Ergodicity of the LLR method for the Density of States

Antonio Rago

Plymouth University

20^{th} June 2017

In collaboration with G. Cossu B. Lucini and R. Pellegrini

Motivations

- A large part of the success of Lattice Gauge Theory is inherently tied with advances in Monte Carlo simulations
- Monte Carlo methods used in Lattice Gauge Theory are importance sampling methods
- Most quantities of interest can be expressed in the path integral formalism as ensemble averages over a positive-definite (and sharply peaked) measure.
- Importance sampling methods are inefficient for
  - Studying systems with strong metastabilities or more generically systems with a rough free action landscape.
  - Direct computations of free energies.
  - Every time exceptional configurations play a role.
  - Studying systems with a sign problem.

An alternative approach to numerical simulations could accelerate progress in those cases (or at least in some of those).

Density of states

Let us consider a Euclidean quantum field theory

$$Z[\beta] = \int [D\phi]\, e^{-\beta S[\phi]}$$

The density of states is defined as

$$\rho(\mathcal{S}) = \int [D\phi]\, \delta(\mathcal{S} - S[\phi])$$

which leads to

$$Z[\beta] = \int \mathrm{d}\mathcal{S}\, \mathrm{e}^{-\beta\mathcal{S}} \rho(\mathcal{S})$$

Density of states

Let us consider a Euclidean quantum field theory

$$Z[\beta] = \int [D\phi]\, e^{-\beta S[\phi]}$$

The density of states is defined as

$$\rho(\mathcal{S}) = \int [D\phi]\, \delta(\mathcal{S} - S[\phi])$$

which leads to

$$Z[\beta] = \int \mathrm{d}\mathcal{S}\, \mathrm{e}^{-\beta \mathcal{S}} \rho(\mathcal{S})$$

The LLR algorithm gives access to a controllable approximation of the $\rho(\mathcal{S})$ featuring strong convergence properties.

[K. Langfeld, B. Lucini, R. Pellegrini and AR, Eur. Phys. J. C **76** (2016)]

The LLR method in a nutshell

1. Divide the action interval in N sub-intervals of amplitude δ_S, each centered at $\mathcal{S}_i = \mathcal{S}_{\min} + (i - \frac{1}{2})\delta_S$
2. In each sub-interval, compute $a_i = \frac{d \ln \rho}{d\mathcal{S}}\Big|_{\mathcal{S}=\mathcal{S}_i} + \mathcal{O}(\delta_S^2)$.

   The a_i are the zero of the stochastic equation: $\langle\langle \mathcal{S} - \mathcal{S}_i \rangle\rangle_i (a_i) = 0$,

   where: $$\langle\langle \mathcal{S} - \mathcal{S}_i \rangle\rangle_i (a) = \frac{1}{\mathcal{N}} \int d\mathcal{S}\, e^{\frac{(\mathcal{S}-\mathcal{S}_i)^2}{2\delta_S^2}} \rho(\mathcal{S}) (\mathcal{S} - \mathcal{S}_i)\, e^{-a\mathcal{S}} .$$

   the zeros can be found using: $a_i^{(n+1)} = a_i^{(n)} + c_n \langle\langle \mathcal{S} - \mathcal{S}_i \rangle\rangle_i (a_i^{(n)})$
3. From the knowledge of the coefficients a_i reconstruct the density of states as

   $$\rho_{\mathsf{LLR}}(\mathcal{S}) = \rho_0 \prod_{i=1}^{k-1} e^{a_i \delta_S} \exp(a_k(\mathcal{S} - \mathcal{S}_k)), \qquad \mathcal{S}_k \leq \mathcal{S} < \mathcal{S}_{k+1}$$
4. Evaluate the partition function and eventually any observables by means of a numerical one dimensional integration

Few remarks

- The LLR algorithm is a first principle method: $\rho(\mathcal{S}) = \rho_{\mathsf{LLR}}(\mathcal{S}) e^{c\delta_S^2}$ almost everywhere (the $\rho(\mathcal{S})$ is supposed to be almost everywhere C_2).
- The above equation shows exponential error suppression: the relative approximation error does not depend on the magnitude of ρ.
  The method works over several orders of magnitude!
- For observables, the convergence to their continuum action values is $\mathcal{O}(\delta_S^2)$
- The method allows us to compute generic observables, and not only observables that can be expressed as a function of the action
- The gaussian support function can be directly embedded in the HMC dynamic through:

$$H[a, p, \phi] = \sum \frac{p^2}{2} + aS[\phi] + \frac{(S[\phi] - \mathcal{S}_k)^2}{\delta_S^2}$$

with force $\quad f_i = -\frac{\partial S}{\partial \phi_i}(a + \frac{1}{\delta_S^2}(S[\phi] - \mathcal{S}_k))$

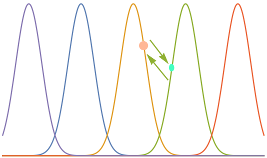
Ergodicity and efficiency

The choice (and width) of the support function can affect the efficiency of the update algorithm and its ergodicity properties.

