An Introduction to Monte Carlo Methods for Lattice Quantum Field Theory

A D Kennedy
University of Edinburgh

Model or Theory?

- QCD is highly constrained by symmetry
- Only very few free parameters
  - Coupling constant g
  - Quark masses m
- Scaling behaviour well understood
  - Approach to continuum limit
  - Approach to thermodynamic (infinite volume) limit

Progress in Algorithms

- Important advances made
  - None so far have had a major impact on machine architecture
- Inclusion of dynamical fermion effects
  - Now routine, but expensive
- Improved actions
- New formulations for lattice fermions
  - Chiral fermions can be defined satisfactorily
  - Need algorithms to use new formalism
  - Very compute intensive

Algorithms

- Hybrid Monte Carlo
  - Full QCD including dynamical quarks
  - Most of the computer time spent integrating Hamilton's equations in fictitious (fifth dimensional) time
  - Symmetric symplectic integrators used
    - Known as leapfrog to its friends
    - Higher-order integration schemes go unstable for smaller integration step sizes

Sparse Linear Solvers

- Non-local dynamical fermion effects require solution of large system of linear equations for each time step
  - Krylov space methods used
    - Conjugate gradients, BiCGStab,…
  - Typical matrix size > $10^7 \times 10^7$
    - 32^4 lattice
    - 3 colours
    - 4 spinor components

Functional Integrals and QFT

- The Expectation value of an operator Ω is defined non-perturbatively by the *Functional Integral* $\langle \Omega \rangle \equiv \frac{1}{Z} \int d\phi \, e^{-S(\phi)} \, \Omega(\phi)$
  - Normalisation constant Z chosen such that $\langle 1 \rangle = 1$
  - The action is $S(\phi)$
  - Defined in Euclidean space-time
  - Lattice regularisation
  - $d\phi$ is the appropriate functional measure $d\phi = \prod_x d\phi_x$
  - Continuum limit: lattice spacing $a \to 0$
  - Thermodynamic limit: physical volume $V \to \infty$

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

$$\int_0^{2\pi} dx \sin x$$

Example

- Let's integrate $\sin x$ over the interval $[0, 2\pi]$
  - The integral is

$$I = \int_0^{2\pi} dx \sin x = 0$$

  - Its variance is

$$V = \int_0^{2\pi} dx \left(\sin x\right)^2 - 0 = \pi$$

  - The error (or standard deviation) of our Monte Carlo estimate is thus $\sigma = \pm 1.7725$ ($\sigma^2 = V$)

Monte Carlo Method

Generate a sequence of random field configurations $(\phi_1, \phi_2, \ldots, \phi_t, \ldots, \phi_N)$ 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 $\langle \Omega \rangle \sim \overline{\Omega} + O\left(\sqrt{\frac{C_2}{N}}\right)$
  - 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 $\omega = \Omega(\phi)$

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

Cumulant Expansion

- Generating function for connected moments

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

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

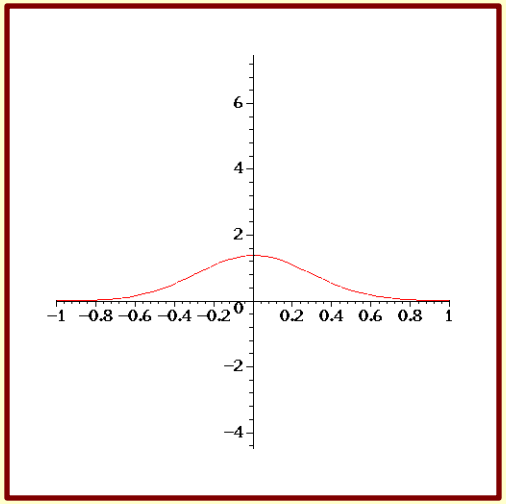
  - The first few *cumulants* are

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

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

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

  - Note that this is an *asymptotic* expansion



Distribution of the Average

- Distribution of the average of *N* samples

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

Connected generating function

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

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

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

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

Central Limit Theorem

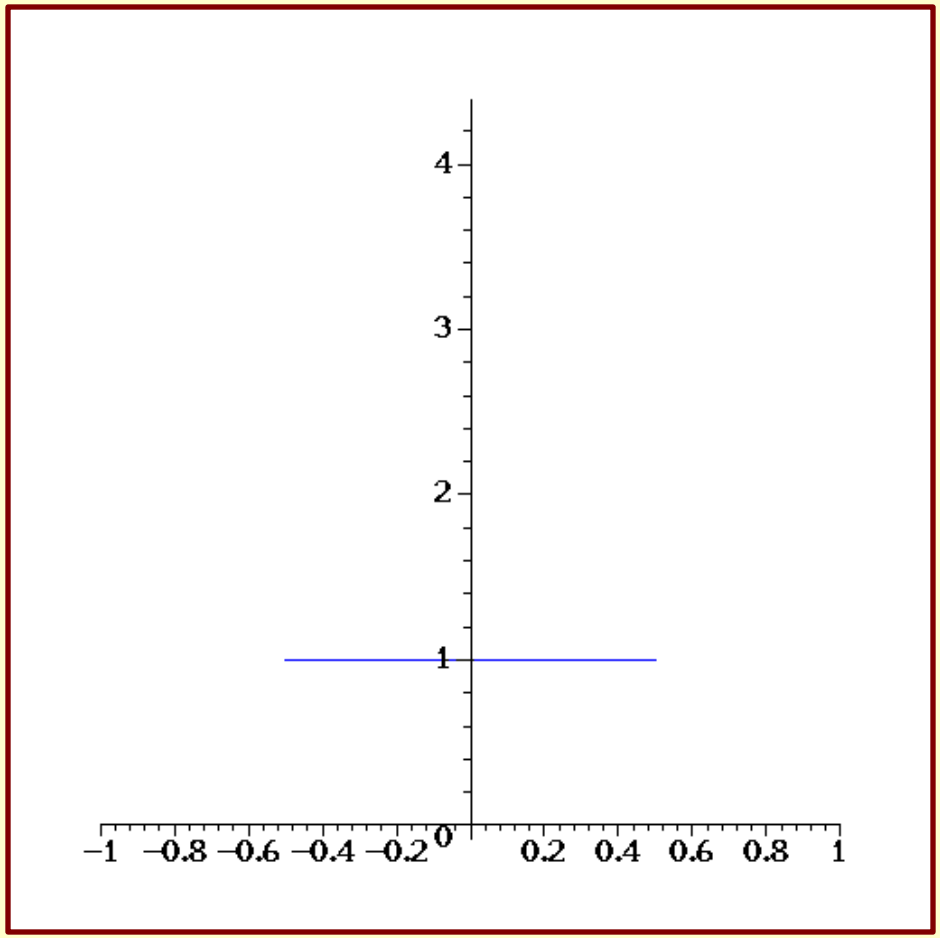
Take inverse Fourier transform to obtain $P_{\bar{\Omega}}$ distribution

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

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

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

Central Limit Theorem



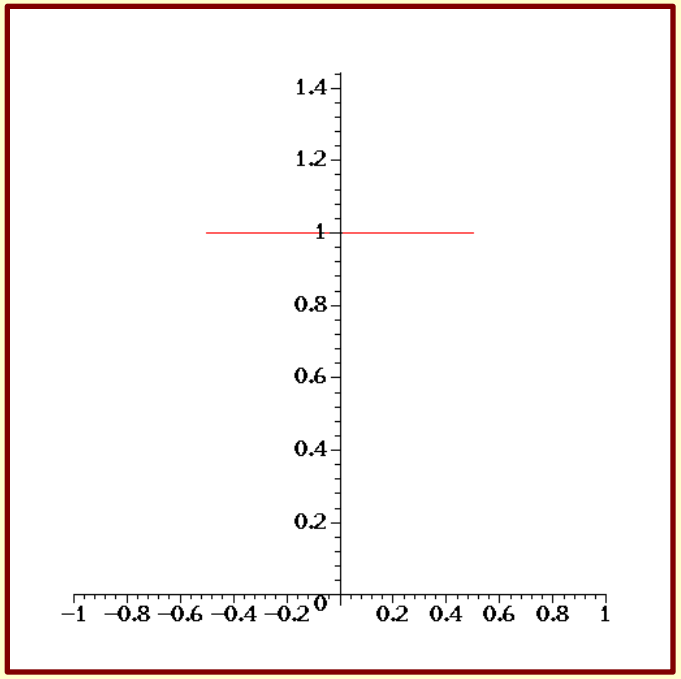
Central Limit Theorem

Re-scale to show convergence to Gaussian distribution

$$P_{\bar{\Omega}}(\bar{\omega}) = F(\xi)\frac{d\xi}{d\bar{\omega}}$$

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

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



Importance Sampling

- Integral $I = \int dx\, f(x)$
- Sample from distribution
  - Probability $0 < \rho(x) \quad a.e.$
  - Normalisation $N = \int dx\, \rho(x) = 1$
- Estimator of integral $I = \int \rho(x) \quad dx \quad \frac{f(x)}{\rho(x)}$
- Estimator of variance

$$V = \int \rho(x)\, dx \left(\frac{f(x)}{\rho(x)} - I \right)^2 = \int dx \frac{f(x)^2}{\rho(x)} - I^2$$

Optimal Importance Sampling

- Minimise variance
  - Constraint *N=1*
  - Lagrange multiplier λ

$$\frac{\delta(V+\lambda N)}{\delta\rho(y)} = -\frac{f(y)^2}{\rho(y)^2} + \lambda = 0$$

- Optimal measure

$$\Rightarrow \rho_{opt}(x) = \frac{|f(x)|}{\int dx \ |f(x)|}$$

- Optimal variance

$$V_{\text{opt}} = \left(\int dx \ |f(x)|\right)^2 - \left(\int dx \ f(x)\right)^2$$

$$\int_0^{2\pi} dx \sin x$$

Example

- Example $f(x) = \sin x$
- Optimal weight $\rho(x) = \frac{1}{4}\left|\sin x\right|$
- Optimal variance

$$V_{\text{opt}} = \left(\int_0^{2\pi} dx \ \left|\sin(x)\right| \right)^2 - \left(\int_0^{2\pi} dx \ \sin(x) \right)^2 = 16$$

$$\int_0^{2\pi} dx \sin x$$

Binning

1 Construct cheap approximation to |sin(*x*)|

2 Calculate relative area within each rectangle

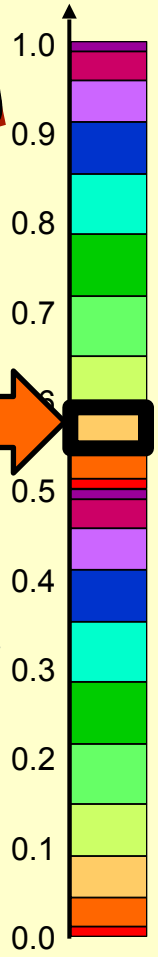
3 Choose a random number uniformly

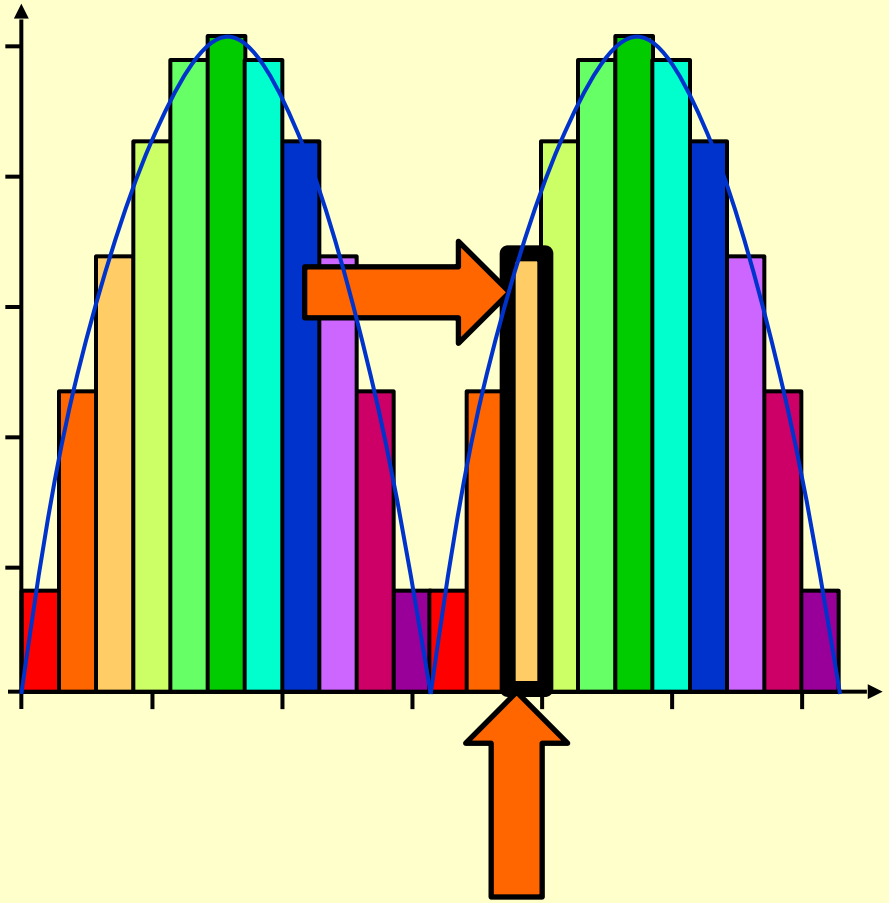
4 Select corresponding rectangle

5 Choose another random number uniformly

6 Select corresponding *x* value within rectangle

7 Compute |sin(*x*)|





$$\int_0^{2\pi} dx \sin x$$

Example

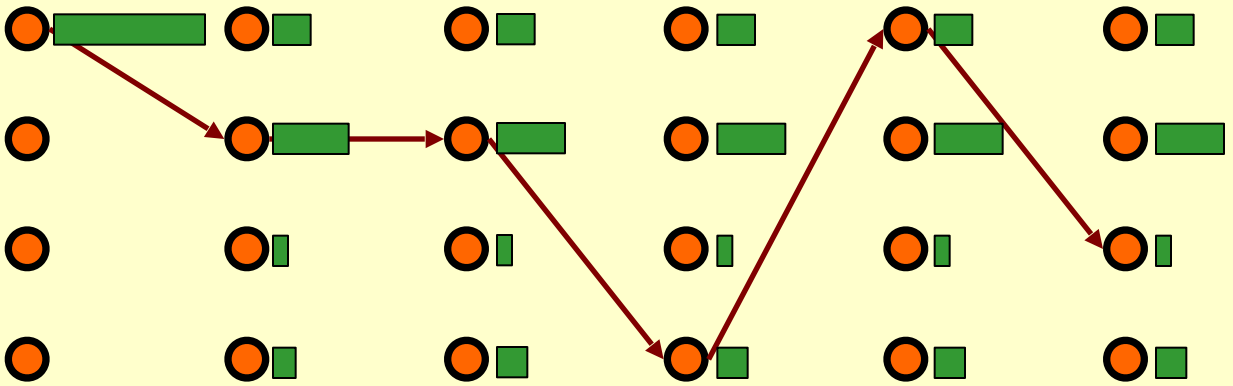
- With 100 rectangles we have *V = 16.02328561*
- … but we can do better*!*

$$I = \int_0^{2\pi} dx \sin x\, \theta\left(\sin x\right) + \int_0^{2\pi} dx \sin x\, \theta\left(-\sin x\right)$$

