Swansea University
Prifysgol Abertawe

Cronfa
Setting Research Free

# Cronfa - Swansea University Open Access Repository

_____

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa28023

_____

_____

THE EUROPEAN
PHYSICAL JOURNAL C

CrossMark

Regular Article - Theoretical Physics

# An efficient algorithm for numerical computations of continuous densities of states

**K. Langfeld**[1,a]**, B. Lucini**[2,b]**, R. Pellegrini**[3,c]**, A. Rago**[1,d]

[1] Centre for Mathematical Sciences, Plymouth University, Plymouth PL4 8AA, UK
[2] College of Science, Swansea University, Singleton Park, Swansea SA2 8PP, UK
[3] School of Physics and Astronomy, University of Edinburgh, Edinburgh EH9 3FD, UK

**Abstract** In Wang–Landau type algorithms, Monte-Carlo updates are performed with respect to the density of states, which is iteratively refined during simulations. The partition function and thermodynamic observables are then obtained by standard integration. In this work, our recently introduced method in this class (the LLR approach) is analysed and further developed. Our approach is a histogram free method particularly suited for systems with continuous degrees of freedom giving rise to a continuum density of states, as it is commonly found in lattice gauge theories and in some statistical mechanics systems. We show that the method possesses an exponential error suppression that allows us to estimate the density of states over several orders of magnitude with nearly constant *relative* precision. We explain how ergodicity issues can be avoided and how expectation values of arbitrary observables can be obtained within this framework. We then demonstrate the method using compact U(1) lattice gauge theory as a show case. A thorough study of the algorithm parameter dependence of the results is performed and compared with the analytically expected behaviour. We obtain high precision values for the critical coupling for the phase transition and for the peak value of the specific heat for lattice sizes ranging from $8^4$ to $20^4$. Our results perfectly agree with the reference values reported in the literature, which covers lattice sizes up to $18^4$. Robust results for the $20^4$ volume are obtained for the first time. This latter investigation, which, due to strong metastabilities developed at the pseudo-critical coupling of the system, so far has been out of reach even on supercomputers with importance sampling approaches, has been performed to high accuracy with modest computational resources. This shows

the potential of the method for studies of first order phase transitions. Other situations where the method is expected to be superior to importance sampling techniques are pointed out.

## Contents

## 1 Introduction and motivations

Monte-Carlo methods are widely used in theoretical physics, statistical mechanics and condensed matter (for an overview, see e.g. [1]). Since the inception of the field [2], most of the applications have relied on importance sampling, which allows us to evaluate stochastically with a controllable error multi-dimensional integrals of localised functions. These

[a] e-mail: kurt.langfeld@plymouth.ac.uk

[b] e-mail: b.lucini@swansea.ac.uk

[c] e-mail: r.pellegrini@ed.ac.uk

[d] e-mail: antonio.rago@plymouth.ac.uk

Springer

methods have immediate applications when one needs to compute thermodynamic properties, since statistical averages of (most) observables can be computed efficiently with importance sampling techniques. Similarly, in lattice gauge theories, most quantities of interest can be expressed in the path integral formalism as ensemble averages over a positive-definite (and sharply peaked) measure, which, once again, provide an ideal scenario for applying importance sampling methods.

However, there are noticeable cases in which Monte-Carlo importance sampling methods are either very inefficient or produce inherently wrong results for well understood reasons. Among those cases, some of the most relevant situations include systems with a sign problem (see [3] for a recent review), direct computations of free energies (comprising the study of properties of interfaces), systems with strong metastabilities (for instance, a system with a first order phase transition in the region in which the phases coexist) and systems with a rough free energy landscape. Alternatives to importance sampling techniques do exist, but generally they are less efficient in standard cases and hence their use is limited to ad hoc situations in which more standard methods are inapplicable. Noticeable exceptions are microcanonical methods, which have experienced a surge in interest in the past 15 years. Most of the growing popularity of those methods is due to the work of Wang and Landau [4], which provided an efficient algorithm to access the density of states in a statistical system with a discrete spectrum. Once the density of states is known, the partition function (and from it all thermodynamic properties of the system) can be reconstructed by performing one-dimensional numerical integrals. Histogram-based straightforward generalisations of the Wang–Landau algorithm to models with a continuum spectrum have been shown to break down even on systems of moderate size [5,6], hence more sophisticate techniques have to be employed, as done for instance in [7], where the Wang–Landau method is used to compute the weights for a multi-canonical recursion (see also [8]).

A very promising method, here referred to as the logarithmic linear relaxation (LLR) algorithm, was introduced in [9]. The potentialities of the method were demonstrated in subsequent studies of systems afflicted by a sign problem [10,11], in the computation of the Polyakov loop probability distribution function in two-colour QCD with heavy quarks at finite density [12] and—rather unexpectedly—even in the determination of thermodynamic properties of systems with a discrete energy spectrum [13].

The main purpose of this work is to discuss in detail some improvements of the original LLR algorithm and to formally prove that expectation values of observables computed with this method converge to the correct result, which fills a gap in the current literature. In addition, we apply the algorithm to the study of compact U(1) lattice gauge theory, a system with

severe metastabilities at its first order phase transition point that make the determination of observables near the transition very difficult from a numerical point of view. We find that in the LLR approach correlation times near criticality grow *at most quadratically* with the volume, as opposed to the exponential growth that one expects with importance sampling methods. This investigation shows the efficiency of the LLR method when dealing with systems having a first order phase transition. These results suggest that the LLR method can be efficient at overcoming numerical metastabilities in other classes of systems with a multi-peaked probability distribution, such as those with rough free energy landscapes (as commonly found, for instance, in models of protein folding or spin glasses).

The rest of the paper is organised as follows. In Sect. 2 we cover the formal general aspects of the algorithm. The investigation of compact U(1) lattice gauge theory is reported in Sect. 3. A critical analysis of our findings, our conclusions and our future plans are presented in Sect. 4. Finally, some technical material is discussed in the appendix. Some preliminary results of this study have already been presented in [14].

## 2 Numerical determination of the density of states

### 2.1 The density of states

Owing to formal similarities between the two fields, the approach we are proposing can be applied to both statistical mechanics and lattice field theory systems. In order to keep the discussion as general as possible, we shall introduce notations and conventions that can describe simultaneously both cases. We shall consider a system described by the set of dynamical variables $\phi$, which could represent a set of spin or field variables and are assumed to be continuous. The action (in the field theory case) or the Hamiltonian (for the statistical system) is indicated by $S$ and the coupling (or inverse temperature) by $\beta$. Since the product $\beta S$ is dimensionless, without loss of generality we will take both $S$ and $\beta$ dimensionless.

We consider a system with a finite volume $V$, which will be sent to infinity in the final step of our calculations. The finiteness of $V$ in the intermediate steps allows us to define naturally a measure over the variables $\phi$, which we shall call $\mathcal{D}\phi$. Properties of the system can be derived from the function

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

which defines the canonical partition function for the statistical system or the path integral in the field theory case. The density of state (which is a function of the value of $S[\phi] = E$) is formally defined by the integral

$$\rho(E) = \int \mathcal{D}\phi \, \delta\Big(S[\phi] - E\Big). \tag{2.1}$$

In terms of $\rho(E)$, $Z$ takes the form

$$Z(\beta) = \int dE \, \rho(E) \, e^{\beta E}.$$

The vacuum expectation value (or ensemble average) of an observable $O$ which is a function of $E$ can be written as[1]

$$\langle \mathcal{O} \rangle = \frac{1}{Z(\beta)} \int dE \, \mathcal{O}(E) \, \rho(E) \, e^{\beta E}. \tag{2.2}$$

Hence, a numerical determination of $\rho(E)$ would enable us to express $Z$ and $\langle O \rangle$ as numerical integrals of known functions in the single variable $E$. This approach is inherently different from conventional Monte-Carlo calculations, which relies on the concept of importance sampling, i.e. the configurations contributing to the integral are generated with probability

$$P_\beta(E) = \rho(E) \, e^{\beta E}/Z(\beta).$$

Owing to this conceptual difference, the method we are proposing can overcome notorious drawbacks of importance sampling techniques.

## 2.2 The LLR method

We will now detail our approach to the evaluation of the density of states by means of a lattice simulation. Our initial assumption is that the density of states is a regular function of the energy that can be always approximated in a finite interval by a suitable functional expansion. If we consider the energy interval $[E_k, E_k + \delta_E]$, under the physically motivated assumption that the density of states is a smooth function in this interval, the logarithm of the latter quantity can be written, using Taylor's theorem, as

$$\ln \rho(E) = \ln \rho\left(E_k + \frac{\delta_E}{2}\right) + \frac{d \ln \rho}{dE}\bigg|_{E=E_k+\delta_E/2}$$
$$\times \left(E - E_k - \frac{\delta_E}{2}\right) + R_k(E), \tag{2.3}$$
$$R_k(E) = \frac{1}{2} \frac{d^2 \ln \rho}{dE^2}\bigg|_{E_k+\delta_E/2} \left(E - E_k - \frac{\delta_E}{2}\right)^2 + \mathcal{O}(\delta_E^3).$$

Thereby, for a given action $E$, the integer $k$ is chosen such that

$$E_k \leq E \leq E_k + \delta_E, \quad E_k = E_0 + k\,\delta_E.$$

---

Our goal will be to devise a numerical method to calculate the Taylor coefficients

$$a_k := \frac{d \ln \rho}{dE}\bigg|_{E=E_k+\delta_E/2} \tag{2.4}$$

and to reconstruct from these an approximation for the density of states $\rho(E)$. By introducing the *intrinsic* thermodynamic quantities, $T_k$ (temperature) and $c_k$ (specific heat) by

$$\frac{d \ln \rho}{dE}\bigg|_{E=E_k+\delta_E/2} = \frac{1}{T_k} = a_k,$$
$$\frac{d^2 \ln \rho}{dE^2}\bigg|_{E=E_k+\delta_E/2} = -\frac{1}{T_k^2 \, c_k} \frac{1}{V}. \tag{2.5}$$

we expose the important feature that the target coefficients $a_k$ are independent of the volume while the correction $R_k(E)$ is of order $\delta_E^2/V$. In all practical applications, $R_k$ will be numerically much smaller than $a_k \delta_E$. For a certain parameter range (i.e., for the correlation length smaller than the lattice size), we can analytically derive this particular volume dependence of the density derivatives. Details are left to the appendix.

