Last Modified Wednesday, 15 June 2011

Monte Carlo Algorithms

A D Kennedy

Tait Institute, School of Physics & Astronomy,
The University of Edinburgh

Tuesday, 14 June 2011

STRONGnet Summer School 2011, Bielefeld

Outline

- Monte Carlo integration
- Markov Chain Monte Carlo
- Autocorrelations
- Hybrid Monte Carlo
- Symplectic Integrators
- Shadow Hamiltonians
- Gauge Fields
- Fermions

Monte Carlo Integration

Monte Carlo Integration

- Monte Carlo integration is based on the identification of *probabilities* with *measures*
- There are much better methods of carrying out low dimensional quadrature
  - All other methods become hopelessly expensive for large dimensions
  - In lattice QFT there is one integration per degree of freedom
    - We are approximating an infinite dimensional functional integral

Monte Carlo Integration

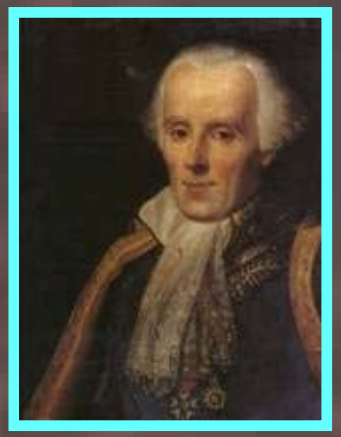
- Generate a sequence of random field configurations $\left(\phi_1, \phi_2, \ldots, \phi_t, \ldots, \phi_N\right)$ chosen from the probability distribution

$$P(\phi_t)\,d\phi_t = \frac{1}{Z} e^{-S(\phi_t)} d\phi_t$$

- Measure the value of Ω on each configuration and compute the average

$$\overline{\Omega} \equiv \frac{1}{N} \sum_{t=1}^{N} \Omega(\phi_t)$$

Central Limit Theorem



- Law of Large Numbers $\langle \Omega \rangle = \lim_{N \to \infty} \overline{\Omega}$
- Central Limit Theorem $\langle \Omega \rangle \sim \overline{\Omega} + O\left(\sqrt{\frac{C_2}{N}}\right)$
  - where the *variance* of the distribution of Ω is $C_2 \equiv \left\langle \left(\Omega - \langle \Omega \rangle \right)^2 \right\rangle$
- The Laplace–DeMoivre Central Limit theorem is an asymptotic expansion for the probability distribution of $\overline{\Omega}$
  - Distribution of values for a single sample $\omega = \Omega(\phi)$

$$P_\Omega(\omega) \equiv \int d\phi \; P(\phi) \; \delta\left(\omega - \Omega(\phi)\right) = \left\langle \delta\left(\omega - \Omega(\phi)\right) \right\rangle$$

Central Limit Theorem

- Generating function for connected moments

$$W_\Omega(k) \equiv \ln \int d\omega \;\; P_\Omega(\omega) \;\; e^{ik\omega}$$

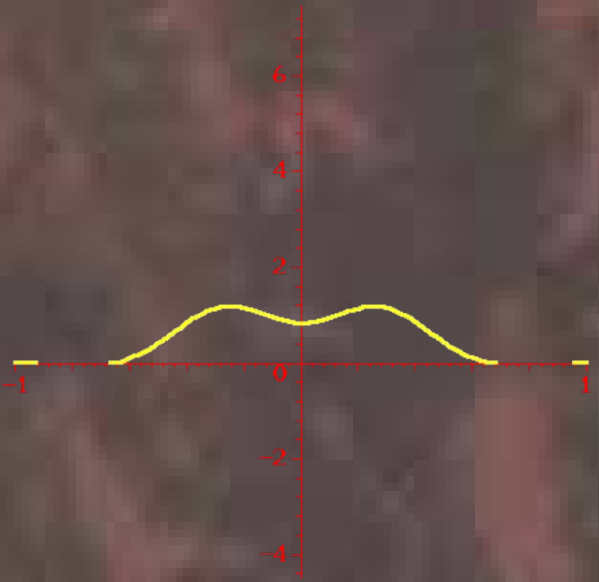
$$= \ln \int d\phi \;\; P(\phi) \;\; e^{ik\Omega(\phi)} = \ln \left\langle e^{ik\Omega} \right\rangle \sim \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} C_n$$

- The first few *cumulants* are

$$C_0 = 0 \qquad C_3 = \left\langle \left(\Omega - \langle \Omega \rangle \right)^3 \right\rangle$$

$$C_1 = \langle \Omega \rangle \qquad C_4 = \left\langle \left(\Omega - \langle \Omega \rangle \right)^4 \right\rangle - 3C_2^2$$

$$C_2 = \left\langle \left(\Omega - \langle \Omega \rangle \right)^2 \right\rangle$$



- Note that this is an *asymptotic* expansion

Central Limit Theorem

- Distribution of the average of N samples

$$P_{\bar{\Omega}}(\bar{\omega}) \equiv \int d\phi_1 \ldots\ d\phi_N \quad P(\phi_1)\ldots \quad P(\phi_N) \quad \delta\left(\bar{\omega} - \frac{1}{N}\sum_{t=1}^{N}\Omega(\phi_t)\right)$$

- Connected generating function

$$W_{\bar{\Omega}}(k) \equiv \ln \int d\bar{\omega} \quad P_{\bar{\Omega}}(\bar{\omega}) \quad e^{ik\bar{\omega}}$$

$$= \ln \int d\phi_1 \ldots d\phi_N \quad P(\phi_1)\ldots P(\phi_N) \quad \exp\left[\frac{ik}{N}\sum_{t=1}^{N}\Omega(\phi_t)\right]$$

$$= \ln\left[\int d\phi \quad P(\phi) \quad e^{ik\Omega(\phi)/N}\right]^N = N \ln\left\langle e^{ik\Omega/N}\right\rangle$$

$$= NW_\Omega\left(\frac{k}{N}\right) \sim \sum_{n=1}^{\infty}\frac{(ik)^n}{n!}\frac{C_n}{N^{n-1}}$$

Central Limit Theorem

- Take inverse Fourier transform to obtain distribution $P_{\bar{\Omega}}$

$$P_{\bar{\Omega}}\left(\bar{\omega}\right) = \frac{1}{2\pi}\int dk \quad e^{W_{\bar{\Omega}}(k)} e^{-ik\bar{\omega}}$$

$$\sim e^{-\frac{C_3}{3!N^2}\frac{d^3}{d\bar{\omega}^3}+\frac{C_4}{4!N^3}\frac{d^4}{d\bar{\omega}^4}-\cdots} \int \frac{dk}{2\pi} \quad e^{ik\langle\Omega\rangle+\frac{1}{2N}(ik)^2 C_2} e^{-ik\bar{\omega}}$$

$$= e^{-\frac{C_3}{3!N^2}\frac{d^3}{d\bar{\omega}^3}+\frac{C_4}{4!N^3}\frac{d^4}{d\bar{\omega}^4}-\cdots} \frac{e^{-\frac{\left(\bar{\omega}-\langle\Omega\rangle\right)^2}{2C_2/N}}}{\sqrt{2\pi\, C_2/N}}$$

Central Limit Theorem

- Re-scale to show convergence to Gaussian distribution

$$P_{\overline{\Omega}}\left(\overline{\omega}\right) = F\left(\xi\right)\frac{d\xi}{d\overline{\omega}}$$

- where $\xi \equiv \left(\overline{\omega} - \left\langle\Omega\right\rangle\right)\sqrt{N}$ and

$$F\left(\xi\right) = \left[1 + \frac{C_3\xi\left(\xi^2 - 3C_2\right)}{6C_2^3\sqrt{N}} + \cdots\right]\frac{e^{-\xi^2/2C_2}}{\sqrt{2\pi C_2}}$$

Asymptotic Expansions

- If $\int_{-\infty}^{\infty} dx\, e^{-\frac{f(x)-f(x_0)}{\varepsilon^2}}$ exists for $\varepsilon \le \varepsilon_0$, where x_0 is the absolute minimum of f, show that

$$\int_{-\infty}^{\infty} dx\, e^{-\frac{f(x)-f(x_0)}{\varepsilon^2}} = \int_{-\infty}^{\infty} dx\, e^{-\frac{1}{\varepsilon^2}\left[\frac{1}{2}f''(x_0)(x-x_0)^2+\cdots\right]} \qquad \xi \equiv \frac{x-x_0}{\varepsilon}$$

$$= \varepsilon \int_{-\infty}^{\infty} d\xi\, e^{-\frac{1}{2}f''(x_0)\xi^2}\left[1+O(\varepsilon)\right]+O\left(e^{-k/\varepsilon}\right)$$

$$= \varepsilon\sqrt{\tfrac{1}{2}\pi f''(x_0)}\left[1+O(\varepsilon)\right]+O\left(e^{-k/\varepsilon}\right)$$

Laplace's Method

$$\int_{-\infty}^{\infty} dx\, e^{-\frac{f(x)-f(x_0)}{\varepsilon^2}} = \int_{x_0-\Delta}^{x_0+\Delta} dx\, e^{-\frac{f(x)-f(x_0)}{\varepsilon^2}} + \int_{|x-x_0|>\Delta} dx\, e^{-\frac{f(x)-f(x_0)}{\varepsilon^2}}$$

$$= \int_{x_0-\Delta}^{x_0+\Delta} dx\, e^{-\frac{f(x)-f(x_0)}{\varepsilon^2}} + O\left(e^{-\frac{K\Delta^2}{\varepsilon^2}}\right) \sim \varepsilon \int_{x_0-\Delta}^{x_0+\Delta} d\xi\, e^{-\frac{1}{2}f''(x_0)\xi^2} \sum_{j=0}^{\infty} c_j \varepsilon^j \xi^{j+2}$$

$$= \varepsilon \sum_{j=0}^{\infty} c_j \varepsilon^j \int_{x_0-\Delta}^{x_0+\Delta} d\xi\, \xi^{j+2} e^{-\frac{1}{2}f''(x_0)\xi^2} \sim \varepsilon \sum_{j=0}^{\infty} c_j \varepsilon^j \int_{-\infty}^{\infty} d\xi\, \xi^{j+2} e^{-\frac{1}{2}f''(x_0)\xi^2}$$

Proof for Laplace's Method

$$f(x) - f(x_0) \geq K\left|x - x_0\right|^2$$

$$f(x) \geq f(x_0 + \delta')$$

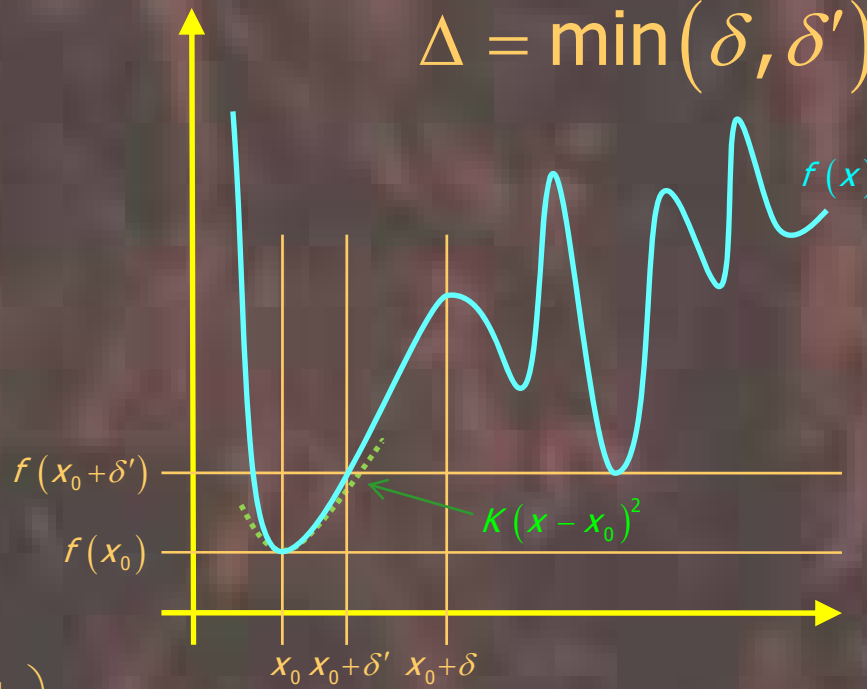
$$f(x) - f(x_0) \geq K\Delta^2$$

$$I_\varepsilon = \int_{x_0+\Delta}^{\infty} dx\, e^{-\frac{f(x)-f(x_0)}{\varepsilon^2}}$$

$$\leq \int_{x_0+\Delta}^{\infty} dx\, e^{\frac{f(x)-f(x_0)}{\varepsilon_0^2}} e^{-\left(\frac{1}{\varepsilon^2}-\frac{1}{\varepsilon_0^2}\right)K\Delta^2}$$

$$\leq I_{\varepsilon_0}\, e^{-\left(\frac{1}{\varepsilon^2}-\frac{1}{\varepsilon_0^2}\right)K\Delta^2}$$

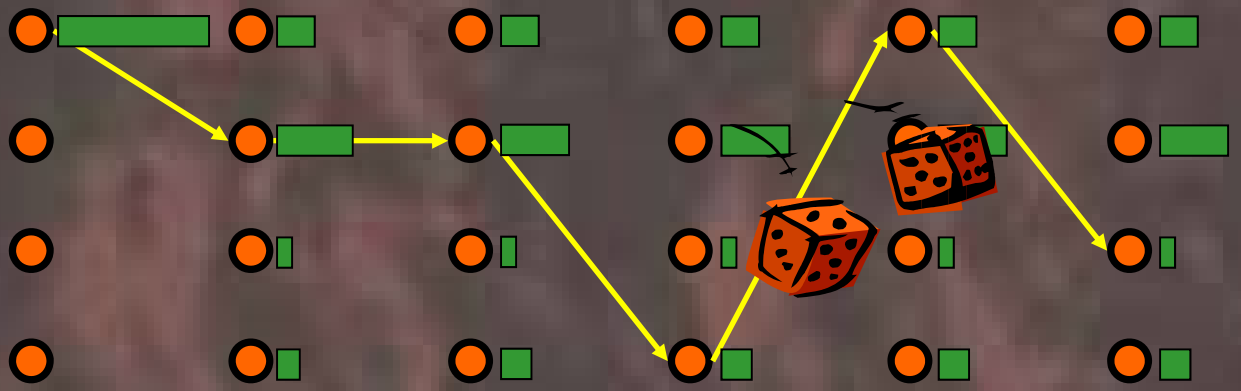
$$\Delta = \min(\delta, \delta')$$



Markov Chain Monte Carlo

Markov Chains

- State space Ω
- (Ergodic) stochastic transitions $P': \Omega \to \Omega$
- Deterministic evolution of probability distribution $P: Q \to Q$
- Distribution converges to unique fixed point $\overline{Q}$



Convergence of Markov Chains

- Define a metric $d(Q_1, Q_2) \equiv \int dx \left| Q_1(x) - Q_2(x) \right|$ on the space of (equivalence classes of) probability distributions
- Prove that $d(PQ_1, PQ_2) \leq (1-\alpha)\, d(Q_1, Q_2)$ with $\alpha > 0$, so the Markov process P is a contraction mapping
- The sequence Q, PQ, P^2Q, P^3Q,... is Cauchy
- The space of probability distributions is *complete*, so the sequence converges to a unique fixed point $\overline{Q} = \lim_{n \to \infty} P^n Q$

Simple but Inadequate Proof

$$d\left(PQ_1, PQ_2\right) = \int dx \left| PQ_1(x) - PQ_2(x) \right|$$

$$= \int dx \left| \int dy\, P(x \leftarrow y) Q_1(y) - \int dy\, P(x \leftarrow y) Q_2(y) \right|$$

$$= \int dx \left| \int dy\, P(x \leftarrow y) \Delta Q(y) \right| \qquad \Delta Q(y) \equiv Q_1(y) - Q_2(y)$$

$$\leq \int dx \int dy\, P(x \leftarrow y) \left| \Delta Q(y) \right| \qquad \left| \int dx\, f(x) \right| \leq \int dx \left| f(x) \right|$$

$$= \int dy \int dx\, P(x \leftarrow y) \left| \Delta Q(y) \right|$$

$$= \int dy \left| \Delta Q(y) \right| = d\left(Q_1, Q_2\right) \qquad \int dx\, P(x \leftarrow y) = 1$$

