Algorithms in Lattice QCD V

Stefan Schaefer

NIC, DESY

Kolkata Lattice Gauge Theory School

What happened so far

Methods for Markov Chain Monte Carlo

Sequence of field configuratoins

$\rightarrow$MC time series of measurements

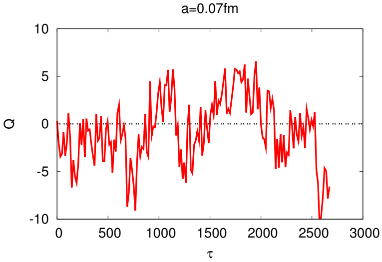
Field updates are expensive $\rightarrow$ limited statistics

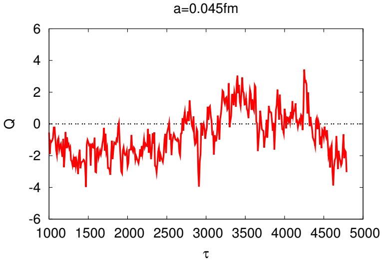
Outline for today

Methods to deal with autocorrelations

Bad start

Topological charge





$a \approx 0.08\text{fm}$

64×32^3

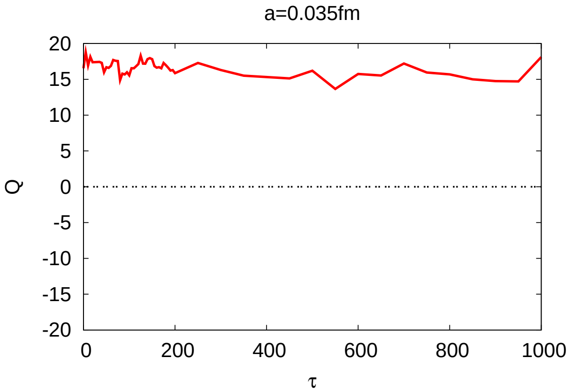
$m_\pi \approx 360\text{MeV}$

$a \approx 0.06\text{fm}$

64×32^3

$m_\pi \approx 460\text{MeV}$

A bad start



$a \approx 0.04$fm

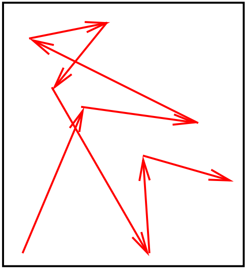
128×64^3

$m_\pi \approx 480$MeV

Markov Chain Monte Carlo

Sequence of field configurations

$$U_1 \to U_2 \to U_3 \to \cdots \to U_N$$



Generated by a **transition probability density**

$$T(U' \leftarrow U) \geq 0 \quad \text{for all } U, U'$$

Stability

$$\int [dU]\, T(U' \leftarrow U)\, P[U] = P[U']$$

Normalization

$$\int [dU']\, T(U' \leftarrow U) = 1$$

Autocorrelations

Sequence of field configurations

$$U_1 \to U_2 \to U_3 \to \cdots \to U_N$$

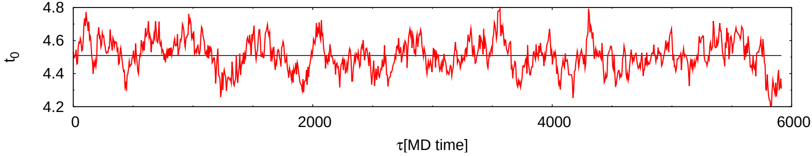
Measurements of observables are correlated

$$A_1 \to A_2 \to A_3 \to \cdots \to A_N$$

Estimates

$$\bar{A} \approx \tilde{A}_N = \frac{1}{N} \sum_{i=1}^{N} A_i$$

How far is this off?