Using the trapezium rule for integration, we find in particular

$$\ln \frac{\rho(E_{k+1} + \delta_E/2)}{\rho(E_k + \delta_E/2)} = \int_{E_k + \frac{\delta_E}{2}}^{E_{k+1} + \frac{\delta_E}{2}} \frac{d \ln \rho}{dE} \, dE$$
$$= \frac{\delta_E}{2} [a_k + a_{k+1}] + \mathcal{O}(\delta_E^3). \tag{2.6}$$

Using this equation recursively, we find

$$\ln \frac{\rho(E_N + \frac{\delta_E}{2})}{\rho(E_0 + \frac{\delta_E}{2})} = \frac{a_0}{2} \delta_E + \sum_{k=1}^{N-1} a_k \delta_E$$
$$+ \frac{a_N}{2} \delta_E + \mathcal{O}(\delta_E^2). \tag{2.7}$$

Note that $N \delta_E = \mathcal{O}(1)$. Exponentiating (2.3) and using (2.7), we obtain

$$\rho(E) = \rho\left(E_N + \frac{\delta_E}{2}\right) \exp\Big\{a_N (E - E_N - \delta_E/2) + \mathcal{O}(\delta_E^2)\Big\} \tag{2.8}$$
$$= \rho_0 \left(\prod_{k=1}^{N-1} e^{a_k \delta_E}\right) \exp\Big\{a_N (E - E_N) + \mathcal{O}(\delta_E^2)\Big\}, \tag{2.9}$$

where we have defined an overall multiplicative constant by

$$\rho_0 = \rho\left(E_0 + \frac{\delta_E}{2}\right) e^{a_0 \delta_E/2}.$$

We are now in the position to introduce the piecewise-linear and continuous approximation of the density of states by

$$\tilde{\rho}(E) = \rho_0 \left( \prod_{k=1}^{N-1} e^{a_k \delta_E} \right) e^{a_N (E-E_N)}, \tag{2.10}$$

with $N$ chosen in such a way that $E_N \leq E < E_N + \delta_E$ for a given $E$. With this definition, we obtain the remarkable identity

$$\rho(E) = \tilde{\rho}(E) \exp\left\{ \mathcal{O}(\delta_E^2) \right\} = \tilde{\rho}(E) \left[ 1 + \mathcal{O}(\delta_E^2) \right], \tag{2.11}$$

which we will extensively use below. We will observe that $\rho(E)$ spans many orders of magnitude. The key observation is that our approximation implements *exponential error suppression*, meaning that $\rho(E)$ can be approximated with nearly constant *relative error* despite that it may reach over thousands of orders of magnitude:

$$1 - \frac{\tilde{\rho}(E)}{\rho(E)} = \mathcal{O}\left(\delta_E^2\right). \tag{2.12}$$

We will now present our method to calculate the coefficients $a_k$. To this aim, we introduce the action restricted and re-weighted expectation values [9] with $a$ being an external variable:

$$\langle\langle W[\phi] \rangle\rangle_k (a) = \frac{1}{\mathcal{N}_k} \int \mathcal{D}\phi \, \theta_{[E_k,\delta_E]}(S[\phi]) \, W[\phi] \, e^{-aS[\phi]}, \tag{2.13}$$

$$\mathcal{N}_k = \int \mathcal{D}\phi \, \theta_{[E_k,\delta_E]} \, e^{-aS[\phi]} = \int_{E_k}^{E_k+\delta_E} dE \, \rho(E) e^{-aE}, \tag{2.14}$$

where we have used (2.1) to express $\mathcal{N}_k$ as an ordinary integral. We also introduced the modified Heaviside function,

$$\theta_{[E_k,\delta_E]}(S) = \begin{cases} 1 & \text{for} \quad E_k \leq S \leq E_k + \delta_E \\ 0 & \text{otherwise} . \end{cases}$$

If the observable only depends on the action, i.e., $W[\phi] = O(S[\phi])$, (2.13) simplifies to

$$\langle\langle O \rangle\rangle_k (a) = \frac{1}{\mathcal{N}_k} \int_{E_k}^{E_k+\delta_E} dE \, \rho(E) \, O(E) \, e^{-aE}, \tag{2.15}$$

Let us now consider the specific action observable

$$\Delta E = S - E_k - \frac{\delta_E}{2}, \tag{2.16}$$

and the solution $a^*$ of the non-linear equation

$$\langle\langle \Delta E \rangle\rangle_k (a = a^*) = 0. \tag{2.17}$$

Inserting $\rho(E)$ from (2.8) into (2.15) and defining $\Delta a = a_k - a$, we obtain

$$\langle\langle \Delta E \rangle\rangle_k (a)$$
$$= \frac{\rho(E_k + \delta_E/2) \int_{E_k}^{E_k+\delta_E} dE \, (E - E_k - \delta_E/2) \, e^{\Delta a (E-E_k)} \, e^{\mathcal{O}(\delta_E^2)}}{\rho(E_k + \delta_E/2) \int_{E_k}^{E_k+\delta_E} dE \, e^{\Delta a (E-E_k)} \, e^{\mathcal{O}(\delta_E^2)}}$$
$$= \frac{\int_{E_k}^{E_k+\delta_E} dE \, (E - E_k - \delta_E/2) \, e^{\Delta a (E-E_k)}}{\int_{E_k}^{E_k+\delta_E} dE \, e^{\Delta a (E-E_k)}} \tag{2.18}$$

Let us consider for the moment the function

$$F(\Delta a) := \frac{\int_{E_k}^{E_k+\delta_E} dE \, (E - E_k - \delta_E/2) \, e^{\Delta a (E-E_k)}}{\int_{E_k}^{E_k+\delta_E} dE \, e^{\Delta a (E-E_k)}}.$$

It is easy to check that $F$ is monotonic and vanishing for $\Delta a = 0$:

$$F'(\Delta a) > 0, \quad F(\Delta a = 0) = 0.$$

Since (2.18) approximates $\langle\langle \Delta E \rangle\rangle_k (a)$ up to $\mathcal{O}(\delta_E^2)$, we conclude that

$$a^* = a_k + \mathcal{O}\left(\delta_E^2\right) = \left. \frac{d \ln \rho}{dE} \right|_{E=E_k+\frac{\delta_E}{2}} + \mathcal{O}\left(\delta_E^2\right). \tag{2.19}$$

The latter equation is at the heart of the LLR algorithm: it details how we can obtain the log-rho derivative by calculating the Monte-Carlo average $\langle\langle \Delta E \rangle\rangle_k (a)$ (using (2.13)) and solving a non-linear equation, i.e. (2.17) with the identification $a^* \equiv a_k$ (justified by the order of our approximation). In the following, we will discuss the practical implementation by addressing two questions: (i) How do we solve the non-linear equation? (ii) How do we deal with the statistical uncertainty since the Monte-Carlo method only provides stochastic estimates for the expectation value $\langle\langle \Delta E \rangle\rangle_k (a)$? Let us start with the standard Newton–Raphson method to answer question (i). Starting from an initial guess $a_k^{(0)}$ for the solution, this method produces a sequence

$$a_k^{(0)} \to a_k^{(1)} \to a_k^{(2)} \to \cdots \to a_k^{(n)} \to a_k^{(n+1)} \cdots,$$

which converges to the true solution $a_k$. Starting from $a_k^{(n)}$ for the solution, we would like to derive an equation that generates a value $a_k^{(n+1)}$ that is even closer to the true solution:

$$\langle\langle \Delta E \rangle\rangle_k \left( a_k^{(n+1)} \right) = \langle\langle \Delta E \rangle\rangle_k \left( a_k^{(n)} \right)$$
$$+ \frac{d}{da} \langle\langle \Delta E \rangle\rangle_k \left( a_k^{(n)} \right) \left( a_k^{(n+1)} - a_k^{(n)} \right) = 0. \tag{2.20}$$

Using the definition of $\langle\langle \Delta E \rangle\rangle_k \left( a^{(n+1)} \right)$ in (2.18) with reference to (2.16) and (2.15), we find

$$\frac{d}{da} \langle\langle \Delta E \rangle\rangle_k (a) = - \left[ \left\langle\!\left\langle \Delta E^2 \right\rangle\!\right\rangle_k (a) - \langle\langle \Delta E \rangle\rangle_k^2 (a) \right]$$
$$=: -\sigma^2(\Delta E; a). \tag{2.21}$$

We thus find for the improved solution:

$$a_k^{(n+1)} = a_k^{(n)} + \frac{\langle\langle \Delta E\rangle\rangle_k (a_k^{(n)})}{\sigma^2(\Delta E; a_k^{(n)})}. \tag{2.22}$$

We can convert the Newton–Raphson recursion into a simpler fixed point iteration if we assume that the choice $a_k^{(n)}$ is sufficiently close to the true value $a_k$ such that

$$\delta_E \left(a_k^{(n)} - a_k\right) \ll 1.$$

Without affecting the precision with which the solution $a_k$ of (2.18) can be obtained, we replace $\sigma^2$ with its first order Taylor expansion around $a = a_k$

$$\sigma^2(\Delta E; a) = \frac{1}{12}\delta_E^2 \left[1 + \mathcal{O}\left(\delta_E \Delta a\right)^2\right] \left[1 + \mathcal{O}(\delta_E)\right]. \tag{2.23}$$

Hence, the Newton–Raphson iteration is given by

$$a_k^{(n+1)} = a_k^{(n)} + \frac{12}{\delta_E^2} \langle\langle \Delta E\rangle\rangle_k (a_k^{(n)}) \tag{2.24}$$

We point out that one fixed point of the above iteration, i.e., $a_k^{(n+1)} = a_k^{(n)} = a_k$, is attained for

$$\langle\langle \Delta E\rangle\rangle_k (a_k) = 0,$$

which, indeed, is the correct solution. We have already shown that the above equation has only one solution. Hence, if the iteration converges at all, it necessarily converges to the true solution. Note that convergence can always be achieved by suitable choice of under-relaxation. We here point out that the solution to question (ii) above will involve a particular type of under-relaxation.

Let us address question (ii) now. We have already pointed out that we have only a stochastic estimate for the expectation value $\langle\langle \Delta E\rangle\rangle_k (a)$ and the convergence of the Newton–Raphson method is necessarily hampered by the inevitable statistical error of the estimator. This problem, however, has been already solved by Robbins and Monroe [15].

For completeness, we shall now give a brief presentation of the algorithm. The starting point is the function $M(x)$, and a constant $\alpha$, such that the equation $M(x) = \alpha$ has a unique root at $x = \theta$. $M(x)$ is only available by stochastic estimation using the random variable $N(x)$:

$$\mathbb{E}[N(x)] = M(x),$$

with $\mathbb{E}[N(x)]$ being the ensemble average of $N(x)$. The iterative root finding problem is of the type

$$x_{n+1} = x_n + c_n (\alpha - N(x_n)) \tag{2.25}$$

where $c_n$ is a sequence of positive numbers sizes satisfying the requirements

$$\sum_{n=0}^{\infty} c_n = \infty \quad \text{and} \quad \sum_{n=0}^{\infty} c_n^2 < \infty \tag{2.26}$$

It is possible to prove that under certain assumptions [15] on the function $M(x)$ the $\lim_{n\to\infty} x_n$ converges in $L^2$ and hence in probability to the true value $\theta$. A major advance in understanding the asymptotic properties of this algorithm was the main result of [15]. If we restrict ourselves to the case

$$c_n = \frac{c}{n+1} \tag{2.27}$$

one can prove that $\sqrt{n}(x_n - \theta)$ is asymptotically normal with variance

$$\sigma_x^2 = \frac{c^2 \sigma_\xi^2}{2\,c\,M'(x) - 1} \tag{2.28}$$

where $\sigma_\xi^2$ is the variance of the noise. Hence, the optimal value of the constant $c$, which minimises the variance is given by

$$c = \frac{1}{M'(\theta)}. \tag{2.29}$$

Adapting the Robbins–Monro approach to our root finding iteration in (2.24), we finally obtain an under-relaxed Newton–Raphson iteration

$$a_k^{(n+1)} = a_k^{(n)} + \frac{12}{\delta_E^2\,(n+1)} \langle\langle \Delta E\rangle\rangle_k (a_k^{(n)}), \tag{2.30}$$

which is optimal with respect to the statistical noise during iteration.