Proof

$$d(PQ_1, PQ_2) = \int dx \left| PQ_1(x) - PQ_2(x) \right|$$

$$= \int dx \left| \int dy\, P(x \leftarrow y) Q_1(y) - \int dy\, P(x \leftarrow y) Q_2(y) \right|$$

$$\Delta Q(y) \equiv Q_1(y) - Q_2(y)$$

$$= \int dx \left| \int dy\, P(x \leftarrow y) \Delta Q(y) \right|$$

$$\theta(y) + \theta(-y) = 1$$

$$= \int dx \left| \int dy\, P(x \leftarrow y) \Delta Q(y) \left[\theta(\Delta Q(y)) + \theta(-\Delta Q(y)) \right] \right|$$

$$\big| |a| - |b| \big| = |a| + |b| - 2\min(|a|, |b|)$$

$$= \int dx \int dy\, P(x \leftarrow y) \left| \Delta Q(y) \right|$$

$$-2 \int dx \min_{\pm} \left| \int dy\, P(x \leftarrow y) \Delta Q(y) \theta(\pm \Delta Q(y)) \right|$$

Proof

$$d(PQ_1, PQ_2) \qquad \int dx\, P(x \leftarrow y) = 1$$

$$= \int dy \left|\Delta Q(y)\right| - 2\int dx \min_{\pm} \left| \int dy\, P(x \leftarrow y)\, \Delta Q(y)\, \theta\left(\pm \Delta Q(y)\right) \right|$$

$$\leq \int dy \left|\Delta Q(y)\right| - 2\int dx \inf_{y} P(x \leftarrow y) \min_{\pm} \left| \int dy\, \Delta Q(y)\, \theta\left(\pm \Delta Q(y)\right) \right|$$

$$\int dy\, \Delta Q(y)\, \theta\left(\Delta Q(y)\right) + \int dy\, \Delta Q(y)\, \theta\left(-\Delta Q(y)\right)$$

$$= \int dy\, \Delta Q(y) = \int dy\, Q_1(y) - \int dy\, Q_2(y) = 1 - 1 = 0$$

$$\Rightarrow \left| \int dy\, \Delta Q(y)\, \theta\left(\pm \Delta Q(y)\right) \right| = \tfrac{1}{2} \int dy \left|\Delta Q(y)\right|$$

$$\leq \int dy \left|\Delta Q(y)\right| - \int dx \inf_{y} P(x \leftarrow y) \int dy \left|\Delta Q(y)\right| \leq (1 - \alpha) d(Q_1, Q_2)$$

$$0 < \alpha = \int dx \inf_{y} P(x \leftarrow y)$$

Banach Fixed-Point Theorem

- We show that the sequence of distributions is Cauchy

$$d(P^m Q, P^n Q) \le \sum_{j=0}^{n-m-1} d(P^{m+j} Q, P^{m+j+1} Q)$$

$$\le \sum_{j=0}^{n-m-1} (1-\alpha)^j \, d(P^m Q, P^{m+1} Q) \le d(P^m Q, P^{m+1} Q) \sum_{j=0}^{\infty} (1-\alpha)^j$$

$$= \frac{d(P^m Q, P^{m+1} Q)}{\alpha} = \frac{(1-\alpha)^m}{\alpha} d(Q, PQ) < \varepsilon$$

  for any $\varepsilon > 0$, provided m is large enough and $\alpha > 0$.

- Hence the sequence converges to the unique fixed point probability distribution $\overline{Q} = \lim_{n \to \infty} P^n Q$

Markov Chains

- Use Markov chains to sample from Q
  - Suppose we can construct an ergodic Markov process P which has distribution Q as its fixed point
  - Start with an arbitrary *state ("field configuration")*
  - Iterate the Markov process until it has converged *("thermalized")*
  - Thereafter, successive configurations will be distributed according to Q
    - But in general they will be correlated
  - To construct P we only need relative probabilities of states
    - Do not know the normalisation of Q
    - Cannot use Markov chains to compute integrals directly
    - We can compute ratios of integrals

Markov Chains

- How do we construct a Markov process with a specified fixed point $\bar{Q}(x) = \int dy\, P(x \leftarrow y)\bar{Q}(y)$?
- Detailed balance $P(y \leftarrow x)\bar{Q}(x) = P(x \leftarrow y)\bar{Q}(y)$
  - Integrate w.r.t. y to obtain fixed point condition
  - *Sufficient* but not *necessary* for fixed point
- Metropolis algorithm $P(x \leftarrow y) = \min\left(1, \frac{\bar{Q}(x)}{\bar{Q}(y)}\right)$
  - Consider cases $\bar{Q}(x) > \bar{Q}(y)$ and $\bar{Q}(x) < \bar{Q}(y)$ separately to obtain detailed balance condition
  - *Sufficient* but not *necessary* for detailed balance
  - Other choices are possible, e.g., $P(x \leftarrow y) = \frac{\bar{Q}(x)}{\bar{Q}(x) + \bar{Q}(y)}$

Markov Chains

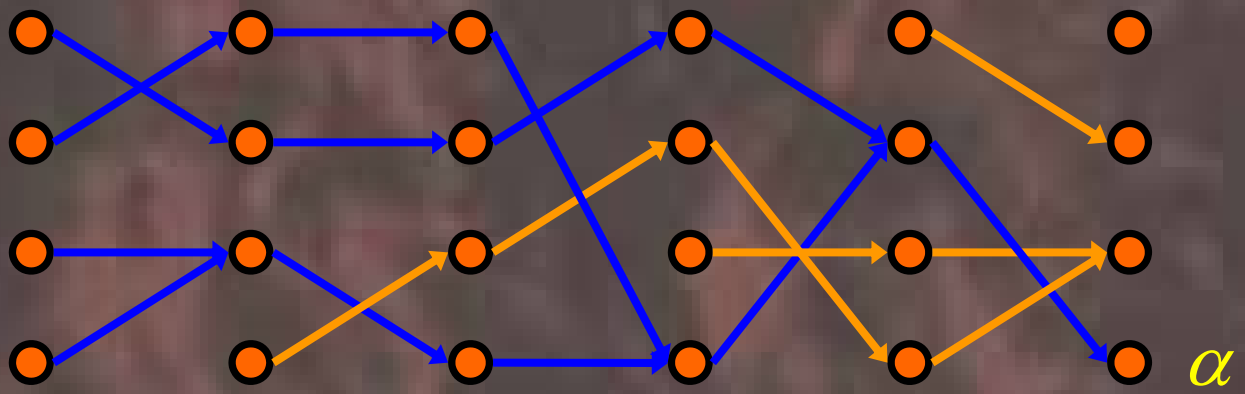
- Composition of Markov steps
  - Let P_1 and P_2 be two Markov steps which have the desired fixed point distribution
    - They need *not* be ergodic
  - Then the composition of the two steps $P_2 \circ P_1$ will also have the desired fixed point
    - And it may be ergodic
  - This trivially generalises to any (fixed) number of steps
    - For the case where P_1 is not ergodic but $(P_1)^n$ is the terminology weakly and strongly ergodic are sometimes used

Markov Chains

- This result justifies "sweeping" through a lattice performing single site updates
  - Each individual single site update has the desired fixed point because it satisfies detailed balance
  - The entire sweep therefore has the desired fixed point, and is ergodic
  - But the entire sweep does *not* satisfy detailed balance
  - Of course it would satisfy detailed balance if the sites were updated in a random order
    - But this is not necessary
    - And it is undesirable because it puts too much randomness into the system

Markov Chains

- Coupling from the Past
  - *Propp and Wilson (1996)*
  - Use *fixed* set of random numbers
  - Flypaper principle: If states coalesce they stay together forever
    - Eventually, all states coalesce to some state α with probability one
    - Any state from $t = -\infty$ will coalesce to α
    - α is a sample from the fixed point distribution



Auto-correlations

Exponential Autocorrelations

- The unique fixed point of an ergodic Markov process corresponds to a unique eigenvector with eigenvalue 1
- All its other eigenvalues must lie within the unit circle
- In particular, the largest subleading eigenvalue is $\left|\lambda_{\max}\right| < 1$
- The eigenvectors $\int dy\, P(x \leftarrow y) u(y) = \lambda u(x)$ satisfy

$$\lambda \int dx\, u(x) = \int dx \int dy\, P(x \leftarrow y) u(y) = \int dy \left[\int dx\, P(x \leftarrow y) \right] u(y) = \int dy\, u(y)$$

so either $\lambda = 1$ or $\int dx\, u(x) = 0$

- Hence we may expand any probability density as $Q = \overline{Q} + \sum_{|\lambda_j|<1} c_j u_j$

$$\left\| P^N Q - \overline{Q} \right\| = \left\| \sum_{|\lambda_j|<1} c_j \lambda_j^N u_j \right\| \le \left|\lambda_{\max}\right|^N \sum_{|\lambda_j|<1} \left|c_j\right| \left\|u_j\right\| = K e^{-N/N_{\exp}}$$

with the exponential autocorrelation time $N_{\exp} \equiv -\dfrac{1}{\ln\left|\lambda_{\max}\right|} > 0$

Integrated Autocorrelations

- Consider the autocorrelation of operator Ω
- Without loss of generality we assume $\langle \Omega \rangle = 0$

$$\left\langle \overline{\Omega}^2 \right\rangle = \frac{1}{N^2} \sum_{t=1}^{N} \sum_{t'=1}^{N} \left\langle \Omega(\phi_t) \Omega(\phi_{t'}) \right\rangle = \frac{1}{N^2} \left\{ \sum_{t=1}^{N} \left\langle \Omega(\phi_t)^2 \right\rangle + 2 \sum_{t=1}^{N-1} \sum_{t'=t+1}^{N} \left\langle \Omega(\phi_t) \Omega(\phi_{t'}) \right\rangle \right\}$$

- Define <u>autocorrelation function</u> $C_\Omega(\ell) \equiv \dfrac{\left\langle \Omega(\phi_{t+\ell}) \Omega(\phi_t) \right\rangle}{\left\langle \Omega(\phi)^2 \right\rangle}$

$$\left\langle \overline{\Omega}^2 \right\rangle = \frac{1}{N} \left\{ \left\langle \Omega^2 \right\rangle + \frac{2}{N} \sum_{\ell=1}^{N-1} (N - \ell) C_\Omega(\ell) \left\langle \Omega^2 \right\rangle \right\}$$

Integrated Autocorrelations

- The autocorrelation function must fall faster that the exponential autocorrelation $\left|C_\Omega(\ell)\right| \le \lambda_{\max}^\ell = e^{-\ell/N_{\exp}}$
- For a sufficiently large number of samples

$$\left\langle \overline{\Omega}^2 \right\rangle = \left\{ 1 + 2\sum_{\ell=1}^{\infty} C_\Omega(\ell) \right\} \frac{\left\langle \Omega^2 \right\rangle}{N} \left[1 + O\left(\frac{N_{\exp}}{N} \right) \right]$$

- Define integrated autocorrelation function $A_\Omega \equiv \sum_{\ell=1}^{\infty} C_\Omega(\ell)$

$$\left\langle \overline{\Omega}^2 \right\rangle = \left\{ 1 + 2A_\Omega \right\} \frac{\left\langle \Omega^2 \right\rangle}{N} \left[1 + O\left(\frac{N_{\exp}}{N} \right) \right]$$


Hybrid
Monte Carlo

Hybrid Monte Carlo

- In order to carry out Monte Carlo computations including the effects of dynamical fermions we would like to find an algorithm which
  - Updates the fields globally
    - Because single link updates are not cheap if the action is not local
  - Take large steps through configuration space
    - Because small-step methods carry out a random walk which leads to critical slowing down with a dynamical critical exponent $z=2$
      - z relates the autocorrelation to the correlation length of the system, $A_\Omega \propto \xi^z$
  - Does not introduce any systematic errors

Hybrid Monte Carlo

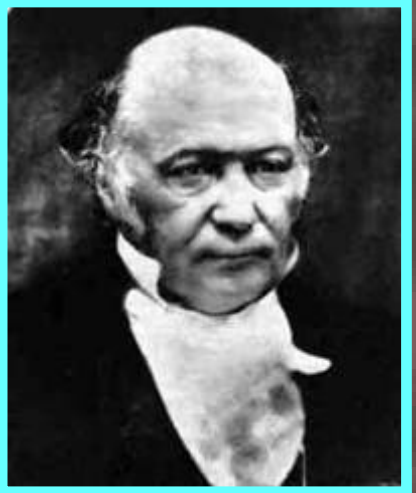
- A useful class of algorithms with these properties is the (Generalised) Hybrid Monte Carlo (HMC) method
  - Introduce a "fictitious" momentum p corresponding to each dynamical degree of freedom q
  - Find a Markov chain with fixed point $\propto \exp[-H(q,p)]$ where H is the "fictitious" Hamiltonian $\frac{1}{2}p^2 + S(q)$
    - The action S of the underlying QFT plays the rôle of the potential in the "fictitious" classical mechanical system
    - This gives the evolution of the system in a fifth dimension, "fictitious" or computer time
  - This generates the desired distribution $\exp[-S(q)]$ if we ignore the momenta p (i.e., the marginal distribution)

Hybrid Monte Carlo

- The HMC Markov chain alternates two Markov steps
  - Molecular Dynamics Monte Carlo (MDMC)
  - (Partial) Momentum Refreshment
- Both have the desired fixed point
- Together they are ergodic

MDMC

- If we could integrate Hamilton's equations exactly we could follow a trajectory of constant fictitious energy
  - This corresponds to a set of equiprobable fictitious phase space configurations
  - Liouville's theorem tells us that this also preserves the functional integral measure $dp \wedge dq$ as required
- Any approximate integration scheme which is reversible and area preserving may be used to suggest configurations to a Metropolis accept/reject test
  - With acceptance probability $\min[1, \exp(-\delta H)]$



MDMC

- We build the MDMC step out of three parts
  - Molecular Dynamics (MD), an approximate integrator $U(\tau):(q,p)\mapsto(q',p')$ which is *exactly*
    - Area preserving, $\det U_* = \det\left[\frac{\partial(q',p')}{\partial(q,p)}\right] = 1$
    - Reversible, $F\circ U(\tau)\circ F\circ U(\tau) = 1$
  - A momentum flip $F: p\mapsto -p$
  - A Metropolis accept/reject step
- The composition of these gives

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \left[F\circ U(\tau) \;\; \vartheta\left(e^{-\delta H} - y\right) + 1 \;\; \vartheta\left(y - e^{-\delta H}\right)\right] \begin{pmatrix} q \\ p \end{pmatrix}$$

  - with y being a uniformly distributed random number in [0,1]

Partial Momentum Refreshment

- This mixes the Gaussian distributed momenta p with Gaussian noise ξ

$$\begin{pmatrix} p' \\ \xi' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \circ F \begin{pmatrix} p \\ \xi \end{pmatrix}$$

  - The Gaussian distribution of p is invariant under F
  - The extra momentum flip F ensures that for small θ the momenta are reversed after a rejection rather than after an acceptance
  - For $\theta = \pi/2$ all momentum flips are irrelevant

Acceptance Rates

- The normalization of the equilibrium distribution is

$$1 = \frac{1}{Z}\int dq\,dp\, e^{-H(q,p)} = \frac{1}{Z}\int dq'\,dp'\, e^{-H(q',p')}$$

$$= \frac{1}{Z}\int dq'\,dp'\, e^{-H(q,p)-\delta H} = \frac{1}{Z}\int dq\,dp\, e^{-H(q,p)-\delta H} = \left\langle e^{-\delta H}\right\rangle$$

since $\delta H \equiv H(q',p') - H(q,p)$ and $dq \wedge dp = dq' \wedge dp'$

- For small δH we have $\left\langle e^{-\delta H}\right\rangle = 1 - \left\langle \delta H\right\rangle + \frac{1}{2}\left\langle \delta H^2\right\rangle + \cdots$, hence $\left\langle \delta H\right\rangle \approx \frac{1}{2}\left\langle \delta H^2\right\rangle$
- Thus if $\delta H = O\left(\delta\tau^n\right)$ we have $\left\langle \delta H\right\rangle = O\left(\delta\tau^{2n}\right)$