- In principle, this algorithm is ergodic: given enough time, it will explore the entire phase space.
- However the probability of visiting states with action far from the peak of the Gaussian will be very small and this will lead to a slow dynamic of the Markov Chain.

- The proposed solution is to simultaneously simulate multiple overlapping intervals with fixed central action and periodically propose a swap of the configurations belgin to two of them with probability:

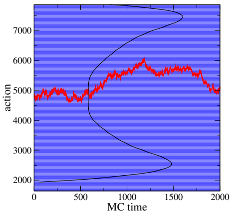
$$P_{sw} = \min(1, \exp(U[a_1, \phi^{(1)}, \mathcal{S}_1] + U[a_2, \phi^{(2)}, \mathcal{S}_2] \\ -U[a_2, \phi^{(1)}, \mathcal{S}_2] - U[a_1, \phi^{(2)}, \mathcal{S}_1]))$$

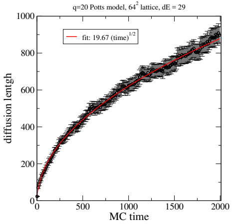


Preserving the detailed balance of action of the entire system.

Ergodicity and efficiency

- Subsequent exchanges allow any of the configuration sequences to travel through all the action intervals, hence overcoming any potential action barrier.
- The method has been already applied to the study of systems with strong metastabilities like the q-state Potts model at large q
  - The hopping between action intervals is reminiscent of a random walk
  - It is possible to associate to the process a diffusion coefficient independent from the action barrier

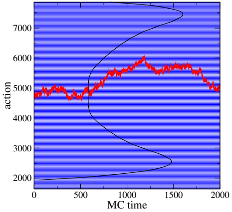


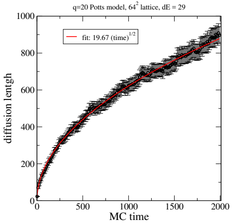


[B. Lucini, W. Fall and K. Langfeld, PoS LATTICE **2016** (2016) 275 [arXiv:1611.00019 [hep-lat]]]

Ergodicity and efficiency

- Subsequent exchanges allow any of the configuration sequences to travel through all the action intervals, hence overcoming any potential action barrier.
- The method has been already applied to the study of systems with strong metastabilities like the q-state Potts model at large q
  - The hopping between action intervals is reminiscent of a random walk
  - It is possible to associate to the process a diffusion coefficient independent from the action barrier





[B. Lucini, W. Fall and K. Langfeld, PoS LATTICE **2016** (2016) 275 [arXiv:1611.00019 [hep-lat]]]

What happens to theories with topologycal sectors
and to generic observables?

Generic observables

- We define an approximation of $\langle \mathcal{B}[\phi] \rangle$ by

$$\langle \mathcal{B}[\phi] \rangle_{\text{app}} = \frac{1}{\mathcal{N}(\beta)} \sum_i \delta_S \, \rho_{\text{LLR}}(\mathcal{S}_i) \, e^{-a_i \mathcal{S}_i} \left\langle\!\!\left\langle \mathcal{B}[\phi] \, e^{(a_i - \beta) S[\phi]} \right\rangle\!\!\right\rangle$$

$$\mathcal{N}(\beta) = \sum_i \delta_S \, \rho_{\text{LLR}}(\mathcal{S}_i) \; e^{-a_i \mathcal{S}_i} \left\langle\!\!\left\langle e^{(a_i - \beta) S[\phi]} \right\rangle\!\!\right\rangle$$

- It is possible to show that

$$\langle \mathcal{B}[\phi] \rangle \; = \; \langle \mathcal{B}[\phi] \rangle_{\text{app}} \Big[1 \; + \; \mathcal{O}(\delta_S^2) \; \Big].$$

Implying that the observable $\langle \mathcal{B}[\phi] \rangle$ can be calculated with a *relative error* of order δ_S^2.

Some care must be payed in evaluating the auto-correlation time of $\langle \mathcal{B}[\phi] \rangle_{\text{app}}$.
This can be done only by considering simultaneously all the contributions of each action interval to the observable at fixed Montecarlo step.