### 2.3 Observables and convergence with $\delta_E$

We have already pointed out that expectation values of observables depending on the action only can be obtained by a simple integral over the density of states (see (2.2)). Here we develop a prescription for determining the values of expectations of more general observables by folding with the numerical density of states and analyse the dependence of the estimate on $\delta_E$.

Let us denote a generic observable by $B(\phi)$. Its expectation value is defined by

$$\langle B[\phi]\rangle = \frac{1}{Z(\beta)} \int \mathcal{D}\phi\, B[\phi]\, e^{\beta S[\phi]} \tag{2.31}$$

In order to relate to the LLR approach, we break up the latter integration into energy intervals:

$$\langle B[\phi] \rangle = \frac{1}{Z(\beta)} \sum_i \int \mathcal{D}\phi \; \theta_{[E_i, \delta_E]} \; B[\phi] \; \mathrm{e}^{\beta S[\phi]}. \quad (2.32)$$

Note that $\langle B[\phi] \rangle$ *does not* depend on $\delta_E$.

We can express $\langle B[\phi] \rangle$ in terms of a sum over double-bracket expectation values by choosing

$$W := B[\phi] \; \exp\{(\beta + a_i)S[\phi]\}$$

in (2.13). Without any approximation, we find

$$\langle B[\phi] \rangle = \frac{1}{Z(\beta)} \sum_i \mathcal{N}_i \; \mathrm{e}^{a_i E_i}$$
$$\times \langle\langle B[\phi] \exp\{\beta S[\phi] + a_i(S[\phi] - E_i)\}\rangle\rangle (E_i), \quad (2.33)$$
$$Z(\beta) = \sum_i \mathcal{N}_i \; \mathrm{e}^{a_i E_i}$$
$$\times \langle\langle \exp\{\beta S[\phi] + a_i(S[\phi] - E_i)\}\rangle\rangle (E_i). \quad (2.34)$$

where $\mathcal{N}_i = \mathcal{N}_i(a_i)$ is defined in (2.14). The above result can be further simplified by using (2.10) and (2.11):

$$\mathcal{N}_i \, \mathrm{e}^{a_i E_i} = \int_{E_i}^{E_i + \delta_E} \mathrm{d}E \; \rho(E) \; \exp\{-a_i(E - E_i)\}$$
$$= \mathrm{e}^{\mathcal{O}(\delta_E^2)} \int_{E_i}^{E_i + \delta_E} \mathrm{d}E \; \tilde{\rho}(E) \; \exp\{-a_i(E - E_i)\}$$
$$= \mathrm{e}^{\mathcal{O}(\delta_E^2)} \tilde{\rho}(E_i) \int_{E_i}^{E_i + \delta_E} \mathrm{d}E = \delta_E \, \tilde{\rho}(E_i) \, \mathrm{e}^{\mathcal{O}(\delta_E^2)}$$
$$= \delta_E \, \tilde{\rho}(E_i) \left[ 1 + \mathcal{O}(\delta_E^2) \right]. \quad (2.35)$$

We now define the approximation to $\langle B[\phi] \rangle$ by

$$\langle B[\phi] \rangle_{\mathrm{app}} = \frac{1}{Z(\beta)} \sum_i \delta_E \, \tilde{\rho}(E_i)$$
$$\times \langle\langle B[\phi] \exp\{\beta S[\phi] + a_i(S[\phi] - E_i)\}\rangle\rangle \quad (2.36)$$
$$Z(\beta) := \sum_i \delta_E \, \tilde{\rho}(E_i)$$
$$\times \langle\langle \exp\{\beta S[\phi] + a_i(S[\phi] - E_i)\}\rangle\rangle. \quad (2.37)$$

Since the double-bracket expectation values do not produce a singularity if $\delta_E \to 0$, i.e.,

$$\lim_{\delta_E \to 0} \langle\langle B[\phi] \exp\{\beta S[\phi] + a_i(S[\phi] - E_i)\}\rangle\rangle = \text{finite},$$

using (2.35), from (2.33) and (2.34) we find that

$$\langle B[\phi] \rangle = \langle B[\phi] \rangle_{\mathrm{app}} + \sum_i \mathcal{O}(\delta_E^3)$$
$$= \langle B[\phi] \rangle_{\mathrm{app}} + \mathcal{O}(\delta_E^2). \quad (2.38)$$

The latter formula together with (2.36) provides access to all types of observables using the LLR method with little more computational resources: Once the Robbins–Monro iteration (2.30) has settled for an estimate of the coefficient $a_k$, the Monte-Carlo simulation simply continues to derive estimators for the double-bracket expectation values in (2.36) and (2.37).

With the further assumption that the double-bracket expectation values are (semi-)positive, an even better error estimate is produced by our approach:

$$\langle B[\phi] \rangle = \langle B[\phi] \rangle_{\mathrm{app}} + \sum_i \mathcal{O}(\delta_E^3)$$
$$= \langle B[\phi] \rangle_{\mathrm{app}} \left[ 1 + \mathcal{O}(\delta_E^2) \right].$$

This implies that the observable $\langle B[\phi] \rangle$ can be calculated with an *relative error* of order $\delta_E^2$. Indeed, we find from (2.33, 2.34, 2.35) that

$$\langle B[\phi] \rangle = \frac{1}{Z(\beta)} \sum_i \delta_E \, \tilde{\rho}(E_i) \, \langle B[\phi] \exp\{\beta S[\phi]$$
$$+ a_i(S[\phi] - E_i)\}\rangle \exp\left\{\mathcal{O}(\delta_E^2)\right\}, \quad (2.39)$$

$$Z(\beta) := \sum_i \delta_E \, \tilde{\rho}(E_i) \, \langle\langle \exp\{\beta S[\phi] + a_i(S[\phi] - E_i)\}\rangle\rangle. \quad (2.40)$$

Thereby, we have used

$$\left| \sum_i a_i \, \exp\left\{c_i \delta_E^2\right\} \right| \leq \sum_i |a_i| \, \left| \exp\{c_i \delta_E^2\} \right|$$
$$\leq \sum_i |a_i| \, \exp\{c_{\max} \delta_E^2\}$$
$$= \exp\{c_{\max} \delta_E^2\} \sum_i a_i$$
$$= \exp\left\{\mathcal{O}(\delta_E^2)\right\} \times \sum_i a_i.$$

The assumption of (semi-)positive double-expectation values is true for many action observables, and possibly also for Wilson loops, whose re-weighted and action restricted double-expectation values might turn out to be positive (as is the case for their standard expectation values). In this case, our method would provide an efficient determination of those quantities. This is important in particular for large Wilson loop expectation values, since they are notoriously difficult to measure with importance sampling methods (see e.g. [16]). We also note that, in order to have an accurate determination of a generic observable, any Monte-Carlo estimate of the double-expectation values must be obtained to good precision dictated by the size of $\delta_E$. A detailed numerical investigation of these and related issues is left to future work.

For the specific case that the observable $B[\phi]$ only depends on the action $S[\phi]$, we circumvent this problem and evaluate the double-expectation values directly. To this aim, we introduce for the general case $\langle\langle W[\phi]\rangle\rangle_k$ the generalised density $w_k(E)$ by

$$\rho(E)\, w_k(E) \;=\; \int \mathcal{D}\phi\; \theta_{[E_k,\delta_E]}(S[\phi])\; W[\phi]\; \delta\Big(E - S[\phi]\Big).$$
(2.41)

We then point out that if $W[\phi]$ is depending on the action only, i.e., $W[\phi] = f(S[\phi])$, we obtain

$$w_k(E) \;=\; f(E)\, \theta_{[E_k,\delta_E]}(E).$$

With the definition of the double-expectation value (2.13), we find

$$\langle\langle W[\phi]\rangle\rangle_k\,(a_k) \;=\; \frac{\int_{E_k}^{E_k+\delta_E} \mathrm{d}E\; \rho(E)\, \mathrm{e}^{-a_k E}\, w_k(E)}{\int_{E_k}^{E_k+\delta_E} \mathrm{d}E\; \rho(E)\, \mathrm{e}^{-a_k E}}.$$
(2.42)

Rather than calculating $\langle\langle W[\phi]\rangle\rangle_k$ by Monte-Carlo methods, we can *analytically* evaluate this quantity (up to order $\mathcal{O}(\delta_E^2)$). Using the observation that for any smooth ($C_2$) function $g$

$$\int_{E_k}^{E_k+\delta_E} \mathrm{d}E\; g(E) \;=\; \delta_E\, g\left(E_k + \frac{\delta_E}{2}\right) \;+\; \mathcal{O}\Big(\delta_E^3\Big),$$

and, using this equation for both numerator and denominator of (2.42), we conclude that

$$\langle\langle W[\phi]\rangle\rangle_k\,(a_k) \;=\; w_k\left(E_k + \frac{\delta_E}{2}\right) \;+\; \mathcal{O}\Big(\delta_E^2\Big).$$
(2.43)

Let us now specialise to the case that is relevant for (2.39) with $B$ depending on the action only:

$$W[\phi] = b\Big(S[\phi]\Big)\, \exp\{\beta S[\phi] + a_i(S[\phi] - E_i)\},$$
$$w_i(E) = b(E)\, \exp\{\beta E + a_i(E - E_i)\}.$$
(2.44)

This leaves us with

$$\langle\langle W[\phi]\rangle\rangle_i\,(a_i) = b\Big(E_i + \frac{\delta_E}{2}\Big)\, \mathrm{e}^{\beta(E_i + \frac{\delta_E}{2})}\, \mathrm{e}^{a_i \frac{\delta_E}{2}} + \mathcal{O}\Big(\delta_E^2\Big).$$
(2.45)

Inserting (2.43) together with (2.44) into (2.36), we find

$$\langle B[\phi]\rangle = \frac{1}{Z(\beta)} \sum_i \delta_E\, \tilde\rho\left(E_i + \frac{\delta_E}{2}\right)$$
$$\times b_i\Big(E_i + \frac{\delta_E}{2}\Big)\, \mathrm{e}^{\beta(E_i + \frac{\delta_E}{2})} \;+\; \mathcal{O}\Big(\delta_E^2\Big),$$
(2.46)

$$Z(\beta) = \sum_i \delta_E\, \tilde\rho\left(E_i + \frac{\delta_E}{2}\right)\, \mathrm{e}^{\beta(E_i + \frac{\delta_E}{2})}.$$
(2.47)

Below, we will numerically test the quality of expectation values obtained by the LLR approach using action observables only, i.e., $B[\phi] = O(S[\phi])$. We will find that we indeed achieve the predicted precision in $\delta_E^2$ for this type of observables (see below Fig. 6).