Thermodynamic Limit

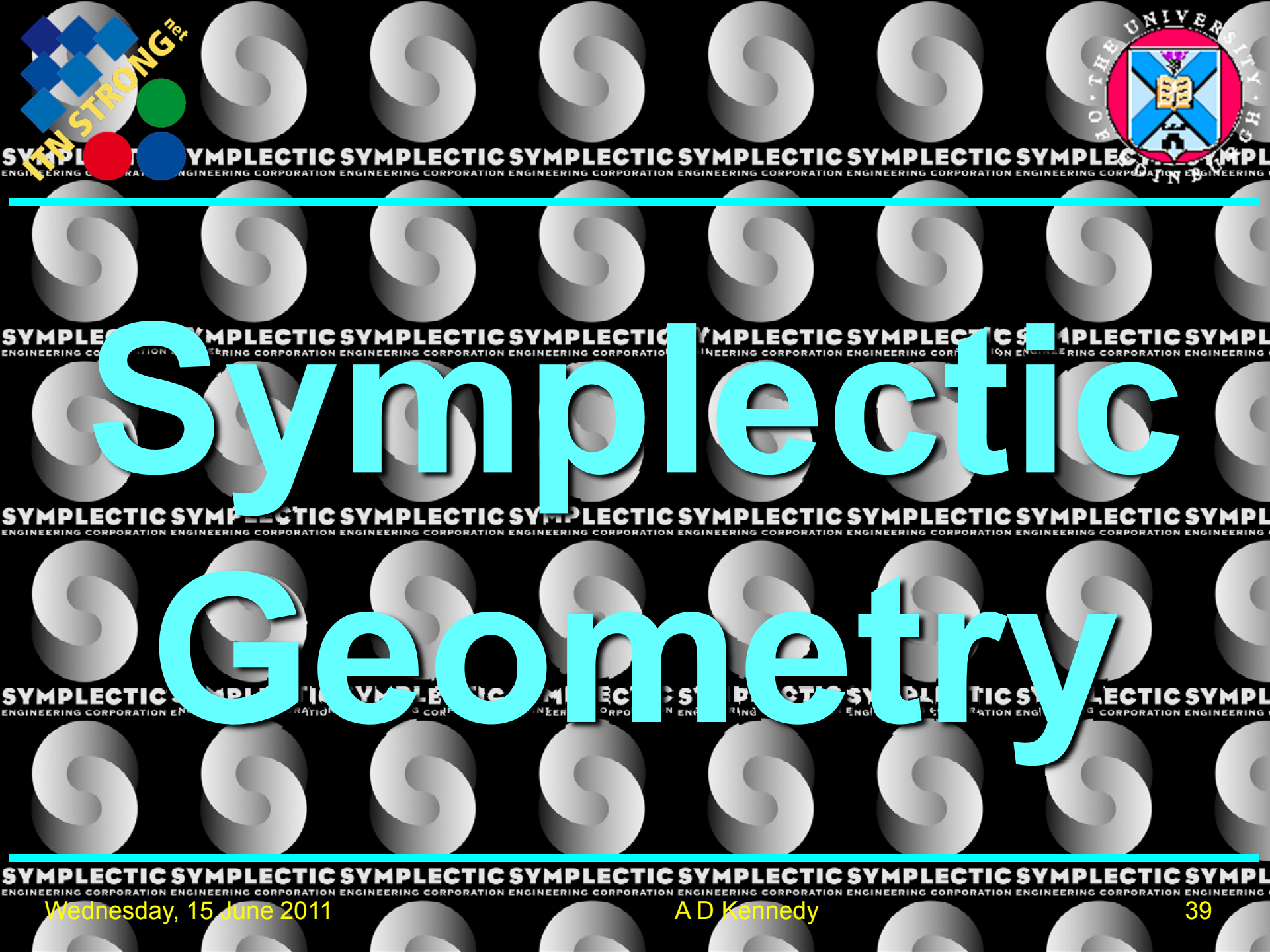
- If box size $L \gg \xi$ (correlation length) cluster decomposition and central limit theorem $\Rightarrow P_{\delta H}(h) = (2\pi V)^{-\frac{1}{2}} e^{-\frac{(h-\langle\delta H\rangle)^2}{2V}}$
- so we must have $V = 2\langle\delta H\rangle$

$$1 = \langle e^{-h}\rangle = \int dh\, P_{\delta H}(h) e^{-h} = \int \frac{dh}{\sqrt{2\pi V}} e^{-\frac{(h-\langle\delta H\rangle)H^2 - 2hV}{2V}} = e^{\frac{1}{2}(V-2\langle\delta H\rangle)}$$

- and thus the average acceptance rate $\langle P_{\text{acc}}\rangle$ is

$$\left\langle \min\left(1, e^{-h}\right)\right\rangle = \frac{1}{2\sqrt{\pi\langle\delta H\rangle}}\left(\int_{-\infty}^{0} dh\, e^{-\frac{(h-\langle\delta H\rangle)^2}{4\langle\delta H\rangle}} + \int_{0}^{\infty} dh\, e^{-\frac{(h-\langle\delta H\rangle)^2}{4\langle\delta H\rangle} - h}\right)$$

$$= \frac{2}{\sqrt{\pi}} \int_{\frac{1}{2}\sqrt{\langle\delta H\rangle}}^{\infty} dh\, e^{-h^2} = \operatorname{erfc}\left(\tfrac{1}{2}\sqrt{\langle\delta H\rangle}\right)$$


Symplectic
Geometry

Hamiltonian Vector Fields

- Classical mechanics is not specified just by a Hamiltonian H but also by a closed fundamental 2-form ω
- For every function (0-form) A this defines a Hamiltonian vector field $\hat{A}$ such that $dA = i_{\hat{A}}\omega$
  - Which just means that for all X $dA(X) = \omega\left(\hat{A}, X\right)$

Hamiltonian Vector Fields

- To be a little less abstract consider the familiar case where $\omega = dq \wedge dp$ and

$$dA \equiv \frac{\partial A}{\partial q} dq + \frac{\partial A}{\partial p} dp; \; X \equiv X_q \frac{\partial}{\partial q} + X_p \frac{\partial}{\partial p}$$

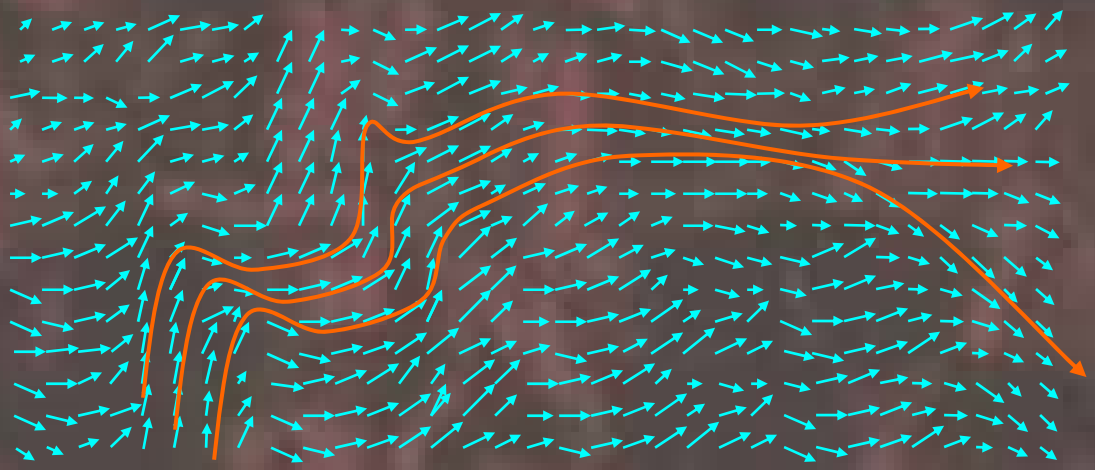
  - so

$$dA(X) = \frac{\partial A}{\partial q} X_q + \frac{\partial A}{\partial p} X_p = \omega\left(\hat{A}, X\right) = \hat{A}_q X_p - \hat{A}_p X_q$$

$$\hat{A} \equiv \hat{A}_q \frac{\partial}{\partial q} + \hat{A}_p \frac{\partial}{\partial p} = \frac{\partial A}{\partial p}\frac{\partial}{\partial q} - \frac{\partial A}{\partial q}\frac{\partial}{\partial p}$$

Classical Trajectories

- Classical trajectories are then integral curves of the Hamiltonian vector field $\hat{H}$ of the Hamiltonian H



$$\dot{f} \equiv \frac{\partial f}{\partial q}\dot{q} + \frac{\partial f}{\partial p}\dot{p} = \hat{H}f$$
$$= \frac{\partial H}{\partial p}\frac{\partial f}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial f}{\partial p}$$
$$\dot{q} = \frac{\partial H}{\partial p};\ \dot{p} = -\frac{\partial H}{\partial q}$$

- In other words, this vector field is always tangent to the classical trajectory

Symplectic Integrators

- We are interested in finding the classical trajectory in phase space of a system described by the Hamiltonian

$$H(q,p) = T(p) + S(q) = \tfrac{1}{2}p^2 + S(q)$$

- Define the corresponding Hamiltonian vector fields (with $\omega = dq \wedge dp$) $\hat{T} \equiv T'(p)\dfrac{\partial}{\partial q}$ and $\hat{S} \equiv -S'(q)\dfrac{\partial}{\partial p}$ so that $\hat{H} = \hat{T} + \hat{S}$

Symplectic Integrators

- Formally the solution of Hamilton's equations with trajectory length τ is the exponential of the Hamiltonian Hamiltonian vector field, $e^{\tau\hat{H}}$
- Since the kinetic energy T is a function only of p and the potential energy S is a function only of q, it follows that the action of $e^{\tau\hat{S}}$ and $e^{\tau\hat{T}}$ may be evaluated trivially (Taylor's theorem!)

$$e^{\tau\hat{T}} : f(q,p) \mapsto f\left(q + \tau T'(p), p\right)$$

$$e^{\tau\hat{S}} : f(q,p) \mapsto f\left(q, p - \tau S'(q)\right)$$

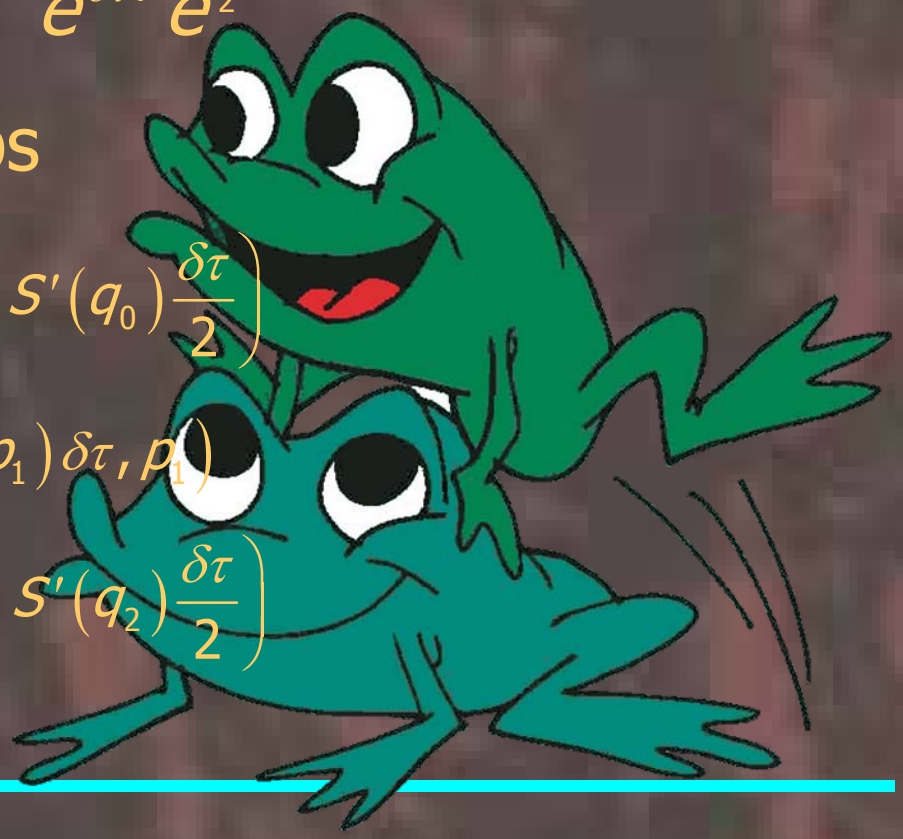
Leapfrog

- The simplest example is the Leapfrog PQP integrator $U_0(\delta\tau) \equiv e^{\frac{1}{2}\delta\tau\hat{S}} e^{\delta\tau\hat{T}} e^{\frac{1}{2}\delta\tau\hat{S}}$
- It consists of three steps

$$f(q_1, p_1) = e^{\frac{1}{2}\delta\tau\hat{S}} f(q_0, p_0) = f\left(q_0, p_0 - S'(q_0)\frac{\delta\tau}{2}\right)$$

$$f(q_2, p_2) = e^{\delta\tau\hat{T}} f(q_1, p_1) = f\left(q_1 + T'(p_1)\delta\tau, p_1\right)$$

$$f(q_3, p_3) = e^{\frac{1}{2}\delta\tau\hat{S}} f(q_2, p_2) = f\left(q_2, p_2 - S'(q_2)\frac{\delta\tau}{2}\right)$$



Langevin Algorithm

- The leapfrog update is
- If we ignore the Metropolis acceptance step (*e.g.*, if we take small enough steps)
- Rescale time step $\varepsilon \equiv \frac{1}{2}\delta\tau^2$ and initial Gaussian noise (momenta) $p_s(0) \equiv \frac{1}{2}\eta(s)\delta\tau$
- We obtain the Langevin equation

$$\left.\begin{aligned} p\left(\frac{\delta\tau}{2}\right) &= p(0) - S'(q(0))\frac{\delta\tau}{2} \\ q(\delta\tau) &= q(0) + p\left(\frac{\delta\tau}{2}\right)\delta\tau \\ p(\delta\tau) &= p\left(\frac{\delta\tau}{2}\right) - S'(q(\delta\tau))\frac{\delta\tau}{2} \end{aligned}\right\}$$

$$\frac{q(\delta\tau)-q(0)}{\frac{1}{2}\delta\tau^2} = -S'(q(0)) + \frac{p(0)}{\frac{1}{2}\delta\tau}$$

$$\boxed{\frac{dq}{ds} = -\frac{\delta S}{\delta q} + \eta}$$

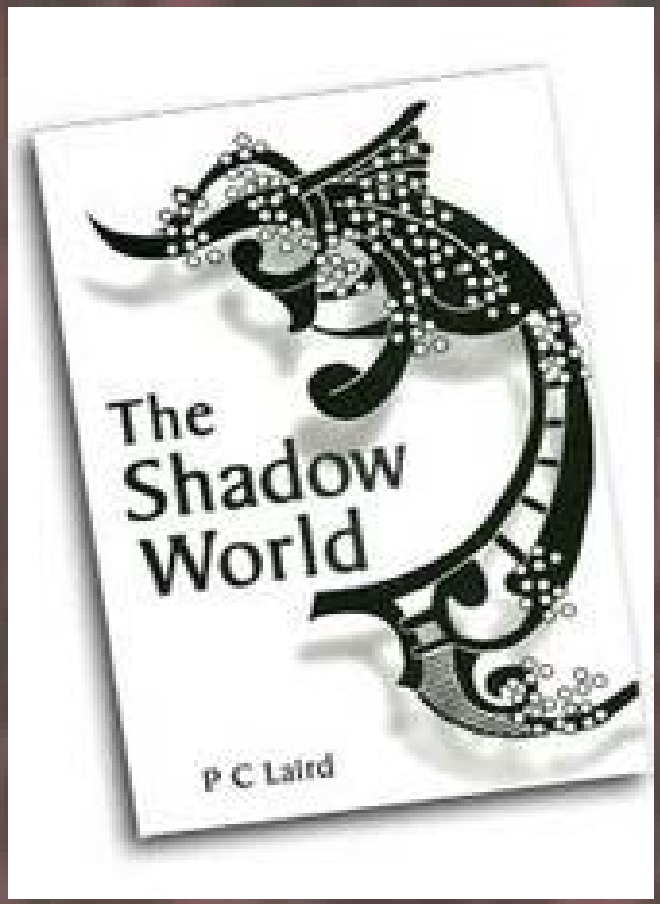
$$\langle p_s(0)\rangle = 0; \langle p_s(0)\,p_{s'}(0)\rangle = \delta_{s,s'}$$

$$\langle \eta(s)\rangle = 0; \langle \eta(s)\eta(s')\rangle = 2\frac{\delta_{s,s'}}{\varepsilon} \to 2\delta(s-s')$$