- For which $V_{\text{opt}} = 0$
- With 100 rectangles we have *V = 0.011642808*

Markov chains

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



Convergence of Markov Chains

- Define a metric $d(Q_1, Q_2) \equiv \int dx \; |Q_1(x) - Q_2(x)|$ 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 $Q, PQ, P^2Q, P^3Q, \ldots$ is Cauchy
- The space of probability distributions is *complete*, so the sequence converges to a unique fixed point distribution $\overline{Q} = \lim_{n \to \infty} P^n Q$

Convergence of Markov Chains

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

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

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

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

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

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

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

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

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

Convergence of Markov Chains

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

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

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

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

$$= \int dy \;\; \Delta Q\left(y\right) = \int dy \;\; Q_1\left(y\right) - \int dy \;\; Q_2\left(y\right) = 1 - 1 = 0$$

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

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

$$0 \leq \alpha \leq 1 - \int dx \;\; \inf_{y} P\left(x \leftarrow y\right) < 1$$

Use of 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
    - Don't know the normalisation of Q
    - Cannot use Markov chains to compute integrals directly
    - We can compute ratios of integrals

Metropolis Algorithm

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

Composite Markov Steps

- 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 $P_1 P_2$ 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 is not ergodic but P^n is the terminology "weakly" and "strongly" ergodic are sometimes used

Site-by-Site Updates

- 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

Exponential Autocorrelations

- The unique fixed point of an ergodic Markov process corresponds to the unique eigenvector with eigenvalue *1*
- All its other eigenvalues must lie within the unit circle
- In particular, the largest subleading eigenvalue is $|\lambda_{\max}|<1$
- This corresponds to the *exponential autocorrelation time* $N_{\exp} \equiv -\frac{1}{\ln|\lambda_{\max}|} > 0$

Integrated Autocorrelations

- Consider autocorrelation of some operator Ω
  - Without loss of generality we may 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 the *autocorrelation function*

$$C_\Omega(\ell) \equiv \frac{\left\langle \Omega(\phi_{t+\ell}) \Omega(\phi_t) \right\rangle}{\left\langle \Omega(\phi)^2 \right\rangle}$$

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

Integrated Autocorrelations

- The autocorrelation function falls 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

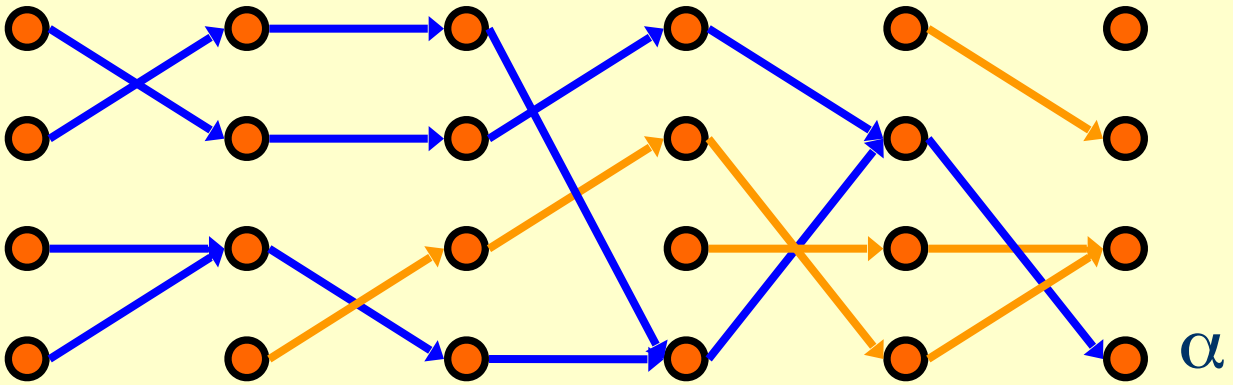
$$\left\langle \overline{\Omega}^2 \right\rangle = \left\{1 + 2\sum_{\ell=1}^{\infty} C_\Omega(\ell)\right\} \frac{\left\langle \Omega^2 \right\rangle}{N} \left[1 + O\left(\frac{N_{\exp}}{N}\right)\right]$$

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

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

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



Continuum Limits

- We are not interested in lattice QFTs per se, but in their continuum limit as $a \to 0$
  - This corresponds to a continuous phase transition of the discrete lattice model
    - For the continuum theory to have a finite correlation length ξa (inverse mass gap) in physical units this correlation length must diverge in lattice units
  - We expect such systems to exhibit universal behaviour as they approach a continuous phase transition
    - The nature of the continuum theory is characterised by its symmetries
    - The details of the microscopic interactions are unimportant at macroscopic length scales
    - Universality is a consequence of the way that the the theory scales as $a \to 0$ while ξa is kept fixed

Continuum Limits

- The nature of the continuum field theory depends on the way that physical quantities behave as the system approaches the continuum limit
- The scaling of the parameters in the action required is described by the renormalisation group equation (RGE)
  - The RGE states the “reparameterisation invariance” of the theory as the choice of scale at which we choose to fix the renormalisation conditions
  - As $a \to 0$ we expect the details of the regularisation scheme (cut off effects, lattice artefacts) to vanish, so the effect of changing a is just an RGE transformation
    - On the lattice the a renormalisation group transformation may be implemented as a “block spin” transformation of the fields
    - Strictly speaking, the renormalisation “group” is a semigroup on the lattice, as blockings are not invertible

Continuum Limits

- The continuum limit of a lattice QFT corresponds to a fixed point of the renormalisation group
  - At such a fixed point the *form* of the action does not change under an RG transformation
- The parameters in the action scale according to a set of *critical exponents*
- All known four dimensional QFTs correspond to trivial (Gaussian) fixed points
  - For such a fixed point the UV nature of the theory may by analysed using perturbation theory
  - Monomials in the action may be classified according to their power counting dimension
    - $d < 4$ relevant (superrenormalisable)
    - $d = 4$ marginal (renormalisable)
    - $d > 4$ irrelevant (nonrenormalisable)

Continuum Limits

- The behaviour of our Markov chains as the system approaches a continuous phase transition is described by its *dynamical critical exponents*
  - these describe how badly the system (model + algorithm) undergo *critical slowing down*
  - the dynamical critical exponent z tells us how the cost of generating an independent configuration grows as the correlation length of the system is taken to ∞, $\text{cost} \propto \xi^z$
  - this is closely related (but not always identical) to the dynamical critical exponent for the exponential or integrated autocorrelations

Global Heatbath

- The ideal generator selects field configurations randomly and independently from the desired distribution
  - It is hard to construct such global heatbath generators for any but the simplest systems
  - They can be built by selecting sufficiently independent configurations from a Markov process…
  - … or better yet using CFTP which guarantees that the samples are truly uncorrelated
  - But this such generators are expensive!

Global Heatbath

- For the purposes of Monte Carlo integration these is no need for the configurations to be completely uncorrelated
  - We just need to take the autocorrelations into account in our estimate of the variance of the resulting integral
  - Using all (or most) of the configurations generated by a Markov process is more cost-effective than just using independent ones
  - The optimal choice balances the cost of applying each Markov step with the cost of making measurements

Local Heatbath

- For systems with a local bosonic action we can build a Markov process with the fixed point distribution $\propto exp(-S)$ out of steps which update a single site with this fixed point
- If this update generates a new site variable value which is completely independent of its old value then it is called a local heatbath
- For free field theory we just need to generate a Gaussian-distributed random variable

Gaussian Generators

- If $\{x_1, x_2, \ldots, x_n\}$ are uniformly distributed random numbers then $\frac{1}{n}\sum_{i=1}^{n} x_i$ is approximately Gaussian by the Central Limit theorem
  - This is neither cheap nor accurate

Gaussian Generators

- If x is uniformly distributed and f is a monotonically increasing function then $f(x)$ is distributed as

$$P(y) = \int dx \;\; \delta\left(y - f(x)\right) = \int \frac{d\xi}{\left|f'\left(f^{-1}(\xi)\right)\right|} \;\; \delta(y - \xi) = \frac{1}{\left|f'\left(f^{-1}(y)\right)\right|}$$

  - Choosing $f(x) = \sqrt{-2\ln x}$ we obtain $P(y) = y \;\; e^{-\frac{1}{2}y^2}$
  - Therefore generate two uniform random numbers x_1 and x_2, set $r = \sqrt{-2\ln x_1}$, $\theta = 2\pi\, x_2$, then $y_1 = r\cos\theta$, $y_2 = r\sin\theta$ are two independent Gaussian distributed random numbers

Gaussian Generators

- Even better methods exist
  - The Rectangle-Wedge-Tail (RWT) method
  - The Ziggurat method
  - These do not require special function evaluations
  - They can be more interesting to implement for parallel computers

Local Heatbath

- For pure gauge theories the field variables live on the links of the lattice and take their values in a representation of the gauge group
  - For SU(2) Creutz gave an exact local heatbath algorithm
    - It requires a rejection test: this is different from the Metropolis accept/ reject step in that one must continue generating candidate group elements until one is accepted
  - For SU(3) the “quasi-heatbath” method of Cabibbo and Marinari is widely used
    - Update a sequence of SU(2) subgroups
    - This is not quite an SU(3) heatbath method…
    - … but sweeping through the lattice updating SU(2) subgroups is also a valid algorithm, as long as the entire sweep is ergodic
    - For a higher acceptance rate there is an alternative SU(2) subgroup heatbath algorithm

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
  - Takes 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 $z=2$
  - 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 q (i.e., the marginal distribution)

Hybrid Monte Carlo

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

Molecular Dynamics

- 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 $dp\,dq$ 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 uniformly distributed in *[0,1]*

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

Special Cases of GHMC

- The Hybrid Monte Carlo (HMC) algorithm is the special case where $\theta=\pi/2$
- $\theta=0$ corresponds to an exact version of the Molecular Dynamics (MD) or Microcanonical algorithm (which is in general non-ergodic)
- The L2MC algorithm of Horowitz corresponds to choosing arbitrary θ but MDMC trajectories of a single leapfrog integration step ($\tau=\delta\tau$). This method is also called Kramers algorithm.
- The Langevin Monte Carlo algorithm corresponds to choosing $\theta=\pi/2$ and MDMC trajectories of a single leapfrog integration step ($\tau=\delta\tau$).

Further Special Cases

- The Hybrid and Langevin algorithms are approximations where the Metropolis step is omitted
- The Local Hybrid Monte Carlo (LHMC) or Overrelaxation algorithm corresponds to updating a subset of the degrees of freedom (typically those living on one site or link) at a time

Langevin algorithm

- Consider the Hybrid Monte Carlo algorithm when take only one leapfrog step.
  - Combining the leapfrog equations of motion we obtain

$$\phi' = \phi + \pi\,\delta\tau - \tfrac{1}{2}\frac{\partial S}{\partial \tau}\delta\tau^2; \quad \pi' = \pi - \left(\frac{\partial S}{\partial \phi} + \frac{\partial S'}{\partial \phi}\right)\frac{\delta\tau}{2}$$

  - Ignore the Monte Carlo step
  - Recall that for $\theta = \frac{\pi}{2}$ the fictitious momenta π are just Gaussian distributed random noise ξ
  - Let $\Delta\tau \equiv \delta\tau^2$, then $\phi' = \phi + \xi\sqrt{\Delta\tau} - \tfrac{1}{2}\frac{\partial S}{\partial \phi}\Delta\tau$ which is the usual form of the Langevin equation

Inexact algorithms

- What happens if we omit the Metropolis test?
- We still have an ergodic algorithm, so there is some unique fixed point distribution $e^{-[S+\Delta S]}$ which will be generated by the algorithm

The condition for this to be a fixed point is

$$e^{-[S(\phi')+\Delta S(\phi')]} = \int [d\phi] e^{-[S(\phi)+\Delta S(\phi)]} \int [d\pi] e^{-\frac{1}{2}\pi^2} \delta\left[\phi' - U(\tau)\phi\right]$$

Langevin Algorithm

- For the Langevin algorithm (which corresponds to $\tau=\delta\tau$, a single leapfrog step) we may expand in powers of $\delta\tau$ and find a solution for

$$\Delta S = \sum_n \Delta S_n \delta\tau^n$$

- The equation determining the leading term in this expansion is the *Fokker-Planck* equation; the general equations are sometimes known as the *Kramers-Moyal* equations

Inexact algorithms

- In the general case we change variables to $\begin{pmatrix}\phi''\\ \pi''\end{pmatrix} = U(\tau)\begin{pmatrix}\phi\\ \pi\end{pmatrix}$
- Whence we obtain

$$e^{-[S(\phi')+\Delta S(\phi')]} = \int [d\phi][d\pi]\; e^{-H(\phi,\pi)-\Delta S(\phi)}\delta[\phi' - U\phi]$$

$$= \int [d\phi''][d\pi'']\; \det U_*\, e^{-(H+\Delta S)\circ U^{-1}}\delta[\phi' - U\phi]$$

area preservation: $\det U_* = 1$

reversibility: $U^{-1} = F\circ U\circ F$

$H\circ F = H$

$\Delta S\circ F = \Delta S$

$F^2 = 1$

$\delta H \equiv H\circ(U-1)$

$\delta\Delta S \equiv \Delta S\circ(U-1)$

$$= \int [d\phi''][d\pi'']\; e^{-H} e^{-(\delta H+\Delta S\circ U)\circ F}\delta[\phi' - \phi'']$$

$$= e^{-[S(\phi')+\Delta S(\phi')]}\int [d\pi'']\; e^{-\frac{1}{2}\pi'^2} e^{-(\delta H+\delta\Delta S)}$$

$$\left\langle e^{-(\delta H+\delta\Delta S)}\right\rangle_\pi = 1$$

Inexact algorithms

- Since H is extensive so is δH, and thus so is the connected generating function (cumulants)

$$e^{-F(\phi)} \equiv \left\langle e^{-\delta H(\phi,\pi)} \right\rangle_{\pi}$$

- We can thus show order by order in $\delta\tau$ that ΔS is extensive too

Langevin Algorithm

- For the Langevin algorithm we have $\langle \delta H \rangle_\pi = O(\delta\tau^4)$ and $\langle \delta \Delta S_n \rangle_\pi = O(\delta\tau^2)$, so we immediately find that $\Delta S = O(\delta\tau^2)$

$$\Delta S_2 = \tfrac{1}{8} \sum_j \left[2 \frac{\partial^2 S}{\partial \phi_j^2} - \left(\frac{\partial S}{\partial \phi_j} \right)^2 \right] = \tfrac{1}{8} \left[2 S_{j\,j} - S_j^2 \right]$$

$$(\Delta S_4)_i = \tfrac{1}{48} \left[S_{ijjkk} - 2 S_{ijjk} S_k + S_{ijk} S_j S_k + 2 S_{ijk} S_{ijk} \right]$$

  - If S is local then so are all the ΔS_n
  - We only have an asymptotic expansion for ΔS, and in general the subleading terms are neither local nor extensive
  - Therefore, taking the continuum limit $a \to 0$ limit for fixed $\delta\tau$ will probably not give exact results for renormalised quantities