### 2.4 The numerical algorithm

So far, we have shown that a piecewise continuous approximation of the density of states that is linear in intervals of sufficiently small amplitude $\delta_E$ allows us to obtain a controlled estimate of averages of observables and that the angular coefficients $a_i$ of the linear approximations can be computed in each interval $i$ using the Robbins–Monro recursion (2.30). Imposing the continuity of $\log \rho(E)$, one can then determine the latter quantity up to an additive constant, which does not play any role in cases in which observables are standard ensemble averages.

The Robbins–Monro recursion can easily be implemented in a numerical algorithm. Ideally, the recurrence would be stopped when a tolerance $\epsilon$ for $a_i$ is reached, i.e. when

$$\left| a_i^{(n+1)} - a_i^{(n)} \right| \;=\; \frac{12\, \left| \Delta E_i(a_i^{(n)}) \right|}{(n+1)\, \delta_E^2} \;\le\; \epsilon\,,$$
(2.48)

with (for instance) $\epsilon$ set to the precision of the computation. When this condition is fulfilled, we can set $a_i = a_i^{(n+1)}$. However, one has to keep into account the fact that the computation of $\Delta E_i$ requires an averaging over Monte-Carlo configurations. This brings into play considerations about thermalisation (which has to be taken into account each time we send $a_i^{(n)} \to a_i^{(n+1)}$), the number of measurements used for determining $\Delta E_i$ at fixed $a_i^{(n)}$ and—last but not least—fluctuations of the $a_i^{(n)}$ themselves.

Following those considerations, an algorithm based on the Robbins–Monro recursion relation should depend on the following input (tuneable) parameters:

- $N_{\mathrm{TH}}$, the number of Monte-Carlo updates in the restricted energy interval before starting to measure expectation values;
- $N_{\mathrm{SW}}$, the number of iterations used for computing expectation values;
- $N_{\mathrm{RM}}$, the number of Robbins–Monro iterations for determining $a_i$;
- $N_B$, number of final values from the Robbins–Monro iteration subjected to a subsequent bootstrap analysis.

The version of the LLR method proposed and implemented in this paper is reported in an algorithmic fashion in the box Algorithm 1. This implementation differs from that provided in [9,10] by the replacement of the originally proposed root

finding procedure based on a deterministic Newton–Raphson like recursion with the Robbins–Monro recursion, which is better suited to the problem of finding zeroes of stochastic equations.

---

**Algorithm 1:** The LLR method as implemented in this work.

**Input**: $N_{SW}$, $N_{TH}$, $N_{RM}$, $N_A$
**Output**: $a_i$ $\forall i$

1  **for** $0 \leq i < (E_{max} - E_{min})/\delta_E$ **do**
2    | Initialise $E_i = E_{min} + i\delta_E$, $a_i^0 = \bar{a}_i$;
3    | **for** $0 \leq n < N_{RM}$ **do**
4       | **for** $k \leq N_{TH} + N_{SW}$ **do**
5          | Evolve the whole system with an importance sampling algorithm for one sweep according to the probability distribution

$$P(E) \propto e^{-a_i^n E}$$

accepting only configuration such that
$E_i \leq E \leq E_i + \delta_E$
6          | **if** $k \geq N_{TH}$ **then**
7             | Compute $E^{(k)}$, the value of the energy in the current configuration $k$;
8       | Compute

$$\langle\langle \Delta E \rangle\rangle_i (a_i^{(n)}) = \frac{1}{N_{SW}} \left( \sum_{j > N_{TH}} E^{(j)} \right) - E_i - \frac{\delta_E}{2}$$

Compute

$$a_i^{(n+1)} = a_i^{(n)} + \frac{12 \langle\langle \Delta E \rangle\rangle_i (a_i^{(n)})}{(n+1)\delta_E^2}$$

9    | Repeat $N_B$ times to produce $N_B$ candidates $a_i$ for a subsequent bootstrap analysis

---

Since the $a_i$ are determined stochastically, a different reiteration of the algorithm with different starting conditions and different random seeds would produce a different value for the same $a_i$. The stochastic nature of the process implies that the distribution of the $a_i$ found in different runs is Gaussian. The generated ensemble of the $a_i$ can then be used to determine the error of the estimate of observables using analysis techniques such as jackknife and bootstrap.

The parameters $E_{min}$ and $E_{max}$ depend on the system and on the phenomenon under investigation. In particular, standard thermodynamic considerations on the infinite volume limit imply that if one is interested in a specific range of temperatures and the studied observables can be written as statistical averages with Gaussian fluctuations, it is possible to restrict the range of energies between the energy that is typical of the smallest considered temperature and the energy that is typical of the highest considered temperature. Determining a reasonable value for the amplitude of the energy interval $\delta_E$ and the other tuneable parameters $N_{SW}$, $N_{TH}$, $N_{RM}$ and $N_A$

requires a modest amount of experimenting with trial values. In our applications we found that the results were very stable for wide ranges of values of those parameters. Likewise, $\bar{a}_i$, the initial value for the Robbins–Monro recursion in interval $i$, does not play a crucial role; when required and possible, an initial value close to the expected result can be inferred inverting $\langle E(\beta) \rangle$, which can be obtained with a quick study using conventional techniques.

The average $\langle\langle \ldots \rangle\rangle$ imposes an update that restricts configurations to those with energies in a specific range. In most of our studies, we have imposed the constraint analytically at the level of the generation of the newly proposed variables, which results in a performance that is comparable with that of the unconstrained system. Using a simple-minded more direct approach, in which one imposes the constraint after the generation of the proposed new variable, we found that in most cases the efficiency of Monte-Carlo algorithms did not drop drastically as a consequence of the restriction, and even for systems like SU(3) (see Ref. [9]) we were able to keep an efficiency of at least 30 % and in most cases no less than 50 % with respect to the unconstrained system.

### 2.5 Ergodicity

Our implementation of the energy restricted average $\langle\langle \cdots \rangle\rangle$ assumes that the update algorithm is able to generate all configurations with energy in the relevant interval starting from configurations that have energy in the same interval. This assumption might be too strong when the update is local[2] in the energy (i.e. each elementary update step changes the energy by a quantity of order one for a system with total energy of order $V$) and there are topological excitations that can create regions with the same energy that are separated by high energy barriers. In these cases, which are rather common in gauge theories and statistical mechanics[3], generally in order to go from one acceptable region to the other one has to travel through a region of energies that is forbidden by an energy-restricted update method such as the LLR. Hence, by construction, in such a scenario our algorithm will get trapped in one of the allowed regions. Therefore, the update will not be ergodic.

In order to solve this problem, one can use an adaptation of the replica exchange method [17], as first proposed in [18]. The idea is that instead of dividing the whole energy interval in contiguous sub-intervals overlapping only in one point (in the following simply referred to as contiguous intervals), one can divide it in sub-intervals overlapping in a finite energy

---

[2] This is for instance the case for the popular heat-bath and Metropolis update schemes.

[3] For instance, in a $d$-dimensional Ising system of size $L^d$, to go from one ground state to the other one needs to create a kink, which has energy growing as $L^{d-1}$.
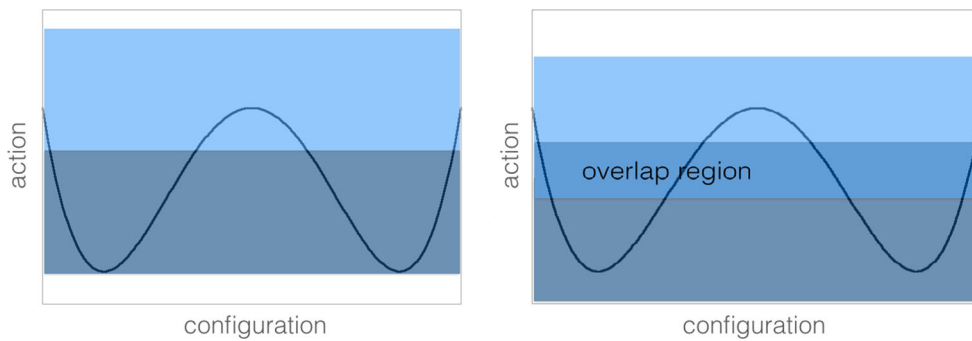
**Fig. 1** *Left* For contiguous energy intervals if a transition between configurations with energy in the same interval requires going through configurations with energy that are outside that interval, the simulation might get trapped in one of the allowed regions. *Right* For overlap-ping energy intervals with replica exchange, the simulation can travel from one allowed region to the other through excursions to the upper interval

region (this case will be referred to as overlapping intervals). With the latter prescription, after a fixed number of iterations of the Robbins–Monro procedure, we can check whether in any pairs of overlapping intervals $(I_1, I_2)$ the energy of both corresponding configurations is in the common region. For pairs fulfilling this condition, we can propose an exchange of the configurations with a Metropolis probability

$$P_{\text{swap}} = \min\left(1, e^{\left(a_{I_1}^{(n)} - a_{I_2}^{(n)}\right)\left(E_{C_1} - E_{C_2}\right)}\right), \qquad (2.49)$$

where $a_{I_1}^{(n)}$ and $a_{I_2}^{(n)}$ are the values of the parameter $a$ at the current $n$-th iterations of the Robbins–Monro procedure, respectively, in intervals $I_1$ and $I_2$ and $E_{C_1}$ ($E_{C_2}$) is the value of the energy of the current configuration $C_1$ ($C_2$) of the replica in the interval $I_1$ ($I_2$). If the proposed exchange is accepted, $C_1 \rightarrow C_2$ and $C_2 \rightarrow C_1$. With repeated exchanges of configurations from neighbour intervals, the system can now travel through all configuration space. A schematic illustration of how this mechanism works is provided in Fig. 1.

As already noticed in [18], the replica exchange step is amenable to parallelisation and hence can be conveniently deployed in calculations on massively parallel computers. Note that the replica exchange step adds another tuneable parameter to the algorithm, which is the number $N_{\text{SWAP}}$ of configurations swaps during the Monte-Carlo simulation at a given Monte-Carlo step. A modification of the LLR algorithm that incorporates this step can easily be implemented.