Autocorrelations

Variance of estimator

$$\langle\!\langle (\tilde{A}_N - \bar{A})^2 \rangle\!\rangle = \frac{1}{N^2} \sum_{i,j=1}^{N} \langle\!\langle (A_i - A)(A_j - A) \rangle\!\rangle$$

For N large, this depends only on the difference in simulation time

$$\langle (\tilde{A}_N - \bar{A})^2 \rangle = \frac{1}{N} \sum_{t=-\infty}^{\infty} \Gamma_A(t)$$

$$\Gamma_A(t) = \langle (A_0 - \bar{A})(A_t - \bar{A}) \rangle$$

Note:
again substitution average over simulations
$\rightarrow$ average in simulation time

Error of the measurement

$$\langle(\tilde{A}-\bar{A})^2\rangle = \frac{1}{N}\sum_{t=-\infty}^{\infty}\Gamma_A(t)\ ; \qquad \Gamma_A(t) \quad = \langle(A_0-A)(A_t-A)\rangle$$

$$= \frac{\mathrm{var}(A)}{N}\sum_{t=-\infty}^{\infty}\rho_A(t)$$

Integrated autocorrelation time

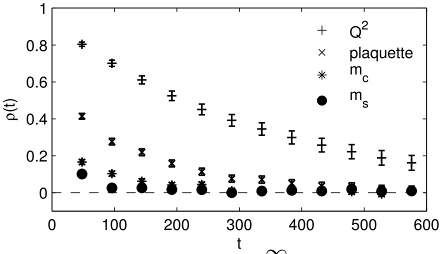
$$\tau_{\mathrm{int}}(A) = \frac{1}{2}+\sum_{t=1}^{\infty}\frac{\Gamma_A(t)}{\Gamma_A(0)} \equiv \frac{1}{2}+\sum_{t=0}^{\infty}\rho_A(t)$$

Error of the measurement

$$\sigma_A = \sqrt{\frac{\mathrm{var}(A)}{N/(2\tau_{\mathrm{int}}(A))}}$$

Measures efficiency of algorithm.
$\rightarrow$ eff. statistics reduced by $2\tau_{\mathrm{int}}$

Measuring autocorrelations



$$\tau_{\rm int}(A) = \frac{1}{2} + \sum_{t=1}^{\infty} \rho_A(t)$$

We only have a limited precision estimate of the integrand.
Summing to $t = \infty$ leads to diverging variance.
$\rightarrow$ need to cut the summation
$\rightarrow$ biased estimate
Need to find a balance between stat. and syst. error.

Error of τ_{int}

$$\tau_{\text{int}}(A) = \frac{1}{2} + \sum_{t=1}^{W} \frac{\Gamma_A(t)}{\Gamma_A(0)}$$

Systematic error

Summation truncated at W
$\rightarrow$ neglect potentially large tail.
Particular problem in presence of slow modes.

Statistical error Madras,Sokal

$$\langle [\tilde{\tau}_{\text{int}}(A,W) - \tau_{\text{int}}(A,W)]^2 \rangle \approx \frac{4}{N}(W + \frac{1}{2} - \tau_{\text{int}}(A))\tau_{\text{int}}(A)^2$$

Infinite variance for $W \rightarrow \infty$.

Criteria for W

All automatic methods are problematic.

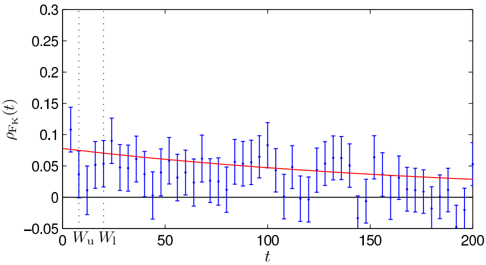
- Cut where $\delta\Gamma > \Gamma$
  $\rightarrow$ large systematic error
- Madras-Sokal criterion
  $\rightarrow$ minimum of sum of systematic and statistical error

$$\frac{\delta\sqrt{\tau}}{\sqrt{\tau}} \propto \min_W \left(e^{-W/\tau} + 2\sqrt{W/N}\right)$$

- ALPHA method (2010)
  Estimate τ_{exp} from various (slow) observables
  Add tail to all other observables before losing signal Γ_A

Attaching the tail

ALPHA collaboration, 2012



Use slow observables and scaling laws to estimate tail.