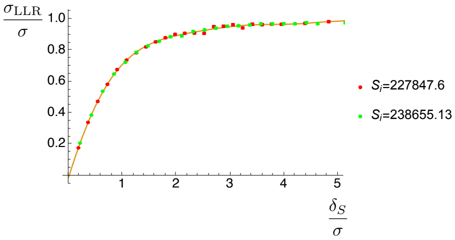
The model

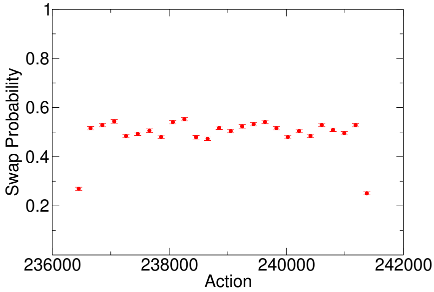
- We studied the pure $SU(3)$ Yang Mills model in $d = 4$.
- Our observables are
  - Topological Charge: evaluated by means of the Wilson Flow.
  - WF Energies: flowed plaquette and clover plaquette.
  - Action Observables: Every observable that can be written as a function of the action.
- Reference parameters

  [S. Schaefer *et al.* [ALPHA Collaboration], Nucl. Phys. B **845** (2011) 93 [arXiv:1009.5228 [hep-lat]].]
- Update algorithm: HMC, 2^{nd} ord Omelyan integrator, $\tau = 1$, acceptance $\sim 98\%$
- Range of actions corresponding to $5.789fm \leq \beta \leq 6.2$ or $0.140fm \geq a \geq 0.068fm$
- Lattice points 16^4, Lattice extension ranging $2.2fm$ and $1.1fm$

Tuning of the replica exchange method

- The swapping probability of two replicas will depend on the size of their overlapping probability distributions $(\mathcal{S}_i, \sigma_{\mathrm{LLR}})$
- The variance of a constrained action can be related to the variance of the action with same central value and no constraints
- We can choose the $\mathcal{S}_i$ and δ_S to obtain a flat swapping probability

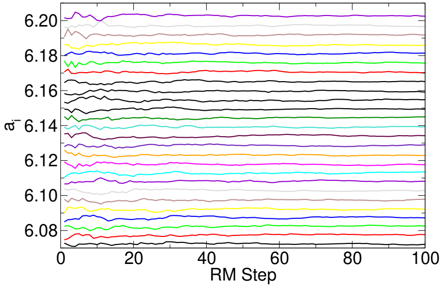




Robbins Monroe Iteration

The iteration $a_i^{(n+1)} = a_i^{(n)} + \frac{12}{\delta_S^2 (n+1)} \langle\langle \mathcal{S} - \mathcal{S}_i \rangle\rangle_i \, (a_i^{(n)})$:

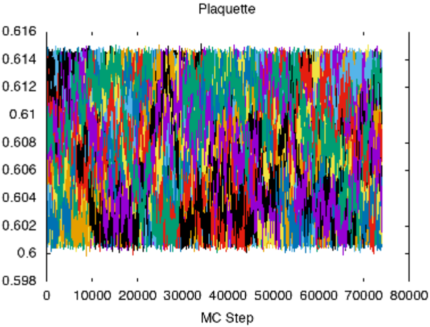
- Will converge to $\lim_{n\to\infty} a_i^{(n)} = a_i$
- $|a_i^{(n)} - a_i| \sim \frac{1}{n^\alpha}$
- The distribution of the $a_i^{(n)}$ is asymptotically normal around a_i

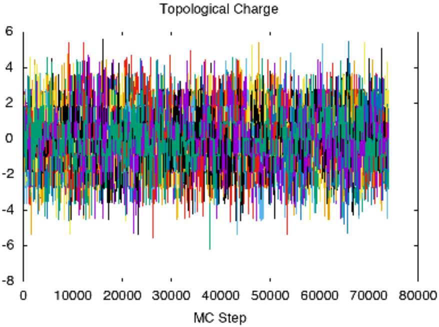


Observables

... a couple of "Pollock" colourful plots

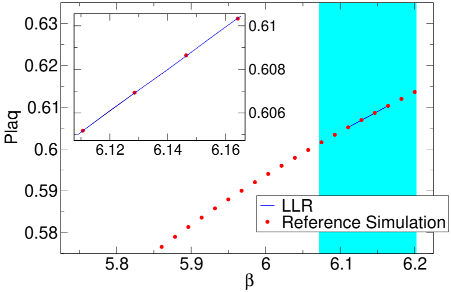
- The colour identifies the history of the trajectory
- The value of the plaquette "roughly" identifies the contribution to a single replica.