Shadow
Hamiltonians

Into the Shadow World

- For each symplectic integrator there exists a Hamiltonian H' which is exactly conserved
- This may be obtained by replacing the commutators $\left[\hat{S},\hat{T}\right]$ in the BCH expansion of $\ln\left(e^{-\hat{S}}e^{-\hat{T}}\right)$ with the Poisson bracket $\{S,T\}$



Poisson Brackets

- Consider the action of a Hamiltonian vector field on a function (0-form)

$$\hat{A}F = dF\left(\hat{A}\right) = i_{\hat{F}}\omega\left(\hat{A}\right) = \omega\left(\hat{F},\hat{A}\right) \equiv \left\{A,F\right\}$$

- We have introduced the Poisson bracket of two functions
- These obey the Jacobi identity

$$\left\{A,\left\{B,C\right\}\right\} + \left\{B,\left\{C,A\right\}\right\} + \left\{C,\left\{A,B\right\}\right\} = 0$$

  - This follows from the closure of the fundamental 2-form $d\omega = 0$
  - It is not trivial: Poisson brackets are not commutators
  - *Functions* form a Lie algebra with PBs as the Lie product

Jacobi Identity

- Invariant definition of exterior derivative $dF(X) = XF$

$$d\theta(X,Y) = X\theta(Y) - Y\theta(X) - \theta([X,Y])$$

$$d\omega(X,Y,Z) = X\omega(Y,Z) + Y\omega(Z,X) + Z\omega(X,Y) - \omega([X,Y],Z) - \omega([Y,Z],X) - \omega([Z,X],Y)$$

- For Hamiltonian vector fields we have

$$\hat{X}\omega(\hat{Y},\hat{Z}) = -\hat{X}\{Y,Z\} = -\{X,\{Y,Z\}\}$$

$$\omega([\hat{X},\hat{Y}],\hat{Z}) = -\omega(\hat{Z},[\hat{X},\hat{Y}]) = -dZ([\hat{X},\hat{Y}]) = -[\hat{X},\hat{Y}]Z = -(\hat{X}\hat{Y} - \hat{Y}\hat{X})Z$$

$$= -\hat{X}\hat{Y}Z + \hat{Y}\hat{X}Z = -\hat{X}\{Y,Z\} + \hat{Y}\{X,Z\} = -\{X,\{Y,Z\}\} + \{Y,\{X,Z\}\}$$

- The condition $d\omega = 0$ gives the Jacobi identity

$$d\omega(\hat{X},\hat{Y},\hat{Z}) = 0 = \{X,\{Y,Z\}\} + \{Y,\{Z,X\}\} + \{Z,\{X,Y\}\}$$

Concrete Poisson Brackets

- To make this more familiar when $\omega = dq \wedge dp$ the Poisson bracket $\{A,B\} \equiv -\omega\left(\hat{A},\hat{B}\right)$ becomes

$$\hat{A} = \frac{\partial A}{\partial p}\frac{\partial}{\partial q} - \frac{\partial A}{\partial q}\frac{\partial}{\partial p}$$

$$\omega\left(\hat{A},\hat{B}\right) = (dq \wedge dp)\left(\frac{\partial A}{\partial p}\frac{\partial}{\partial q} - \frac{\partial A}{\partial q}\frac{\partial}{\partial p}, \frac{\partial B}{\partial p}\frac{\partial}{\partial q} - \frac{\partial B}{\partial q}\frac{\partial}{\partial p}\right)$$

$$\{A,B\} = \frac{\partial A}{\partial p}\frac{\partial B}{\partial q} - \frac{\partial A}{\partial q}\frac{\partial B}{\partial p}$$

Hamilton's Equations (again)

- To make this really concrete consider the action of the Hamiltonian Hamiltonian vector field on an arbitrary function f that we saw earlier

$$\dot{f} \equiv \frac{\partial f}{\partial q}\dot{q} + \frac{\partial f}{\partial p}\dot{p} = \hat{H}f = \{H, f\}$$

$$\{H, f\} = \frac{\partial H}{\partial p}\frac{\partial f}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial f}{\partial p}$$

Commutators

- So far this is just a fancy (and complicated) way of rewriting Hamilton's equations, but now we derive a surprising new result
- To derive it consider

$$\left[\hat{A},\hat{B}\right]F = \left(\hat{A}\hat{B} - \hat{B}\hat{A}\right)F = \hat{A}\{B,F\} - \hat{B}\{A,F\}$$

$$= \{A,\{B,F\}\} - \{B,\{A,F\}\} = \{\{A,B\},F\} = \widehat{\{A,B\}}F$$

- The commutator of Hamiltonian vector fields is itself a Hamiltonian vector field

Baker—Campbell—Hausdorff (BCH) Formula

- If A and B belong to any (non-commutative) algebra then $e^{\hat{A}}e^{\hat{B}} = e^{\hat{A}+\hat{B}+\hat{\delta}}$, where $\hat{\delta}$ is constructed from commutators of $\hat{A}$ and $\hat{B}$
  - *i.e.*, $\hat{\delta}$ is in the Free Lie Algebra generated by $\hat{A}$ and $\hat{B}$
  - More precisely, $\ln\left(e^{\hat{A}}e^{\hat{B}}\right) = \sum_{n\geq 1}\widehat{c_n}$ where $\widehat{c_1} = \hat{A}+\hat{B}$ and

$$(n+1)\widehat{c_{n+1}} = -\tfrac{1}{2}\left[\widehat{c_n}, \hat{A}-\hat{B}\right] + \sum_{m=0}^{\lfloor n/2\rfloor}\frac{B_{2m}}{(2m)!}\sum_{\substack{k_1,\ldots,k_{2m}\geq 1\\ k_1+\cdots+k_{2m}=n}}\left[\widehat{c_{k_1}},\left[\ldots,\left[\widehat{c_{k_{2m}}},\hat{A}+\hat{B}\right]\ldots\right]\right]$$

Symplectic Integrators

- Explicitly, the first few terms are

$$\ln\left(e^{\hat{A}}e^{\hat{B}}\right)=\left\{\hat{A}+\hat{B}\right\}+\tfrac{1}{2}\left[\hat{A},\hat{B}\right]+\tfrac{1}{12}\left\{\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]-\left[\hat{B},\left[\hat{A},\hat{B}\right]\right]\right\}-\tfrac{1}{24}\left[\hat{B},\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]$$
$$+\tfrac{1}{720}\left\{\begin{array}{l}-\left[\hat{A},\left[\hat{A},\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\right]-4\left[\hat{B},\left[\hat{A},\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\right]-6\left[\left[\hat{A},\hat{B}\right],\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\\+4\left[\hat{B},\left[\hat{B},\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\right]-2\left[\left[\hat{A},\hat{B}\right],\left[\hat{B},\left[\hat{A},\hat{B}\right]\right]\right]+\left[\hat{B},\left[\hat{B},\left[\hat{B},\left[\hat{A},\hat{B}\right]\right]\right]\right]\end{array}\right\}+\cdots$$

- In order to construct reversible integrators we use symmetric symplectic integrators
- The following identity follows directly from the BCH formula

$$\ln\left(e^{\hat{A}/2}e^{\hat{B}}e^{\hat{A}/2}\right)=\left\{\hat{A}+\hat{B}\right\}+\tfrac{1}{24}\left\{\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]-2\left[\hat{B},\left[\hat{A},\hat{B}\right]\right]\right\}$$
$$+\tfrac{1}{5760}\left\{\begin{array}{l}7\left[\hat{A},\left[\hat{A},\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\right]+28\left[\hat{B},\left[\hat{A},\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\right]+12\left[\left[\hat{A},\hat{B}\right],\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\\+32\left[\hat{B},\left[\hat{B},\left[\hat{A},\left[\hat{A},\hat{B}\right]\right]\right]\right]-16\left[\left[\hat{A},\hat{B}\right],\left[\hat{B},\left[\hat{A},\hat{B}\right]\right]\right]+8\left[\hat{B},\left[\hat{B},\left[\hat{B},\left[\hat{A},\hat{B}\right]\right]\right]\right]\end{array}\right\}+\cdots$$

Symplectic Integrators

- From the BCH formula we find that the PQP symmetric symplectic integrator is given by

$$\left(e^{\frac{1}{2}\delta\tau\hat{S}}e^{\delta\tau\hat{T}}e^{\frac{1}{2}\delta\tau\hat{S}}\right)^{\tau/\delta\tau}$$

$$= \left(\exp\left[\left(\hat{T}+\hat{S}\right)\delta\tau - \frac{1}{24}\left(\left[\hat{S},\left[\hat{S},\hat{T}\right]\right]+2\left[\hat{T},\left[\hat{S},\hat{T}\right]\right]\right)\delta\tau^3 + O\left(\delta\tau^5\right)\right]\right)^{\tau/\delta\tau}$$

$$= \exp\left[\tau\left(\hat{T}+\hat{S}-\frac{1}{24}\left(\left[\hat{S},\left[\hat{S},\hat{T}\right]\right]+2\left[\hat{T},\left[\hat{S},\hat{T}\right]\right]\right)\delta\tau^2 + O\left(\delta\tau^4\right)\right)\right]$$

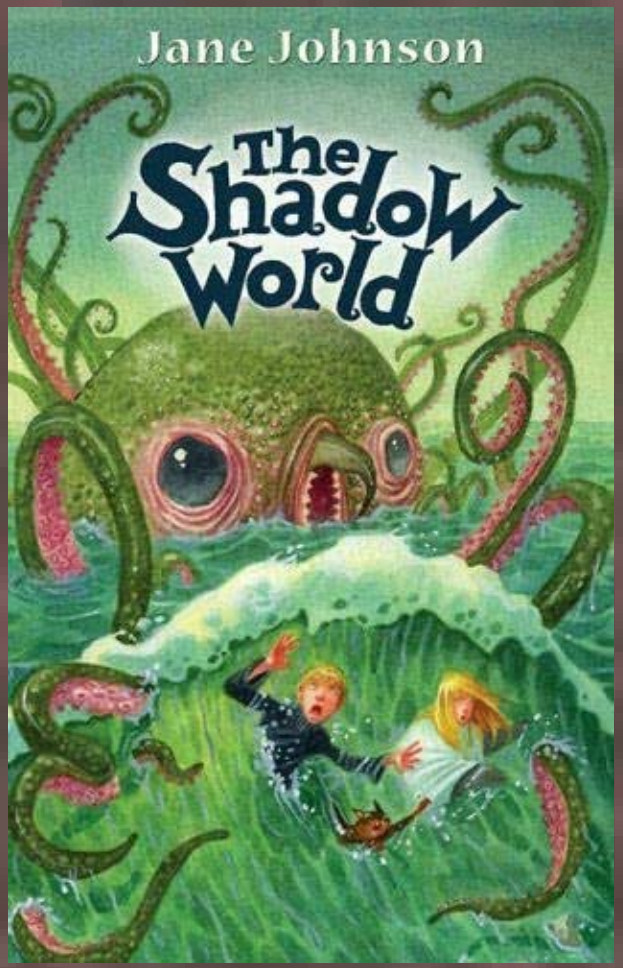
$$= e^{\tau\widehat{H}'(\delta\tau)} = e^{\tau\widehat{H}+O\left(\delta\tau^2\right)} = e^{\tau\widehat{H}} + O\left(\delta\tau^2\right)$$

- In addition to conserving energy to $O(\delta\tau^2)$ such symmetric symplectic integrators are manifestly area preserving and reversible

Shadow Hamiltonian

- But more significantly the PQP integrator follows the integral curves of $\widehat{H}'(\delta\tau)$ exactly
- And $\widehat{H}'(\delta\tau)$ is constructed from commutators of the Hamiltonian vector fields $\widehat{S}$ and $\widehat{T}$
- Therefore it is the Hamiltonian vector field of the corresponding combination of Poisson brackets $H'(\delta\tau)$
- This is called the shadow Hamiltonian

Leapfrog Shadow Hamiltonian



- For the PQP integrator we have

$$H'_{PQP}(\delta\tau) = T + S + \tfrac{\delta\tau^2}{24}\Big[\{S,\{S,T\}\} - 2\{T,\{S,T\}\}\Big]$$
$$+ \tfrac{\delta\tau^4}{5760}\begin{bmatrix} 7\{S,\{S,\{S,\{S,T\}\}\}\} \\ +28\{T,\{S,\{S,\{S,T\}\}\}\} \\ +12\{\{S,T\}\{S,\{S,T\}\}\} \\ +32\{T,\{T,\{S,\{S,T\}\}\}\} \\ -16\{\{S,T\}\{T,\{S,T\}\} \\ +8\{T,\{T,\{T,\{S,T\}\}\}\} \end{bmatrix} + \cdots$$

Leapfrog Shadow Hamiltonian

- Evaluating the Poisson brackets with $\omega = dq \wedge dp$ gives

$$H'_{PQP}(\delta\tau) = H + \frac{\delta\tau^2}{24}\left\{ 2p^2 S'' - S'^2 \right\}$$

$$+ \frac{\delta\tau^4}{720}\left\{ -p^4 S^{(4)} + 6p^2\left(S'S''' + 2S''^2\right) - 3S'^2 S'' \right\} + O\left(\delta\tau^6\right)$$

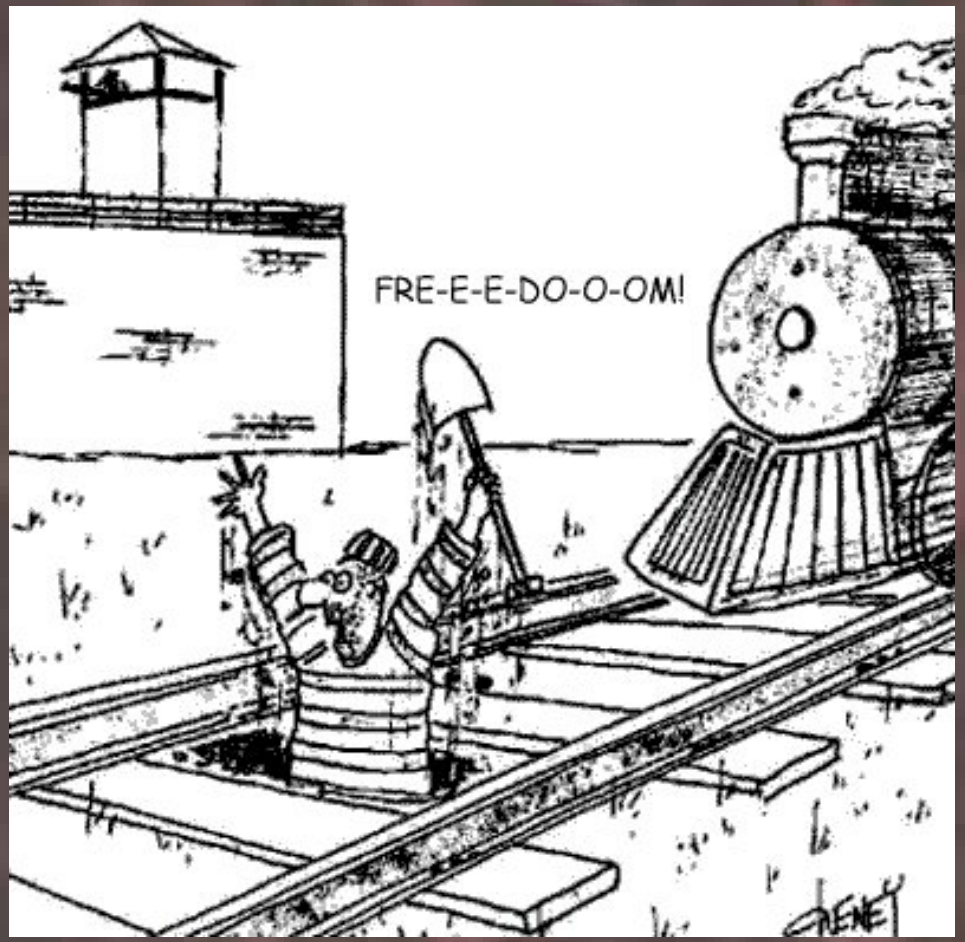
- Note that H'_{PQP} cannot be written as the sum of a p-dependent kinetic term and a q-dependent potential term
  - So, sadly, it is not possible to construct an integrator that conserves the Hamiltonian we started with

What use are Shadows?

