful name.

What is HMC?

- HMC introduces a "fictitious" Hamiltonian $H(q,p) = T(p) + V(q)$ defined on the "positions" (or fields) q and a set of equally fictitious momenta p. The kinetic energy is chosen to be $T = p^2/2$ (changing the coefficient is equivalent to rescaling time).
- It constructs an ergodic Markov Chain with the fixed-point distribution e^{-H} over the phase space (q,p).
- It does this by alternating two steps
  - Selecting new momenta from a Gaussian heatbath, leaving the position unchaged.
  - Approximately integrating Hamilton's equations for a reasonably long time τ using an integrator that is exactly reversible and area-preserving (such as leapfrog), and then performing a Metropolis accept/reject step.
  - Both steps have e^{-H} as their fixed point distribution, but they are certainly not ergodic individually.

Cost and Tuning

- The fictitious momenta have nothing to do with any physical momenta in the system. People sometimes try to use the "real" Hamiltonian dynamics instead, but it is not clear why this should be advantageous.
- In free field theory the optimal trajectory length is of the order of the correlation length of the field theory under study. Both theoretical analysis and empirical measurements indicate that using shorter trajectories is a false economy.
- When the Hamiltonian is extensive, so using an integrator which has errors of $\mathcal{O}(\delta\tau^n)$ requires $V\delta\tau^n$ to be held constant. The volume scaling of the cost is therefore $V^{1+1/n}$ at fixed trajectory length = correlation length. For leapfrog $n = 4$, so the cost grows as $V^{5/4}$ independent of the dimension of the lattice.
- For large volumes using higher order symplectic integrators becomes worthwhile, but this makes the parameter-tuning problem harder.

What is HMC good for?

- What problems doesn't it solve?
  - Sign problems: i.e., when there are huge cancellations between different regions of the integration region.
  - Such sign problems can be solved by reformulating the theory, but this is very model-dependent (no general solution).
  - There are interesting results using the complex Langevin algorithm (but I do not really understand why).
  - Is there a suitable algorithm for a Quantum Computer?
- What problems does it solve?
  - It reduces autocorrelations.
  - There is a large reduction in cost for interacting theories compared with Langevin or other small step or local update methods.
  - For free field theory it reduces the dynamical critical exponent z from 2 to 1.
  - For interacting theories this is also true in practice, even if perhaps not in theory (more later...).

Fermions

- In principle we could keep track of the sign and explicitly integrate over Grassmann-valued fields. The problem is that the resulting integral has a "sign problem".
- For renormalizable field theories (without too much improvement) fermion fields only occur quadratically in the action, and can be integrated to give the determinant of the discrete Dirac operator.

$$\langle\Omega\rangle \propto \int dU\, d\psi\, d\bar{\psi}\, \Omega(\bar{\psi},\psi,U) e^{-S(U)-\bar{\psi}\mathcal{M}(U)\psi} \propto \int dU\, \det\mathcal{M}(U)\, \bar{\Omega}(U) e^{-S(U)}$$

with

$$\bar{\Omega}(U) = \left[\Omega\left(\frac{\delta}{\delta\eta},\frac{\delta}{\delta\bar{\eta}},U\right) e^{\bar{\eta}\mathcal{M}^{-1}(U)\eta}\right]_{\bar{\eta}=\eta=0},$$

where the fermion kernel $\mathcal{M} = \not{D} + m$ (or $(\not{D}+m)^\dagger(\not{D}+m)$ in practice).

- The determinant is real and positive, $\det\mathcal{M} = \det\gamma_5\mathcal{M}\gamma_5 = \det\mathcal{M}^\dagger$.

Pseudofermions

- We can replace the determinant with an integral over pseudofermion fields with kernel $\mathcal{M}^{-1}$,

$$\langle\Omega\rangle \propto \int dU\, d\bar{\phi}\, d\phi\, \bar{\Omega}(U) e^{-S(U)-\bar{\phi}\mathcal{M}^{-1}(U)\phi}.$$

- For heavy fermions this works well, but for light fermions the step size $\delta\tau \to 0$ in order to accept anything.
- For a long time this was thought to be due to large fermionic forces coming from "exceptional" gauge field backgrounds ("blame the usual suspects").