Hybrid Algorithm

- For the Hybrid algorithm $\langle \delta H \rangle_\pi = O(\delta\tau^2)$ and $\langle \delta \Delta S_n \rangle_\pi = O(\delta\tau^0)$ so again we find $\Delta S = O(\delta\tau^2)$
  - We have made use of the fact that H' is conserved and differs from H by $O(\delta\tau^2)$

Inexact algorithms

- What effect do such systematic errors in the distribution have on measurements?
  - Large errors will occur where observables are discontinuous functions of the parameters in the action, *e.g., near phase transitions*
  - If we want to improve the action so as to reduce lattice discretisation effects and get a better approximation to the underlying continuum theory, then we have to ensure that for the appropriate n $\delta\tau^n \ll a^2$
  - Step size errors need <u>not</u> be irrelevant

Local HMC

- Consider the Gaussian model defined by the free field action $S(\phi) = \frac{1}{2} \sum_x \left\{ \sum_{\mu=1}^{d} \left(\partial_\mu \phi(x) \right)^2 + m^2 \phi^2 \right\}$
  - Introduce the Hamiltonian $H = \frac{1}{2}\pi^2 + S(\phi)$ on "fictitious" phase space.
    - The corresponding equations of motion for the single site x $\ddot{\phi}_x = -\omega^2 \phi_x + F_x$ where $\omega^2 \equiv 2d + m^2$ and $F_x \equiv \sum_{\|x-y\|=1} \phi(y)$
  - The solution in terms of the Gaussian distributed random initial momentum π_x and the initial field value ϕ_x is

$$\phi_x(t) = \phi_x \cos\omega\tau + \frac{1-\cos\omega\tau}{\omega^2} F_x + \frac{\pi_x}{\omega} \sin\omega\tau$$

Local HMC

- Identify $\zeta \equiv 1-\cos\omega\tau$ and $\pi_x \equiv \xi$ to get the usual Adler overrelaxation update

$$\phi'(x) = (1-\zeta)\phi(x) + \frac{\zeta F}{\omega^2} + \frac{\sqrt{\zeta(2-\zeta)}}{\omega}\xi$$

- For gauge theories various overrelaxation methods have been suggested
  - *Hybrid Overrelaxation:* this alternates a heatbath step with many overrelaxation steps with $\zeta=2$
  - *LHMC:* this uses an analytic solution for the equations of motion for the update of a single U(1) or SO(2) subgroup at a time. In this case the equations of motion may be solved in terms of elliptic functions

Classical Mechanics on Group Manifolds

- Gauge fields take their values in some Lie group, so we need to define classical mechanics on a group manifold which preserves the group-invariant Haar measure
  - A Lie group G is a smooth manifold on which there is a natural mapping $L: G \times G \rightarrow G$ defined by the group action
  - This induces a map called the *pull-back* of L on the cotangent bundle defined by

$$L^* : G \times F \rightarrow F \qquad L_g^* f = f \circ L_g$$

$$L_* : G \times TG \rightarrow TG \qquad \left(L_{g*} v\right)(f) = v\left(L_g^* f\right)$$

$$L^* : G \times T^*G \rightarrow T^*G \qquad \left(L_g^* \theta\right)(v) = \theta\left(L_{g*} v\right)$$

    - F is the space of *0 forms*, which are smooth mappings from G to the real numbers $F : G \rightarrow \Re$

Classical Mechanics on Group Manifolds

- A form is *left invariant* if $L^{*}\theta = \theta$
- The tangent space to a Lie group at the origin is called the *Lie algebra*, and we may choose a set of basis vectors $\{e_i(0)\}$ which satisfy the commutation relations $[e_i, e_j] = \sum_k c_{ij}^k e_k$ where c_{ij}^k are the *structure constants* of the algebra
- We may define a set of left invariant vector fields on TG by $e_i(g) \equiv L_{g*} e_i(0)$
- The corresponding left invariant dual forms $\{\theta_i\}$ satisfy the *Maurer-Cartan* equations $d\theta^i = -\frac{1}{2}\sum_{jk} c_{jk}^i \ \theta^j \wedge \theta^k$
- We may therefore define a closed *symplectic 2 form* which globally defines an invariant Poisson bracket by

$$\omega \equiv -d\sum_i \theta^i p^i = \sum_i \left(\theta^i \wedge dp^i - p^i d\theta^i\right) = \sum_i \left(\theta^i \wedge dp^i + \tfrac{1}{2} p^i c_{jk}^i \theta^j \wedge \theta^k\right)$$

Classical Mechanics on Group Manifolds

- 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 h such that $dH(y) = \omega(h, y) \quad \forall y \in TG$
  - The classical trajectories $\sigma_t = (Q_t, P_t)$ are then the *integral curves* of h: $\dot{\sigma}_t = h(s_t)$
  - In the natural basis we have

$$h = \sum_i \left(h^i e_i + \overline{h}^i \frac{\partial}{\partial p^i} \right) \qquad y = \sum_i \left(y^i e_i + \overline{y}^i \frac{\partial}{\partial p^i} \right)$$

$$dH(y) = \sum_i \left(e_i(H) y^i + \frac{\partial H}{\partial p^i} \overline{y}^i \right) = \omega(h, y) = \sum_i \left(h^i \overline{y}^i - y^i \overline{h}^i + p^i \sum_{jk} c^i_{jk} h^j y^k \right)$$

Classical Mechanics on Group Manifolds

- Equating coefficients of the components of y we find

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

- The equations of motion in the local coordinate basis $e_j = \sum_j e^i_j \, \partial / \partial q^j$ are therefore

$$\dot{Q}^j_t = \sum_i \frac{\partial H}{\partial p^i} e^j_i, \quad \dot{P}^j_t = \sum_k \left(\sum_i c^i_{kj} P^i_t \frac{\partial H}{\partial p^k} - e^k_j \frac{\partial H}{\partial q^k} \right)$$

- Which for a Hamiltonian of the form $H = f(p^2) + S(q)$ reduce to

$$\dot{Q}^j_t = \sum_i \frac{\partial H}{\partial p^i} e^j_i, \quad \dot{P}^j_t = -\sum_k e^k_j \frac{\partial H}{\partial q^k}$$

Classical Mechanics on Group Manifolds

- In terms of constrained variables
  - The representation of the generators is $U(q) = e^{\sum_i q^i T_i}$

$$T_i = \left.\frac{\partial U(g)}{\partial g^i}\right|_{g=0} = e_i(g)U(g)\Big|_{g=0}$$

  - From which it follows that $e_i(U) = UT_i$
  - And for the Hamiltonian $H = \frac{1}{2}\sum_i \left(p^i\right)^2 + S(U)$ leads to the equations of motion

$$\dot{U} = PU$$

$$\dot{P} = \sum_{iab} T_i \left[\frac{\partial S}{\partial U_{ab}}\left(UT_i\right)_{ab} - \frac{\partial S}{\partial U^\dagger_{ab}}\left(T_i U^\dagger\right)_{ab}\right] = -T\left[S'(U)U\right]$$

  - For the case $G = SU(n)$ the operator T is the projection onto traceless antihermitian matrices

Classical Mechanics on Group Manifolds

- Discrete equations of motion
  - We can now easily construct a discrete *PQP* symmetric integrator from these equations

$$P\left(\tfrac{1}{2}\delta\tau\right) = P(0) - T\left[S'\left(U(0)\right)U(0)\right]\tfrac{1}{2}\delta\tau$$

$$U(\delta\tau) = \exp\left[P\left(\tfrac{1}{2}\delta\tau\right)\delta\tau\right]U(0)$$

$$P(\delta\tau) = P\left(\tfrac{1}{2}\delta\tau\right) - T\left[S'\left(U(\delta\tau)\right)U(\delta\tau)\right]\tfrac{1}{2}\delta\tau$$

  - The exponential map from the Lie algebra to the Lie group may be evaluated exactly using the Cayley-Hamilton theorem
    - All functions of an $n \times n$ matrix M may be written as a polynomial of degree $n - 1$ in M
    - The coefficients of this polynomial can be expressed in terms of the invariants (traces of powers) of M

Symplectic Integrators

- Reversible and area-preserving integrator for Hamiltonian $H(q,p) = T(p) + S(q) = \frac{1}{2}p^2 + S(q)$

$$\exp\left(\tau \frac{d}{dt}\right) = \exp\left(\tau \left\{\frac{\partial p}{\partial t}\frac{\partial}{\partial p} + \frac{\partial q}{\partial t}\frac{\partial}{\partial q}\right\}\right)$$

$$= \exp\left(\tau \left\{-\frac{\partial H}{\partial q}\frac{\partial}{\partial p} + \frac{\partial H}{\partial p}\frac{\partial}{\partial q}\right\}\right) = e^{\tau \hat{H}}$$

$$= \exp\left(\tau \left\{-S'(q)\frac{\partial}{\partial p} + T'(p)\frac{\partial}{\partial q}\right\}\right)$$

Symplectic Integrators

- The operators $Q \equiv T'(p)\frac{\partial}{\partial q}$ and $P \equiv -S'(q)\frac{\partial}{\partial p}$ are easily exponentiated (Taylor's theorem)

$$e^{tQ} : f(q,p) \mapsto f\left(q + tT'(p), p\right)$$

$$e^{tP} : f(q,p) \mapsto f\left(q, p - tS'(q)\right)$$

- The symmetric symplectic QPQ integrator applies these steps iteratively

$$U_{QPQ}(\tau) \equiv \left(e^{\frac{1}{2}\delta\tau Q} e^{\delta\tau P} e^{\frac{1}{2}\delta\tau Q}\right)^{\tau/\delta\tau}$$

BCH Formula

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

$$c_{n+1} = \frac{1}{n+1}\left\{-\tfrac{1}{2}[c_n, A-B] + \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{B_m}{(2m)!} \sum_{\substack{k1,\ldots,k_{2m}\geq 1 \\ k_1+\cdots+k_m=n}} [c_{k_1},[\ldots,[c_{k_{2m}} A + B]\ldots]]\right\}$$

  - The B_m are Bernouilli numbers

BCH Formula

Explicitly, the first few terms are

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

Symmetric Symplectic Integrator

- In order to construct reversible integrators we use symmetric symplectic integrators

The following identity follows directly from the BCH formula

$$\ln\left(e^{A/2}e^{B}e^{A/2}\right)=\left\{A+B\right\}+\tfrac{1}{24}\left\{\left[A,\left[A,B\right]\right]-2\left[B,\left[A,B\right]\right]\right\}$$

$$+\tfrac{1}{5760}\left\{\begin{array}{l}7\left[A,\left[A,\left[A,\left[A,B\right]\right]\right]\right]+28\left[B,\left[A,\left[A,\left[A,B\right]\right]\right]\right]\\+12\left[\left[A,B\right],\left[A,\left[A,B\right]\right]\right]+32\left[B,\left[B,\left[A,\left[A,B\right]\right]\right]\right]\\-16\left[\left[A,B\right],\left[B,\left[A,B\right]\right]\right]+8\left[B,\left[B,\left[B,\left[A,B\right]\right]\right]\right]\end{array}\right\}+\cdots$$

Symmetric Symplectic Integrator

- The BCH formula tells us that the QPQ integrator has evolution

$$
\begin{aligned}
U_{QPQ}(\tau) &\equiv \left(e^{\frac{1}{2}\delta\tau Q} e^{\delta\tau P} e^{\frac{1}{2}\delta\tau Q} \right)^{\tau/\delta\tau} \\
&= \left(\exp\left[(Q+P)\delta\tau - \tfrac{1}{24}\left([Q,[Q,P]] + 2[P,[Q,P]] \right)\delta\tau^3 + O(\delta\tau^5) \right] \right)^{\tau/\delta\tau} \\
&= \exp\left[\tau \left((Q+P)\delta\tau - \tfrac{1}{24}\left([Q,[Q,P]] + 2[P,[Q,P]] \right)\delta\tau^3 + O(\delta\tau^4) \right) \right] \\
&= e^{\tau \hat{H}'_{QPQ}}
\end{aligned}
$$

QPQ Integrator

- The QPQ integrator therefore <u>exactly</u> conserves the Hamiltonian H'_{QPQ}, where

$$\hat{H}'_{QPQ} \equiv \frac{\partial H'_{QPQ}}{\partial p}\frac{\partial}{\partial q} - \frac{\partial H'_{QPQ}}{\partial q}\frac{\partial}{\partial p}$$

So $$H'_{QPQ} = H + \tfrac{1}{24}\left\{-p^2 S'' + 2S'^2\right\}\delta\tau^2$$
$$+ \tfrac{1}{5760}\left\{7p^4 S^{(4)} - 24p^2\left(3S'S''' + S''^2\right) + 96S'^2 S''\right\}\delta\tau^4 + O\left(\delta\tau^6\right)$$

  - Note that H'_{QPQ} cannot be written as the sum of a p-dependent kinetic term and a q-dependent potential term

QPQ Integrator

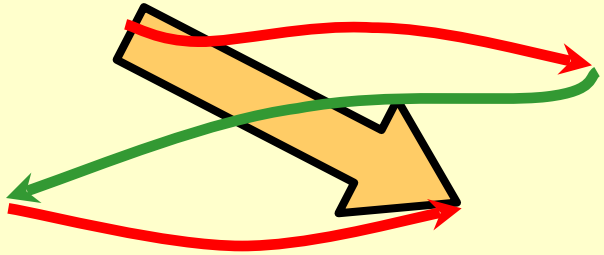
- Observe that for any trajectory length, even $\tau \gg 1/\delta\tau$

$$\delta H \equiv H \circ \left[U_{QPQ}(\tau) - 1\right] = \left(H - H'_{QPQ}\right) \circ \left[U_{QPQ}(\tau) - 1\right] = O\left(\delta\tau^2\right)$$

Campostrini Wiggles

- From the form of the evolution operator $U_0(\delta\tau) = e^{\delta\tau \hat{H}} + R_0 \delta\tau^3 + O(\delta\tau^5)$ Campostrini noted that we can easily write a higher-order integrator

$$U_0(\varepsilon) U_0(-\sigma\varepsilon) U_0(\varepsilon) = e^{\varepsilon(2-\sigma)\hat{H}} + R_0(2-\sigma^3)\varepsilon^3 + O(\varepsilon^5)$$



  - The leading error vanishes if we choose $\sigma = \sqrt[3]{2}$
  - The total step size is unchanged if $\varepsilon = \delta\tau / (2-\sigma)$
  - This trick may be applied to recursively to obtain arbitrarily high-order symplectic integrators

Instabilities

- Consider the simplest leapfrog scheme for a single simple harmonic oscillator with frequency ω
  - For a single step we have

$$\begin{pmatrix} q(t+\delta\tau) \\ p(t+\delta\tau) \end{pmatrix} = \begin{pmatrix} 1-\frac{1}{2}(\omega\delta\tau)^2 & \omega\delta\tau \\ -\omega\delta\tau + \frac{1}{4}(\omega\delta\tau)^3 & 1-\frac{1}{2}(\omega\delta\tau)^2 \end{pmatrix} \begin{pmatrix} q(t) \\ p(t) \end{pmatrix}$$