Detailed balance

Detailed balance

$$T(U' \leftarrow U)\,P[U] = P[U']\,T(U' \leftarrow U)$$

Implies stability

$$\int [dU] T(U' \leftarrow U) P[U] = \int [dU] P[U'] T(U' \leftarrow U) = P[U']$$

Elementary steps frequently fulfill this condition.

As a consequence we have a **symmetric matrix** M

$$M(U' \leftarrow U) = P[U']^{-1/2} T(U' \leftarrow U) P[U]^{1/2}$$

Detailed balance

Detailed balance

$$T(U' \leftarrow U)\, P[U] = P[U']\, T(U' \leftarrow U)$$

Associated symmetric matrix M

$$M(U' \leftarrow U) = P[U']^{-1/2} T(U' \leftarrow U) P[U]^{1/2}$$

If η eigenvector of T

$$\xi(U) = P^{-1/2}(U)\eta(U)$$

is eigenvector of M **with the same eigenvalue** λ.

Spectral decomposition

$$M = \sum_i \lambda_i \, \xi_i \xi_i^\dagger$$

Autocorrelation

Spectral decomposition

$$
\begin{aligned}
\Gamma_A(t) &= \langle\, (A_t - \bar{A})\,(A_0 - \bar{A})\,\rangle \\
&= \int [dU][dU'] \delta A(U')\, T^t(U' \leftarrow U)\, \delta A(U)\, P[U] \\
&= \int [dU][dU'] P^{1/2}[U'] \delta A(U') M^t(U' \leftarrow U) \delta A(U) P^{1/2}[U] \\
&= \sum_{n>0} (\lambda_n)^t\, [c_n(A)]^2
\end{aligned}
$$

With “matrix elements”

$$
c_n(A) = \int [dU] \xi_n(U) [P[U]]^{1/2} (A(U) - \bar{A})
$$

Spectral representation

$$\begin{aligned}\Gamma_A(t) &= \sum_n (\lambda_n)^t \, [c_n(A)]^2 \\ &= \sum_n \mathrm{sign}\lambda_n \, e^{-t/\tau_n} \, [c_n(A)]^2\end{aligned}$$

- $\tau_n = 1/\log|\lambda_n| > 0$
- For the analysis of algorithms it is useful to think of Monte Carlo time t as a fifth dimension.
- Autocorrelation function is a 2pt function.
- time constants $\tau_n \to$ inverse masses
- Slowest decay $\tau_1 \to$ exponential AC time

Comments

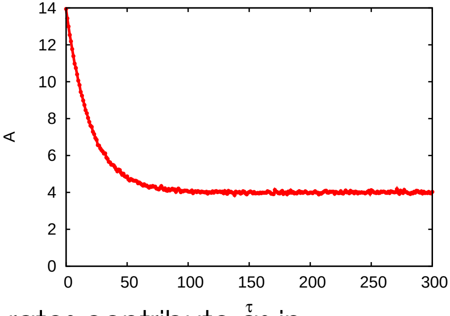
$$\Gamma_A(t) = \sum_n e^{-t/\tau_n} c_n^2(A)$$

- τ_n depend only on algorithm
- Matrix elements c_n depend on observable.
- All observables affected by slow modes.

Length of a simulation

- Simulation must have length of at least $\mathrm{O}(100) \times \tau_1$.
- $\tau_{\mathrm{int}}(A)$ can be much smaller than τ_1
- Danger of
  - Incomplete thermalization.
  - Bias.
  - Wrong estimate of autocorrelations.

Thermalization



Same decay rates contribute as in $\tau_{\rm int}$
different initial distribution/matrix elements

$$\int [dU][dU']P^{1/2}[U']\delta A(U')M^t(U' \leftarrow U)\delta A(U)\frac{P_0[U]}{P^{1/2}[U]}$$
$$= \sum_{n>0}(\lambda_n)^t\,[c_n(A)][c_n^{(0)}(A)]$$

Opportunity to learn about largest τ_1.