Integrator Instability

- But this was not the case. The culprit was that we are using a single pseudofermion field to estimate the fermionic force, and it is therefore very noisy.
- The Markov process still has the correct fixed point distribution even when the integrator becomes unstable, but the autocorrelation becomes zero because nothing is accepted.
- The fundamental problem is that the large fluctuations in the estimator of the fermionic force cause the gauge field integrator to become (exponentially) unstable.
- The solution is to use several pseudofermions to estimate the fermionic force.
- Ideally, each of the n pseudofermions should each contribute contribute $\det \mathcal{M}^{1/n}$ to the functional integral.

Multiple Pseudofermions

- There are several ways of doings this
  - 
    Hasenbusch: several pseudofermions with larger masses. Several cheap heavier pseudofermions plus a more costly correction term to ensure the correct determinant.
  - 
    Lüscher: domain decomposition. Partition lattice into even and odd blocks together with a "small" residual correction, each with their own pseudofermions. For light enough fermions the residual part will domainate.
  - Rational HMC (RHMC): use rational approximation for $\mathcal{M}^{1/n}$. Requires the solution of a large linear system for each pole in the rational approximation for each pseudofermion.
- In practice some combination of the techniques usually works best.

RHMC

- RHMC makes use of multishift Krylov space solvers.
- It is remarkable that all the coefficients in the Chebyshev optimal rational approximation $x^{1/n} \approx \sum_j \frac{\alpha_j}{x+\beta_j}$ are positive, so the method is numerically stable.
- Usually only two or three pseudofermions are required, and 10 to 20 poles are required to achieve full floating-point accuracy over the spectrum of $\mathcal{M}$.
- Using these simple tricks the integrator step size can be increased by about an order of magnitude before the fermionic force (rather than the noise of its estimator) causes the instability to be triggered.
- This has allowed lattice QCD computations to be carried out at the physical π mass, thus eliminating the need for chiral perturbation theory extrapolation in the mass.

"I expect chiral perturbation theory to apply below the pion mass"

— Jürg Gasser

Symplectic Structure

- Hamiltonian systems have an innate symplectic symmetry.
  - Consider the exterior derivative of the Lagrangian

$$dL(q,\dot{q}) = \frac{\partial L}{\partial q}dq + \frac{\partial L}{\partial \dot{q}}d\dot{q} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\,dq\right) = \frac{d}{dt}(p\,dq)$$

    using the Euler–Lagrange equations and defining $p \equiv \partial L/\partial\dot{q}$.
  - Hence $d^2L = 0$ requires that the fundamental 2-form $\omega \equiv -d(p\,dq) = dq \wedge dp$ is time independent (Darboux theorem). It must also be closed, $d\omega = 0$.
- This allows us to associate a vector field $\widehat{A}$ with any 0-form (function) A on phase space, $dA(X) = \omega(\widehat{A}, X)$ for any vector field X.
- $\widehat{A}$ is called a Hamiltonian vector field, even if A is not the Hamiltonian.
- Classical trajectories are integral curves of $\widehat{H}$, where H is the Hamiltonian.

Poisson Brackets

- The definition of the exterior derivative of a 0-form is $dA(X) = X\,A$, hence $\boxed{\widehat{A}\,B = dB(\widehat{A}) = \omega(\widehat{B},\widehat{A}) \equiv \{A,B\}}$ where $\{A,B\}$ is the Poisson bracket of two phase space functions. Clearly $\{A,B\} = -\{B,A\}$.
- According to Jacobi (1840)

  > *«. . . quelques remarques sur la plus profonde découverte de M. Poisson, mais qui, je crois, n'a pas été bien comprise ni par Lagrange, ni par les nombreux géomètres qui l'ont citée, ni par son auteur lui-même.»*

- For non-Francophones we may thank Google for the following translation

  > *". . . a few remarks on the most profound discovery of Mr. Fish, which, I think, was not well understood by either Lagrange or by many others who have cited it, nor by the author himself."*