  - The eigenvalues of this matrix are

$$1-\frac{1}{2}(\omega\delta\tau)^2 \pm i\omega\delta\tau\sqrt{1-\frac{1}{4}(\omega\delta\tau)^2} = e^{\pm i\cos^{-1}\left(1-\frac{1}{2}(\omega\delta\tau)^2\right)}$$

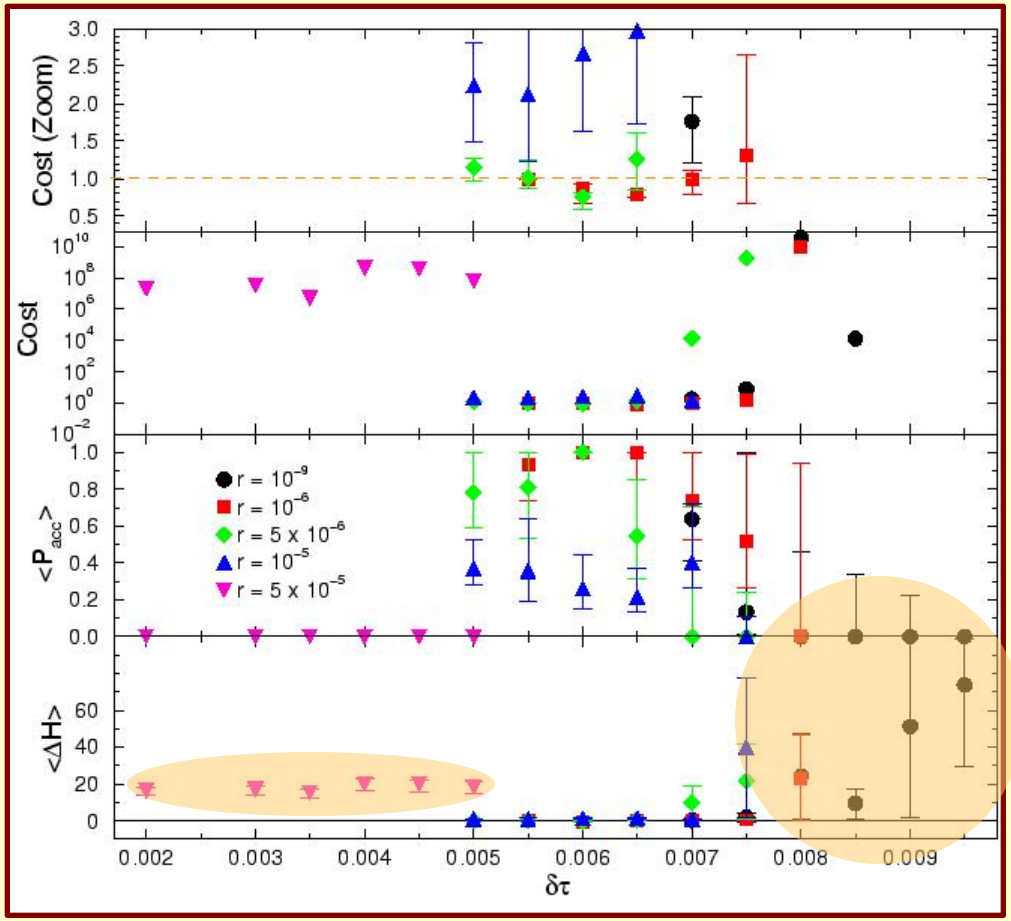
  - For $\omega\delta\tau > 2$ the integrator becomes unstable

Instabilities

- The orbits change from ellipses to hyperbolae
- The energy diverges exponentially in τ instead of oscillating
- For bosonic systems $\delta H \gg 1$ for such a large integration step size
- For light dynamical fermions there seem to be a few "modes" which have a large force due to the small eigenvalues of the Dirac operator, and this force plays the rôle of the frequency ω
- For these systems $\delta\tau$ is limited by stability rather than by the Metropolis acceptance rate

Instabilities

- Some full QCD measurements on large lattices with quite light quarks
  - large CG residual: *always unstable*
    - small CG residual: unstable for $\delta\tau > 0.0075$
    - intermediate CG residual: unstable for $\delta\tau > 0.0075$



$16^3 \times 32$ lattice, $\kappa = 0.1355, \beta = 5.2, c_{SW} = 2.0171$

Dynamical Fermions

- Fermion fields are Grassmann valued
  - Required to satisfy the spin-statistics theorem
  - Even “classical” Fermion fields obey anticommutation relations
  - Grassmann algebras behave like negative dimensional manifolds

Grassmann Algebras

- Linear space spanned by generators $\{\theta_1, \theta_2, \theta_3, \ldots\}$ with coefficients $a, b, \ldots$ in some field (usually the real or complex numbers)
- Algebra structure defined by nilpotency condition for elements of the linear space $\alpha^2 = 0$
  - There are many elements of the algebra which are not in the linear space (e.g., $\theta_1\theta_2$)
  - Nilpotency implies anticommutativity $\alpha\beta + \beta\alpha = 0$
    - $0 = \alpha^2 = \beta^2 = (\alpha + \beta)^2 = \alpha^2 + \alpha\beta + \beta\alpha + \beta^2 = \alpha\beta + \beta\alpha = 0$
  - Anticommutativity implies nilpotency $2\alpha^2 = 0$ (unless the coefficient field has characteristic 2, i.e., $2 = 0$)

Grassmann Algebras

- Grassmann algebras have a natural *grading* corresponding to the number of generators in a given product
  - $deg(1) = 0,\ deg(\theta_i) = 1,\ deg(\theta_i\theta_j) = 2,\ldots$
  - All elements of the algebra can be decomposed into a sum of terms of definite grading
  - The *parity* transform of α is $P(\alpha) = (-1)^{\deg(\alpha)}\alpha$
- A natural *antiderivation* is defined on a Grassmann algebra
  - Linearity: $d(a\alpha + b\beta) = a\,d\alpha + b\,d\beta$
  - Anti-Leibniz rule: $d(\alpha\beta) = (d\alpha)\beta + P(\alpha)(d\beta)$

Grassmann Integration

- Definite integration on a Grassmann algebra is defined to be the same as derivation
  - Hence change of variables leads to the *inverse* Jacobian $$\int d\theta_1' \cdots d\theta_n' = \int d\theta_1 \cdots d\theta_n \quad \det\left(\frac{\partial \theta_i}{\partial \theta_j'}\right)$$
- Gaussians over Grassmann manifolds
  - Like all functions, Gaussians are polynomials over Grassmann algebras $$e^{\sum_{ij} \theta_i a_{ij} \theta_j} = \prod_{ij} e^{\theta_i a_{ij} \theta_j} = \prod_{ij} \left(1 + \theta_i a_{ij} \theta_j\right)$$
  - There is no reason for this function to be positive even for real coefficients

Grassmann Integration

- Gaussian integrals over Grassmann manifolds

$$\int d\theta_1 \cdots d\theta_n \quad e^{\pm \sum_{ij} \theta_i a_{ij} \theta_j} = \pm \mathrm{Pf}\left(a_{ij}\right)$$

  - Where $\mathrm{Pf}(a)$ is the *Pfaffian,* $\mathrm{Pf}(a)^2 = \det(a)$
  - If we separate the Grassmann variables into two "conjugate" sets we find the more familiar result

$$\int d\theta_1 \cdots d\theta_n d\bar{\theta}_1 \cdots d\bar{\theta}_n \quad e^{\pm \sum_{ij} \bar{\theta}_i a_{ij} \theta_j} = \pm \det\left(a_{ij}\right)$$

  - Despite the notation, this is a purely *algebraic* identity
  - It does not require the matrix $a > 0$, unlike its bosonic analogue

Fermion Determinant

- 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 integrate out the fermion fields to obtain the fermion 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

Fermionic Observables

- 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

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

Dynamical Fermions

- Including the determinant as part of the observable to be measured is not feasible

$$\langle \Omega \rangle = \frac{\langle \det M(\phi) \;\; \Omega(\phi) \rangle_{S_B}}{\langle \det M(\phi) \rangle_{S_B}}$$

  - The determinant is extensive in the lattice volume, thus we get poor importance sampling

Pseudofermions

- Represent the fermion determinant as a bosonic Gaussian integral with a non-local kernel $\det M(\phi) \propto \int d\bar{\chi}\, d\chi \quad e^{-\bar{\chi} M^{-1}(\phi)\chi}$
  - 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

$$\left(\det M\left(\phi\right)\right)^2 = \det\left(M\left(\phi\right)M^\dagger\left(\phi\right)\right) \propto \int d\overline{\chi}\, d\chi \quad e^{-\overline{\chi}\left(M^\dagger M\right)^{-1}\chi}$$

  - 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

Equations of Motion

The equations for motion for the boson (gauge) fields are $\dot{\phi} = \pi$ and

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

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

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

Linear Solver Accuracy

- 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
  - We usually take “exactly” to be synonymous with “to machine precision”

Krylov Spaces

- One of the main reasons why dynamical fermion lattice calculations are feasible is the existence of very effective numerical methods for solving large sparse systems of linear equations
- Family of iterative methods based on Krylov spaces
  - Conjugate Gradients (CG, CGNE)
  - BiConjugate Gradients (BiCG, BiCGstab, BiCGγ_5)

Krylov Spaces

- These are often introduced as exact methods
  - They require $O(V)$ iterations to find the solution
  - They do not give the exact answer in practice because of rounding errors
  - They are more naturally thought of as methods for solving systems of linear equations in an (almost) ∞-dimensional linear space
  - This is what we are approximating on the lattice anyway

BiCG on a Banach Space

- We want to solve the equation $Ax = b$ on a Banach space
  - This is a normed linear space
  - The norm endows the space with a topology
  - The linear space operations are continuous in this topology

We solve the system in the Krylov subspace

$$K_n = \text{span}\left(b, Ab, A^2b, A^3b, \ldots, A^nb\right)$$

BiCG on a Banach Space

There is no concept of "orthogonality" in a Banach space, so we also need to introduce a dual Krylov space of linear functionals on the Banach space

$$\hat{K}_n = \text{span}\left(\hat{b}, A^\dagger \hat{b}, A^{\dagger 2}\hat{b}, A^{\dagger 3}\hat{b}, \ldots, A^{\dagger n}\hat{b}\right)$$

- The vector $\hat{b}$ is arbitrary
- The *adjoint* of a linear operator is defined by

$$\left(A^\dagger \hat{b}\right)(x) = \hat{b}\left(Ax\right) \quad \forall x$$

BiCG on a Banach Space

- We construct bi-orthonormal bases

$$K_n \qquad r_0 = b$$

$$R \equiv \begin{pmatrix} r_0 & r_1 & \cdots & r_{n-1} \end{pmatrix}$$

$$\gamma\, r_n = \left(1 - R\hat{R}^\dagger\right) A r_{n-1}$$

$$\hat{K}_n \qquad \hat{r}_0 = \hat{b}$$

$$\hat{R} \equiv \begin{pmatrix} \hat{r}_0 & \hat{r}_1 & \cdots & \hat{r}_{n-1} \end{pmatrix}$$

$$\hat{\gamma}\, \hat{r}_n = \left(1 - \hat{R}R^\dagger\right) A^\dagger \hat{r}_{n-1}$$

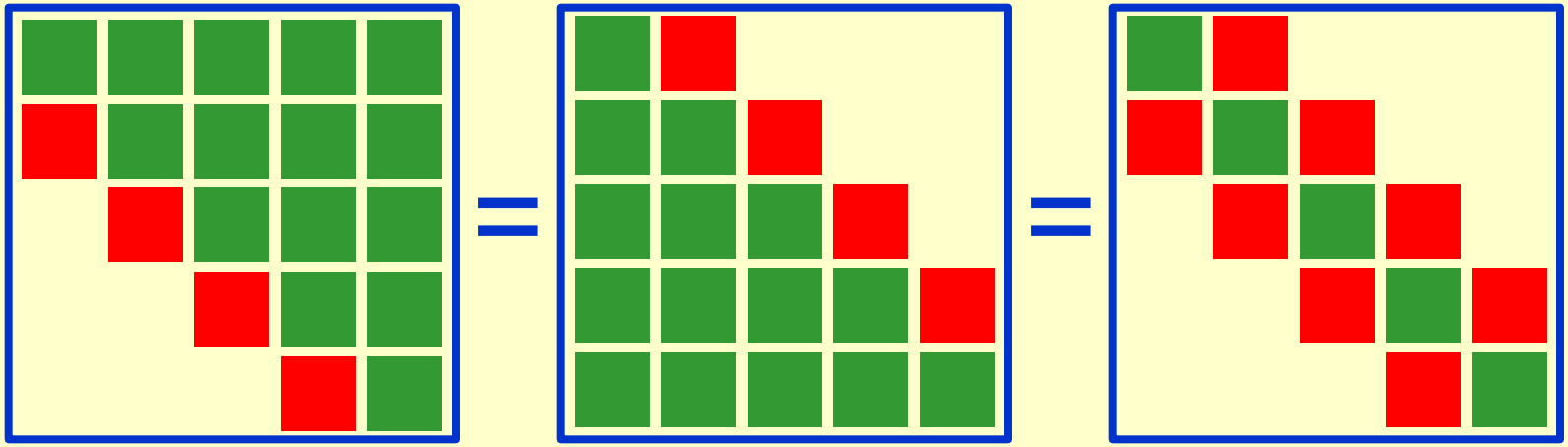
$$\hat{R}^\dagger R = 1$$

$$\hat{R}^\dagger A R = T$$

Galerkin condition (projectors)
Bi-orthogonality and normalisation
Short (3 term) recurrence
Lanczos tridiagonal form
BiCGstab: minimise norm $||r_n||$

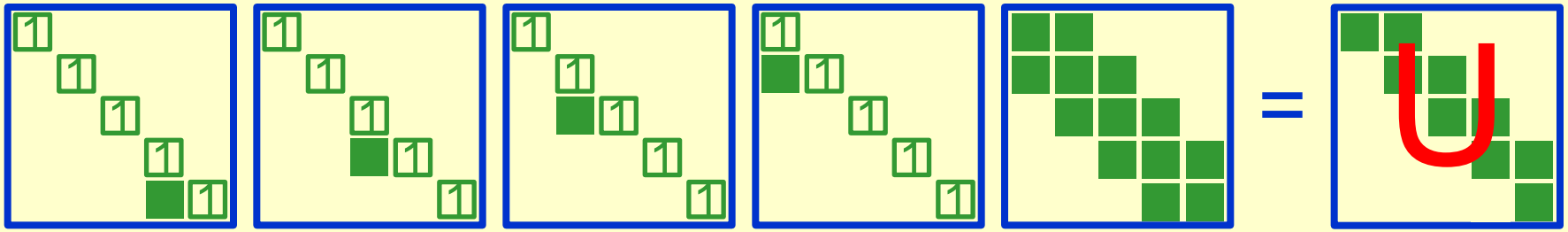
Tridiagonal Systems

- The problem is now reduced to solving a fairly small tridiagonal system
  - Hessenberg and Symmetric → Tridiagonal