### 2.6 Reweighting with the numerical density of states

In order to screen our approach outlined in Sects. 2.2 and 2.3 for ergodicity violations and to propose an efficient procedure to calculate any observable once an estimate for the density of states has been obtained, as an alternative to the replica exchange method discussed in the previous section,

we here introduce an importance sampling algorithm with re-weighting with respect to the estimate $\tilde{\rho}$. This algorithm features short correlation times even near critical points. Consider for instance a system described by the canonical ensemble. We define a modified Boltzmann weight $W_B(E)$ as follows:

$$W_B(E) = \begin{cases} e^{\beta_1 E + c_1} & \text{for } E < E_{\min} ; \\ 1/\tilde{\rho}(E) & \text{for } E_{\min} \le E \le E_{\max} ; \\ e^{\beta_2 E + c_2} & \text{for } E > E_{\max} . \end{cases} \qquad (2.50)$$

Here $E_{\min}$ and $E_{\max}$ are two values of the energy that are far from the typical energy of interest $E$:

$$E_{\min} \ll E \ll E_{\max} . \qquad (2.51)$$

If conventional Monte-Carlo simulations can be used for numerical studies of the given system, we can choose $\beta_1$ and $\beta_2$ from the conditions

$$\langle E(\beta_i) \rangle = E_i , \qquad i = 1, 2 . \qquad (2.52)$$

If importance sampling methods are inefficient or unreliable, $\beta_1$ and $\beta_2$ can be chosen to be the micro-canonical $\beta_\mu$ corresponding respectively to the density of states centred in $E_{\min}$ and $E_{\max}$. These $\beta_\mu$ are outputs of our numerical determination $\tilde{\rho}(E)$. The two constants $c_1$ and $c_2$ are determined by requiring continuity of $W_B(E)$ at $E_{\min}$ and at $E_{\max}$:

$$\lim_{E \to E_{\min}^-} W_B(E) = \lim_{E \to E_{\min}^+} W_B(E) \quad \text{and} \quad \lim_{E \to E_{\max}^-} W_B(E)$$
$$= \lim_{E \to E_{\max}^+} W_B(E) . \qquad (2.53)$$

Let $\rho(E)$ be the correct density of state of the system. If $\tilde{\rho}(E) = \rho(E)$, then for $E_{\min} \le E \le E_{\max}$

$$\rho(E) W_B(E) = 1 , \qquad (2.54)$$

and a Monte-Carlo update with weights $W_B(E)$ drives the system in configuration space following a random walk in the energy. In practice, since $\tilde{\rho}(E)$ is determined numerically, upon normalisation

$$\rho(E)W_B(E) \simeq 1 , \tag{2.55}$$

and the random walk is only approximate. However, if $\tilde{\rho}(E)$ is a good approximation of $\rho(E)$, possible free energy barriers and metastabilities of the canonical system can be successfully overcome with the weights (2.50). Values of observables for the canonical ensemble at temperature $T = 1/\beta$ can be obtained using re-weighting:

$$\langle O(\beta) \rangle = \frac{\langle Oe^{\beta E}(W_B(E))^{-1} \rangle_W}{\langle e^{\beta E}(W_B(E))^{-1} \rangle_W} , \tag{2.56}$$

where $\langle \ \rangle$ denotes average over the canonical ensemble and $\langle \ \rangle_W$ average over the modified ensemble defined in (2.50). The weights $W_B(E)$ guarantee ergodic sampling with small auto-correlation time for the configurations with energies $E$ such that $E_{\min} \leq E \leq E_{\max}$, while suppressing to energy $E \ll E_{\min}$ and $E \gg E_{\max}$. Hence, as long as for a given $\beta$ of the canonical system $\overline{E} = \langle E \rangle$ and the energy fluctuation $\Delta\overline{E} = \sqrt{\langle E^2 \rangle - \langle E \rangle^2}$ are such that

$$E_{\min} \ll \overline{E} - \Delta\overline{E} \quad \text{and} \quad \overline{E} + \Delta\overline{E} \ll E_{\max} , \tag{2.57}$$

the re-weighting (2.56) does not present any overlap problem. The role of $E_{\min}$ and $E_{\max}$ is to restrict the approximate random walk only to energies that are physically interesting, in order to save computer time. Hence, the choice of $E_{\min}$, $E_{\max}$ and of the corresponding $\beta_1$, $\beta_2$ do not need to be fine-tuned, the only requirement being that Eq. (2.57) hold. These conditions can be verified a posteriori. Obviously, choosing the smallest interval $E_{\max} - E_{\min}$ where the conditions (2.57) hold optimises the computational time required by the algorithm. The weights (2.56) can easily be imposed using a metropolis or a biased metropolis [19]. Again, due to the absence of free energy barriers, no ergodicity problems are expected to arise. This can be checked by verifying that in the simulation there are various *tunnellings* (i.e. round trips) between $E_{\min}$ and $E_{\max}$ and that the frequency histogram of the energy is approximately flat between $E_{\min}$ and $E_{\max}$. Reasonable requirements are to have $\mathcal{O}(100 - 1000)$ tunnellings and an histogram that is flat within 15–20 %. These criteria can be used to confirm that the numerically determined $\tilde{\rho}(E)$ is a good approximation of $\rho(E)$. The flatness of the histogram is not influenced by the $\beta$ of interest in the original multi-canonical simulation. This is particularly important for first order phase transitions, where traditional Monte-Carlo algorithms have a tunnelling time that is exponentially suppressed with the volume of the system. Since the modified

ensemble relies on a random walk in energy, the tunnelling time between two fixed energy densities is expected to grow only as the square root of the volume.

This procedure of using a modified ensemble followed by re-weighting is inspired by the multi-canonical method [20], the only substantial difference being the recursion relation for determining the weights. Indeed for U(1) lattice gauge theory a multi-canonical update for which the weights are determined starting from a Wang–Landau recursion is discussed in [7]. We also note that the procedure used here to restrict ergodically the energy interval between $E_{\min}$ and $E_{\max}$ can easily be implemented also in the replica exchange method analysed in the previous subsection.

# 3 Application to compact U(1) lattice gauge theory

## 3.1 The model

Compact U(1) lattice gauge theory is the simplest gauge theory based on a Lie group. Its action is given by

$$S = \beta \sum_{x,\mu<\nu} \cos(\theta_{\mu\nu}(x)) , \tag{3.1}$$

where $\beta = 1/g^2$, with $g^2$ the gauge coupling, $x$ is a point of a d-dimensional lattice of size $L^d$ and $\mu$ and $\nu$ indicate two lattice directions, with index from 1 to $d$ (for simplicity, in this work we shall consider only the case $d = 4$), $\theta_{\mu\nu}$ plays the role of the electromagnetic field tensor: if we associate the compact angular variable $\theta_\mu(x) \in [-\pi; \pi[$ with the link stemming from $i$ in direction $\hat{\mu}$,

$$\theta_{\mu\nu}(x) = \theta_\mu(x) + \theta_\nu(x+\hat{\mu}) - \theta_\mu(x+\hat{\nu}) - \theta_\nu(x) . \tag{3.2}$$

The path integral of the theory is given by

$$Z = \int \mathcal{D}\theta_\mu \, e^S, \quad \mathcal{D}\theta_\mu = \prod_{x,\mu} \frac{d\theta_\mu(x)}{2\pi} , \tag{3.3}$$

the latter identity defining the Haar measure of the U(1) group.

The connection with the general framework of lattice gauge theories is better elucidated if we introduce the link variable

$$U_\mu(x) = e^{i\theta_\mu(x)} . \tag{3.4}$$

With this definition, $S$ can be rewritten as

$$S = \beta \sum_{x,\mu<\nu} \mathcal{R}e \, P_{\mu\nu}(x) , \tag{3.5}$$

with

$$P_{\mu\nu}(x) = U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^*(x+\hat{\nu})U_\nu^*(x)$$

the plaquette variable, and $U_\mu^*(x)$ is the complex conjugate of $U_\mu(x)$. Working with the variables $U_\mu(x)$ allows us to show immediately that $S$ is invariant under U(1) gauge transformations, which act as

$$U_\mu(x) \mapsto \Lambda^*(x)\, U_\mu(x)\, \Lambda(x+\hat{\mu})\,, \qquad \Lambda(x) = e^{i\lambda(x)}\,, \tag{3.6}$$

with $\lambda(x) \in [-\pi;\ \pi[$ a function defined on lattice points.

The connection with U(1) gauge theory in the continuum can be shown by introducing the lattice spacing $a$ and the non-compact gauge field $a\, A_\mu(x) = \theta_\mu(x)/g$, so that

$$U_\mu(x) = e^{i\, g a\, A_\mu(x)}\,. \tag{3.7}$$

Taking $a$ small and expanding the cosine leads us to

$$S = -\frac{1}{4} a^4 \sum_{x,\mu,\nu} \left( \Delta_\mu A_\nu(x) - \Delta_\nu A_\mu(x) \right)^2 + \mathcal{O}(a^6) + \text{constant}, \tag{3.8}$$

with $\Delta_\mu$ the forward difference operator. In the limit $a \to 0$, we finally find

$$S \simeq -\frac{1}{4} \int \mathrm{d}^4 x\, F_{\mu\nu}(x)^2, \tag{3.9}$$

with $F_{\mu\nu}$ being the usual field strength tensor. This shows that in the *classical* $a \to 0$ limit $S$ becomes the Euclidean action of a free gas of photons, with interactions being related to the neglected lattice corrections. It is worth to remark that this classical continuum limit is not the continuum limit of the full theory. In fact, this classical continuum limit is spoiled by quantum fluctuations. These prevent the system from developing a second order transition point in the $a \to 0$ limit, which is a necessary condition to be able to remove the ultraviolet cutoff introduced with the lattice discretisation. The lack of a continuum limit is related to the fact that the theory is strongly coupled in the ultraviolet. Despite the non-existence of a continuum limit for compact U(1) lattice gauge theory, this lattice model is still interesting, since it provides a simple realisation of a weakly first order phase transition. This bulk phase transition separates a confining phase at low $\beta$ (whose existence was pointed out by Wilson [21] in his seminal work on lattice gauge theory) from a deconfined phase at high $\beta$, with the transition itself occurring at a critical value of the coupling $\beta_c \simeq 1$. Rather unexpectedly at first side, importance sampling Monte-Carlo studies of this phase transitions turned out to be demanding and not immediate to interpret, with the order of the transition having been debated for a long time (see e.g. [22–31]). The issue was cleared only relatively recently, with investigations that made a crucial use of supercomputers [32,33]. What makes the transition difficult

to observe numerically is the role played in the deconfinement phase transition by magnetic monopoles [34], which condense in the confined phase [34,35].

The existence of topological sectors and the presence of a transition with exponentially suppressed tunnelling times can provide robust tests for the efficiency and the ergodicity of our algorithm. This motivates our choice of compact U(1) for the numerical investigation presented in this paper.

### 3.2 Simulation details

The study of the critical properties of U(1) lattice gauge theory is presented in this section. In order to test our algorithm, we investigated the behaviour of specific heat as a function of the volume. This quantity has been carefully investigated in previous studies, and as such provides a stringent test of our procedure. In order to compare data across different sizes, our results will be often provided normalised to the number of plaquette $6L^4 = 6V$.

We studied lattices of size ranging from $8^4$ to $20^4$ and for each lattice size we computed the density of states $\rho(E)$ in the interval $E_{\min} \le E \le E_{\max}$ (see Table 1). The rationale behind the choice of the energy region is that it must be centred around the critical energy and it has to be large enough to study all the critical properties of the theory, i.e. every observable evaluated has to have support in this region and have virtually no correction coming from the choice of the energy boundaries.

We divided the energy interval in steps of $\delta_E$ and for each of the sub-interval we have repeated the entire generation of the log-linear density of states function and evaluation of the observables $N_B = 20$ times to create the bootstrap samples for the estimate of the errors. The values of the other tuneable parameters of the algorithm used in our study are reported in Table 1. An example determination of one of the $a_i$ is reported in Fig. 2. The plot shows the rapid convergence to the asymptotic value and the negligible amplitude of residual fluctuations. Concerning the cost of the simulations, we found that accurate determinations of observables can be obtained with modest computational resources compared to those needed in investigations of the system with importance sampling methods. For instance, the most costly simulation presented here, the investigation of the $20^4$ lattice, was performed on 512 cores of Intel Westmere processors in about five days. This needs to be contrasted with the fact that in the early 2000s only lattices up to $18^4$ could be reliably investigated with importance sampling methods, with the largest sizes requiring supercomputers [32,33].