- An integrator becomes unstable when the BCH (asymptotic) expansion for its shadow fails to converge
  - In which case there is no (real) conserved shadow Hamiltonian
- Use the shadow to tune an integrator
  - Reduce large contributions to the shadow Hamiltonian
  - Optimize the integrator by minimizing $\Delta H \equiv H - H'$?
    - Not quite, as we shall see later

Gauge Theories

- But first there are a few details that we shouldn't overlook
- Can we compute Poisson brackets and shadow Hamiltonians for gauge fields and fermions?



Inexact Algorithms

- Let us take a small detour to consider what happens the an HMC-like algorithm if we omit the Metropolis step.
- If such an algorithm is ergodic it has a unique fixed point

$$\begin{aligned} e^{-(S(q')+\Delta S(q'))} &= \int dq\, e^{-(S(q)+\Delta S(q))} \int dp\, e^{-\frac{1}{2}p^2}\, \delta(q'-q'') \\ &= \int dq\, dp\, e^{-H(q,p)-\Delta S(q)}\, \delta(q'-q'') \\ &= \int dq''\, dp'' \left(\det U_*\right)^{-1} e^{-(H+\Delta S)\circ U^{-1}}\, \delta(q'-q'') \end{aligned}$$

where the MD evolution is $U:(q,p)\mapsto(q'',p'')$
and the Jacobian is $\det \frac{\partial(q'',p'')}{\partial(q,p)} = \det U_* = \mathrm{tr}\,\ln\, U_*$

  - We assume that the momenta a completely refreshed from a Gaussian heatbath before each trajectory
  - Momentum distribution is not Gaussian at the end of a trajectory

Inexact Algorithms

- If we define the quantities

  $\bar{\delta} : \Omega \mapsto \Omega \circ U^{-1} - \Omega \circ F \circ U \circ F$ violation of reversibility

  $\delta : \Omega \mapsto \Omega \circ U \circ F - \Omega \circ F$ change in Ω over a trajectory

  then we obtain $\left\langle e^{-\left(\delta+\bar{\delta}\right)(H+\Delta S)-\mathrm{tr}\ln U_*} \right\rangle_p = 1$

- This may be expanded *as an asymptotic series* in the integration step size $\delta\tau$ to obtain an expression for ΔS
  - This shows that ΔS is a power series in $\delta\tau$ up to terms like $e^{-C/\delta\tau}$, where the constant C is not necessarily real
  - The form of ΔS is not known in closed form for long trajectories


ITN STRONG net
THE UNIVERSITY OF EDINBURGH
Gauge Fields

Review

Symplectic 2-form

Hamiltonian vector field

Equations of motion

Poisson bracket

Flat Manifold

$$dq \wedge dp$$

$$\hat{H} = \frac{\partial H}{\partial p}\frac{\partial}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial}{\partial p}$$

$$\dot{q} = \frac{\partial H}{\partial p}$$

$$\{A, B\} = \frac{\partial A}{\partial p}\frac{\partial B}{\partial q} - \frac{\partial A}{\partial q}\frac{\partial B}{\partial p}$$

General

$$\omega : d\omega = 0$$

$$\{\hat{A}, \hat{B}\} = -\omega(\hat{A}, \hat{B})$$

Darboux theorem: All manifolds are locally flat

Lie Group Manifolds

- A Lie Group is a differential manifold with a globally well-defined smooth multiplication $h : g \mapsto hg$
- This induces a smooth map on 0-forms (functions) $h_* : f \mapsto f \circ h$, *i.e.*, $h_* f(g) = f(hg)$
- A left-invariant 0-form satisfies $f = h_* f$ for any h, so it is specified by its value at the origin $f(1) = g_* f(1) = f(g)$ — a constant, not very interesting
- Similarly, we have smooth induced map on vector fields (linear differential operators) $h^* : v \mapsto v \circ h_*$
- A left-invariant vector field satisfies $v = h^* v \; (\forall h)$ $h^* v(f) = v(h_* f) = v(f \circ h)$

Lie Group Manifolds

- An frame of vectors $\{e_j\}$ at the origin can therefore be extended to a frame of left-invariant vector fields over the whole group $e_i(h) = h^* e_i(0)$
  - For a classical matrix group this is quite intuitive: each element g of the group is associated with the frame which is obtained from that at the origin by the action of g
  - While there is a globally well-defined basis of left-invariant frame vector fields, but there are no globally well-defined coordinates
- Their commutators satisfy $\left[e_i, e_j\right] = \sum_k c_{ij}^k e_k$ with coefficients called structure constants

Maurer—Cartan Equations

- The dual left invariant forms $\{\theta^i\}$ with $\theta^i(e_j) = \delta^i_j$ satisfy the Maurer—Cartan equations $d\theta^i = -\frac{1}{2}\sum_{jk} c^i_{jk}\,\theta^j \wedge \theta^k$

$$
\begin{aligned}
d\theta^i(e_j, e_k) &= e_j\theta^i(e_k) - e_k\theta^i(e_i) - \theta^i\left([e_j, e_k]\right) \\
&= e_j\delta^i_k - e_k\delta^i_j - \theta^i\left(c^\ell_{jk} e_\ell\right) = -c^i_{jk} \\
&= -\tfrac{1}{2}\sum_{j',k'} c^i_{j'k'}\theta^{j'} \wedge \theta^{k'}(e_j, e_k)
\end{aligned}
$$

- The Maurer—Cartan forms also provide the group-invariant Haar measure $d\Omega \equiv \theta_1 \wedge \theta_2 \wedge \ldots \wedge \theta_n$

Fundamental 2-form

- We can invent any Classical Mechanics we want…
- So we may therefore define the closed fundamental 2-form ω which globally defines an invariant Poisson bracket by

$$\omega \equiv -d\sum_i \theta^i p^i = \sum_i \left(\theta^i \wedge dp^i - p^i d\theta^i \right)$$

$$= \sum_i \left(\theta^i \wedge dp^i + \tfrac{1}{2} p^i c^i_{jk} \theta^j \wedge \theta^k \right)$$

Hamiltonian Vector Field

- We may now follow the usual procedure to find the equations of motion
- Introduce a Hamiltonian function (0-form) H on the cotangent bundle (phase space) over the group manifold
- Define a vector field $\hat{H}$ such that $dH = i_{\hat{H}}\omega$

$$\hat{H} = \sum_i \left(\frac{\partial H}{\partial p^i} e_i + \left[\sum_{jk} c^k_{ji} p^k \frac{\partial H}{\partial p^j} - e_i(H) \right] \frac{\partial}{\partial p^i} \right)$$

Poisson Brackets

- For any Hamiltonian vector field $\hat{A}$

$$\hat{A} = \sum_i \left(\frac{\partial A}{\partial p^i} e_i + \left[\sum_{jk} c^k_{ji} p^k \frac{\partial A}{\partial p^j} - e_i(A) \right] \frac{\partial}{\partial p^i} \right)$$

- So for $H(q,p) = T(p) + S(q)$ we have vector fields

$$\hat{S} = -\sum_i e_i(S) \frac{\partial}{\partial p^i}$$

$$\hat{T} = \sum_i \left(\frac{\partial T}{\partial p^i} e_i + \left[\sum_{jk} c^k_{ji} p^k \frac{\partial T}{\partial p^j} \right] \frac{\partial}{\partial p^i} \right)$$

$$= \sum_i \left(p^i e_i + \sum_{jk} c^k_{ji} p^k p^j \frac{\partial}{\partial p^i} \right) \text{ if } T(p) = \frac{p^2}{2}$$

More Poisson Brackets

- We thus compute the lowest-order Poisson bracket

$$\{S,T\} = -\omega\left(\hat{S},\hat{T}\right)$$

$$= -\left(\theta^i \wedge dp^i + \tfrac{1}{2} p^i c^i_{jk} \theta^j \wedge \theta^k\right)\left(\hat{S},\hat{T}\right) = -p^i e_i(S)$$

- and the Hamiltonian vector corresponding to it

$$\widehat{\{S,T\}} = \sum_i \left(\frac{\partial\{S,T\}}{\partial p^i} e_i + \left[\sum_{jk} c^k_{ji} p^k \frac{\partial\{S,T\}}{\partial p^j} - e_i\left(\{S,T\}\right) \right] \frac{\partial}{\partial p^i} \right)$$

$$= -e_i(S) e_i + \left[-c^k_{ji} p^k e_j(S) + p^j e_i e_j(S) \right] \frac{\partial}{\partial p^i}$$

Even More Poisson Brackets

$$\{S,\{S,T\}\} = e_i(S)e_i(S)$$

$$\{T,\{S,T\}\} = -p^i p^j e_i e_j(S)$$

$$\{T,\{T,\{S,\{S,T\}\}\}\} = 2p^i p^j \left[e_i e_j e_k(S)e_k(S) + e_i e_k(S)e_j e_k(S)\right]$$

$$\{\{S,T\},\{T,\{S,T\}\}\} = c^i_{jk} p^i p^\ell e_j(S)\left[e_k e_\ell(S) + e_\ell e_k(S)\right] +$$

$$+ p^i p^j \left[e_k(S)e_k e_i e_j(S) - \left[e_k e_i(S) + e_i e_k(S)\right]e_k e_j(S)\right]$$

$$\{T,\{S,\{S,\{S,T\}\}\}\} = 0$$

$$\{\{S,T\},\{S,\{S,T\}\}\} = -2e_i(S)e_j(S)e_i e_j(S)$$

$$\{T,\{T,\{T,\{S,T\}\}\}\} = -p^i p^j p^k p^\ell e_i e_j e_k e_\ell(S)$$

$$\{S,\{S,\{S,\{S,T\}\}\}\} = 0$$

Computing Poisson Brackets

- These are quite complicated (some might say disgusting) objects to compute on the lattice
  - Even for the simplest Wilson gauge action
  - They consists sums of complicated lattice loops with momenta inserted in various places
- Fortunately there is a recursive way of computing them which is tractable even for more complicated gauge actions
  - It involves inserting previously computed Lie-algebra-valued fields living on links into the loops in the action using a "tower" algorithm

Fermion Poisson Brackets

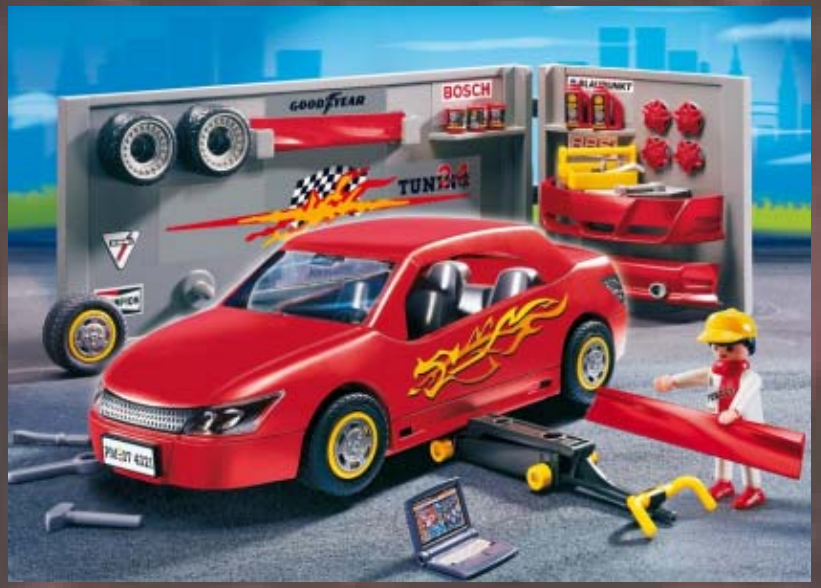
- Fermions are easy to include in the formalism: we only need a few extra linear equation solves

$$S_F(U) = \phi^\dagger \mathcal{M}^{-1}(U)\phi = \mathrm{Tr}\left[\mathcal{M}^{-1}(U)\phi \otimes \phi^\dagger\right]$$

$$\frac{\partial \mathcal{M}^{-1}}{\partial U} = -\mathcal{M}^{-1}\frac{\partial \mathcal{M}}{\partial U}\mathcal{M}^{-1}$$

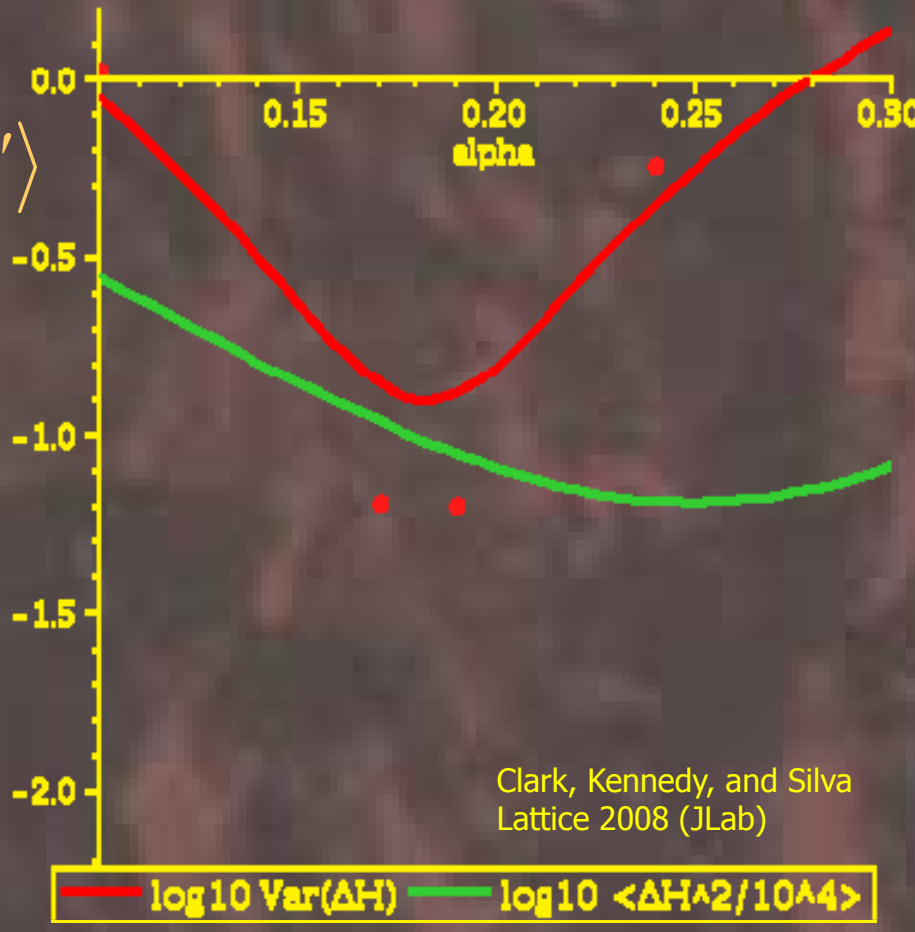
Tuning Your Integrator

- For *any* (symmetric) symplectic integrator the conserved Hamiltonian is constructed from the same Poisson brackets
- A procedure for tuning such integrators is
  - Measure the Poisson brackets during an HMC run
  - Optimize the integrator (number of pseudofermions, step-sizes, order of integration scheme, etc.) offline using these measured values



What to Tune

- As I said a while ago, minimizing $\langle \Delta H \rangle \equiv \langle H - H' \rangle$ is not a good choice
- It is much better to minimize the variance of ΔH
  - This is a function of two sets of quantities
    - The ensemble-averaged Poisson brackets
    - The integrator parameters



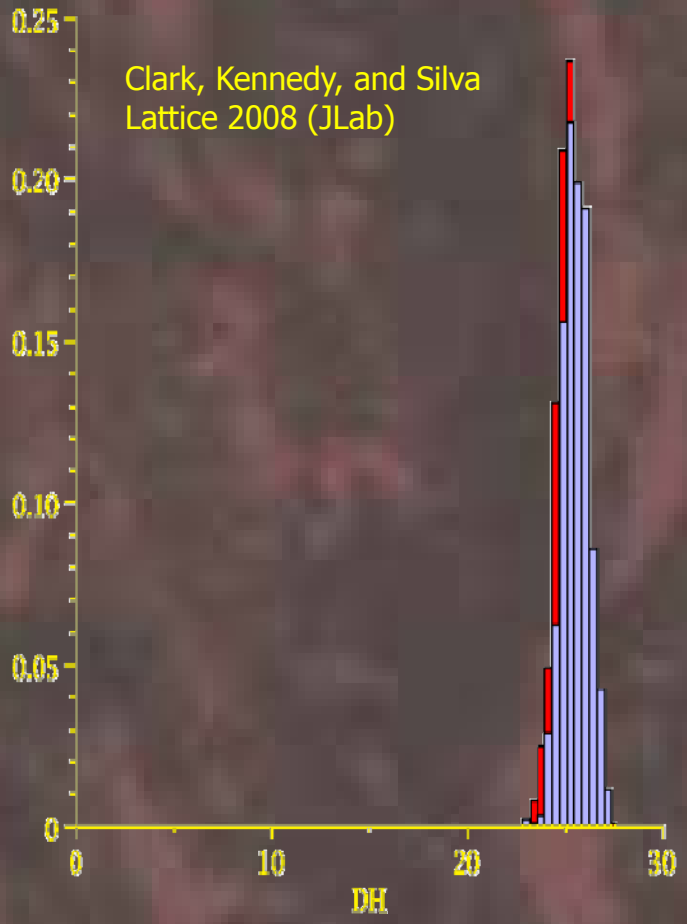
Clark, Kennedy, and Silva
Lattice 2008 (JLab)

Why Minimize the Variance?