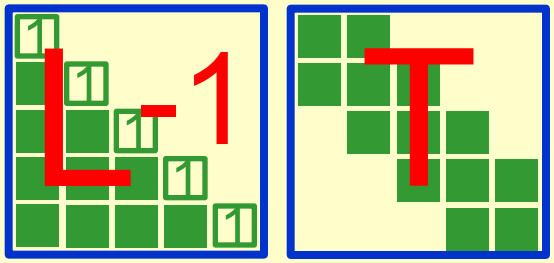
LU Decomposition

- Reduce matrix to triangular form
  - Gaussian elimination (LU decomposition)



$$T = LU$$

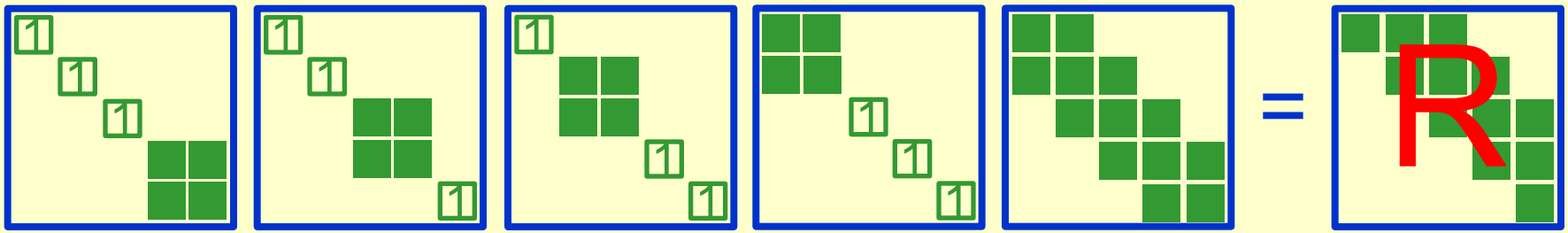
$$T^{-1} = U^{-1}L^{-1}$$



$$x = A^{-1}b \qquad R\hat{R}^{\dagger}x = R\hat{R}^{\dagger}A^{-1}R\hat{R}^{\dagger}b = RT^{-1}e_0 = \left(RU^{-1}\right)\left(L^{-1}e_0\right)$$

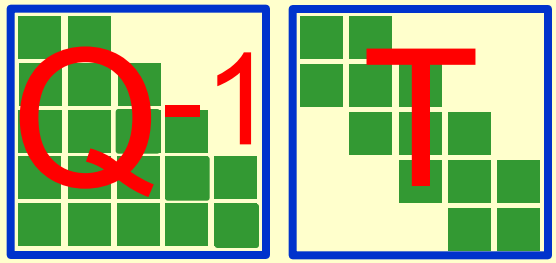
QR Decomposition

- Reduce matrix to triangular form
  - Givens rotations (QR factorisation)



$$T = QR$$

$$T^{-1} = R^{-1}Q^{-1} = R^{-1}Q^{\dagger}$$



Krylov Solvers

- Convergence is measured by the *residual*
  $r_n = ||b - Ax_n||$
  - Does not decrease monotonically for BiCG
  - Better for BiCGstab
  - Bad breakdown for unlucky choice of starting form
  - LU might fail if zero pivot occurs
  - QR is more stable
- In a Hilbert space there is an inner product
  - Which is relevant if A is symmetric (Hermitian)
  - In this case we get the CG algorithm
  - No bad breakdown (solution is in Krylov space)
  - $A>0$ required only to avoid zero pivot for LU

Random Number Generators

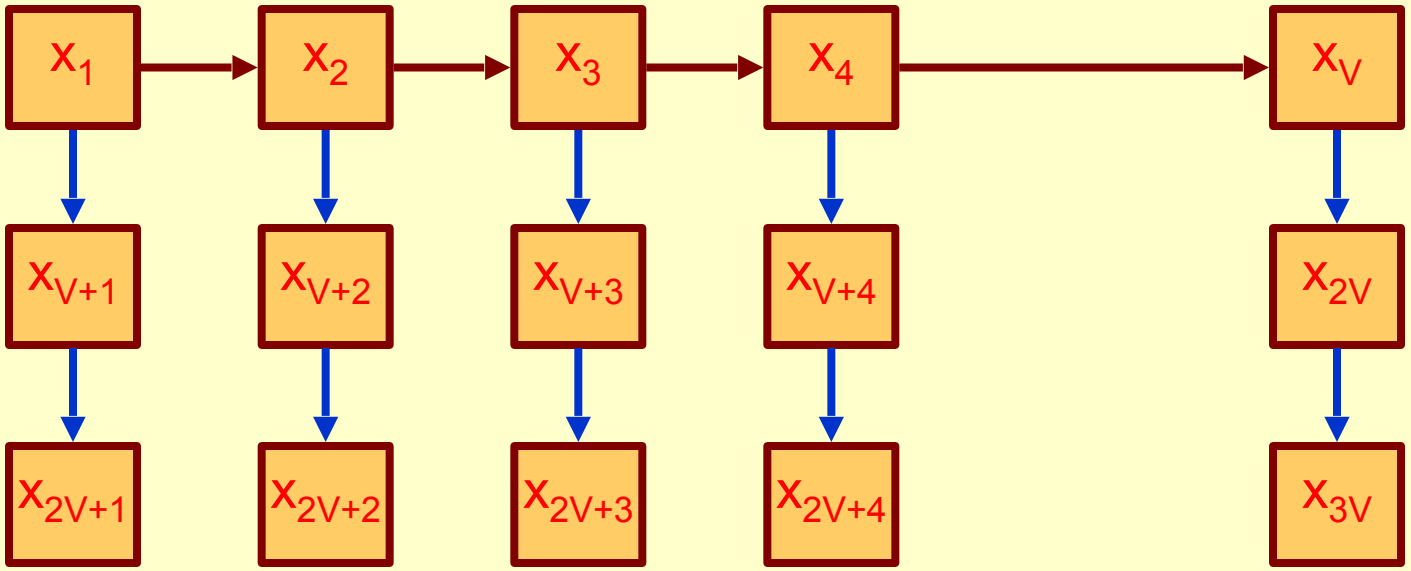
- Pseudorandom number generators
  - Random numbers used infrequently in Monte Carlo for QFT
    - Compared to spin models
  - Linear congruential generator $x_{n+1} \equiv a x_n + b \pmod{m}$
    - For suitable choice of *a, b,* and *m* (*see, e.g., Donald Knuth, Art of Computer Programming*)
    - Usually chose *b* = *0* and *m* = power of 2
    - Seem to be good enough in practice
    - Problems for spin models if *m* is too small a multiple of *V*

Parallel Random Numbers

$$x_{n+V} \equiv a' x_n + b' \pmod{m}$$

$$a' \equiv a^{V} \pmod{m}$$

$$b' \equiv \sum_{j=0}^{V-1} a^j b \pmod{m}$$



Acceptance Rates

We can compute the average Metropolis acceptance rate $\left\langle P_{\text{acc}} \right\rangle$

- Area preservation implies that $\left\langle e^{-\delta H} \right\rangle = 1$

$$\frac{1}{Z}\int [d\phi][d\pi]\ e^{-H} = \frac{1}{Z}\int [d\phi'][d\pi']\ e^{-H'} = \frac{1}{Z}\int [d\phi][d\pi]\ e^{-H} e^{-\delta H}$$

- The probability distribution of δH has an asymptotic expansion as the lattice volume $V \to \infty$

$$P_{\delta H}(\xi) = \frac{1}{Z}\int [d\phi][d\pi]\ e^{-H}\delta(\xi - \delta H) \sim \frac{1}{\sqrt{4\pi \left\langle \delta H \right\rangle}} \exp\left[-\frac{\left(\xi - \left\langle \delta H \right\rangle\right)^2}{4\left\langle \delta H \right\rangle} \right]$$

- The average Metropolis acceptance rate is thus

$$\left\langle P_{\text{acc}} \right\rangle \sim \operatorname{erfc}\left(\tfrac{1}{2}\sqrt{\left\langle \delta H \right\rangle} \right) = \operatorname{erfc}\left(\sqrt{\tfrac{1}{8}\left\langle \delta H^2 \right\rangle} \right)$$

Acceptance Rates

- The acceptance rate is a function of the variable $x = V\delta\tau^{4n+4}$ for the nth order Campostrini "wiggle" used to generate trajectories with $\tau = 1$
- For single-step trajectories $\tau = \delta\tau$ the acceptance rate is a function of $x = V\delta\tau^{4n+6}$

Cost and Dynamical Critical Exponents

- To a good approximation, the cost C of a Molecular Dynamics computation is proportional to the total time for which we have to integrate Hamilton's equations.
  - The cost per independent configuration is then

$$C_\Omega \propto \frac{\tau}{\delta\tau}\left(1 + 2A_\Omega\right)$$

  - Note that the cost depends on the particular operator under consideration
  - The optimal trajectory lengths obtained by minimising the cost as a function of the parameters τ, $\delta\tau$, and θ of the algorithm

Cost and Dynamical Critical Exponents

- While the cost depends upon the details of the implementation of the algorithm, the way that it scales with the correlation length ξ of the system is an intrinsic property of the algorithm; $C_\Omega \propto \xi^z$ where z is the *dynamical critical exponent.*
  - For local algorithms the cost is independent of the trajectory length, $C_\Omega \propto 1 + 2A_\Omega$, and thus minimising the cost is equivalent to minimising A_Ω
  - Free field theory analysis is useful for understanding and optimising algorithms, especially if our results do not depend on the details of the spectrum

Cost and Dynamical Critical Exponents

- HMC
  - For free field theory we can show that choosing $\tau \propto \xi$ and $\theta = \pi/2$ gives $z=1$
  - This is to be compared with $z=2$ for constant τ
  - The optimum cost for HMC is thus

$$C_\Omega \propto V \left(V \delta\tau^4 \right)^{1/4} \left(\xi / \delta\tau \right) = V^{5/4} \xi$$

  - For nth order Campostrini integration (if it were stable) the cost is

$$C_\Omega \propto V \left(V \delta\tau^{4n+4} \right)^{\frac{1}{4n+4}} \left(\xi / \delta\tau \right) = V^{\frac{4n+5}{4n+4}} \xi$$

Cost and Dynamical Critical Exponents

- LMC
  - For free field theory we can show that choosing $\tau=\delta\tau$ and $\theta=\pi/2$ gives $z=2$
  - The optimum cost for LMC is thus

$$C_\Omega \propto V\left(V\delta\tau^6\right)^{1/3}\left(\xi/\delta\tau\right)^2 = V^{4/3}\xi^2$$

  - For nth order Campostrini integration (if it were stable) the cost is

$$C_\Omega \propto V\left(V\delta\tau^{4n+6}\right)^{\frac{1}{2n+3}}\left(\xi/\delta\tau\right)^2 = V^{\frac{2n+4}{2n+3}}\xi^2$$

Cost and Dynamical Critical Exponents

- L2MC (Kramers)
  - In the approximation of unit acceptance rate we find that setting $\tau = \delta\tau$ and suitably tuning θ we can arrange to get $z = 1$
  - However, if $\langle P_{\mathrm{acc}} \rangle < 1$ then the system will carry out a random walk backwards and forwards along a trajectory because the momentum, and thus the direction of travel, must be reversed upon a Metropolis rejection.

Cost and Dynamical Critical Exponents

- A simple-minded analysis is that the average time between rejections must be $O(\xi)$ to achieve $z = 1$
- This time is approximately $\sum_{n=0}^{\infty} \langle P_{\rm acc} \rangle^n \left(1 - \langle P_{\rm acc} \rangle\right) n \quad \delta\tau = \dfrac{\langle P_{\rm acc} \rangle \delta\tau}{1 - \langle P_{\rm acc} \rangle}$
- For small $\delta\tau$ we have $1 - \langle P_{\rm acc} \rangle = \mathrm{erf} \sqrt{kV\delta\tau^6}$, hence we are required to scale $\delta\tau$ so as to keep $V\delta\tau^4\xi^2$ fixed
- This leads to a cost for L2MC of

$$C \propto V \left(V\delta\tau^4\xi^2\right)^{1/4} \left(\xi / \delta\tau\right) = V^{5/4}\xi^{3/2}$$

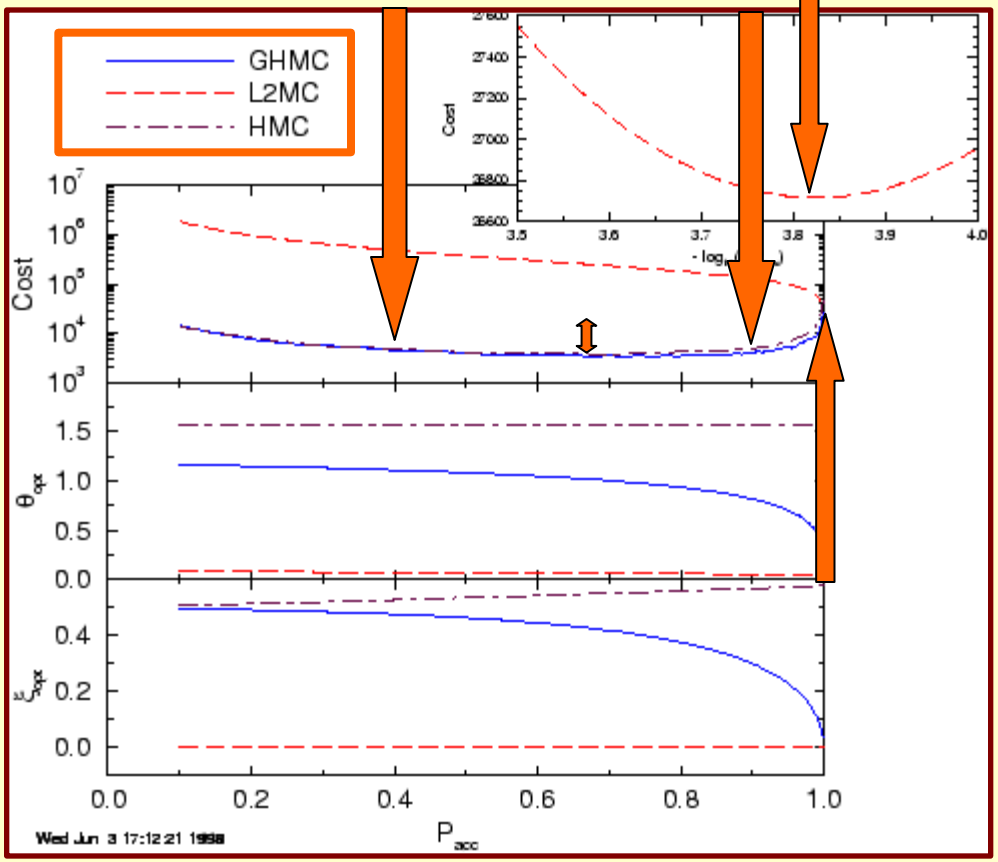
- Or using nth order Campostrini integration

$$C \propto V \left(V\delta\tau^{4n+4}\xi^2\right)^{\frac{1}{4n+4}} \left(\xi / \delta\tau\right) = V^{\frac{4n+5}{4n+4}}\xi^{\frac{2n+3}{2n+2}}$$