Poisson Brackets

- The definition of the exterior derivative of a 0-form is $dA(X) = X\,A$, hence $\boxed{\widehat{A}\,B = dB(\widehat{A}) = \omega(\widehat{B},\widehat{A}) \equiv \{A,B\}}$ where $\{A,B\}$ is the Poisson bracket of two phase space functions. Clearly $\{A,B\} = -\{B,A\}$.
- According to Jacobi (1840)

  > *«. . . quelques remarques sur la plus profonde découverte de M. Poisson, mais qui, je crois, n'a pas été bien comprise ni par Lagrange, ni par les nombreux géomètres qui l'ont citée, ni par son auteur lui-même.»*

- For non-Francophones we may thank Google for the following translation

  > *". . . a few remarks on the most profound discovery of Mr. Fish, which, I think, was not well understood by either Lagrange or by many others who have cited it, nor by the author himself."*

Jacobi Identity

- The definition of the exterior derivative of a 2-form is

$$\begin{aligned} d\omega(X,Y,Z) &= X\omega(Y,Z) + Y\omega(Z,X) + Z\omega(Y,X) \\ &\quad -\omega([X,Y],Z) - \omega([Y,Z],X) - \omega([Z,X],Y) \end{aligned}$$

  where $[X,Y]$, the commutator of two vector fields, is iteself a vector field.
- For Hamiltonian vector fields $\widehat{A}\omega(\widehat{B},\widehat{C}) = -\widehat{A}\{B,C\} = -\{A,\{B,C\}\}$.
- Moreover, $\omega([\widehat{A},\widehat{B}],\widehat{C}) = -\omega(\widehat{C},[\widehat{A},\widehat{B}]) = -dC([\widehat{A},\widehat{B}]) = -(\widehat{A}\widehat{B}-\widehat{B}\widehat{A})C = -\widehat{A}\{B,C\} + \widehat{B}\{A,C\} = -\{A,\{B,C\}\} + \{B,\{A,C\}\}$.
- Combining these we find that Poisson brackets satisfy the Jacobi identity

$$\boxed{\{A,\{B,C\}\} + \{B,\{C,A\}\} + \{C,\{A,B\}\} = 0.}$$

- Hence functions on phase space form a Lie algebra with the Poisson bracket as the Lie product.

BCH Formula

- Using the Jacobi identity we find that $\left[\widehat{A},\widehat{B}\right]C = \left\{\widehat{A},\left\{\widehat{B},\widehat{C}\right\}\right\} - \left\{\widehat{B},\left\{\widehat{A},\widehat{C}\right\}\right\} = \left\{\left\{\widehat{A},\widehat{B}\right\},\widehat{C}\right\} = \widehat{\left\{A,B\right\}}C$ for all C, so we get the remarkable result that commutator of two Hamiltonian vector fields is a Hamiltonian vector field: $\boxed{\left[\widehat{A},\widehat{B}\right] = \widehat{\left\{A,B\right\}}.}$
- The Baker–Campbell–Hausdorff (BCH) formula (due to F. Schur, according to Bourbaki) states that $e^A e^B = e^C$ where $C = \sum_{n\geq 1} C_n$ with $C_1 = A + B$ and

$$(\mathrm{n}+1)\,\mathrm{C}_{\mathrm{n}+1} = -\,\mathrm{ad}\,\mathrm{C}_{\mathrm{n}}(\mathrm{A}-\mathrm{B}) + \sum_{\mathrm{m}=0}^{\lfloor \mathrm{n}/2\rfloor} \frac{\mathrm{B}_{2\mathrm{m}}}{(2\mathrm{m})!} \sum_{\substack{\mathrm{k}_1,\ldots,\mathrm{k}_{2\mathrm{m}}\geq 1 \\ \mathrm{k}_1+\cdots+\mathrm{k}_{2\mathrm{m}}=\mathrm{n}}} \mathrm{ad}\,\mathrm{C}_{\mathrm{k}_1}\ldots\mathrm{ad}\,\mathrm{C}_{\mathrm{k}_{2\mathrm{m}}}(\mathrm{A}+\mathrm{B}),$$

