Overcoming Ergodicity Problems of the Hybrid Monte Carlo Method using Radial Updates

Finn Temmen,[a,*] **Evan Berkowitz,**[a] **Anthony Kennedy,**[c] **Thomas Luu,**[a,b] **Johann Ostmeyer**[b] **and Xinhao Yu**[c]

[a]*Institute for Advanced Simulation (IAS-4),*
Forschungszentrum Jülich, Germany

[b]*Helmholtz-Institut für Strahlen- und Kernphysik and Bethe Center for Theoretical Physics,*
Rheinische Friedrich-Wilhelms-Universität Bonn, Germany

[c]*School of Physics and Astronomy,*
The University of Edinburgh, Scotland, UK

E-mail: f.temmen@fz-juelich.de, e.berkowitz@fz-juelich.de,
Tony.Kennedy@ed.ac.uk, t.luu@fz-juelich.de, ostmeyer@hiskp.uni-bonn.de,
X.Yu-34@sms.ed.ac.uk

Despite its many advantages, the sensible application of the Hybrid Monte Carlo (HMC) method is often hindered by the presence of large - or even infinite - potential barriers. These potential barriers partition the configuration space into distinct sectors, which leads to ergodicity violations and biased measurements of observables. In this work, we address this problem by augmenting the HMC method with a multiplicative Metropolis-Hastings update in a so-called "radial direction" of the fields, which enables jumps over the aforementioned potential barriers at comparably low computational cost. The effectiveness of this approach is demonstrated for the Hubbard model, formulated in a non-compact space by means of a continuous Hubbard-Stratonovich transformation. Our numerical results show that the radial updates successfully resolve the ergodicity violation, while simultaneously reducing autocorrelations.

The 41st International Symposium on Lattice Field Theory (LATTICE2024)
28 July - 3 August 2024
Liverpool, UK

---

*Speaker

1. Introduction

The Hybrid Monte Carlo (HMC) method [1] is one of the most successful tools in the simulation of lattice field theories. However, despite its many advantages, its application often faces challenges due to manifolds of vanishing fermion determinant and the concomitant emergence of infinite potential barriers. These potential barriers lead to diverging force terms in the molecular dynamics (MD) evolution and, when separating regions in configuration space, result in an ergodicity violation of the algorithm. A prime example for such a system is the Hubbard model, formulated in a non-compact space by means of a continuous Hubbard-Stratonovich (HS) transformation. In the adopted formulation, the occurring fermion determinant vanishes on manifolds of codimension 1 and causes a formal ergodicity problem in HMC simulations [2]. Thus, a sensible application of HMC necessitates the development of strategies to circumvent the potential barriers, which, in the context of the Hubbard model, has been discussed in great detail in [3]. In this work, we propose another method, which interleaves the HMC simulation with so-called *radial updates*. Radial updates refer to multiplicative Metropolis-Hastings (MH) updates in a radial direction of the non-compact fields, that enables jumps over the aforementioned potential barriers at comparably low computational cost.

In the following sections, we commence by introducing the Hubbard model in Section 2 and discuss ergodicity violations of the HMC due to the emergence of potential barriers. Afterwards, in Section 3, we define the radial updates, before applying them to the simulation of the Hubbard model in Section 4. Finally, in Section 5, we conclude with a summary and provide an outlook on potential future avenues.

2. Hubbard model

The Hubbard model is commonly used to describe strongly-correlated electrons in a variety of condensed matter systems. We use the formulation of the Hubbard model in the so-called particle/hole basis, where the Hubbard Hamiltonian on a spatial lattice with N_x sites is given by

$$H = H_K + H_U = -\kappa \sum_{\langle x,y \rangle} \left(a_x^\dagger a_y - b_x^\dagger b_y \right) + \frac{U}{2} \sum_x \left(a_x^\dagger a_x - b_x^\dagger b_x \right)^2. \tag{1}$$

The Hamiltonian contains a nearest neighbor hopping term with the hopping parameter κ and an on-site interaction with interaction strength U. The fermionic operator $a_x^\dagger$, (a_x) creates (annihilates) a spin-$\uparrow$ electron at the lattice site x. On the other hand, the operator $b_x^\dagger$, (b_x) creates (annihilates) a spin-$\downarrow$ electron-hole at the lattice site x. The partition function and expectation value of an observable O are given by the thermal traces

$$Z = \mathrm{tr}\left(e^{-\beta(H_K + H_U)} \right) \quad \text{and} \quad \langle O \rangle = Z^{-1} \mathrm{tr}\left(O e^{-\beta(H_K + H_U)} \right). \tag{2}$$