- A more careful free field theory analysis leads to the same conclusions

Cost and Dynamical Critical Exponents

- Optimal parameters for GHMC
  - (Almost) analytic calculation *in free field theory*
  - Minimum cost for GHMC appears for acceptance probability in the range 40%-90%
  - Very similar to HMC
  - Minimum cost for L2MC (Kramers) occurs for acceptance rate very close to 1
  - And this cost is much larger

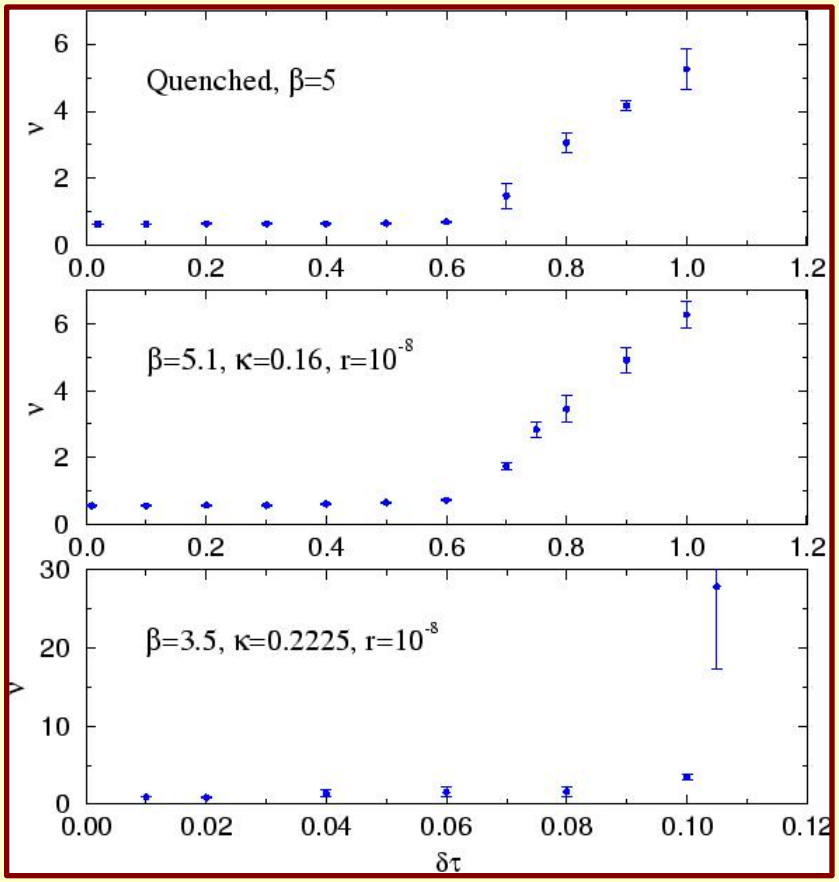


Reversibility

- 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 might do this if we want to use (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
  - Liapunov exponents characterise the divergence of nearby trajectories
  - The instability in the integrator occurs when $\delta H \gg 1$
    - Zero acceptance rate anyhow

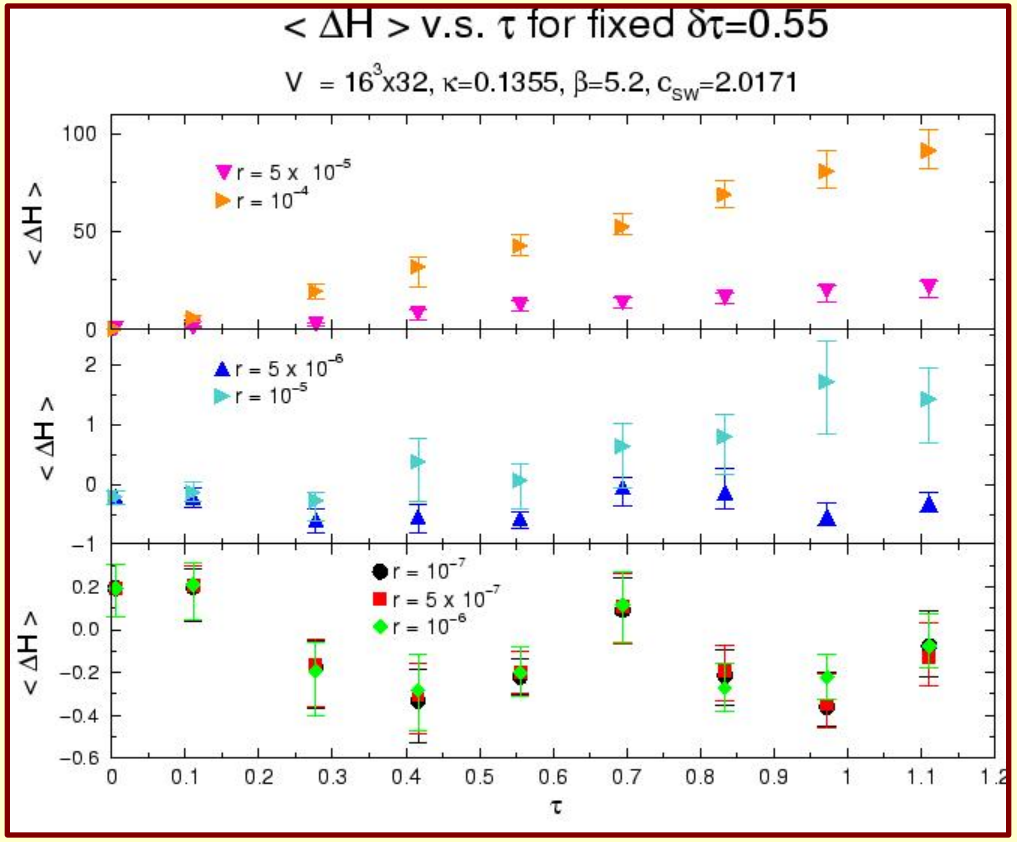


Reversibility

- In QCD the Liapunov exponents appear to scale with β as the system approaches the continuum limit $\beta \rightarrow \infty$
  - $\nu\xi = constant$
  - This can be interpreted as saying that the Liapunov 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

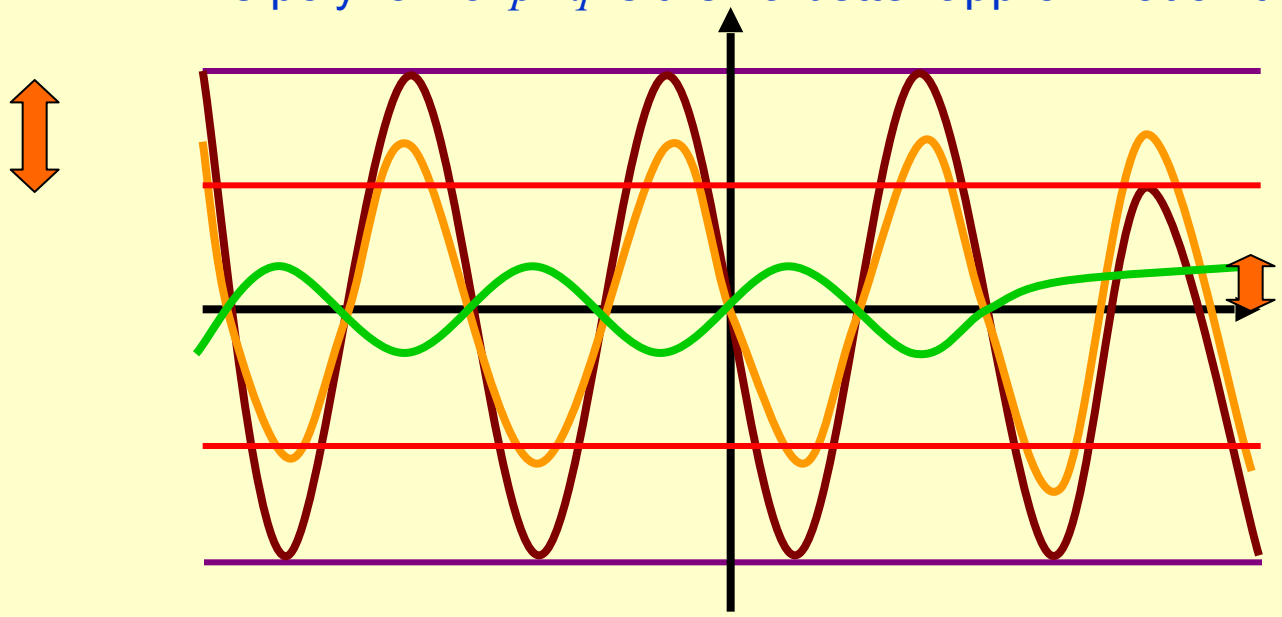


Chebyshev Approximation

- What is the best polynomial approximation $p(x)$ to a continuous function $f(x)$ for x in $[0,1]$?
  - Weierstrass' theorem: any continuous function can be arbitrarily well approximated by a polynomial
    - Bernstein polynomials: $p_n(x) \equiv \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} t^n (1-t)^{n-k}$
  - Minimise the appropriate norm $\|p-f\|_n = \left(\int_0^1 dx \; |p(x) - f(x)|^n \right)^{1/n}$ where $n \geq 1$
  - Chebyshev's theorem
    - The error $|p(x) - f(x)|$ reaches its maximum at exactly $d+2$ points on the unit interval
    - There is always a unique polynomial of any degree d which minimises $\|p-f\|_\infty = \max_{0 \leq x \leq 1} |p(x) - f(x)|$

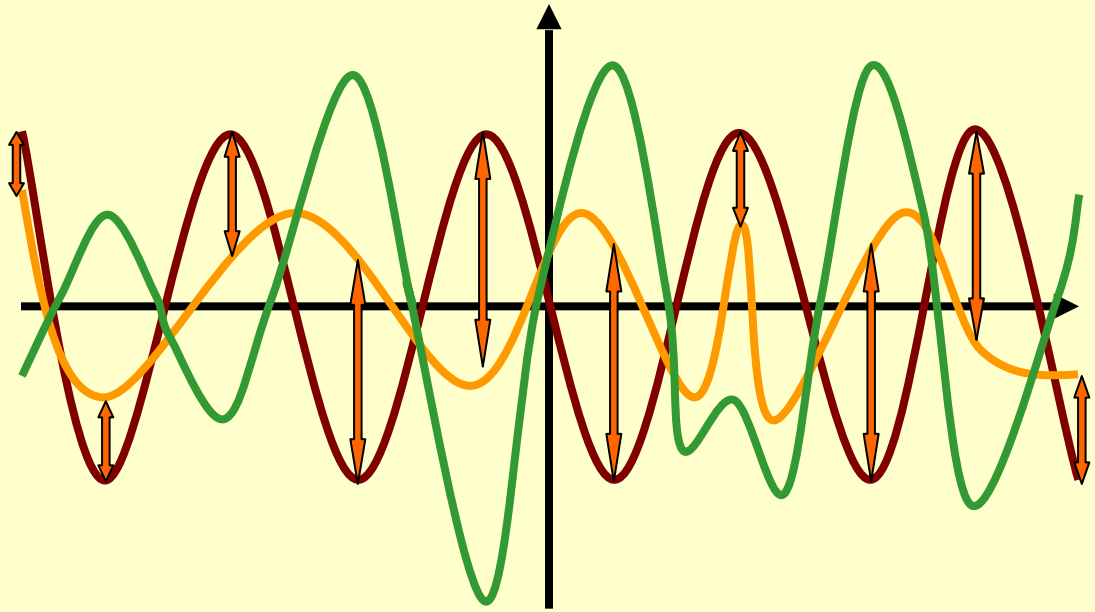
Chebyshev Approximation

- 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



Chebyshev Approximation

- 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 of $|p(x) - f(x)|$
  - Therefore $p' - p$ must have $d+1$ zeros on the unit interval
  - Thus $p' - p = 0$ as it is a polynomial of degree d



Chebyshev Approximation

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

Chebyshev Approximation

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

Chebyshev Approximation

- A realistic example of a rational approximation is

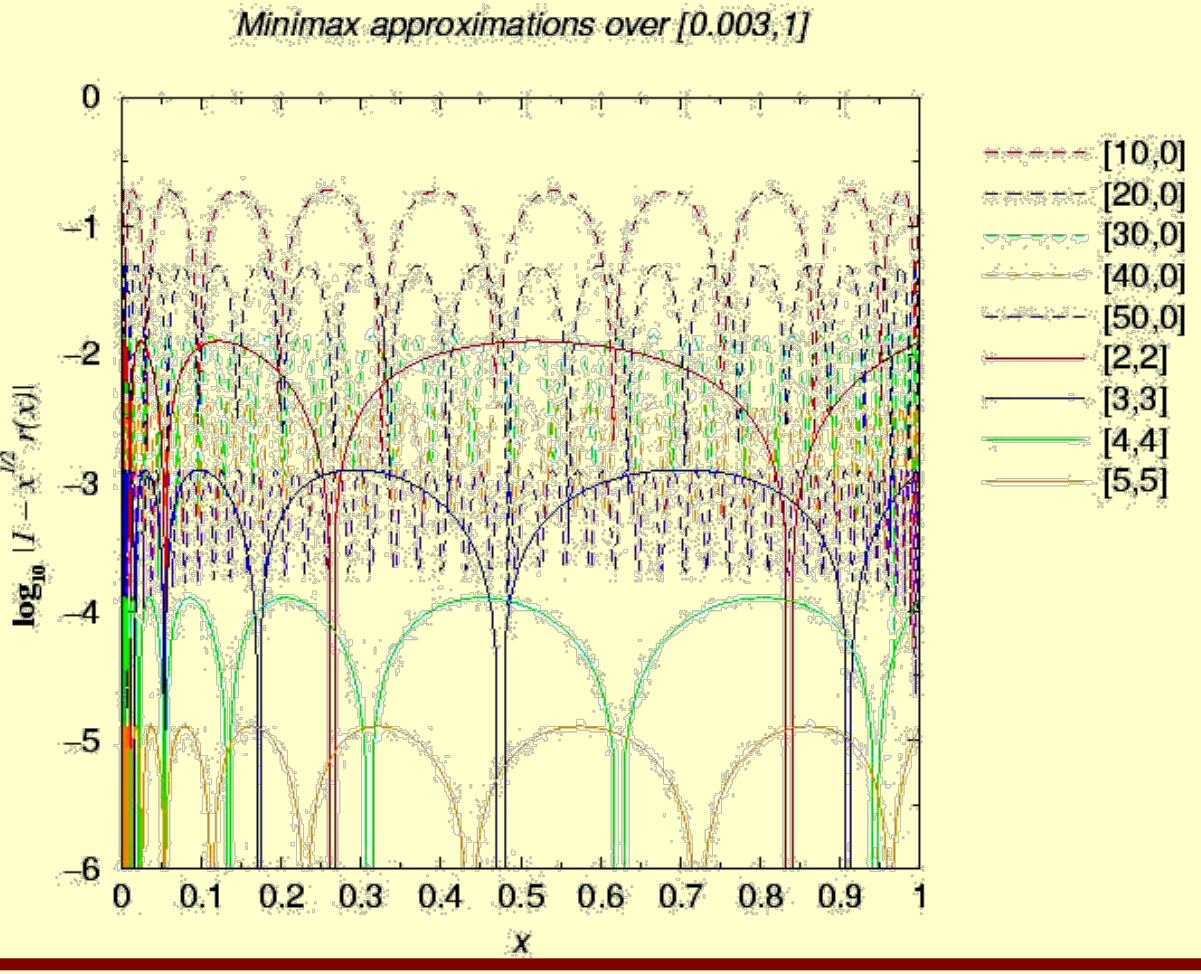
$$\frac{1}{\sqrt{x}} \approx 0.3904603901 \frac{(x + 2.3475661045)(x + 0.1048344600)(x + 0.0073063814)}{(x + 0.4105999719)(x + 0.0286165446)(x + 0.0012779193)}$$

  - 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 multiple mass linear equation solver, thus reducing the cost significantly.
- The partial fraction expansion of the rational function above is

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

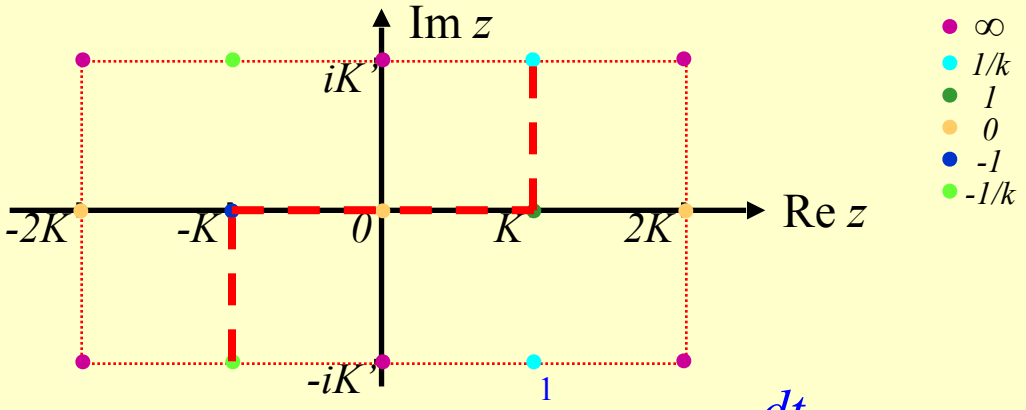
  - This appears to be numerically stable.