Observables

Need to look at observables with large $c_i(A_k)$, i small.
Hunt for slow quantities.

Noise can cover up auto-correlations (η Gaussian noise)

$$A \to A + c\eta \qquad \Rightarrow \qquad \Gamma(t) \to \Gamma(t) + c^2 \delta_{t,0}$$

$$\Rightarrow \qquad \tau_{\text{int}}(A) \to \tau_{\text{int}}(A) \frac{\text{var}(A)}{\text{var}(A) + c^2}$$

Look at low-noise observables

Take into consideration expected **scaling properties**

$$\tau_{\text{int}} \propto \frac{1}{a^2} \text{ for } a \to 0$$

Wilson flow

LÜSCHER'10, LÜSCHER&WEISZ'11

- Smoothing with **gradient flow** at fixed flow time $t = t_0$.

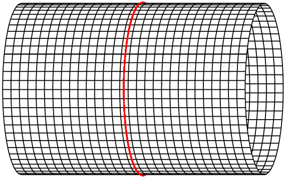
$$\partial_t V_t(x,\mu) = -g_0^2 \left[\partial_{x,\mu} S(V_t)\right] V_t(x,\mu); \quad V_t(x,\mu)|_{t=} = U(x,\mu)$$

- Gaussian smoothing over $r \sim \sqrt{8t}$.
- Renormalized quantities with continuum limit.
- Smooth observables $\rightarrow$ long autocorrelations.

$$\overline{E} = -\frac{a^3}{2L^3} \sum_{\vec{x}} \operatorname{tr} G_{\mu\nu} G_{\mu\nu}\Big|_{x_0=T/2}$$

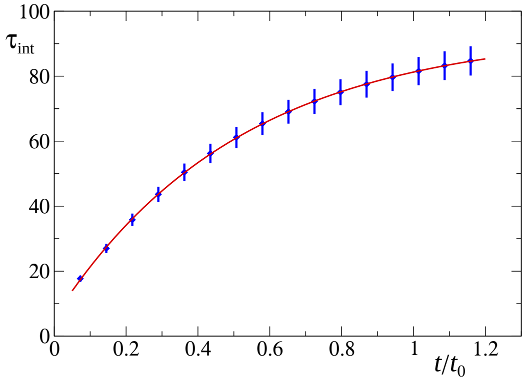
$$\overline{Q} = -\frac{a^3}{32\pi^2} \sum_{\vec{x}} \operatorname{tr} \tilde{G}_{\mu\nu} G_{\mu\nu}\Big|_{x_0=T/2}$$

$$Q = -\frac{a^4}{32\pi^2} \sum_{x} \operatorname{tr} \tilde{G}_{\mu\nu} G_{\mu\nu}$$



Effect of the smoothing

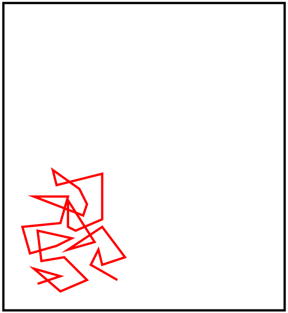
Autocorrelation time of $\bar{E}$ vs smoothing range (a=0.05fm).



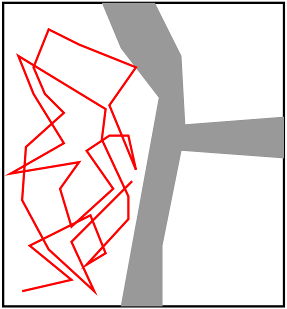
- $\sqrt{8t}$ smoothing radius $\rightarrow t = t_0$ smoothing over $r \approx r_0$
- τ_{int} saturates with $\tau_{\text{int}} = 93 + ae^{-ct}$.

Dangers

- Algorithm is slow.
- Detectable by measuring autocorrelations.



