Introduction to the HMC

KAPCTEH YPbAX

Helmholtz-Institut für Strahlen- und Kernphysik
Universität Bonn

Helmholtz-School Dubna 2011

Outline

Lectures and Tutorials

1. Lectures: introduce the Theory
   - basic HMC algorithm and Schwinger model
   - algorithm improvements
   - recent developments
2. Tutorials with Pavel Buividovic: you can practice
   - example: Schwinger model
   - template code provided
   - online tutorial with step-by-step instructions
   - based on the lecture

Motivation

- Lattice QCD: solve high dimensional integral

$$\mathcal{Z}_{\rm QCD} = \int \mathcal{D}U\, \mathcal{D}\bar{\psi}\, \mathcal{D}\psi\, e^{-S_G[U] - \bar{\psi}\ D[U]\ \psi} \propto \int \mathcal{D}U\ \det(D[U]) e^{-S_G[U]}$$

- determinant can be represented by bosonic fields:

$$\det(D) \propto \int \mathcal{D}\phi^\dagger\ \mathcal{D}\phi\ e^{-\phi^\dagger\ D^{-1}\ \phi}$$

ϕ fields also called pseudo-fermion fields

- can deal with $D^{-1}\ \phi$, but: non-local

Markov-Chain Monte Carlo

- stochastic method to solve the generic integral

$$\langle\, O \,\rangle = \int \mathcal{D}x \; O(x)\, e^{-S(x)}$$

- by generating a Markov-Chain $\{x_1, x_2, \ldots\}$ distributed as

$$e^{-S(x)}$$

- then

$$\langle\, O \,\rangle \approx \frac{1}{N}\sum_{i=1}^{N} O(x_i)$$

  with statistical error:

$$\delta\, O \propto 1/\sqrt{N}$$

- how to generate such a chain $\{x_1, x_2, \ldots\}$?

Metropolis Monte-Carlo algorithm

1. start with arbitrary x
2. chose a test x' with probability $P(x')$
   $P(x) > 0 \; \forall x$
3. accept x' with probability

$$P_A(x \to x') = \min\{1, \exp[-\Delta S = -(S(x') - S(x))]\}$$

4. continue with step 2

Fulfils detailed balance condition (easy exercise)

$$\exp(-S(x))P(x \to x') = \exp(-S(x'))P(x' \to x)$$

Metropolis Monte Carlo

- how to generate the proposal x'?
  1. chose x' randomly
     completely uncorrelated to previous x

  $\Rightarrow$ expect large $\Delta S \Rightarrow$ low acceptance $\Rightarrow$ large autocorrelation

  2. use $x' = x + \delta x$ with random but small δx
     δx can be tuned for ΔS to be small

  $\Rightarrow$ large autocorrelation

- if computation of ΔS is very expensive (like for QCD)
  both choices turn out to be not feasible

- desired: a global update combined with large acceptance

The Hamiltonian Monte Carlo (Hybrid Monte Carlo)

[Duane, Kennedy, Pendleton, Roweth, 1987]

- Introduce p_i conjugate to fundamental fields x_i and a Hamiltonian

$$\mathcal{H} = \frac{1}{2}\sum_i p_i^2 + S(x)$$

- $\mathcal{H}$ is conserved under Hamilton's EoM

$$\dot{x}_i = \frac{\partial \mathcal{H}}{\partial p_i} = p_i \,, \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial x_i} = -\frac{\partial \mathcal{S}}{\partial x_i}$$

$\Rightarrow$ use Hamilton's EoM for global update (molucular dynamics):

$$(p, x) \quad \rightarrow \quad (p', x')$$

- Accept with probability

$$P_A(\mathcal{H} \rightarrow \mathcal{H}') = \min\{1, \exp(\mathcal{H}(p, x) - \mathcal{H}(p', x')\}$$

- Energy conservation guarantees large acceptance!