Chebyshev Approximation



Золотарев Formula

- Elliptic function are doubly periodic analytic functions
- Jacobi elliptic function $z = \int_0^{\mathrm{sn}\, z} \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}$
  - Real period $4K$, $K=\mathrm{K}(k)$
  - Imaginary period $2iK'$, $K'=\mathrm{K}(k')$, $k^2+k'^2=1$

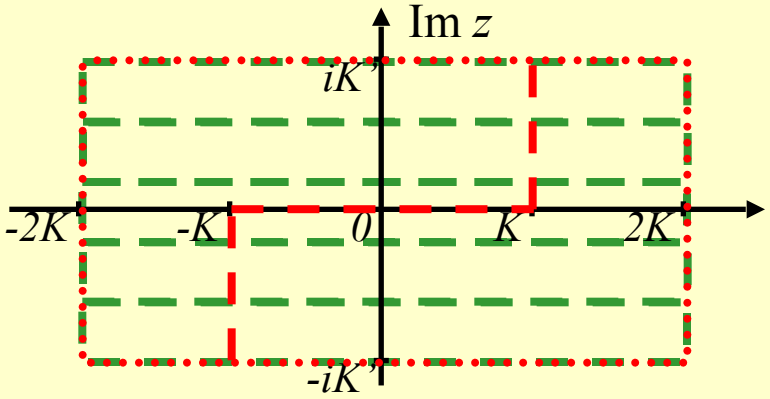


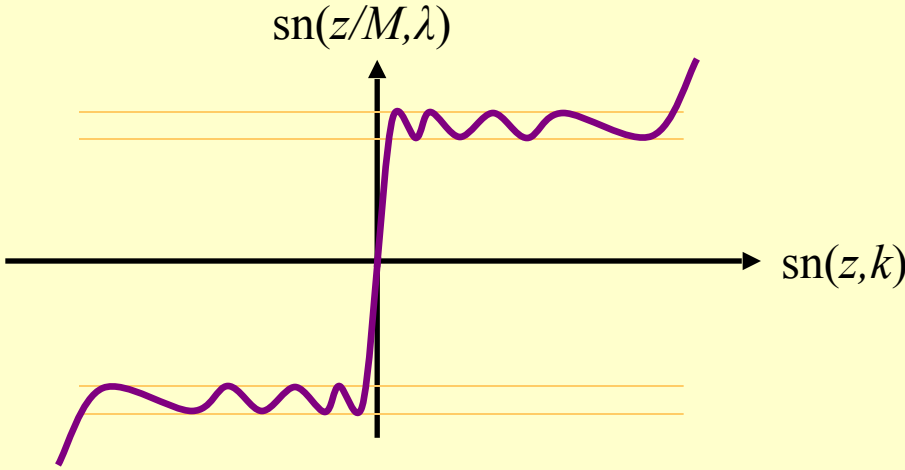
- Complete elliptic integral $\mathrm{K}(k) = \int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}$

Золотарев Formula

- All analytic functions with the same periods may be expressed as a rational function of $\mathrm{sn}(z,k)$
- Modular transformations
- Divide imaginary period by n

$$\mathrm{sn}\left(\frac{z}{M},\lambda\right)=\mathrm{sn}\left(z,k\right)\frac{1}{M}\prod_{m=1}^{\left|\frac{n}{2}\right|}\frac{1-\dfrac{\mathrm{sn}\left(z,k\right)^2}{\mathrm{sn}\left(2iK'm/n,k\right)^2}}{1-\dfrac{\mathrm{sn}\left(z,k\right)^2}{\mathrm{sn}\left(2iK'\left(m-\frac{1}{2}\right)/n,k\right)^2}}$$





Multibosons

- Chebyshev polynomial approximations were introduced by Lüscher in his *Multiboson* method

$$\det\left[M\right] \approx \det\left[P\left(M\right)^{-1}\right] \propto \det\left[\prod_i \left(M-\lambda_i\right)^{-1}\right]$$
$$= \prod_i \det\left[M-\lambda_i\right]^{-1} \propto \prod_i \int\left[d\varphi_i\right] e^{-\varphi^\dagger\left(M-\lambda_i\right)\varphi}$$

  - No solution of linear equations is required
  - One must store n scalar fields
  - The dynamics of multiboson fields is stiff, so very small step sizes are required for gauge field updates

Reweighting

- Making Lüscher's method exact
  - One can introduce an accept/reject step
  - One can reweight the configurations by the ratio $\det[M]\ \det[P(M)]$
    - This factor is close to one if $P(M)$ is a good approximation
    - If one only approximates $1/x$ over the interval $[\varepsilon, 1]$ which does not cover the spectrum of M, then the reweighting factor may be large
    - It has been suggested that this will help to sample configurations which would otherwise be suppressed by small eigenvalues of the Dirac operator

Noisy methods

- Since the GHMC algorithm is correct for any reversible and area-preserving mapping, it is often useful to use some modified MD process which is cheaper
  - Accept/reject step must use the true Hamiltonian, of course
  - Tune the parameters in the MD Hamiltonian to maximise the Metropolis acceptance rate
  - Add irrelevant operators?
    - A different kind of “improvement”

Noisy Inexact Algorithms

- Replace the force $-S'$ in the leapfrog equations of motion by a noisy estimator $\hat{F}$ such that $\left\langle \hat{F} \right\rangle_{\eta} = -S'$
  - Choose the noise η independently for each step
  - The equation for the shift in the fixed point distribution is now $\left\langle e^{-(\delta H + \delta \Delta S)} \right\rangle_{\pi,\eta} = 1$
  - For the noisy Langevin algorithm we have $\left\langle \delta H \right\rangle_{\pi,\eta} = O\left(\delta\tau^2\right)$ and $\left\langle \delta \Delta S_n \right\rangle_{\pi,\eta} = O\left(\delta\tau^2\right)$, but since $\left\langle e^{-\delta H} \right\rangle_{\pi,\eta} = O\left(\delta\tau^4\right)$ we obtain $\Delta S = O\left(\delta\tau^2\right)$
  - For the noisy Hybrid algorithm $\left\langle \delta H \right\rangle_{\pi,\eta} = O\left(\delta\tau\right)$ and $\left\langle \delta \Delta S_n \right\rangle_{\pi,\eta} = O\left(\delta\tau^0\right)$, so $\Delta S = O\left(\delta\tau\right)$

Noisy Inexact Algorithms

- This method is useful for non-local actions, such as staggered fermions with $n_f \neq 4$ flavours
  - Here the action may be written as $\frac{1}{4} n_f \operatorname{tr} \ln M$, and the force is $-n_f \operatorname{tr}\left(M^{-1} M'\right)$
    - A cheap way of estimating such a trace noisily is to use the fact that $\operatorname{tr} Q = \left\langle \sum_{ij} \eta_i Q_{ij} \eta_j \right\rangle_\eta$
  - Using a non area preserving irreversible update step the noisy Hybrid algorithm can be adjusted to have $\left\langle \delta H \right\rangle_{\pi,\eta} = O\left(\delta\tau^2\right)$, and thus $\Delta S = O\left(\delta\tau^2\right)$ too
    - This is the *Hybrid R algorithm*
    - One must use Gaussian noise (Z_2 noise does not work)
  - Campostrini's integration scheme may be used to produce higher order Langevin and Hybrid algorithms with $\Delta S = O\left(\delta\tau^{2n}\right)$ for arbitrary n, but this is not applicable to the noisy versions

Kennedy—Kuti Algorithm

- Suppose it is prohibitively expensive to evaluate $R \equiv Q(\phi')/Q(\phi)$, but that we can compute an unbiased estimator for it cheaply, $\left\langle \hat{R}(\eta) \right\rangle_\eta = R$
- If we look carefully at the proof that the Metropolis algorithm satisfies detailed balance we see that the ratio R is used for two quite different purposes
  - It is used to give an ordering to configurations: $\phi' < \phi$ if $R < 1$, that is, if $Q(\phi') < Q(\phi)$
  - It is used as the acceptance probability if $\phi' < \phi$
- We can produce a valid algorithm using the noisy estimator $\hat{R}$ for the latter rôle just by choosing another ordering for the configurations
  - A suitable ordering is provided by any "cheap" function f such that the set of configurations for which $f(\phi') = f(\phi)$ has measure zero

Kennedy—Kuti Algorithm

- We now define the acceptance probability as

$$P(\phi \to \phi') = \left(\lambda_+ + \lambda_- \hat{R}\right) \theta\left[f(\phi) - f(\phi')\right] + \left(\lambda_- + \lambda_+ \hat{R}\right) \theta\left[f(\phi') - f(\phi)\right]$$

  - Where $\lambda_\pm$ are to be chosen so that $0 \le P(\phi \to \phi') \le 1$
    - Taking $\lambda_+ = \frac{1}{2}, \lambda_- = 0$ is often a convenient choice
  - Detailed balance is easily established by considering the cases $f(\phi) > f(\phi')$ and $f(\phi) < f(\phi')$ separately
  - In practice the condition $0 \le P(\phi \to \phi') \le 1$ can rarely be satisfied exactly, but usually the number of violation can be made (exponentially) small

Noisy Fermions

- An interesting application is when we have a non-local fermionic determinant in the fixed point distribution, $\det M(\phi) = \exp \mathrm{tr} \ln \left[M(\phi) \right]$
- In this case we need to produce an unbiased estimator of

$$R = \frac{\det M(\phi')}{\det M(\phi)} = \exp \mathrm{tr} \ln \left[\{ \mathrm{M} + \Delta \mathrm{M} \} M^{-1} \right]$$

$$= \exp \mathrm{tr} \ln \left[1 + \Delta M M^{-1} \right] = \exp \mathrm{tr} \ln \left[1 + \Delta Q \right]$$

Noisy Fermions

- It is easy to construct unbiased estimators for $\mathrm{tr}\left[\Delta Q^n\right]$
  - Let η be a vector of random numbers whose components are chosen independently from a distribution with mean zero and variance one (Gaussian or Z_2 noise, for example)
  - Set $\hat{Q}_n \equiv \sum_{ij} \eta_i \left(\Delta Q^n\right)_{ij} \eta_j$, then $\left\langle \hat{Q}_n \right\rangle_\eta = \mathrm{tr}\left[\Delta Q^n\right]$
  - Furthermore, if $\hat{Q}_n$ and $\hat{Q}'_{n'}$ are independent, then

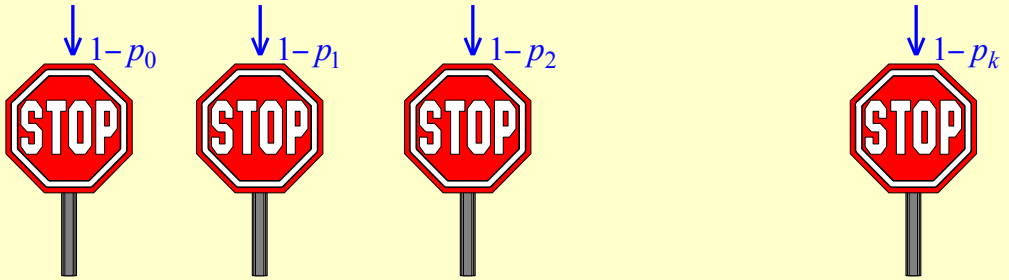
$$\left\langle \hat{Q}_n \hat{Q}'_{n'} \right\rangle_\eta = \mathrm{tr}\left[\Delta Q^n\right] \mathrm{tr}\left[\Delta Q^{n'}\right]$$

- As long as all the eigenvalues of ΔQ lie within the unit circle then $\mathrm{tr}\ln\left[1+\Delta \mathrm{Q}\right] = -\sum_{n=1}^{\infty} \frac{1}{n} \mathrm{tr}\left[\left(-\Delta \mathrm{Q}\right)^n\right]$
  - Similarly the exponential can be expanded as a power series

Bhanot–Kennedy algorithm

- In order to obtain an unbiased estimator for R we sum these series stochastically
  - Suppose $f(x) = 1 + \sum_{n=1}^{\infty} f_n x^n$ where $0 \le p_n \equiv f_{n+1}/f_n \le 1$
    - Our series can be transformed into this form
  - This can be "factored" as $f(x) = 1 + p_0 \left(x + p_1 \left(x^2 + p_2 \left(x^3 + \cdots \right. \right. \right.$
  - And may be summed stochastically using the following scheme

$$\hat{f}(x) = 1 \xrightarrow{p_0} +x \xrightarrow{p_1} +x^2 \xrightarrow{p_2} \cdots \xrightarrow{p_{k-1}} +x^k \xrightarrow{p_k} \cdots$$



“Kentucky” Algorithm

- This gets rid of the systematic errors of the Kennedy—Kuti method by eliminating the violations
  - Promote the noise to the status of dynamical fields
  - Suppose that we have an unbiased estimator

$$\left\langle \hat{R}(\phi,\eta)\right\rangle_{\eta} = \int [d\eta]\; p(\eta)\hat{R}(\phi,\eta) = \det\left[M(\phi)\right]$$

$$\langle\Omega\rangle = \frac{1}{Z}\int [d\phi]\; e^{-S(\phi)} \det\left[M(\phi)\right]\; \Omega(\phi) = \frac{1}{Z}\int [d\phi]\; e^{-S(\phi)} \left\langle \hat{R}(\phi,\eta)\right\rangle_{\eta} \Omega(\phi)$$

$$= \frac{1}{Z}\int [d\phi][d\eta]\; e^{-S(\phi)} p(\eta)\hat{R}(\phi,\eta)\Omega(\phi)$$

$$= \frac{1}{Z}\int [d\phi][d\eta]\; e^{-S(\phi)} p(\eta)\left|\hat{R}(\phi,\eta)\right| \mathrm{sgn}\left(\hat{R}(\phi,\eta)\right)\Omega(\phi)$$

$$= \left\langle \mathrm{sgn}\left(\hat{R}(\phi,\eta)\right)\Omega(\phi)\right\rangle_{\phi,\eta}$$

“Kentucky” Algorithm

- The ϕ and η fields can be updated by alternating Metropolis Markov steps
- There is a sign problem only if the Kennedy—Kuti algorithm would have many violations
- If one constructs the estimator using the Kennedy—Bhanot algorithm then one will need an infinite number of noise fields in principle
  - Will these ever reach equilibrium?