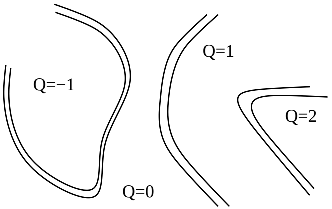
- There are barriers in field space.
- Hard to detect.



Topological charge

$$Q = -\frac{1}{32\pi^2} \int d\,x\, \epsilon_{\mu\nu\rho\sigma} \mathrm{tr}\, F_{\mu\nu} F_{\rho\sigma}$$

- In continuum limit, disconnected **topological sectors** emerge.
- The probability of configurations "in between" sectors drops rapidly. M. Lüscher, '10
- Simulations get stuck in one sector.

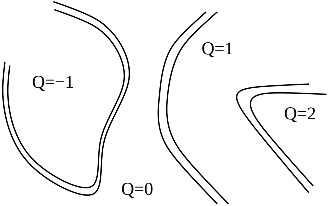


Topological charge

- **Tunneling is a cut-off effect**.
- Quasi continuous algorithms will not cure it.
- Problem for interpretation of data.
- Fixed topology introduces finite volume effects.

$$\langle A \rangle = \langle A \rangle_{Q=Q_0} \cdot \{1 + \mathcal{O}(V^{-1})\}$$

- Prevents simulations on fine lattices.



Topological charge

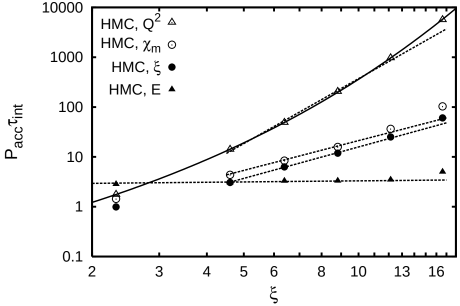
Best recipe: Avoid large autocorrelations τ_n.

Special case: Topological charge

In the continuum, topological sectors form.
Consequence of the periodic boundary conditions.

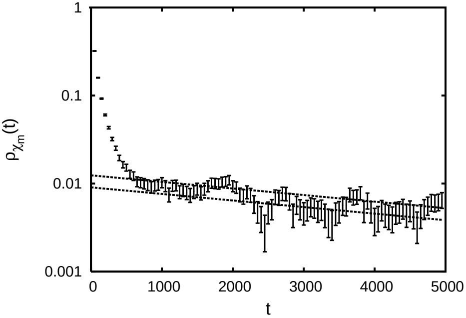
Happens very quickly as $a \to 0$.

Engel, S.'10



Example from CP^{N-1} model

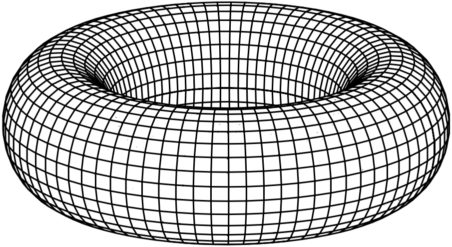
Topological charge AC dominates other observables.



Solution:
Use setup without topological sectors
$\rightarrow$ Open boundary conditions.

Boundary conditions

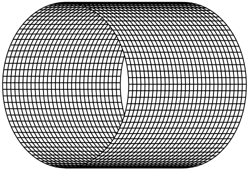
- Periodic boundary conditions do not let charge flow out of the volume.
- Field space is disconnected in continuum.



Open boundary conditions

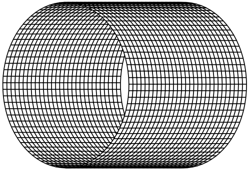
Proposed solution

- open boundary condition in time direction
  $\rightarrow$ same transfer matrix, same particle spectrum
- periodic boundary condition in spatial directions
  $\rightarrow$ momentum projection possible



Open boundary conditions

- Lattices of size $T \times L^3$.
- Neumann boundary conditions in time.