- As the system wanders through phase space H' is constant, so

$$\delta H = H_f - H_i = \delta \Delta H$$

- We hypothesize that the distribution of ΔH is essentially sampled independently and randomly at the start and end of each equilibrium trajectory
- Therefore we want to minimize the variance of this distribution



Simplest Integrators

Integrator	Update Steps	Shadow Hamiltonian
PQP	$e^{\hat{S}/2} e^{\hat{T}} e^{\hat{S}/2}$	$T + S - \frac{\varepsilon^2}{24}\left(\{S,\{S,T\}\} + 2\{T,\{S,T\}\}\right)$
QPQ	$e^{\frac{\hat{T}\varepsilon}{2}} e^{\hat{S}\varepsilon} e^{\frac{\hat{T}\varepsilon}{2}}$	$T + S + \frac{\varepsilon^2}{24}\left(2\{S,\{S,T\}\} + \{T,\{S,T\}\}\right)$
Omelyan *SST*	$e^{\frac{\hat{S}\varepsilon}{6}} e^{\frac{\hat{T}\varepsilon}{2}} e^{\frac{2\hat{S}\varepsilon}{3}} e^{\frac{\hat{T}\varepsilon}{2}} e^{\frac{\hat{S}\varepsilon}{6}}$	$T + S + \frac{\varepsilon^2}{72}\{S,\{S,T\}\}$
Omelyan *TST*	$e^{\frac{(3-\sqrt{3})\hat{S}\varepsilon}{6}} e^{\frac{\hat{T}\varepsilon}{2}} e^{\frac{\hat{S}\varepsilon}{\sqrt{3}}} e^{\frac{\hat{T}\varepsilon}{2}} e^{\frac{(3-\sqrt{3})\hat{S}\varepsilon}{6}}$	$T + S + \frac{\sqrt{3}-2}{24}\varepsilon^2\{\mathrm{T},\{\mathrm{S},\mathrm{T}\}\}$

Campostrini Integrators

- Campostrini found an ingenious way of constructing integrators with errors of arbitrarily high order
  - Start with an integrator with errors of order n in the integration step size $X_n(\varepsilon) \equiv e^{\varepsilon \hat{H}'(\varepsilon)} = e^{\varepsilon \hat{H}} \left[1 + \varepsilon^n \hat{\Delta} + O\left(\varepsilon^{n+2}\right) \right]$
  - Construct the "wiggle" sandwiching a backward step of this integrator between two forward ones

    $$X_n(\varepsilon) X_n(-\sigma\varepsilon) X_n(\varepsilon) = e^{\varepsilon(2-\sigma)\hat{H}} \left[1 + \varepsilon^n \left(2 - \sigma^n\right) \hat{\Delta} + O\left(\varepsilon^{n+2}\right) \right]$$
  - Eliminate the leading order error by choosing $\sigma = \sqrt[n]{2}$ and adjust the step size by setting $\delta\tau = \varepsilon(2-\sigma)$ to obtain $X_{n+2}(\delta\tau) = X_n(\varepsilon) X_n(-\sigma\varepsilon) X_n(\varepsilon) = e^{\delta\tau \hat{H}} \left[1 + O\left(\delta\tau^{n+2}\right) \right]$

Campostrini Integrators

Integrator	Campostrini
Update Steps	$$e^{\frac{\sqrt[3]{4}+2\sqrt[3]{2}+4}{12}\hat{T}\varepsilon}e^{\frac{\sqrt[3]{4}+2\sqrt[3]{2}+4}{6}\hat{S}\varepsilon}\times e^{\frac{-\sqrt[3]{4}-2\sqrt[3]{2}+2}{12}\hat{T}\varepsilon}e^{-\frac{\sqrt[3]{4}+2\sqrt[3]{2}+1}{3}\hat{S}\varepsilon}e^{\frac{-\sqrt[3]{4}-2\sqrt[3]{2}+2}{12}\hat{T}\varepsilon}\times e^{\frac{\sqrt[3]{4}+2\sqrt[3]{2}+4}{6}\hat{S}\varepsilon}e^{\frac{\sqrt[3]{4}+2\sqrt[3]{2}+4}{12}\hat{T}\varepsilon}$$
Shadow Hamiltonian	$$T+S+\frac{\left(\begin{array}{l}-\left(40\sqrt[3]{4}+40\sqrt[3]{2}+48\right)\{S,\{S,\{S,\{S,T\}\}\}\}+\left(20\sqrt[3]{2}+32\right)\{T,\{T,\{S,\{S,T\}\}\}\}\\+\left(60\sqrt[3]{4}+80\sqrt[3]{2}+104\right)\{\{S,T\},\{T,\{S,T\}\}\}+\left(-20\sqrt[3]{4}+8\right)\{T,\{S,\{S,\{S,T\}\}\}\}\\+\left(180\sqrt[3]{4}+240\sqrt[3]{2}+312\right)\{\{S,T\},\{S,\{S,T\}\}\}+\left(5\sqrt[3]{2}+8\right)\{T,\{T,\{T,\{S,T\}\}\}\}\end{array}\right)}{34560}\varepsilon^4$$

Force-Gradient Integrators

- An interesting observation is that the Poisson bracket $\{S,\{S,T\}\}$ depends only of q
- We may therefore evaluate the integrator $e^{\overbrace{\{S,\{S,T\}\}}\delta\tau^3}$ explicitly
- The force for this integrator involves second derivatives of the action
- Using this type of step we can construct efficient Force-Gradient (Hessian) integrators

Force-Gradient Integrators

Integrator	Update Steps	Shadow Hamiltonian
Force-Gradient 1	$e^{\frac{\hat{T}\varepsilon}{6}} e^{\frac{3\hat{S}\varepsilon}{8}} e^{\frac{\hat{T}\varepsilon}{3}} \times e^{\frac{48\hat{S}\varepsilon - \widehat{\{S,\{S,T\}\}}\varepsilon^3}{192}} \times e^{\frac{\hat{T}\varepsilon}{3}} e^{\frac{3\hat{S}\varepsilon}{8}} e^{\frac{\hat{T}\varepsilon}{6}}$	$T+S$ $+\frac{\left(\begin{array}{l} 2259\{S,\{S,\{S,\{S,T\}\}\}\}+4224\{T,\{T,\{S,\{S,T\}\}\}\} \\ +768\{\{S,T\},\{T,\{S,T\}\}\}+5616\{T,\{S,\{S,\{S,T\}\}\}\} \\ +3024\{\{S,T\},\{S,\{S,T\}\}\}+896\{T,\{T,\{T,\{S,T\}\}\}\} \end{array}\right)}{6635520}\varepsilon^4$
Force-Gradient 2	$e^{\frac{\hat{S}\varepsilon}{6}} e^{\frac{\hat{T}\varepsilon}{2}} \times e^{\frac{48\hat{S}\varepsilon - \widehat{\{S,\{S,T\}\}}\varepsilon^3}{72}} \times e^{\frac{\hat{T}\varepsilon}{2}} e^{\frac{\hat{S}\varepsilon}{6}}$	$T+S$ $-\frac{\left(\begin{array}{l} 41\{S,\{S,\{S,\{S,T\}\}\}\}+126\{T,\{T,\{S,\{S,T\}\}\}\} \\ +72\{\{S,T\},\{T,\{S,T\}\}\}+84\{T,\{S,\{S,\{S,T\}\}\}\} \\ +36\{\{S,T\},\{S,\{S,T\}\}\}+54\{T,\{T,\{T,\{S,T\}\}\}\} \end{array}\right)}{155520}\varepsilon^4$

Multiple timescales

- Split Hamiltonian into pieces $H(q,p) = T(p) + S_1(q) + S_2(q)$
- $\widehat{T} \equiv T'(p)\dfrac{\partial}{\partial q}$ and $\widehat{S}_i \equiv -S_i'(q)\dfrac{\partial}{\partial p}$, so $\widehat{H} = \widehat{T} + \widehat{S}_1 + \widehat{S}_2$
- Introduce a symmetric symplectic integrator of the form $U_{\mathrm{SW}}(\delta\tau)^{\tau/\delta\tau} = \left(\left[e^{\frac{1}{2n_2}\delta\tau \widehat{S}_2} \right]^{n_2} \left[e^{\frac{1}{2n_1}\delta\tau \widehat{S}_1} \right]^{n_1} e^{\delta\tau \widehat{T}} \left[e^{\frac{1}{2n_1}\delta\tau \widehat{S}_1} \right]^{n_1} \left[e^{\frac{1}{2n_2}\delta\tau_2} \right]^{n_2} \right)^{\tau/\delta\tau}$
  - If $\dfrac{\left\|\widehat{S}_1\right\|}{2n_1} \approx \dfrac{\left\|\widehat{S}_2\right\|}{2n_2} \approx \left\|\widehat{P}\right\|$ then the instability in the integrator is tickled equally by each sub-step
  - This helps if the most expensive force computation does not correspond to the largest force

Fermion Fields

Pseudofermions

- Direct simulation of Grassmann fields is not feasible
  - The problem is *not* that of manipulating anticommuting values in a computer
  - It is that $e^{-S_F} = e^{-\bar{\psi} M \psi}$ is not positive, and thus we get poor importance sampling
- We therefore integrate out the fermion fields to obtain the determinant $\int d\psi d\bar{\psi}\, e^{-\bar{\psi} M \psi} \propto \det(M)$
  - ψ and $\bar{\psi}$ always occur quadratically
  - The overall sign of the exponent is unimportant

Pseudofermions

- Any operator Ω can be expressed solely in terms of the bosonic fields

$$\Omega'(\phi) = \Omega\left(\phi, \frac{\delta}{\delta\psi}, \frac{\delta}{\delta\bar{\psi}}\right) e^{-\bar{\psi}M(\phi)\psi}\Big|_{\psi=\bar{\psi}=0}$$

  - *E.g.,* the fermion propagator is

$$G_\psi(x,y) = \left\langle \psi(x)\bar{\psi}(y)\right\rangle = M^{-1}(x,y)$$

Pseudofermions

- Including the determinant as part of the observable to be measured is not feasible $\left\langle \Omega \right\rangle = \dfrac{\left\langle \det M(\phi)\,\Omega(\phi) \right\rangle_{S_B}}{\left\langle \det M(\phi) \right\rangle_{S_B}}$
  - The determinant is extensive in the lattice volume, thus again we get poor importance sampling
- Represent the fermion determinant as a bosonic Gaussian integral with a non-local kernel $\det M(\phi) \propto \int d\bar{\chi} d\chi\, e^{-\bar{\chi} M^{-1}(\phi) \chi}$
- The fermion kernel must be positive definite (all its eigenvalues must have positive real parts) otherwise the bosonic integral will not converge
- The new bosonic fields are called pseudofermions

Pseudofermions

- It is usually convenient to introduce two flavours of fermion and to write $\left(\det M(\phi)\right)^2 = \det\left(M(\phi)M^\dagger(\phi)\right) \propto \int d\bar{\chi} d\chi\, e^{-\bar{\chi}\left(M^\dagger M\right)^{-1}\chi}$
  - This not only guarantees positivity, but also allows us to generate the pseudofermions from a global heatbath by applying $M^\dagger$ to a random Gaussian distributed field
- The equations for motion for the boson (gauge) fields are

$$\dot{\phi} = \pi$$

$$\dot{\pi} = -\frac{\partial S_B(\phi)}{\partial \phi} - \chi^\dagger \frac{\partial}{\partial \phi}\left(M^\dagger M\right)^{-1}\chi = -\frac{\partial S_B(\phi)}{\partial \phi} + \left[\left(M^\dagger M\right)^{-1}\chi\right]^\dagger \frac{\partial}{\partial \phi}\left(M^\dagger M\right)\left[\left(M^\dagger M\right)^{-1}\chi\right]$$

- The evaluation of the pseudofermion action and the corresponding force then requires us to find the solution of a (large) set of linear equations $\left(M^\dagger M\right)^{-1}\chi = \psi$

Pseudofermions

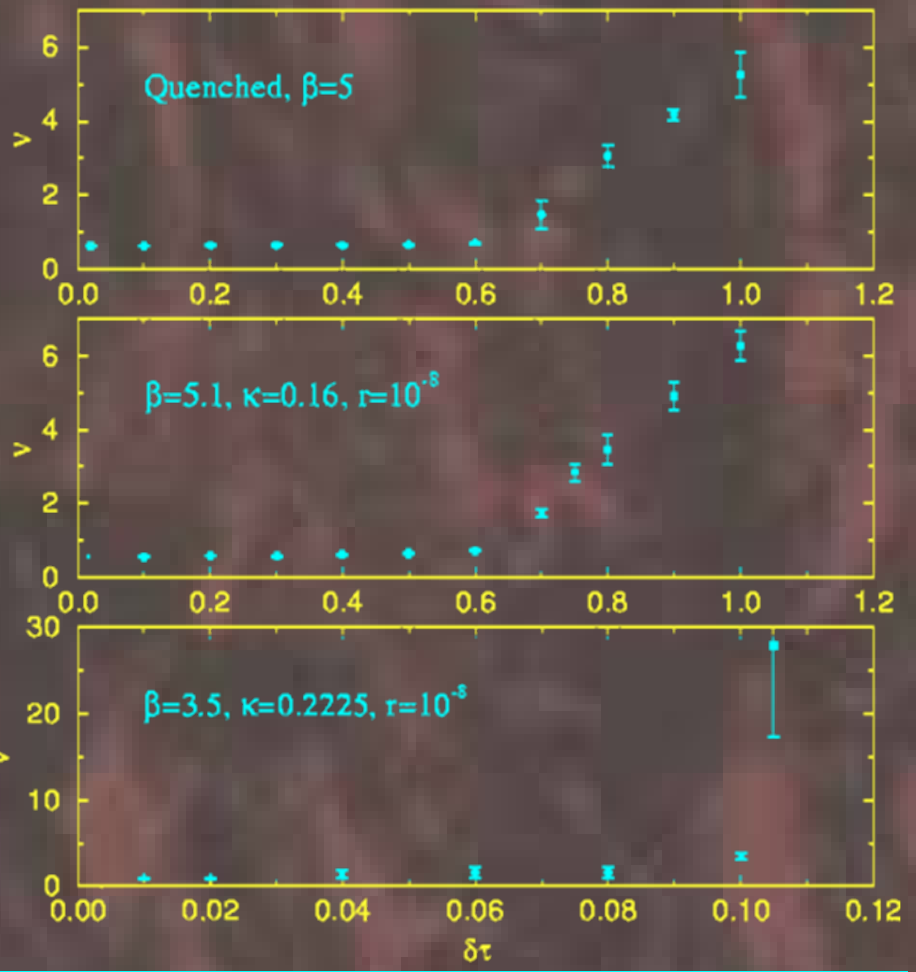
- It is not necessary to carry out the inversions required for the equations of motion exactly
  - There is a trade-off between the cost of computing the force and the acceptance rate of the Metropolis MDMC step
- The inversions required to compute the pseudofermion action for the accept/reject step does need to be computed exactly, however
  - We usually take "exactly" to by synonymous with "to machine precision"