Detailed Balance for HMC

- Need to proof detailed balance

$$e^{-S(x)}P(x \to x') = e^{-S(x')}P(x' \to x)$$

- $P(x \to x')$ is a convolution of

$$P(x \to x') = \int \mathcal{D}p\, \mathcal{D}p'\, P_{\mathrm{G}}(p)\, P_{\mathrm{MD}}[(x,p) \to (x',p')] P_A(\mathcal{H} \to \mathcal{H}')$$

with (x',p') fixed given (x,p) and

$$P_{\mathrm{G}}(p) = \exp\left\{-\sum_i p_i^2\right\}, \qquad P_{\mathrm{G}}(p)\, e^{-S(x)} = e^{-\mathcal{H}(x,p)}$$

- we require molecular dynamics (MD) integration to be reversible

$$P_{\mathrm{MD}}[(x,p) \to (x',p')] = P_{\mathrm{MD}}[(x',-p') \to (x,-p)]$$

Detailed Balance for HMC

- $\mathcal{H}$ is quadratic in p

$$\mathcal{H}(x,p) = \mathcal{H}(x,-p)$$

- and we have the identity

$$\begin{aligned}
\exp(-\mathcal{H})P_A[(x,p) \to (x',p')] = \exp(-\mathcal{H})\min\{1, \exp(\mathcal{H} - \mathcal{H}')\} \\
= \min\{\exp(-\mathcal{H}), \exp(-\mathcal{H}')\} \\
= \exp(-\mathcal{H}')\min\{\exp(\mathcal{H}' - \mathcal{H}), 1\} \\
= \exp(-\mathcal{H}')P_A[(x',p') \to (x,p)]
\end{aligned}$$

Detailed Balance for HMC

- using all these we obtain

$$
\begin{aligned}
e^{-S(x)}\, P(x \to x') = \\
&= \int \mathcal{D}p\, \mathcal{D}p'\, e^{-\mathcal{H}(x,p)}\, P_{\mathrm{MD}}[(x,p) \to (x',p')] P_A(\mathcal{H} \to \mathcal{H}') \\
&= \int \mathcal{D}p\, \mathcal{D}p'\, e^{-\mathcal{H}(x',-p')}\, P_{\mathrm{MD}}[(x',-p') \to (x,-p)] \times \\
&\quad \times P_A(\mathcal{H}(x',-p') \to \mathcal{H}(x,-p))
\end{aligned}
$$

- change of variables $-p' \to p'$ and $-p \to p$

$$
\begin{aligned}
e^{-S(x)}\, P(x \to x') = \int \mathcal{D}p\, \mathcal{D}p'\, e^{-\mathcal{H}(x',p')}\, P_{\mathrm{MD}}[(x',p') \to (x,p)] \times \\
\times P_A(\mathcal{H}' \to \mathcal{H}) \\
= e^{-S(x')}\, P(x' \to x) \qquad \text{q.e.d.}
\end{aligned}
$$

Detailed Balance for HMC

from the proof one learns

- MD must be reversible
- measure must invariant

$$\mathcal{D}p \times \mathcal{D}p' = \mathcal{D}(-p) \times \mathcal{D}(-p')$$

  (area preserving)
- if $\mathcal{H}$ is conserved, the $P_A = 1$

in practice we use

- a numerical integration scheme
- accept/reject step corrects for discretisation errors

$\Rightarrow$ need to find a reversible and area preserving integration scheme

Symplectic Integrators

- by linking together x and p in $z = (x, p)$ we can write

$$\dot{z} = \mathbf{J} \cdot \frac{\partial \mathcal{H}(z)}{\partial z}$$

- with symplectic matrix

$$\mathbf{J} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

- symplectic mans intertwined (see **J**)

- time evolution $z(t_0) \to z(t)$ represents a canonical transformation $\mathbf{A}(t_0, t)$

$$z(t) = \mathbf{A} \cdot z(t_0)$$

Symplectic Integrators

- such a transformation conserves the energy
- but the symplectic form

$$s(z_1, z_2) \equiv z_1^T \, \mathbf{J} \, z_2$$

  is conserved under this mapping

- geometrically:
  the area of the parallelogram spanned by $z_{1,2}$ is preserved
- for the harmonic oscillator you can easily show

$$z_1(t_0)^T \, \mathbf{J} \, z_2(t_0) = z_1(t)^T \, \mathbf{J} \, z_2(t)$$

  by writing down the mapping **A**.