The application of the HMC method requires transitioning from the thermal traces (2) to a path integral formulation, which is achieved by first discretizing the inverse temperature β into N_t time slices with lattice spacing $\Delta_t = \beta/N_t$ and performing a second order Suzuki-Trotter decomposition. This introduces an error of $O(\Delta_t^2)$ and thus necessitates taking the continuum limit $N_t \to \infty$

to recover the exact expression. Applying a continuous HS transformation then decouples the many-body interactions at the cost of introducing a non-compact auxiliary bosonic field ϕ. Finally, inserting fermionic coherent states and integrating out the fermionic degrees of freedom leads to the Hubbard action

$$S[\phi] = \frac{1}{2U\Delta_t} \sum_{t,x} \phi_{tx}^2 - \log\left(\det M[\phi|\kappa] \det M[-\phi|-\kappa]\right), \tag{3}$$

with the fermion matrix

$$M[\phi|\kappa]_{tx,t'y} = \delta_{t,t'}\delta_{x,y} - \left(e^{\kappa h}\right)_{xy} e^{i\phi_{tx}} \mathcal{B}_{t'}\delta_{t',t+1}. \tag{4}$$

Here, $h = \Delta_t \delta_{\langle z,z'\rangle}$ is the hopping matrix and $\mathcal{B}_t$ encodes the anti-periodic boundary conditions with $\mathcal{B}_t = +1$ for $0 < t' < N_t$ and $\mathcal{B}_0 = -1$. For a more detailed derivation of the Hubbard action (3) and similar formulations we refer the reader to [3, 4].

It was shown in [2, 3] that the fermion matrix (4) vanishes on manifolds with codimension 1, i.e. on manifolds with dimension $d - 1$ in the d-dimensional configuration space. This gives rise to infinite potential barriers that separate regions in configuration space, and if the evolution of the molecular dynamics equations attempts to pass them, the force term $F[\phi] = -\frac{\partial S}{\partial \phi}$ diverges. Therefore, the evolution is always repelled when approaching the potential barrier and the algorithm can not cross over into the separated region, resulting in an ergodicity violation. Thus, to obtain correct results, strategies for circumventing the potential barriers have to be developed and recent approaches are e.g. the complex reformulation [5–7], coarse MD integration [3], using an operator with the same continuum limit but worse symmetry properties [3], or utilizing novel generative machine learning architectures [8]. In this work, we adopt the approach of using a non-Hamiltonian MH update to enable jumps over the potential barriers. Specifically, we employ so-called radial updates, which will be introduced in the next section.

3. Radial Updates

Radial updates were proposed in Ref. [9] and optimised in Ref. [10]. Here, they refer to a multiplicative Metropolis-Hastings (MH) update of a non-compact bosonic field $\phi = (\phi_1, \ldots, \phi_d)$ that generates a new proposal by scaling the radius in field space

$$R = \sqrt{\sum_{i=1}^{d} \phi_i^2}. \tag{5}$$

The radial updates are a special case of the updates used in more general multiplicative MH algorithms, such as the Random dive MH algorithm [11] and the Transformation-based Markov chain Monte Carlo method [12]. However, instead of devising a new multiplicative MH algorithm, we propose to augment a standard HMC simulation with intermediate radial updates. This combined algorithm addresses the ergodicity problems of the standalone HMC at low computational cost, while also maintaining its favorable properties. The radial update procedure is defined as follows:

1. Given the initial configuration $\phi = (\phi_1, \ldots, \phi_d)$, an update variable γ is sampled from a zero-centered normal distribution $\mathcal{N}(\gamma|\mu = 0, \sigma_R^2)$ with standard deviation σ_R.

2. A new configuration ϕ' is generated by multiplying the initial configuration ϕ with e^γ, i.e.

$$\phi' = (e^\gamma \phi_1, \ldots, e^\gamma \phi_d). \tag{6}$$

This amounts to rescaling the radius of the initial configuration to $R' = e^\gamma R$, giving rise to the term *radial update*. The scaling factor e^γ follows a log-normal distribution with a median of one, ensuring that increases and decreases in the radius are proposed with equal probability.