Reversibility

- Are HMC trajectories reversible and area preserving in practice?
  - The only fundamental source of irreversibility is the rounding error caused by using finite precision floating point arithmetic
    - For fermionic systems we can also introduce irreversibility by choosing the starting vector for the iterative linear equation solver time-asymmetrically
      - We do this if we to use a Chronological Inverter, which takes (some extrapolation of) the previous solution as the starting vector
    - Floating point arithmetic is not associative
    - It is more natural to store compact variables as scaled integers (fixed point)
      - Saves memory
      - Does not solve the precision problem

Reversibility

- Data for SU(3) gauge theory and QCD with heavy quarks show that rounding errors are amplified exponentially
  - The underlying continuous time equations of motion are chaotic
  - Ляпунов exponents characterise the divergence of nearby trajectories
  - The instability in the integrator occurs when $\delta H \gg 1$
    - Zero acceptance rate anyhow

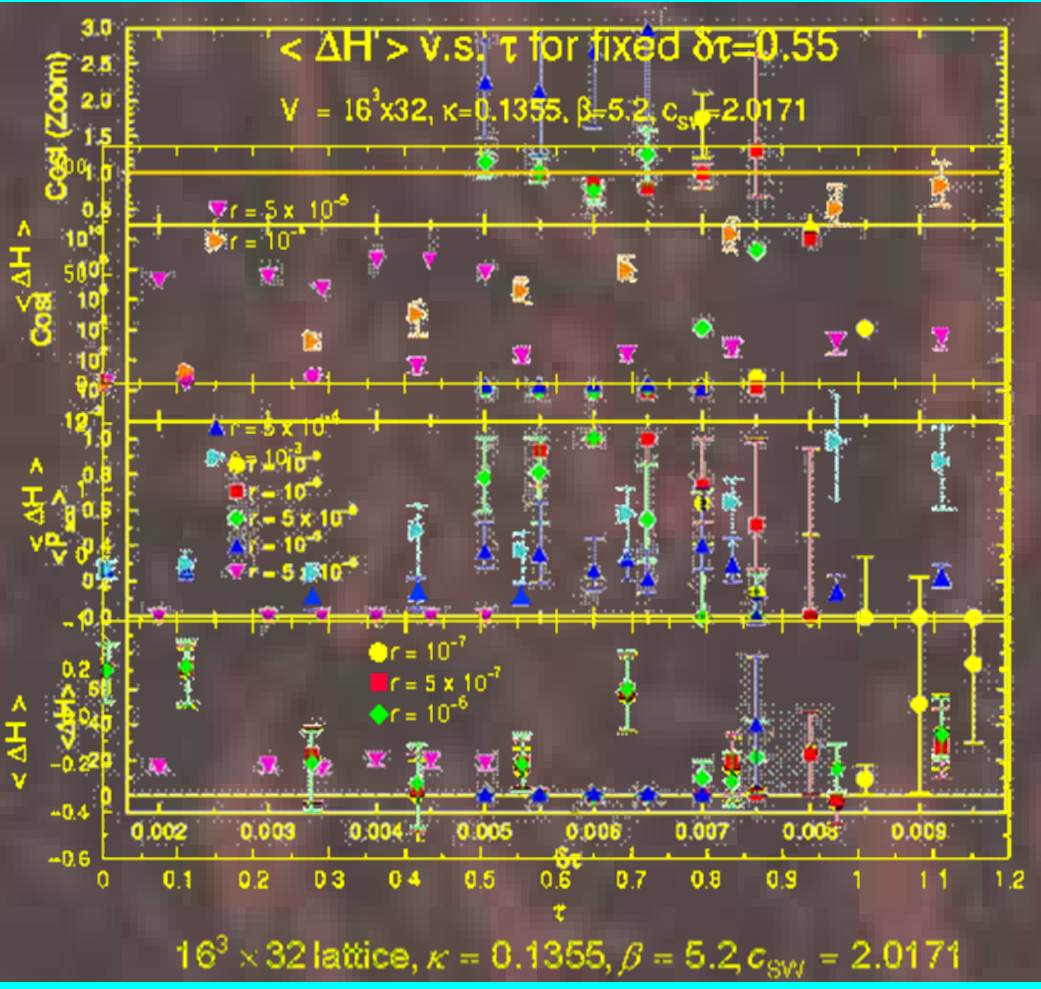


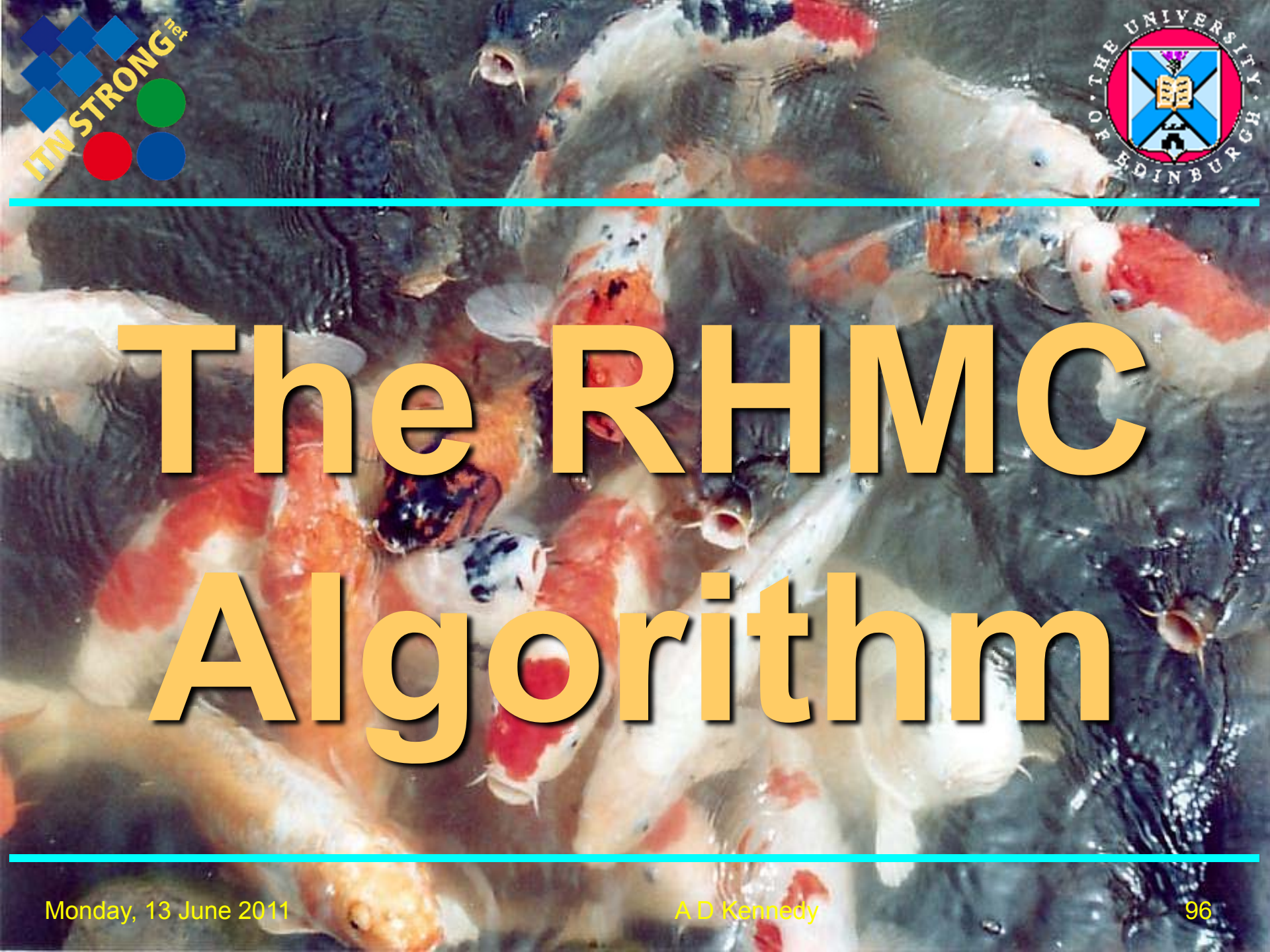
Reversibility

- In QCD the Ляпунов exponents appear to scale with β as the system approaches the continuum limit $\beta \to \infty$
  - $\nu\xi$ = *constant*
  - This can be interpreted as saying that the Ляпунов exponent characterises the chaotic nature of the continuum classical equations of motion, and is not a lattice artefact
  - Therefore we should not have to worry about reversibility breaking down as we approach the continuum limit
  - *Caveat*: data is only for small lattices, and is not conclusive

Reversibility

- Data for QCD with lighter dynamical quarks
  - Instability occurs close to region in $\delta\tau$ where acceptance rate is near one
    - May be explained as a *few* "modes" becoming unstable because of large fermionic force
  - Integrator goes unstable if too poor an approximation to the fermionic force is used





The RHMC Algorithm

Polynomial approximation

- What is the best polynomial approximation *p(x)* to a continuous function *f(x)* for *x* in [*0,1*] ?
  - Best with respect to the appropriate norm

$$\|p - f\|_n = \left(\int_0^1 dx \, |p(x) - f(x)|^n \right)^{1/n}$$

where $n \geq 1$

Weierstraß' theorem

- Taking n $\to\infty$ this is the minimax norm

$$\left\| p - f \right\|_\infty = \min_p \max_{0 \le x \le 1} \left| p(x) - f(x) \right|$$

- Weierstraß: Any continuous function can be arbitrarily well approximated by a polynomial

Бернштейне polynomials

- The explicit solution is provided by Бернштейне polynomials

$$p_n(x) \equiv \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^n (1-x)^{n-k}$$

Чебышев's theorem

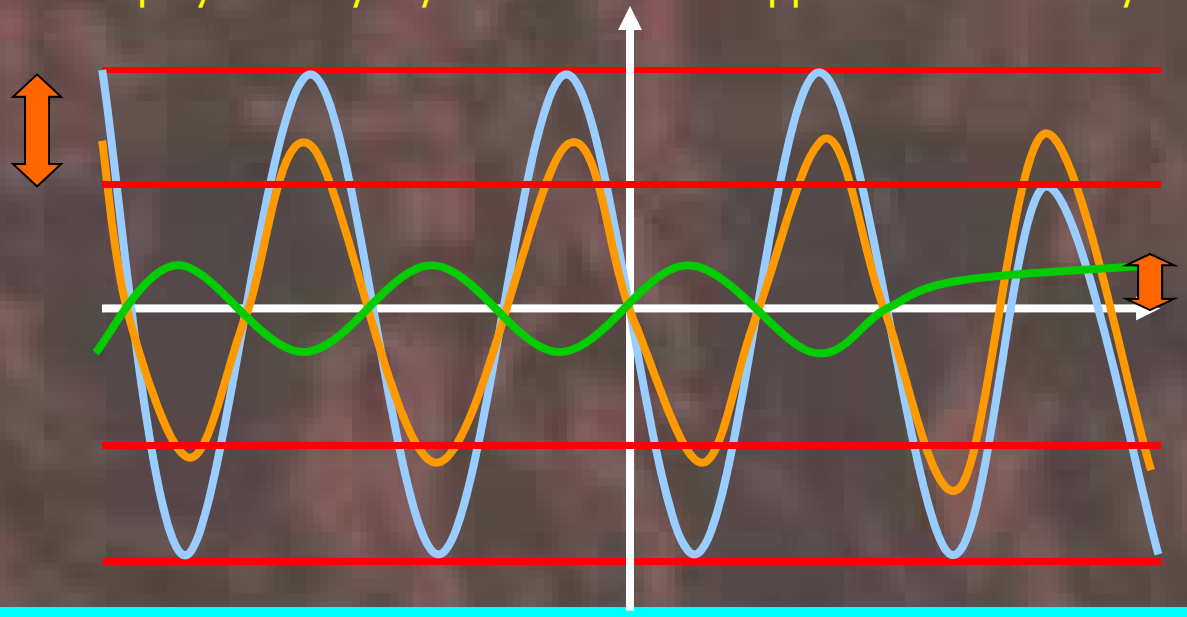
- Чебышев: There is always a unique polynomial of any degree d which minimises

$$\|p - f\|_\infty = \max_{0 \le x \le 1} \left| p(x) - f(x) \right|$$

- The error |p(x) - f(x)| reaches its maximum at exactly d+2 points on the unit interval

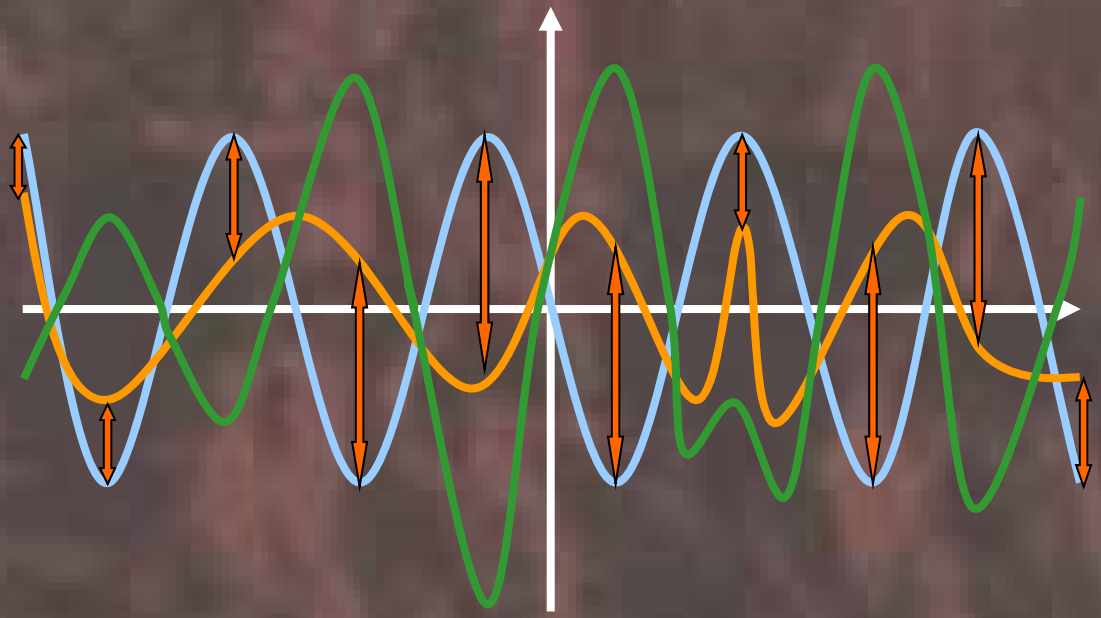
Чебышев's theorem: Necessity

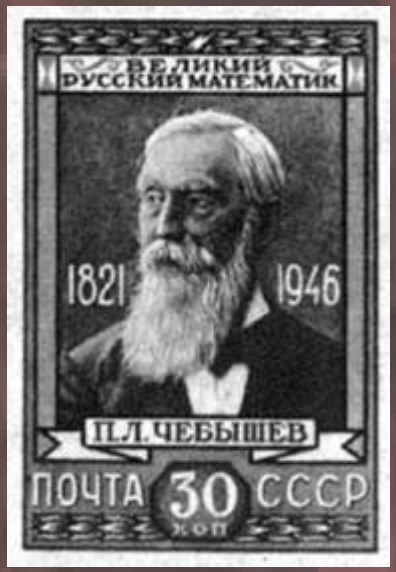
- Suppose *p-f* has less than *d+2* extrema of equal magnitude
- Then at most *d+1* maxima exceed some magnitude
- This defines a "gap"
- We can construct a polynomial *q* of degree *d* which has the opposite sign to *p-f* at each of these maxima (Lagrange interpolation)
- And whose magnitude is smaller than the "gap"
- The polynomial *p+q* is then a better approximation than *p* to *f*



Чебышев's theorem: Sufficiency

- Suppose there is a polynomial $\|p' - f\|_\infty \leq \|p - f\|_\infty$
- Then $|p'(x_i) - f(x_i)| \leq |p(x_i) - f(x_i)|$ at each of the $d+2$ extrema
- Therefore $p' - p$ must have $d+1$ zeros on the unit interval
- Thus $p' - p = 0$ as it is a polynomial of degree d





Чебышев polynomials

- Convergence is often exponential in *d*
  - The best approximation of degree *d-1* over [*-1,1*] to x^d is
$$p_{d-1}(x) \equiv x^d - \left(\frac{1}{2}\right)^{d-1} T_d(x)$$
    - Where the Чебышев polynomials are
$$T_d(x) = \cos\left(d \cos^{-1}(x)\right)$$
    - The notation is an old transliteration of Чебышев !
  - The error is $\left\| x^d - p_d(x) \right\|_\infty = \left(\frac{1}{2}\right)^{d-1} \left\| T_d(x) \right\|_\infty = 2e^{-d \ln 2}$

Чебышев rational functions

- Чебышев's theorem is easily extended to rational approximations
  - Rational functions with nearly equal degree numerator and denominator are usually best
  - Convergence is still often exponential
  - Rational functions usually give a much better approximation

- A simple (but somewhat slow) numerical algorithm for finding the optimal Чебышев rational approximation was given by Ремез

Чебышев rationals: Example

- A realistic example of a rational approximation is

$$\frac{1}{\sqrt{x}} \approx 0.3904603901 \frac{(x + 2.3475661045)(x + 0.1048344600)(x + 0.0073063814)}{(x + 0.4105999719)(x + 0.0286165446)(x + 0.0012779193)}$$

  - This is accurate to within almost *0.1%* over the range [*0.003,1*]

- Using a partial fraction expansion of such rational functions allows us to use a multishift linear equation solver, thus reducing the cost significantly.