$\Rightarrow$ s is conserved if $\mathbf{A}^T \mathbf{J} \mathbf{A} = \mathbf{J}$

Symplectic Integrators

- is this useful for a numerical integration scheme?
- yes! (surprise) one can show:
  symplectic integrators do conserve a Hamiltonian $\mathcal{H}_s$ different from, but close to the given Hamiltonian $\mathcal{H}$

$\Rightarrow$ consequence: $\Delta\mathcal{H} = \mathcal{H}_s - \mathcal{H}$ depends only on step size $\Delta\tau$, not on the length of the integration

- simplest example and exercise for you:

$$x_{n+1} = x_n + \Delta\tau\, p_n \qquad p_{n+1} = p_n - \Delta\tau \frac{\partial\mathcal{H}}{\partial x_{n+1}}$$

is symplectic and conserves for the harmonic oscillator

$$\mathcal{H}_s = p^2/2 + x^2/2 + \Delta\tau\, p\, x/2$$

exactly!

Symplectic Integrators

- however, the simple example is not reversible
- but the leap-frog integration scheme
- Discrete updates for time step $\Delta\tau$

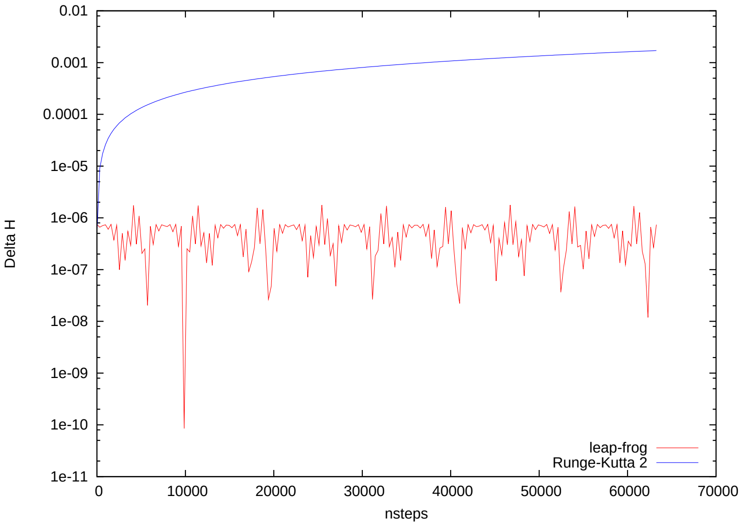
$$
\begin{aligned}
T_{\mathrm{x}}(\Delta\tau): \quad x &\quad \longrightarrow \quad x' = x + \Delta\tau p \\
T_{\mathrm{p}}(\Delta\tau): \quad p &\quad \longrightarrow \quad p' = p - \Delta\tau \frac{\partial \mathcal{H}}{\partial x}
\end{aligned}
$$

- basic Leap Frog time evolution step

$$
T = T_{\mathrm{p}}(\Delta\tau/2)\, T_{\mathrm{x}}(\Delta\tau)\, T_{\mathrm{p}}(\Delta\tau/2)
$$

- trajectory of length τ: $N_{\mathrm{MD}} = \tau/\Delta\tau$ successive applications of T
- $\Delta\mathcal{H}$ independent of τ!

Symplectic Integrators



Integration Errors

- how does $\Delta\mathcal{H}$ scale with $\Delta\tau$?
- introduce time evolution operator $\exp\{\Delta\tau\hat{\mathcal{H}}\}$ with

$$\hat{\mathcal{H}}\, f(p,q) \equiv -\{\mathcal{H}, f\} = \frac{\partial\mathcal{H}}{\partial p}\frac{\partial f}{\partial x} - \frac{\partial\mathcal{H}}{\partial x}\frac{\partial f}{\partial p}$$

- write $\mathcal{H} = T(p) + S(x)$
- the leap-frog scheme has time evolution

$$e^{\Delta\tau/2\,\hat{S}}\, e^{\Delta\tau\,\hat{T}}\, e^{\Delta\tau/2\,\hat{S}} =$$
$$= \exp\{\Delta\tau(\hat{\mathcal{H}} + \Delta\tau^2([[\hat{S},\hat{T}],\hat{S}] + [[\hat{S},\hat{T}],\hat{T}]) + \mathcal{O}(\Delta\tau^3)\}$$

using the Baker-Campbell-Hausdorff formula

$\Rightarrow\ \Delta\mathcal{H} = \mathcal{O}(\Delta\tau^2)$

Summary basic HMC algorithm

1. generate momenta p_i randomly from Gaussian distribution

$$P \sim e^{-p^2/2}$$

   and compute initial Hamiltonian $\mathcal{H}$.