  where B_{n} are Bernoulli numbers and $(\mathrm{ad}\,\mathrm{X})\,\mathrm{Y} = \left[\mathrm{X},\mathrm{Y}\right]$.
- For Hamiltonian vector fields this gives $\mathrm{e}^{\widehat{\mathrm{A}}}\mathrm{e}^{\widehat{\mathrm{B}}} = \mathrm{e}^{\widehat{\mathrm{C}}}$ where $\mathrm{C} = \sum_{\mathrm{n}\geq 1} \mathrm{C}_{\mathrm{n}}$ where the C_{n} are 0-forms given by the same formula, but this time for functions rather than vector fields and with $(\mathrm{ad}\,\mathrm{A})\,\mathrm{B} = \left\{\mathrm{A},\mathrm{B}\right\}$.

Integrators

- Remember that classical trajectories were integral curves of the Hamiltonian Hamiltonian vector field. For $\mathrm{H}(\mathrm{q},\mathrm{p}) = \mathrm{T}(\mathrm{p}) + \mathrm{V}(\mathrm{q})$ (for particle physics $\mathrm{V} = \mathrm{S}$) we have, at least locally where $\omega = \mathrm{dq} \wedge \mathrm{dp}$, $\widehat{\mathrm{H}} = \widehat{\mathrm{T}} + \widehat{\mathrm{S}}$ where $\widehat{\mathrm{T}} = \mathrm{T}'(\mathrm{p})\,\frac{\partial}{\partial \mathrm{q}}$ and $\widehat{\mathrm{V}} = -\mathrm{V}'(\mathrm{q})\,\frac{\partial}{\partial \mathrm{p}}$.
- We cannot integrate $\widehat{\mathrm{H}}$ exactly, but We can construct an exact discrete integrator for $\widehat{\mathrm{T}}$

$$\mathrm{e}^{\delta\tau\widehat{\mathrm{T}}}\mathrm{f}(\mathrm{q},\mathrm{p}) = \sum_{\mathrm{n}\geq 0} \frac{\delta\tau^{\mathrm{n}}\mathrm{T}'(\mathrm{p})^{\mathrm{n}}}{\mathrm{n}!}\frac{\partial^{\mathrm{n}}\mathrm{f}}{\partial \mathrm{q}^{\mathrm{n}}}(\mathrm{q},\mathrm{p}) = \mathrm{f}\Big(\mathrm{q} + \mathrm{T}'(\mathrm{p})\delta\tau, \mathrm{p}\Big)$$

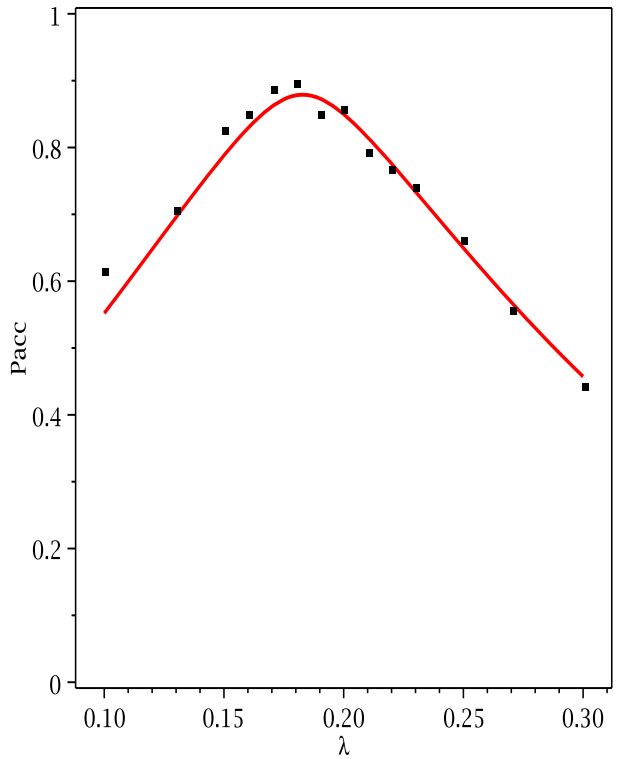
  using Taylor's theorem, and likewise for $\mathrm{e}^{\delta\tau\widehat{\mathrm{V}}}\mathrm{f}(\mathrm{q},\mathrm{p}) = \mathrm{f}\Big(\mathrm{q}, \mathrm{p} - \mathrm{V}'(\mathrm{q})\delta\tau\Big)$.
- This only converges for small enough $\delta\tau$, of course.
- In fact it is really only an asymptotic expansion: if we consider the evolution of the probability distribution $\mathrm{e}^{-\mathrm{H}(\mathrm{q},\mathrm{p})}$ of phase space points, then q or p may be arbitrarily large with exponentially small probability.