- The partial fraction expansion of the rational function above is

$$\frac{1}{\sqrt{x}} \approx 0.3904603901 + \frac{0.0511093775}{x + 0.0012779193} + \frac{0.1408286237}{x + 0.0286165446} + \frac{0.5964845033}{x + 0.4105999719}$$

  - This appears to be numerically stable.

Polynomials versus rationals

- Золотарев's formula has L_∞ error $\Delta \leq e^{\frac{n}{\ln \varepsilon}}$
- Optimal L_2 approximation with weight $\frac{1}{\sqrt{1-x^2}}$ is

$$\sum_{j=0}^{n} \frac{(-)^j 4}{(2j+1)\pi} T_{2j+1}(x)$$

  - This has L_2 error of $O(1/n)$
- Optimal L_∞ approximation cannot be too much better (or it would lead to a better L_2 approximation)

Non-linearity of CG solver

- Suppose we want to solve $A^2x=b$ for Hermitian A by CG
  - It is better to solve $Ax=y$, $Ay=b$ successively
  - Condition number $\kappa(A^2) = \kappa(A)^2$
  - Cost is thus $2\kappa(A) < \kappa(A^2)$ in general
- Suppose we want to solve $Ax=b$
  - Why don't we solve $A^{1/2}x=y$, $A^{1/2}y=b$ successively?
- The square root of A is uniquely defined if $A>0$
  - This is the case for fermion kernels
- All this generalises trivially to n^{th} roots
  - No tuning needed to split condition number evenly
- How do we apply the square root of a matrix?

Rational matrix approximation

- Functions on matrices
  - Defined for a Hermitian matrix by diagonalisation
  - $H = U D U^{-1}$
  - $f(H) = f(U D U^{-1}) = U f(D) U^{-1}$
- Rational functions do not require diagonalisation
  - $\alpha H^m + \beta H^n = U (\alpha D^m + \beta D^n) U^{-1}$
  - $H^{-1} = U D^{-1} U^{-1}$
- Rational functions have nice properties
  - Cheap (relatively)
  - Accurate

No Free Lunch Theorem

- We must apply the rational approximation with each CG iteration
  - $M^{1/n} \approx r(M)$
  - The condition number for each term in the partial fraction expansion is approximately $\kappa(M)$
  - So the cost of applying $M^{1/n}$ is proportional to $\kappa(M)$
  - Even though the condition number $\kappa(M^{1/n})=\kappa(M)^{1/n}$
  - And even though $\kappa(r(M))=\kappa(M)^{1/n}$
- So we don't win this way…

Pseudofermions

- We want to evaluate a functional integral including the fermionic determinant det M
- We write this as a bosonic functional integral over a pseudofermion field with kernel M^{-1}

$$\det M \propto \int d\phi^* \, d\phi \, e^{-\phi^* M^{-1} \phi}$$

Multipseudofermions

- We are introducing extra noise into the system by using a single pseudofermion field to sample this functional integral
  - This noise manifests itself as fluctuations in the force exerted by the pseudofermions on the gauge fields
  - This increases the maximum fermion force
  - This triggers the integrator instability
  - This requires decreasing the integration step size
- A better estimate is $\det M = [\det M^{1/n}]^n$

$$\det M^{\frac{1}{n}} \propto \int d\phi^* \, d\phi \, e^{-\phi^* M^{-\frac{1}{n}} \phi}$$

Hasenbusch's method

- Start with the Wilson fermion kernel $M = 1 - \kappa H$
- Introduce the quantity $M' = 1 - \kappa' H$
- Use the (associative) identity $M = M' M'^{-1} M$
- Introduce separate pseudofermions for each determinant $\det M = \det M' \det\left(M'^{-1} M\right)$
- Adjust κ' to minimise the cost
- Easily generalises
  - More than two pseudofermions
  - Wilson-clover action

Violation of NFL Theorem

- Let's try using our n^{th} root trick to implement multipseudofermions
  - Condition number $\kappa(r(M))=\kappa(M)^{1/n}$
  - So maximum force is reduced by a factor of $n\kappa(M)^{(1/n)-1}$
  - This is a good approximation if the condition number is dominated by a few isolated tiny eigenvalues
  - This is so in the case of interest
- Cost reduced by a factor of $n\kappa(M)^{(1/n)-1}$
  - Optimal value $n_{\text{opt}} \approx \ln \kappa(M)$
  - So optimal cost reduction is $(e \ln\kappa) / \kappa$
- This works!

Rational Hybrid Monte Carlo

- RHMC algorithm for fermionic kernel $\left(\mathcal{M}^\dagger\mathcal{M}\right)^{\frac{1}{2n}}$
  - Generate pseudofermion from Gaussian heatbath

$$P(\xi) \propto e^{-\frac{1}{2}\xi^\dagger\xi} \qquad \chi = \left(\mathcal{M}^\dagger\mathcal{M}\right)^{\frac{1}{4n}}\xi$$

$$P(\chi) \propto \int_{-\infty}^{\infty} d\xi\, e^{-\frac{1}{2}\xi^\dagger\xi}\delta\left(\chi - \left(\mathcal{M}^\dagger\mathcal{M}\right)^{\frac{1}{4n}}\xi\right) \propto e^{-\frac{1}{2}\chi^\dagger\left(\mathcal{M}^\dagger\mathcal{M}\right)^{-\frac{1}{2n}}\chi}$$

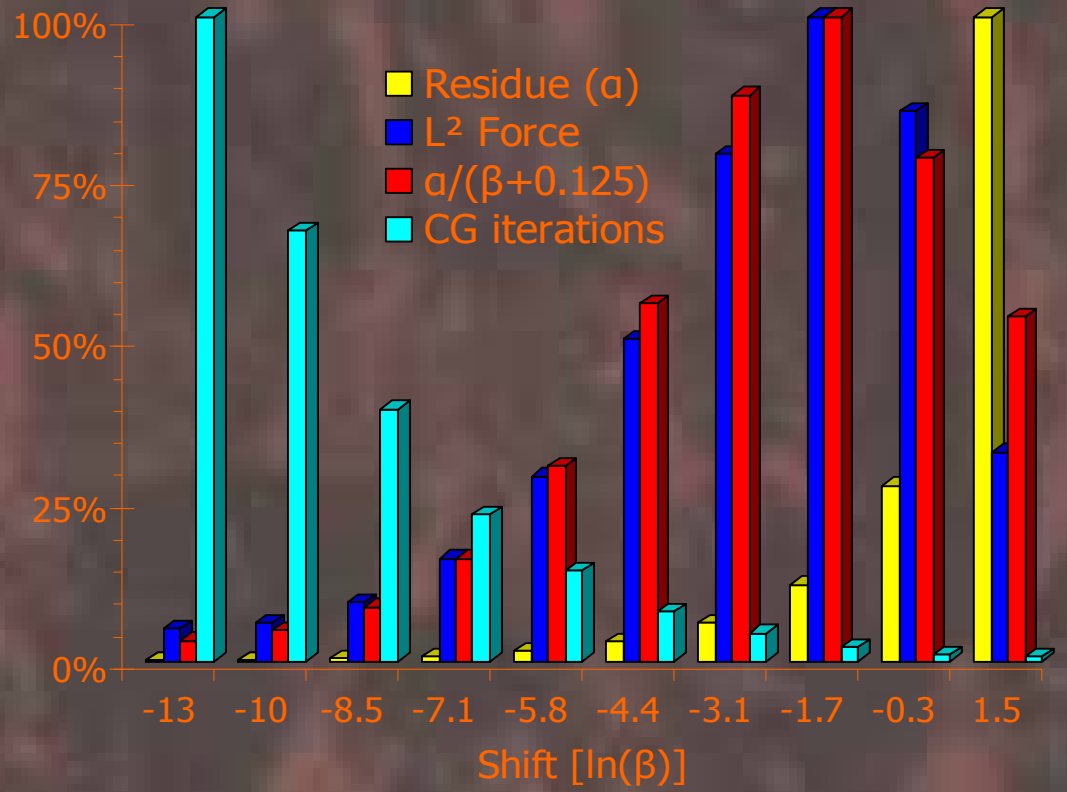
  - Use accurate rational approximation $r(x) \approx \sqrt[4n]{x}$
  - Use less accurate approximation for MD, $\tilde{r}(x) \approx \sqrt[2n]{x}$
    - $\tilde{r}(x) \neq r(x)^2$, so there are no double poles
  - Use accurate approximation for Metropolis acceptance step

Rational Hybrid Monte Carlo

- Reminders
  - Apply rational approximations using their partial fraction expansions
  - Denominators are all just shifts of the original fermion kernel
    - All poles of optimal rational approximations are real and positive for cases of interest (Miracle #1)
    - Only simple poles appear (by construction!)
  - Use multishift solver to invert all the partial fractions using a single Krylov space
    - Cost is dominated by Krylov space construction, at least for *O(20)* shifts
  - Result is numerically stable, even in 32-bit precision
    - All partial fractions have positive coefficients (Miracle #2)
  - MD force term is of the usual form for each partial fraction
  - Applicable to any kernel

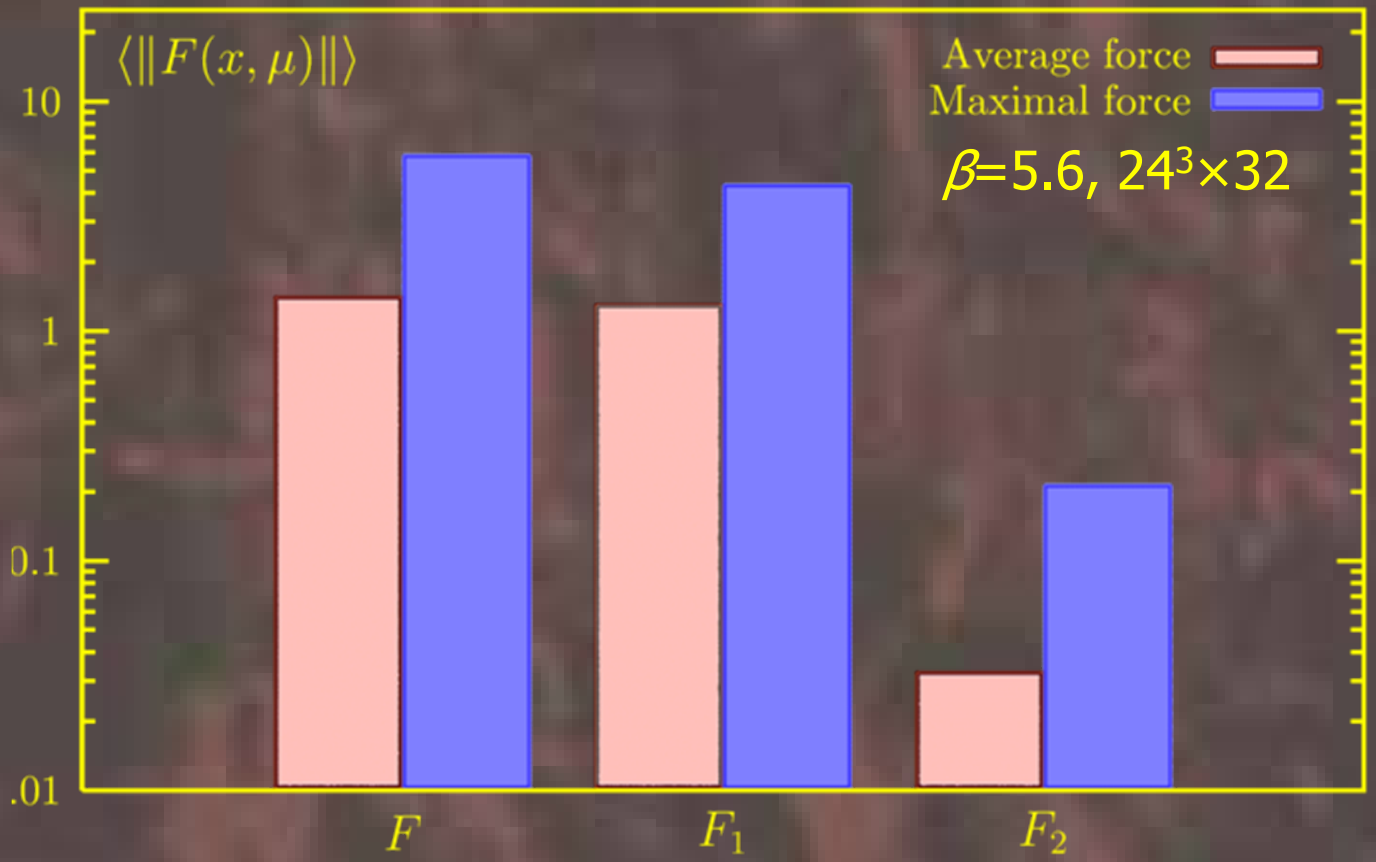
Multipseudofermions with multiple timescales

- Semiempirical observation: The largest force from a single pseudofermion does not come from the smallest shift
  - For example, look at the numerators in the partial fraction expansion we exhibited earlier
- Make use of this by using a coarser timescale for the more expensive smaller shifts



$$\frac{1}{\sqrt{x}} \approx 0.3904603901 + \frac{0.0511093775}{x + 0.0012779193} + \frac{0.1408286237}{x + 0.0286165446} + \frac{0.5964845033}{x + 0.4105999719}$$

L_2 versus L_∞ Force Norms



Wilson fermion forces (from Urbach et. al.)

Conclusions (RHMC)

- Advantages of RHMC
  - Exact
    - No step-size errors; no step-size extrapolations
  - Significantly cheaper than the *R* algorithm
  - Allows easy implementation of Hasenbusch (multipseudofermion) acceleration
  - Further improvements possible
    - Such as multiple timescales for different terms in the partial fraction expansion
- Disadvantages of RHMC
  - Costly for FG integrators (numerous right hand sides)