2. Integrate the equations of motion

$$\dot{x}_i = \frac{\partial \mathcal{H}}{\partial p_i} = p_i \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial x_i} = -\frac{\partial S}{\partial x_i} \quad \forall i$$

   by means of the leap-frog integration scheme

3. the Hamiltonian is conserved up to $\mathcal{O}(\Delta\tau^2)$

4. compute final Hamiltonian $\mathcal{H}'$ and accept/reject

$$P_A = \min\{1, \exp(-\Delta\mathcal{H})\}$$

   to correct for discretisation errors

Some Diagnostics

Things one can use to test an implementation

- if you get acceptance something must be correct
  unless $\Delta\tau$ too small
- check that $\Delta\mathcal{H}$ scales with $\Delta\tau^2$
- perform a reversibility test
  by integrating forward and backward (reverse time)
- one can show

$$\langle \exp(-\Delta\mathcal{H}) \rangle = 1$$

  useful to check

Some Diagnostics

- when to start measuring?
  - $N \longrightarrow \infty$ is not possible

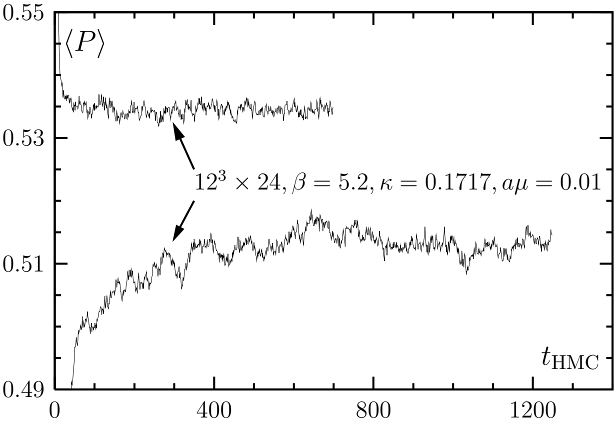
  $\Rightarrow$ we have to equilibrate N_{therm} updates
  - there is no sound theoretical tool for N_{therm}
  - N_{therm} is different for different observables!
- start from several initial configurations until they merge

$\Rightarrow$ expensive

- monitor the moving average until it does not change
- monitor history

Bad Example

This is not just in theory...! Take care!



[Farchioni et. al, Eur.Phys.J. C39 (2005)]

Schwinger Model

- our model for the tutorials:
  QED in 2 dimensions with $N_f = 2$ dynamical fermions

- we use a two-dimensional lattice with extend $L_x \times L_t$

- label the sites with $n = t * L_x + x$

- we use periodic boundary conditions for fermion and gauge fields in both directions (for simplicity only)

$\Rightarrow$ the fermionic fields should have anti-periodic b.c.

- the link variables $U_{n,\mu}$ connect sites n and $n + \hat{\mu}$
- they are $U(1)$ phase factors

$$U_{n,\mu} = \exp\{iA_{n,\mu}\}\,, \qquad A_{n,\mu} \in [-\pi, \pi[$$

Schwinger Model

- lattice action looks identical to QCD

$$S = \beta \sum_P \left[1 - \frac{1}{2}(U_P - U_P^\dagger) \right] + \phi^\dagger \frac{1}{MM^\dagger} \phi = S_G + S_F$$

- with plaquette variable

$$U_P \equiv U_{n,\mu} U_{n+\hat{\mu},\nu} U^\dagger_{n+\hat{\nu},\nu} U^\dagger_{n,\nu}$$

- n is site index and $\mu, \nu \in \{x, t\}$ the directions
- M is the Wilson Dirac operator