- Gauge fields

$$F_{0k}|_{x_0=0} = F_{0k}|_{x_0=T} = 0, \quad k = 1, 2, 3$$

- Fermion fields

$$P_+\psi(x)|_{x_0=0} = P_-\psi(x)|_{x_0=T} = 0 \qquad P_\pm = \frac{1}{2}(1 \pm \gamma_0)$$

$$\bar{\psi}(x)P_-|_{x_0=0} = \bar{\psi}(x)P_+|_{x_0=T} = 0$$

Critical slowing down

Continuum limit = continuous phase transition

Universal dynamical critical behavior

How do $\Gamma(t)$ and τ_{int} scale as $a \to 0$?

Free field expectation

$$\tau_{\text{int}} \propto a^{-1}$$

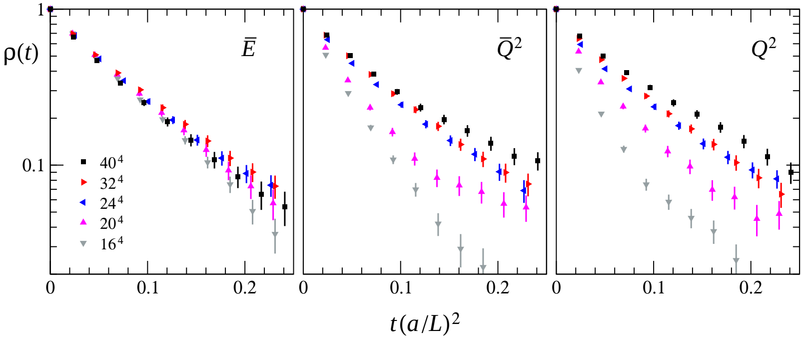
broken by interactions Lüscher, S '11

Expected Langevin (random walk) scaling

$$\tau_{\text{int}} \propto a^{-2}$$

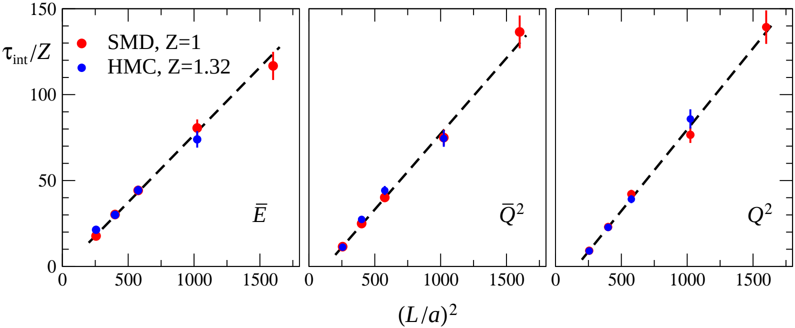
Scaling towards continuum limit

Autocorrelation function vs scaled MC time



- Energy (on time slice) shows very good scaling.
- Large cut-off effects in topological observables.

Scaling towards continuum limit: τ_{int} vs a^{-2}



- HMC and $\mathrm{SMD}_{0.3}$ show same scaling up to a constant. $\rightarrow$ universal behavior
- Topological observables well described by $\tau_{\mathrm{int}} = c_1 + c_2/a^2$
- **Also Q^2 and $\bar{Q}^2$ show a^2 scaling for $a \rightarrow 0$.**

What can we expect?

Experience

- Improved Wilson fermions, Iwasaki gauge action.
- 64×32^3 lattice, $a = 0.09$ fm
- **physical light and strange quark mass**, $m_\pi L = 2$
- $\tau_{\text{int}}(E) \sim \mathrm{O}(20)$

Estimate

- Twice larger lattice for $m_\pi L = 4$, $L \approx 6$ fm.
- Run length $100 \cdot \tau_{\text{int}}(E) = 2000 \cdot (a/0.09\text{fm})^{-2}$.

$$\text{cost} = 3\,\text{Tflops} \cdot \text{years} \cdot (a/0.09\text{fm})^{-7}$$

- $a = 0.045$ fm still cost 400 Tflops·years.