Shadow Hamiltonians

- Using the BCH formula we may construct integrators for $\widehat{\mathrm{H}}$ to any desired order in the step size $\delta\tau$.
- What is perhaps more surprising is that this integrator, while it only approximately conserves H, exactly conserves a nearby Shadow Hamiltonian $\tilde{\mathrm{H}}$, namely that built out of Poisson brackets using the BCH formula.
- The first application of this result is that HMC trajectories conserve energy to $\mathcal{O}(\delta\tau^{\mathrm{n}})$ for arbitrarily long trajectory lengths, because $\tilde{\mathrm{H}}(\mathrm{q},\mathrm{p}) - \mathrm{H}(\mathrm{q},\mathrm{p}) = \mathcal{O}(\delta\tau^{\mathrm{n}})$ everywhere in phase space.

Tuning HMC

- Our second application is that we can use the Shadow to tune the parameters of our symplectic integrator.
- The idea is that the Poisson brackets occuring in the BCH formula are extensive quantities, and therefore average to constants.
- We therefore perform one HMC run and measure all the relevant Poisson brackets, we then minimize the variance of the distribution of $\mathrm{H} - \tilde{\mathrm{H}}$ by adjusting the integrator parameters. This is a simple task that is computationally cheap.
- A formalism for constructing Poisson brackets for gauge fields and pseudofermions has been developed.



Force-Gradient Integrators

- Since the kinetic energy is just $\mathrm{p}^2/2$ some of the Poisson brackets, such as $\{\mathrm{V},\{\mathrm{V},\mathrm{T}\}\}$ are momentum-independent.
- We may therefore integrate the corresponding Hamiltonian vector field $\widehat{\{\mathrm{V},\{\mathrm{V},\mathrm{T}\}\}}$ exactly, giving us another integrator step to play with.
- Using such Force-Gradient integrators have been used to speed up state-of-the-art lattice computations by a factor of about three.
- Improved integrators will become increasingly important as we are able to study larger lattices.

Non-Renormalizability

- The continuum and continuous time limit of the the Langevin algorithm may reformulated as a stochastic field theory. (Zwanziger, Zinn-Justin)
- This allows a theoretical estimate of the scaling behaviour with respect to the lattice spacing a.
- Lüscher and Schaefer have done the same for he generalized HMC algorithm.
  - The "generalized" HMC algorithm (due to Horowitz) mixes a small amount of Gaussian noise at every step, rather than between long trajectories.
  - In practice this does not help because it has to use a tiny rejection rate to avoid oscillating back and forth along its trajectory.
  - This makes no difference in the continuous time limit, however.
- They find that the resulting stochastic field theory is not perturbatively renormalizable, because it cannot be rotated to Euclidian space.
- It is not entirely clear what this signifies, but they conjecture that it may mean that HMC is in the same universality class as the Langevin algorithm, and thus has a dynamical critical exponent of $\mathrm{z} = 2$.

Conclusions

- HMC and varients thereof are the method of choice for almost all lattice field theory computations.
- HMC is widely used in several fields outside of physics.
- HMC works well for systems with light fermions.
- It does not solve the sign problem.
- It does not work for complex actions (at least to the extent that they have a serious sign problem).
- The use of Shadow Hamiltonians can help automate the tuning of integrator parameters, but the method is not yet widely used.