Cost of Noisy Algorithms

- Inexact algorithms
  - These have only a trivial linear volume dependence, with $z = 1$ for Hybrid and $z = 2$ for Langevin
- Noisy inexact algorithms
  - The noisy trajectories deviate from the true classical trajectory by a factor of $1 + O(\delta\tau)$ for each step, or $1 + O\left(\sqrt{\delta\tau\xi}\right)$ for a trajectory of $\xi/\delta\tau$ steps
    - This will not affect the integrated autocorrelation function as long as $\delta\tau \quad \xi \ll 1$
    - Thus the noisy Langevin algorithm should have $C \propto V\xi^2$
    - Thus the noisy Hybrid algorithm should have $C \propto V(\delta\tau\xi)(\xi/\delta\tau) = V\xi^2$

Cost of Noisy Algorithms

- Noisy exact algorithms
  - These algorithms use noisy estimators to produce an (almost) exact algorithm which is applicable to non-local actions
    - A straightforward approach leads to a cost of $C \propto V\left(V\delta\tau^2\right)\left(\xi/\delta\tau\right)^2 = V^2\xi$ for exact noisy Langevin
      - It is amusing to note that this algorithm should not care what force term is used in the equations of motion
    - An exact noisy Hybrid algorithm is also possible, and for it

$$C \propto V\left(V\delta\tau\right)\left(\xi/\delta\tau\right) = V^2\xi$$

Cost of Noisy Algorithms

- These results apply only to operators (like the magnetisation) which couple sufficiently strongly to the slowest modes of the system. For other operators, like the energy in ≥ 2 dimensions, we can even obtain $z = 0$

Cost of Noisy Algorithms

- Summary
  - Too little noise increases critical slowing down because the system is too weakly ergodic
  - Too much noise increases critical slowing down because the system takes a drunkard's walk through phase space
  - To attain $z = 1$ for any operator (and especially for the exponential autocorrelation time) one must be able to tune the amount of noise suitably

PHMC

- Polynomial Hybrid Monte Carlo algorithm
  - Instead of using Chebyshev polynomials in the multiboson algorithm, Frezzotti & Jansen and deForcrand suggested using them directly in HMC
  - Polynomial approximation to $1/x$ are typically of order 40 to 100 at present
  - Numerical stability problems with high order polynomial evaluation
  - Polynomials must be factored
  - Correct ordering of the roots is important
  - Frezzotti & Jansen claim there are advantages from using reweighting

RHMC

- Rational Hybrid Monte Carlo algorithm
  - The idea is similar to PHMC, but uses rational approximations instead of polynomial ones
  - Much lower orders required for a given accuracy
  - Evaluation simplified by using partial fraction expansion and multiple mass linear equation solvers
  - $1/x$ is already a rational function, so RHMC reduces to HMC in this case
  - Can be made exact using noisy accept/reject step

Ginsparg–Wilson fermions

- Is it possible to have chiral symmetry on the lattice without doublers if we only insist that the symmetry holds on shell?
  - Such a transformation should be of the form *(Lüscher)* $\psi \to e^{\alpha\gamma_5\left(1-\frac{1}{2}aD\right)}\psi\,;\ \overline{\psi} \to \overline{\psi}e^{\alpha\gamma_5\left(1-\frac{1}{2}aD\right)}$
  - For it to be a symmetry the Dirac operator must be invariant $D \to e^{\alpha\left(1-\frac{1}{2}aD\right)\gamma_5} D e^{\alpha\gamma_5\left(1-\frac{1}{2}aD\right)} = D$
  - For a small transformation this implies that
$$\left(1-\tfrac{1}{2}aD\right)\gamma_5 D + D\gamma_5\left(1-\tfrac{1}{2}aD\right) = 0$$
- Which is the *Ginsparg-Wilson* relation
$$\gamma_5 D + D\gamma_5 = aD\gamma_5 D$$

Neuberger's Operator

- We can find a solution of the Ginsparg-Wilson relation as follows
  - Let the lattice Dirac operator to be of the form

$$aD = 1 + \gamma_5 \hat{\gamma}_5; \ \hat{\gamma}_5^\dagger = \hat{\gamma}_5; \ aD^\dagger = 1 + \hat{\gamma}_5 \gamma_5 = \gamma_5 aD \gamma_5$$

  - This satisfies the GW relation if $\hat{\gamma}_5^2 = 1$
  - And it must also have the correct continuum limit

$$D \to \not{\partial} + m \Rightarrow \hat{\gamma}_5 = \gamma_5 \left[a\left(\not{\partial} + m \right) - 1 \right] + O\left(a^2 \right)$$

- Both of these conditions are satisfied if we define *(Neuberger)*

$$\hat{\gamma}_5 = \gamma_5 \frac{D_W - 1}{\sqrt{\left(D_W - 1 \right)^\dagger \left(D_W - 1 \right)}} = \mathrm{sgn}\left[\gamma_5 \left(D_w - 1 \right) \right]$$

Neuberger's Operator

- There are many other possible solutions
- The discontinuity is necessary
- This operator is local *(Lüscher)*
  - At least if we constrain the fields such that the plaquette < 1/30
  - By local we mean a function of fast decrease, as opposed to *ultralocal* which means a function of compact support

Computing Neuberger's Operator

- Use polynomial approximation to Neuberger's operator
  - High degree polynomials have numerical instabilities
  - For dynamical GW fermions this leads to a PHMC algorithm
- Use rational approximation
  - Optimal rational approximations for $\mathrm{sgn}(x)$ are know in closed form (Zolotarev)
  - Requires solution of linear equations just to apply the operator
  - For dynamical GW fermions this leads to an RHMC algorithm
  - Requires *nested* linear equation solvers in the dynamical case
  - Nested solvers can be avoided at the price of a much more ill-conditioned system
  - Attempts to combine inner and outer solves in one Krylov space method

Computing Neuberger's Operator

- Extract low-lying eigenvectors explicitly, and evaluate their contribution to the Dirac operator directly
  - Efficient methods based on Ritz functional
  - Very costly for dynamical fermions if there is a finite density of zero eigenvalues in the continuum limit (Banks—Casher)
  - Might allow for noisy accept/reject step if we can replace the step function with something continuous (so it has a reasonable series expansion)
- Use better approximate solution of GW relation instead of Wilson operator
  - E.g., a relatively cheap "perfect action"

Update Ordering for LHMC

Consider the update of a single site x using the LHMC algorithm

$$\phi'(x) = (1-\zeta)\phi(x) + \frac{\zeta F}{\omega^2} + \frac{\sqrt{\zeta(2-\zeta)}}{\omega}\pi$$

- The values of ϕ at all other sites are left unchanged
- This may be written in matrix form as $\phi' = M_x\phi + P_x\pi$ where

$$(M_x)_{yz} = \delta_{yz} + \delta_{xy}\zeta\left[-\delta_{yz} + \frac{1}{\omega^2}\sum_{\mu}\left(\delta_{y+\hat{\mu},z} + \delta_{y-\hat{\mu},z}\right)\right]$$

$$(P_x)_{yz} = \delta_{yz}\delta_{xy}\frac{\sqrt{\zeta(2-\zeta)}}{\omega}$$

Update Ordering for LHMC

- For a complete sweep through the lattice consisting of exactly $V = L^d$ updates in some order $x_1, x_2, \ldots, x_V$, we have

$$\phi \mapsto M_{x_1}\phi + P_{x_1}\pi \mapsto M_{x_2}\left(M_{x_1}\phi + P_{x_1}\pi\right)\phi + P_{x_2}\pi \mapsto \cdots \mapsto M\phi + P\pi$$

  - It is convenient to introduce the Fourier transformed fields $\tilde{\phi}$ and momenta $\tilde{\pi}$
  - The corresponding Fourier transformed single site update matrix is

$$\left(\tilde{M}_x\right)_{pq} = \delta_{pq} + \frac{1}{V} e^{-\frac{2\pi i (p-q)\cdot x}{L}} \zeta \left[\frac{2}{\omega^2} \sum_{\mu} \cos\frac{2\pi}{L} q_\mu - 1\right]$$

Update Ordering for LHMC

- We want to compute integrated autocorrelation functions for operators like the magnetic susceptibility $\chi = \tilde{\phi}_0^2 - \left\langle \tilde{\phi}_0^2 \right\rangle$
  - The behaviour of the autocorrelations of linear operators such as the magnetisation $\tilde{\phi}_0$ can be misleading
- In order to do this it is useful to consider quadratic operators as linear operators on quadratic monomials in the fields, $\tilde{\Phi}_{pq} \equiv \tilde{\phi}_p \tilde{\phi}_q$, so $\chi = \tilde{\Phi}_{00} - \left\langle \tilde{\Phi}_{00} \right\rangle$
  - These quadratic monomials are updated according to
$$\tilde{\Phi}'_{pq} = \tilde{\phi}'_p \tilde{\phi}'_q = \left(M\tilde{\phi} + P\tilde{\pi} \right)_p \left(M\tilde{\phi} + P\tilde{\pi} \right)_q$$
  - So, after averaging over the momenta we have $\left\langle \tilde{\Phi}' \right\rangle_\pi = M^Q \tilde{\Phi} + P^Q$
  - Where $P^Q_{pq} = \sum_r P_{pr} P_{qr}$ and $M^Q_{pq,rs} = M_{pr} M_{qs}$

Update Ordering for LHMC

- The integrated autocorrelation function for χ is

$$1 + A_\chi = \sum_{t=0}^{\infty} C_\chi(t) = \sum_{t=0}^{\infty} \frac{\langle \chi(t)\chi(0) \rangle}{\langle \chi^2 \rangle}$$

$$= \sum_{t=0}^{\infty} \frac{\sum_{pq} \left[\left(M^Q \right)^t \right]_{00,pq} \left[\langle \tilde{\Phi}_{pq} \tilde{\Phi}_{00} \rangle - \langle \tilde{\Phi}_{pq} \rangle \langle \tilde{\Phi}_{00} \rangle \right]}{\langle \tilde{\Phi}_{00}^2 \rangle - \langle \tilde{\Phi}_{00} \rangle^2}$$

$$= \sum_{t=0}^{\infty} \left[\left(M^Q \right)^t \right]_{00,00} = \left[\left(1 - M^Q \right)^{-1} \right]_{00,00}$$

Update Ordering for LHMC

- Random updates
  - The update matrix for a sweep after averaging over the V independent choices of update sites is just

$$\tilde{M}^Q = \prod_{i=1}^{V} \left[\frac{1}{V} \sum_{z_i=1}^{V} \tilde{M}^Q_{z_i} \right] = \left[\frac{1}{V} \sum_{z=1}^{V} \tilde{M}^Q_z \right]^V$$

  - We thus find that $1 + A_\chi$

$$= \left[\left(1 - \tilde{M}^Q \right)^{-1} \right]_{00,00} = \frac{1}{1 - e^{-\frac{2\zeta m^2}{m^2 + 2d}}} + O\left(\frac{1}{V}\right) = \frac{d}{\zeta m^2} + O\left(m^0\right) + O\left(\frac{1}{V}\right)$$

  - Which tells us that $z = 2$ for any choice of the overrelaxation parameter ζ

Update Ordering for LHMC

- Even/Odd updates
  - In a single site update the new value of the field at x only depends at the old value at x and its nearest neighbours, so even sites depend only on odd sites and vice versa
    - The update matrix for a sweep in Fourier space thus reduces to a 2×2 matrix coupling $\tilde{\phi}_p$ and $\tilde{\phi}_{p+\frac{1}{2}(L,\ldots,L)}$
    - $\tilde{M}^Q$ thus becomes a 3×3 matrix in this case
  - The integrated autocorrelation function is

$$A_\chi = \frac{d(2-\zeta)}{2\zeta m^2} + \frac{7\zeta^2 - 16\zeta + 8}{9\zeta(2-\zeta)} + O(m^4)$$

    - This is minimised by choosing $\zeta = 2 - \frac{m}{\sqrt{d}} + O(m^2)$, for which

$$A_\zeta = \frac{\sqrt{d}}{m} - \frac{1}{2} + O(m)$$

    - Hence $z = 1$

Update Ordering for LHMC

- Lexicographical updates
  - This is a scheme in which $\phi_{x-\hat{\mu}}$ is always updated before ϕ_x
    - Except at the edges of the lattice
  - A single site update depends on the new values of the neighbouring sites in the $+\mu$ directions and the old values in the $-\mu$ directions
  - So we may write the sweep update implicitly as $\phi' = O\phi + N\phi' + \bar{P}\pi$
    - Where

$$O_{xy} = (1-\zeta)\delta_{xy} + \frac{\zeta}{\omega^2}\sum_{\mu}\delta_{x,y-\hat{\mu}} + \frac{1}{L}R_{xy} \quad N_{xy} = \frac{\zeta}{\omega^2}\sum_{\mu}\delta_{x,y+\hat{\mu}} - \frac{1}{L}R_{xy}$$

$$R_{xy} = \frac{L\zeta}{\omega^2}\sum_{\mu}\left(\delta_{x,y+\hat{\mu}}\delta_{x_\mu 1} - \delta_{x,y-\hat{\mu}}\delta_{x_\mu L}\right) \quad \bar{P} = \frac{\sqrt{\zeta(2-\zeta)}}{\omega}$$

Update Ordering for LHMC

- The sweep update matrices are easily found explicitly

$$M = (1 - N)^{-1} O \qquad P = (1 - N)^{-1} \bar{P}$$

- The Fourier transforms $\tilde{O}$ and $\tilde{N}$ are diagonal up to surface terms $\tilde{R}$, which we expect to be suppressed by a factor of $1/L$
- We thus obtain $1 + A_\chi = \dfrac{\left(m^2 + 2d - \zeta\right)^2}{\zeta\left(2 - \zeta\right)m^2\left(m^2 + 2d\right)} + O\left(\dfrac{1}{L}\right)$
  - Which is minimised by the choice $\zeta = \dfrac{m^2 + 2d}{m^2 + d}$
  - For which $A_\chi = 0$ and hence $z = 0$
- This can be achieved for most operators
  - Though with different values of ζ
  - The dynamical critical exponent for the exponential autocorrelation time is still one at best