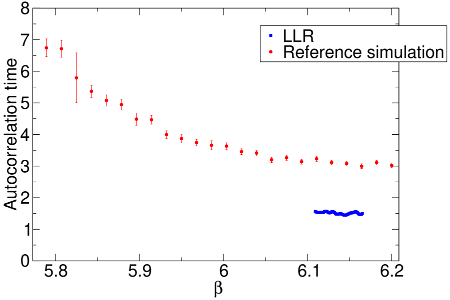




Action Observables

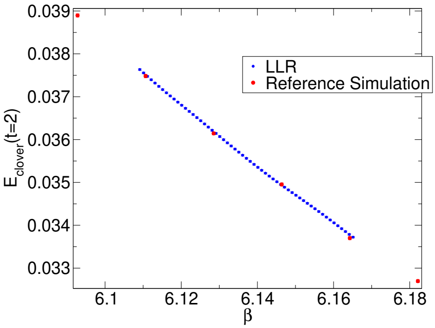
- We have compared our findings with the results of normal HMC simulations
- Autocorrelation time is evaluated with the Madras Sokal alghorithm
- $\sim 1.1 \cdot 10^5$ configuration per β for the Reference Simulations
- $\sim 0.7 \cdot 10^5$ configuration per $\mathcal{S}_i$ for the LLR Simulations

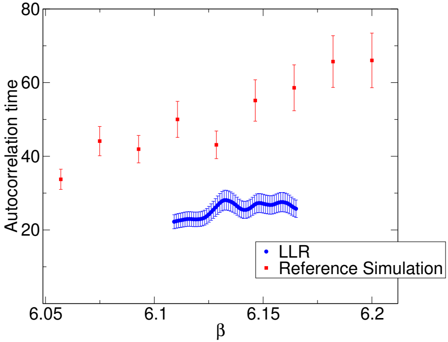




Observables: $E_{clover}(t=2)$

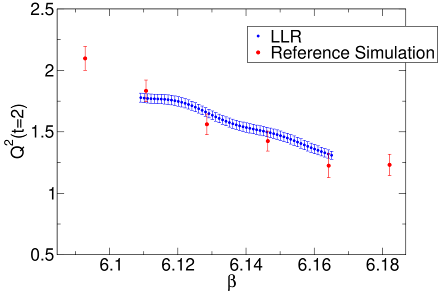
- We have compared our findings with the results of normal HMC simulations
- Reminder: $\langle \mathcal{B}[\phi] \rangle = \frac{1}{\mathcal{N}(\beta)} \sum_i \delta_S \, \rho_{\text{LLR}}(\mathcal{S}_i) \, e^{-a_i \mathcal{S}_i} \, \langle\langle \mathcal{B}[\phi] \, e^{(a_i - \beta) S[\phi]} \rangle\rangle$
- Autocorrelation time is evaluated with the Madras Sokal alghorithm directly on the replica weighted operator
- $\sim 1.1 \cdot 10^5$ configuration per β for the Reference Simulations
- $\sim 0.7 \cdot 10^5$ configuration per $\mathcal{S}_i$ for the LLR Simulations

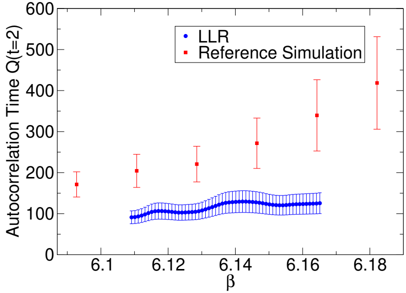




Observables: Topological Charge

- We have compared our findings with the results of normal HMC simulations
- Reminder: $\langle \mathcal{B}[\phi] \rangle = \frac{1}{\mathcal{N}(\beta)} \sum_i \delta_S \, \rho_{\mathsf{LLR}} \left(\mathcal{S}_i\right) e^{-a_i \mathcal{S}_i} \; \langle\!\langle \mathcal{B}[\phi] \, e^{(a_i - \beta) S[\phi]} \rangle\!\rangle$
- Autocorrelation time is evaluated with the Madras Sokal alghorithm directly on the replica weighted operator
- $\sim 1.1 \cdot 10^5$ configuration per β for the Reference Simulations
- $\sim 0.7 \cdot 10^5$ configuration per $\mathcal{S}_i$ for the LLR Simulations





Summary and Perspectives

- We studied the ergodicity properties of the LLR method analyzing the autocorrelation time of different observables
- The LLR method is naturally suited for parallel tempering techniques
- The observables show an improvement in autocorrelation time over normal MC measurement.
- The simulations are performed at a single choice of lattice points (but multiple beta values), further studies are needed to evaluate the continuum limit of the observables.
- The full cost balance for a single beta needs to address the dependance of the autocorrelation time with the boundary of the action span.