One of our first analyses was a screening for potential ergodicity violations with the LLR approach. As detailed in Sect. 2.5, these can emerge for LLR simulations using contiguous intervals as is the case for the U(1) study reported in this paper. To this aim, we calculated the action expecta-

**Table 1** Values of the tuneable parameters of the LLR algorithm used in our numerical investigation, in the last column we report the total number of global MC steps needed to perform the entire investigation

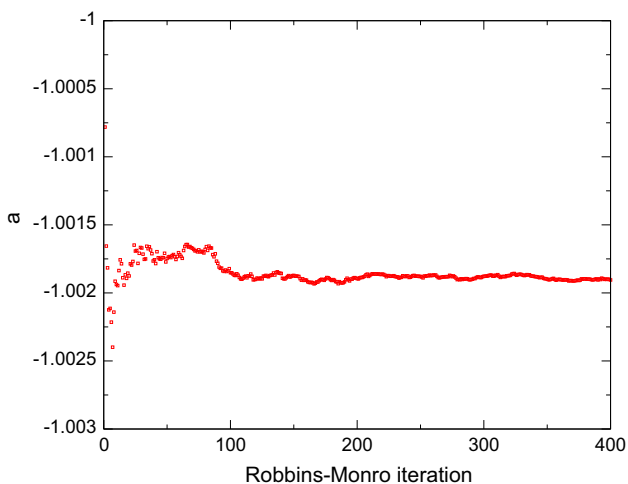| L | $E_{\min}/(6V)$ | $E_{\max}/(6V)$ | $N_{SW}$ | $N_{RM}$ | $(E_{\max} - E_{\min})/\delta_E$ | MC Steps |
|---|---|---|---|---|---|---|
| 8 | 0.5722222 | 0.67 | 250 | 600 | 512 | $7.7\,10^7$ |
| 10, 12, 14, 16, 18, 20 | 0.59 | 0.687777 | 200 | 400 | 512 | $4.1\,10^7$ |

tion value $\langle E \rangle$ for a $12^4$ lattice for several values using the LLR method and using the re-weighting with respect to the estimate $\tilde{\rho}$. Since the latter approach is conceptually free of ergodicity issues, any violations by the LLR method would be flagged by discrepancy. Our findings are summarised in Fig. 3 and the corresponding table. We find good agreement for the results from both methods. This suggests that topological objects do not generate energy barriers that trap our algorithm in a restricted section of configuration space. Said



**Fig. 2** Estimated $a_i$ as a function of the Robbins–Monro iteration, on a $20^4$ lattice and for action $E/(6V) = 0.59009548$ at the centre of the interval with $\delta_E/V = 1.91 \times 10^{-4}$

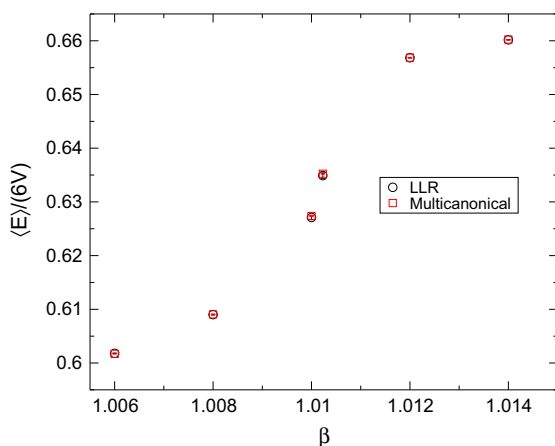in other words, for this system the LLR method using contiguous intervals seems to be ergodic.

### 3.3 Volume dependence of $\log \tilde{\rho}$ and computational cost of the algorithm

As a first investigation we have performed a study of the scaling properties of the $a_i$ as a function of the volume. In Fig. 4 we show the behaviour of the $a_i$ with the lattice volume. The estimates are done for a fixed $\delta_E/V$, where the chosen value for the ratio fulfils the request that within the errors all our observables are not varying for $\delta_E \to 0$ (we report on the study of $\delta_E \to 0$ in Sect. 3.5). As is clearly visible from the plot, the data are scaling towards an infinite volume estimate of the $a_i$ for fixed energy density.

As mentioned before, the issue facing importance sampling studies at first order phase transitions are connected with tunnelling times that grow exponentially with the volume. With the LLR method, the algorithmic cost is expected to grow with the size of the system as $V^2$, where one factor of $V$ comes from the increase of the size and the other factor of $V$ comes from the fact that one needs to keep the width of the energy interval per unit of volume $\delta_E/V$ fixed, as in the large-volume limit only intensive quantities are expected to determine the physics. One might wonder whether this apparently simplistic argument fails at the first order phase transition point. This might happen if the dynamics is such that a slowing down takes place at criticality. In the case of compact U(1), for the range of lattice sizes studied here, we



| | $\langle E \rangle/(6V)$ | |
|---|---|---|
| $\beta$ | (LLR) | (re-weighting) |
| 1.006 | 0.60180(6) | 0.601714(8) |
| 1.008 | 0.60901(8) | 0.60906(1) |
| 1.010 | 0.6271(4) | 0.6274(1) |
| 1.01023 | 0.6349(5) | 0.6352(1) |
| 1.012 | 0.65685(5) | 0.65687(1) |
| 1.014 | 0.66018(4) | 0.660230(5) |

**Fig. 3** Comparison between the plaquette computed with the LLR algorithm (see Sect. 2.2) and via re-weighting with respect to the estimate $\tilde{\rho}$ (see Sect. 2.6) for a $L = 12$ lattice.
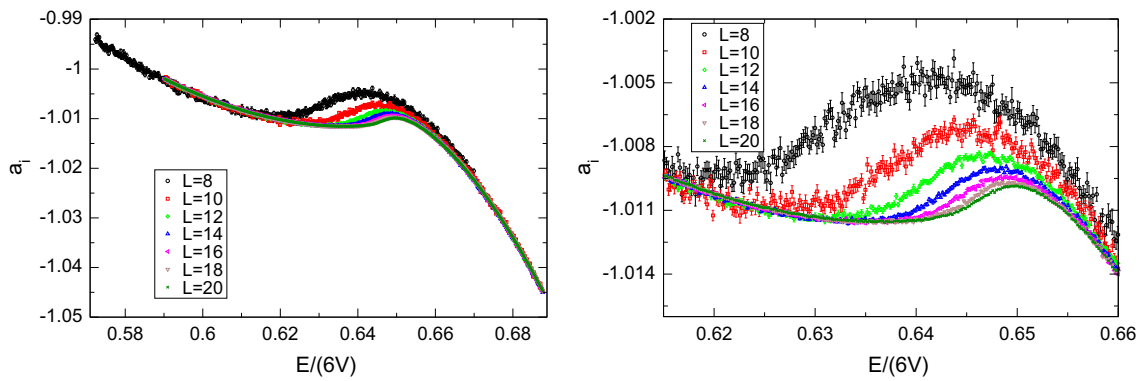
**Fig. 4** Estimate of $a_i$ as a function of the energy density for various volumes. The *right panel* is a zoom of the interesting region

**Table 2** $\beta_c(L)$ evaluated with the LLR algorithm and reference data from [32]

| L | $\beta_c(L)$ present method | $\beta_c(L)$ reference values |
|---|---|---|
| 8 | 1.00744(2) | 1.00741(1) |
| 10 | 1.00939(2) | 1.00938(2) |
| 12 | 1.010245(1) | 1.01023(1) |
| 14 | 1.010635(5) | 1.01063(1) |
| 16 | 1.010833(4) | 1.01084(1) |
| 18 | 1.010948(2) | 1.010943(8) |
| 20 | 1.011006(2) | |

**Table 3** Estimates of $\beta_c$ for various choices of the fit parameters. In bold the best fits

| $L_{\min}$ | $k_{\max}$ | $\beta_c$ | $\chi^2_{\mathrm{red}}$ |
|---|---|---|---|
| **14** | **1** | **1.011125(3)** | **0.91** |
| 12 | 1 | 1.011121(3) | 2.42 |
| **12** | **2** | **1.011129(4)** | **0.67** |
| 10 | 1 | 1.011116(5) | 7.44 |
| **10** | **2** | **1.011127(3)** | **0.60** |
| 8 | 1 | 1.011093(5) | 90.26 |
| **8** | **2** | **1.011126(2)** | **0.62** |

have found that the computational cost of the algorithm is compatible with a quadratic increase with the volume.

### 3.4 Numerical investigation of the phase transition

Using the density of states it is straightforward to evaluate, by direct integration (see Sect. 2.3), the expectation values of any power of the energy and evaluate thermodynamical quantities like the specific heat

$$C_V(\beta) = \langle E^2(\beta)\rangle - \langle E(\beta)\rangle^2. \tag{3.10}$$

As usual we define the pseudo-critical coupling $\beta_c(L)$ such as the coupling at which the peak of the specific heat occurs for a fixed volume. The peak of the specific heat has been located using our numerical procedure and the error bars are computed using the bootstrap method. Our results are summarised in Table 2 with a comparison with the values in [32]. Once again, the agreement testifies to the good ergodic properties of the algorithm.

Using our data it is possible to make a precise estimate of the infinite volume critical beta by means of a finite size scaling analysis. The finite size scaling of the pseudo-critical coupling is given by

$$\beta_c(L) = \beta_c + \sum_{k=1}^{k_{max}} B_k L^{-4k}, \tag{3.11}$$

where $\beta_c$ is the critical coupling. We fit our data with the function in Eq. (3.11); the results are reported in Table 3.

Another quantity easily accessible is the latent heat. This quantity can be related to the height of the peak of the specific heat at the critical temperature through:

$$\frac{C_L(\beta_c(L))}{6L^4} = \frac{G^2}{4} + \sum_{k=1}^{k_{max}} C_k L^{-4k}, \tag{3.12}$$

where $G$ is the latent heat. Our results for this observable are reported in Table 4. We fit the result with Eq. (3.12); see Table 5.