Wilson Fermions

- the Wilson Dirac operator

$$M_{n\alpha,m\beta} = (m_0 + 2r)\delta_{nm}\delta_{\alpha\beta} \\ - \frac{1}{2}\sum_{\mu}\left[(r-\gamma_\mu)_{\alpha\beta}U_{n,\mu}\delta n, m - \hat{\mu} + (r+\gamma_\mu)_{\alpha\beta}U^\dagger_{m,\mu}\delta_{n,m+\hat{\mu}}\right]$$

- in $d = 2$ dimensions the γ-matrices are

$$\gamma_1 = \sigma_1\,, \quad \gamma_2 = \sigma_2\,, \quad \gamma_5 = \sigma_3$$

  with Pauli matrices σ_i
- they fulfil

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$$

- M is γ_5 hermitian

$$M^\dagger = \gamma_5 M \gamma_5$$

HMC for the Schwinger Model

- with $N_f = 2$ flavours of Wilson fermions

$$\det \begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix} = \det(MM) = \det(M\gamma_5 M\gamma_5) = \det(MM^\dagger)$$

$\Rightarrow$ $MM^\dagger \equiv Q^2$ is positive definite (with $Q = M\gamma_5$ and $Q = Q^\dagger$)

$\Rightarrow$ fermionic action is real

- and fermion weight is Gaussian

$$\exp\{-\phi^\dagger \frac{1}{Q^2}\phi\} = \exp\{R^\dagger R\}\,, \qquad \phi = QR$$

$\Rightarrow$ Can generate R from Gaussian distribution and compute ϕ by applying Q

HMC for the Schwinger Model

- what about the derivative with respect to $A_{n,\mu}$?
- $\partial S_G / \partial A_{n,\mu}$ is simple
- the pseudo-fermion action is slightly more involved
- the variation for an inverse matrix

$$\delta(A^{-1}) = A^{-1}\delta(A)A^{-1}$$

$\Rightarrow$ so, for S_F

$$\delta S_F \;=\; \phi^\dagger \frac{1}{Q^2}\delta(Q^2)\frac{1}{Q^2}\phi \;\equiv\; \eta^\dagger \delta(Q^2)\eta$$

- with

$$\eta \;\equiv\; \frac{1}{Q^2}\phi$$

HMC for the Schwinger Model

introduce conjugate momenta $p_{n,\mu}$ for every angle $A_{n,\mu}$

1. generate $p_{n,\mu}$ Gaussian distributed
2. generate R Gaussian distributed
3. compute $\phi = Q\ R$
4. MD update with EoM

$$\eta = (Q^2)^{-1}\phi$$
$$\dot{A}_{n,\mu} = p_{n,\mu}$$
$$\dot{p}_{n,\mu} = -\frac{\partial S_G}{\partial A_{n,\mu}} + \eta^\dagger \frac{\partial (Q^2)}{\partial A_{n,\mu}} \eta$$

   using the leap-frog algorithm (ϕ unchanged)

5. accept/reject step with

$$\mathcal{H}(A',p') = \sum p^2/2 + S_G(A) + R^\dagger R$$
$$\mathcal{H}(A',p') = \sum p^2/2 + S_G(A') + R'^\dagger R'\,, \qquad R' = (Q(A'))^{-1}\phi$$

HMC for the Schwinger Model

- Q^2 or Q must be inverted on a source
  - in each timestep for $\eta = (Q^2)^{-1}\phi$
  - in the acceptance step for $R' = (Q(A'))^{-1}\phi$

- so in total N_{MD} inversions per trajectory

- typically the conjugate gradient (CG) method is used

$\Rightarrow$ requires $\mathcal{O}(1000)$ applications of Q^2 per inversion (depending on lattice spacing, mass, etc...)

Exercises

- for the symplectic integrator (see before!) use

$$\mathcal{H}(x,p) = p^2/2 + x^2/2$$

  and show that the integrator is symplectic and $\mathcal{H}_s$ is conserved

- show that

$$\eta^\dagger \frac{\partial(Q^2)}{\partial A_{n,\mu}} \eta = 2\mathrm{Re}\left[\eta^\dagger \frac{\partial Q}{\partial A_{n,\mu}} Q \eta\right]$$

- compute

$$\frac{\partial Q}{\partial A_{n,\mu}}$$

  explicitly

- the slides are available at

  `http://www.itkp.uni-bonn.de/~urbach/urbach1.pdf`