The latent heat can be obtained also from the knowledge of the location of the peaks of the probability density at $\beta_c$ (of infinite volume), indeed in this case the latent heat is equal to energy gap between the peaks. This direct measure can be used as crosscheck of the previous analysis. In the language of the density of states the probability density is simply given by

$$P_\beta(E) = \frac{1}{Z}\rho(E)e^{\beta E}. \tag{3.13}$$

**Table 4** $C_V(\beta_c(L))$ evaluated with the LLR algorithm and reference data from [32]. Results for a $20^4$ lattice have never been reported before in the literature

| L | $C_V/(6V)$ peak present work | $C_V/(6V)$ peak from [32] |
|---|---|---|
| 8 | 0.000551(2) | 0.000554(1) |
| 10 | 0.000384(2) | 0.000385(1) |
| 12 | 0.0002971(11) | 0.000298(1) |
| 14 | 0.0002537(8) | 0.000254(1) |
| 16 | 0.0002272(7) | 0.000226(2) |
| 18 | 0.0002097(5) | 0.000211(2) |
| 20 | 0.0002007(4) | |

**Table 5** Estimates of $G$ for various choices of the fit parameters. In bold the best fits

| $L_{min}$ | $k_{max}$ | $G$ | $\chi^2_{red}$ |
|---|---|---|---|
| 14 | 1 | 0.02712(9) | 4.6 |
| 12 | 1 | 0.0273(2) | 31 |
| **12** | **2** | **0.02688(7)** | **1.4** |
| 10 | 1 | 0.0276(2) | 74 |
| 10 | 2 | 0.02710(12) | 9.7 |
| **10** | **3** | **0.02681(9)** | **1.4** |
| 8 | 1 | 0.0281(4) | 335 |
| 8 | 2 | 0.02731(15) | 26 |
| 8 | 3 | 0.02703(11) | 6.7 |
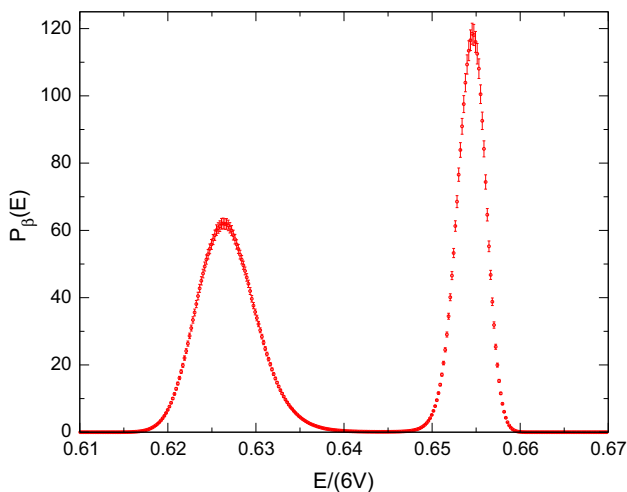


**Fig. 5** Probability density for $L = 20$ at $\beta_c$. The probability is plotted at $\beta_c$ of infinite volume hence the peaks are not of equal height

We have performed the study of the location in energy of the two peaks of $P_{\beta_c}(E)$ (an example is displayed in Fig. 5) and we have reported them in Table 6. Also in this case we have performed a finite size scaling analysis to extract the infinite volume behaviour:

$$E_i(L)/(6V) = \epsilon_i + a_i e^{-b_i L}. \quad (3.14)$$

**Table 6** Location of the peak of the probability density in the two meta-stable phases

| L | $E_1/(6V)$ | $E_2/(6V)$ |
|---|---|---|
| 12 | 0.6263(5) | 0.65580(14) |
| 14 | 0.6264(2) | 0.65532(5) |
| 16 | 0.6272(2) | 0.65512(4) |
| 18 | 0.6274(4) | 0.65495(6) |
| 20 | 0.6275(2) | 0.65491(7) |

**Table 7** Values of $\delta_E$ used to perform the study of the discretisation effects. The other simulation parameters are kept identical to the one reported in Table 1

| L | $(E_{max} - E_{min})/\delta_E$ |
|---|---|
| 10 | 8, 16, 32, 64, 128, 512 |
| 12 | 8, 16, 20, 32, 64, 128, 512 |
| 14 | 16, 32, 64, 512 |
| 16 | 16, 32, 64, 128, 512 |

A fit of the values in Table 6 yields $\chi^2_{red,1} = 0.67$, $\epsilon_1 = 0.6279(9)$ and $\chi^2_{red,2} = 0.2$, $\epsilon_2 = 0.65485(4)$. The latent heat can be evaluated as $G = \epsilon_2 - \epsilon_1 = 0.0270(9)$, which is in perfect agreement with the estimates obtained by studying the scaling of the specific heat.

### 3.5 Discretisation effects

In this section we want to address the dependence of our observables from the size of energy interval $\delta_E$. In order to quantify this effect we study the dependence of the peak of the specific heat $C_{v,peak}$ with $\delta_E$ for various lattice sizes, namely 8, 10, 12, 14, 16. In Table 7 we report the lattice sizes and the corresponding $\delta_E$ used to perform such investigation. For each pair of $\delta_E$ and volume reported we have repeated all our simulations and analysis with the same simulation parameters reported in Table 1.

The choice of the specific heat as an observable for such investigation can easily be justified: we found that specific heat is much more sensible to the discretisation effects with respect to other simpler observables such as the plaquette expectation value. In Fig. 6 we report an example of such study relative to $L = 8$.

We can confirm that all our data are scaling with quadratic law in $\delta_E$ consistent with our findings in Sect. 2.3. Indeed by fitting our data with a form

$$C_V(\beta_c(L), \delta_E) = C_V(\beta_c(L)) + 6V b_{dis} \delta_E^2, \quad (3.15)$$

we found $\chi^2_{red} \sim 1$ for all lattice sizes we investigated. We report in Table 8 the values of $b_{dis}$. Note that the numerical values used in our finite size scaling analysis of the peak of
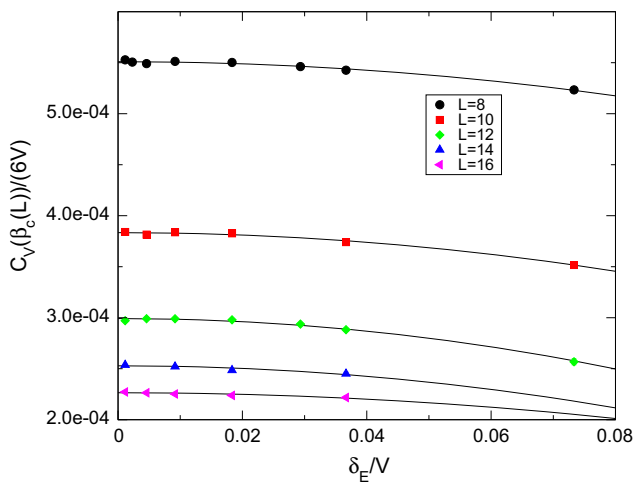
**Fig. 6** The peak of the $C_V(\beta_C(L))$ as a function $\delta_E$

**Table 8** The coefficient $b_{dis}$ for different lattice sizes

| $L$ | $b_{dis}$ |
| --- | --- |
| 8 | $-3.1(2)\ 10^{-10}$ |
| 10 | $-5.9(4)\ 10^{-11}$ |
| 12 | $-1.8(1)\ 10^{-11}$ |
| 14 | $-4(1)\ 10^{-12}$ |
| 16 | $-9(3)\ 10^{-13}$ |

$C_V$ presented in the previous section are compatible with the results extrapolated to $\delta_E = 0$ obtained here.

## 4 Discussion, conclusions and future plans

The density of states $\rho(E)$ is a measure of the number of configurations on the hyper-surface of a given action $E$. Knowing the density of states relays the calculation of the partition function to performing an ordinary integral. Wang–Landau type algorithms perform Markov chain Monte-Carlo updates with respect to $\rho$ while improving the estimate for $\rho$ during simulations. The LLR approach, first introduced in [9], uses a non-linear stochastic equation (see (2.17)) for this task and is particularly suited for systems with continuous degrees of freedom. To date, the LLR method has been applied to gauge theories in several publications, e.g. [10–12,14], and it has turned out in practice to be a reliable and robust method. In the present paper, we have thoroughly investigated the foundations of the method and have presented high-precision results for the U(1) gauge theory to illustrate the excellent performance of the approach.

Two key features of the LLR approach are:

(i) It solves an *overlap* problem in the sense that the method can specifically target the action range that is of particular

importance for an observable. This range might easily be outside the regime for which standard MC methods would be able to produce statistics.

(ii) It features *exponential error* suppression: although the density of states $\rho$ spans many orders of magnitude, $\tilde{\rho}$, the density of states defined from the linear approximation of its log, has a nearly constant *relative* error (see Sect. 2.2) and the numerical determination of $\tilde{\rho}$ preserves this level of accuracy.

We point out that feature (i) is not exclusive of the LLR method, but is quite generic for multi-canonical techniques [20], Wang–Landau type updates [4] or hybrids thereof [7].

Key ingredient for the LLR approach is the double-bracket expectation value [9] (see (2.13)). It appears as a standard Monte-Carlo expectation value over a finite action interval of size $\delta_E$ and with the density of states as a re-weighting factor. The derivative of the density of states $a(E)$ emerges from an iteration involving these Monte-Carlo expectation values. This implies that their statistical errors interfere with the convergence of the iteration. This might introduce a bias preventing the iteration to converge to the true derivative $a(E)$. We resolved this issue by using the Robbins–Monro formalism [15]: we showed that a particular type of under-relaxation produces a normal distribution of the determined values $a(E)$ with the mean of this distribution coinciding with the correct answer (see Sect. 2.2).

In this paper, we also addressed two concerns, which were raised in the wake of the publication of Ref. [9]:

(1) The LLR simulations restrict the Monte-Carlo updates to a finite action interval and might therefore be prone to *ergodicity* violations.

(2) The LLR approach seems to be limited to the calculation of *action dependent observables only*.

To address the first issue, we have proposed in Sects. 2.5 and 2.6 two procedures that are conceptually free of ergodicity violations. The first method is based upon the replica exchange method [17,18]: using overlapping action ranges during the calculation of the double-bracket expectation values offers the possibility to exchange the configurations of neighbouring action intervals with appropriate probability (see Sect. 2.5 for details). The second method is a standard Monte-Carlo simulation but with the inverse of the estimated density of states, i.e., $\tilde{\rho}^{-1}(E)$, as re-weighting factor. The latter approach falls into the class of ergodic Monte-Carlo update techniques and is not limited by a potential overlap problem: if the estimate $\tilde{\rho}$ is close to the true density $\rho$, the Monte-Carlo simulation is essentially a random walk in configuration space sweeping the action range of interest.

To address issue (2), we first point out that the latter re-weighting approach produces a sequence of configurations that can be used to calculate *any* observable by averaging with the correct weight. Second, we have developed in Sect. 2.2 the formalism to calculate any observable by a suitable sum over a combination of the density of states and double-bracket expectation values involving the observable of interest. We were able to show that the order of convergence (with the size $\delta_E$ of the action interval) for these observables is the same as for $\rho$ itself (i.e., $\mathcal{O}(\delta_E^2)$).

In view of the features of the density of states approach, our future plans naturally involve investigations that either are enhanced by the direct access to the partition function (such as the calculation of thermodynamical quantities) or that are otherwise hampered by an overlap problem. These, most notably, include complex action systems such as cold and dense quantum matter. The LLR method is very well equipped for this task since it is based upon Monte-Carlo updates with respect to the positive (and real) estimate of the density of states and features an exponential error suppression that might beat the resulting overlap problem. Indeed, a strong sign problem was solved by LLR techniques using the original degrees of freedom of the $Z_3$ spin model [10,11]. We are currently extending these investigations to other finite density gauge theories. QCD at finite densities for heavy quarks (HDQCD) is work in progress. We have plans to extend the studies to finite density QCD with moderate quark masses.

## Appendix A: Reference scale and volume scaling

Here, we will present further details on the scaling of the density of states $\rho(E)$ with the volume $V$ of our system. To this aim, we will work in the regime of a finite correlation length $\xi$ such that the volume $V \gg \xi^4$. In the case of particle physics, $\xi$ is a multiple of the inverse mass of the lightest excitation of the theory. In this subsection, we do not address the case of a correlation length comparable or larger than the size of the system, as it might occur near a second order phase transition.

Under these assumptions, the total action appears as a sum over uncorrelated contributions:

$$E = \sum_{i=1}^{v} e_i, \quad v = V/\xi^4, \tag{A.1}$$

where the dimensionless variable $v$ is the volume in units of the (physical) correlation length. To ease the notation, we will assume that the densities $\rho$ and $\tilde{\rho}$ are normalised to one. Taking advantage of the above observation, we can introduce the probability distribution $\tilde{\rho}(e_i)$ for the uncorrelated domains:

$$\rho(E) = \int \prod_{i=1}^{v} de_i \, \delta\left(E - \sum_{k=1}^{v} e_k\right) \tilde{\rho}(e_1)\dots\tilde{\rho}(e_v). \tag{A.2}$$

Representing the $\delta$-function as a Fourier integral, we find

$$\rho(E) = \int \frac{d\alpha}{2\pi} \int \prod_{i=1}^{v} de_i \, e^{-i\alpha E} e^{i\alpha e_1}\dots e^{i\alpha e_v} \tilde{\rho}(e_1)\dots\tilde{\rho}(e_v)$$
$$= \int \frac{d\alpha}{2\pi} e^{-i\alpha E} \left\langle e^{i\alpha e}\right\rangle^v. \tag{A.3}$$

The latter equation is the starting point for a study of moments and cumulants of the action expectation values and their scaling with the volume.

Cumulants of the action $E$ are defined by

$$(E^n)_c = (-i)^n \frac{d^n}{d\beta^n} \ln \int dE \, e^{i\beta E} \rho(E) \Big|_{\beta=0}. \tag{A.4}$$

Inserting (A.3) into (A.4), performing the $E$ and the $\alpha$ integration leaves us with

$$(E^n)_c = (-i)^n \frac{d^n}{d\beta^n} \ln \left\langle \exp\{i\beta e\}\right\rangle^v \Big|_{\beta=0} = v \, (e^n)_c, \tag{A.5}$$

where the volume independent cumulants are defined by

$$(e^n)_c = (-i)^n \frac{d^n}{d\beta^n} \ln \left\langle \exp\{i\beta e\}\right\rangle \Big|_{\beta=0}. \tag{A.6}$$

We here make the important observation that *all* cumulants are proportional to the "volume" $v$ rather than powers of it. Re-summing (A.6), i.e. using the identity

$$\sum_n \frac{i^n \alpha^n}{n!}(e^n)_c = \ln \left\langle \exp\{i\alpha e\}\right\rangle,$$

we find for $\rho(E)$ in (A.3)

$$\rho(E) = \int \frac{d\alpha}{2\pi} e^{-i\alpha E} \exp\left\{v \sum_n \frac{(i\alpha)^n}{n!}(e^n)_c\right\}. \tag{A.7}$$

We perform the $\alpha$-integral by using the expansion

$$\rho(E) = \exp\left\{v \sum_{n=3}^{\infty} \frac{(e^n)_c}{n!} \left(-\frac{d}{dE}\right)^n\right\} \rho_0(E), \qquad (A.8)$$

$$\rho_0(E) = \int \frac{d\alpha}{2\pi} e^{-i\alpha E} \exp\left\{i\, v(e)_c\, \alpha - v\frac{(e^2)_c}{2}\alpha^2\right\}$$

$$= \frac{1}{\sqrt{2\pi v (e^2)_c}} \exp\left\{-\frac{v}{2(e^2)_c}[E/v - (e)_c]^2\right\}. \tag{A.9}$$

In next-to-leading order, we obtain (up to an additive constant):

$$\frac{1}{v}\ln\rho(E) \approx -\frac{[(e)_c - E/v]^2}{2(e^2)_c}$$
$$-\frac{(e^3)_c}{6(e^2)_c^3}[(e)_c - E/v]^3 + \mathcal{O}(1/v). \tag{A.10}$$

Hence, we find for the inverse temperature $a_k$

$$a_k = \left.\frac{d\ln\rho}{dE}\right|_{E=E_k}$$
$$\approx \frac{(e)_c - E_k/v}{(e^2)_c} + \frac{(e^3)_c}{2(e^2)_c^3}\left[(e)_c - E_k/v\right]^2. \tag{A.11}$$

We therefore confirm that $a_k$ is an intrinsic quantity, i.e., volume independent. The curvature of $\ln\rho$ at $E = E_K$ is given by

$$\left.\frac{d^2\ln\rho}{dE^2}\right|_{E=E_k} \approx -\frac{1}{v}\left[\frac{1}{(e^2)_c} + \frac{(e^3)_c}{(e^2)_c^3}\left[(e)_c - E_k/v\right]\right]. \tag{A.12}$$

We therefore confirm the key thermodynamic assumptions in (2.5) by explicit calculation:

$$\left.\frac{d\ln\rho(E)}{dE}\right|_{E=E_k} = a_k = \mathcal{O}(1),$$
$$\left.\frac{d^2\ln\rho(E)}{dE^2}\right|_{E=E_k} = \mathcal{O}(1/v). \tag{A.13}$$

# References

1. D. Landau, K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2014)
2. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)
3. G. Aarts, Recent developments at finite density on the lattice. PoS CPOD2014, 012 (2014). arXiv:1502.1850
4. F. Wang, P. Landau, Efficient, multiple-range random walk algorithm to calculate the den sity of states. Phys. Rev. Lett. **86**, 2050–2053 (2001)
5. J. Xu, H.-R. Ma, Density of states of a two-dimensional $xy$ model from the Wang–Landau algorithm. Phys. Rev. E **75**, 041115 (2007)
6. S. Sinha, S.K. Roy, Performance of Wang–Landau algorithm in continuous spin models and a case study: modified XY-model. Phys. Lett. A **373**, 308–314 (2009)
7. B.A. Berg, A. Bazavov, Non-perturbative U(1) gauge theory at finite temperature. Phys. Rev. D **74**, 094502 (2006). arXiv:hep-lat/0605019
8. A. Bazavov, B.A. Berg, Program package for multicanonical simulations of U(1) lattice gauge theory. Second version. Comput. Phys. Commun. **184**, 1075–1076 (2013)
9. K. Langfeld, B. Lucini, A. Rago, The density of states in Gauge theories. Phys. Rev. Lett. **109**, 111601 (2012). arXiv:1204.3243
10. K. Langfeld, B. Lucini, Density of states approach to dense quantum systems, Phys. Rev. D **90**(9), 094502 (2014). arXiv:1404.7187
11. C. Gattringer, P. Törek, Density of states method for the Z(3) spin model, Phys. Lett. B **747**, 545–550 (2015). doi:10.1016/j.physletb.2015.06.017. arXiv:1503.4947
12. K. Langfeld, J.M. Pawlowski, Two-color QCD with heavy quarks at finite densities, Phys. Rev. D **88**(7), 071502 (2013). arXiv:1307.0455
13. M. Guagnelli, Sampling the density of states. arXiv:1209.4443
14. R. Pellegrini, K. Langfeld, B. Lucini, A. Rago, The density of states from first principles, PoS LATTICE2014 229 (2015). arXiv:1411.0655
15. H. Robbins, S. Monro, A stochastic approximation method. Ann. Math. Stat. **22**(09), 400–407 (1951)
16. M. Luscher, P. Weisz, Locality and exponential error reduction in numerical lattice gauge theory. JHEP **09**, 010 (2001). arXiv:hep-lat/0108014
17. R.H. Swendsen, J.-S. Wang, Replica monte carlo simulation of spin-glasses. Phys. Rev. Lett. **57**, 2607–2609 (1986)
18. T. Vogel, Y.W. Li, T. Wüst, D.P. Landau, Scalable replica-exchange framework for Wang–Landau sampling. Phys. Rev. E **90**, 023302 (2014)
19. A. Bazavov, B.A. Berg, Heat bath efficiency with metropolis-type updating. Phys. Rev. D **71**, 114506 (2005). arXiv:hep-lat/0503006
20. B. Berg, T. Neuhaus, Multicanonical ensemble: a new approach to simulate first order phase transitions. Phys. Rev. Lett. **68**, 9–12 (1992). arXiv:hep-lat/9202004
21. K.G. Wilson, Confinement of quarks. Phys. Rev. D **10**, 2445–2459 (1974)
22. M. Creutz, L. Jacobs, C. Rebbi, Monte Carlo study of Abelian lattice gauge theories. Phys. Rev. D **20**, 1915 (1979)
23. B. Lautrup, M. Nauenberg, Phase transition in four-dimensional compact QED. Phys. Lett. B **95**, 63–66 (1980)
24. G. Bhanot, The nature of the phase transition in compact QED. Phys. Rev. D **24**, 461 (1981)
25. J. Jersak, T. Neuhaus, P. Zerwas, U(1) lattice gauge theory near the phase transition. Phys. Lett. B **133**, 103 (1983)
26. V. Azcoiti, G. Di Carlo, A. Grillo, Approaching a first order phase transition in compact pure gauge QED. Phys. Lett. B **268**, 101–105 (1991)
27. G. Bhanot, T. Lippert, K. Schilling, P. Uberholz, First order transitions and the multihistogram method. Nucl. Phys. B **378**, 633–651 (1992)
28. C. Lang, T. Neuhaus, Compact U(1) gauge theory on lattices with trivial homotopy group. Nucl. Phys. B **431**, 119–130 (1994). arXiv:hep-lat/9407005
29. W. Kerler, C. Rebbi, A. Weber, Monopole currents and Dirac sheets in U(1) lattice gauge theory. Phys. Lett. B **348**, 565–570 (1995). arXiv:hep-lat/9501023
30. J. Jersak, C. Lang, T. Neuhaus, NonGaussian fixed point in four-dimensional pure compact U(1) gauge theory on the lattice. Phys. Rev. Lett. **77**, 1933–1936 (1996). arXiv:hep-lat/9606010
31. I. Campos, A. Cruz, A. Tarancon, A Study of the phase transition in 4-D pure compact U(1) LGT on toroidal and spherical lattices. Nucl. Phys. B **528**, 325–354 (1998). arXiv:hep-lat/9803007

32. G. Arnold, T. Lippert, K. Schilling, T. Neuhaus, Finite size scaling analysis of compact QED. Nucl. Phys. Proc. Suppl. **94**, 651–656 (2001). arXiv:hep-lat/0011058

33. G. Arnold, B. Bunk, T. Lippert, K. Schilling, Compact QED under scrutiny: it's first order. Nucl. Phys. Proc. Suppl. **119**, 864–866 (2003). arXiv:hep-lat/0210010

34. J. Fröhlich, P.A. Marchetti, Soliton quantization in lattice field theories. Commun. Math. Phys. **112**(2), 343–383 (1987)

35. A. Di Giacomo, G. Paffuti, A Disorder parameter for dual superconductivity in gauge theories. Phys. Rev. D **56**, 6816–6823 (1997). arXiv:hep-lat/9707003