3. The new configuration ϕ' is used as a trial configuration in a Metropolis acceptance test with acceptance probability

$$\alpha_R = \min\left(1, e^{-\Delta S + d\gamma}\right), \tag{7}$$

where $\Delta S = S[\phi'] - S[\phi]$ is the difference in action. Additionally, due to the multiplicative nature of the radial updates, the acceptance probability comprises a factor of $e^{d\gamma}$, stemming from the Jacobian of the transformation. Notably, in this setup, the acceptance probability is independent of the normal distribution used to sample γ, and therefore independent of σ_R (except indirectly).

The radial updates satisfy the detailed balance condition, and therefore the combined algorithm of radial updates and HMC does as well. Furthermore, in the combined algorithm, the radial updates resolve possible ergodicity problems of the HMC by enabling jumps over potential barriers. In practice, employing radial updates requires additional design choices in comparison to a standalone HMC simulation. These include selecting the standard deviation σ_R and determining the frequency with which each type of step is applied. Throughout this work, we will specify the latter by defining the ratio of HMC steps to the number of radial update steps in a combined update step.

4. Results

In this section, we apply the combined algorithm of HMC and radial updates, as introduced in Section 3, to simulate the Hubbard model (3). Specifically, we demonstrate the algorithm's effectiveness in overcoming ergodicity issues caused by potential barriers, as discussed in Section 2. To achieve this, we examine the 2-site model, investigating autocorrelation times and the optimization of the additional parameter σ_R. We then analyze how the method scales with growing dimensionality by increasing the number of time slices N_t. The simulations were performed using the *Nanosystem Simulation Library* [13] and the data analysis was performed using the *comp-avg* tool [14].

4.1 Restoring ergodicity in the 2-Site model

We begin by examining a 2-site model on a single time slice, which allows for a direct visualization of the configurations $\phi = (\phi_1, \phi_2)$ and the ergodicity problems posed by the infinite potential barriers. We simulate the model using $U = 18$, $\kappa = 1$ and $\beta = 1$, employing both the standalone HMC and HMC augmented with radial updates. In this and in all following simulations, the MD integration is performed using the Leapfrog integrator with the trajectory length set to

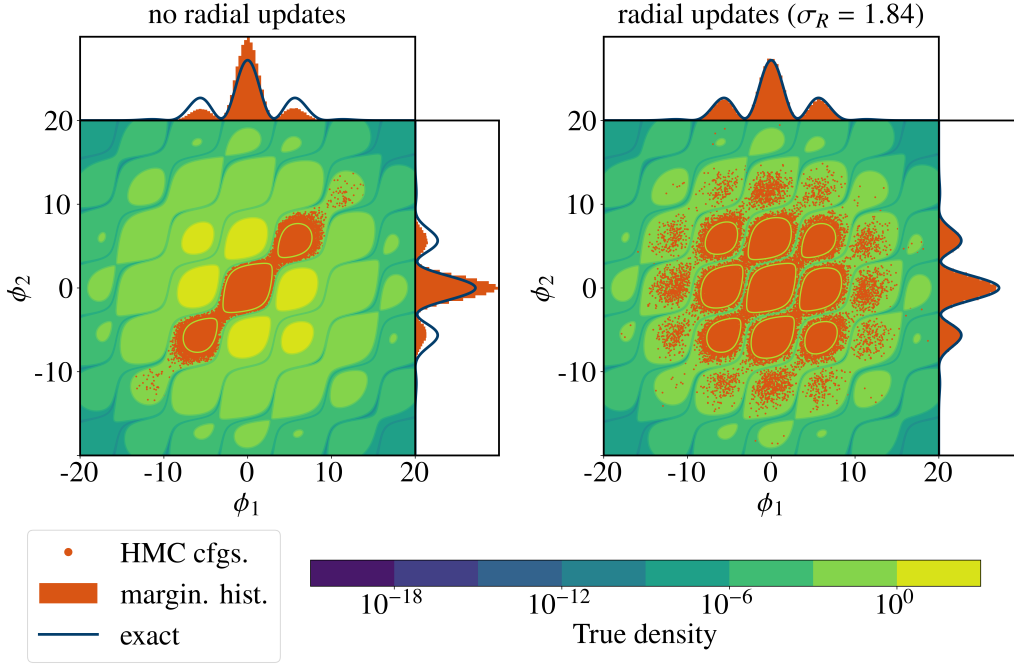$$T = \frac{\pi}{2}\sqrt{U\beta/N_t}, \tag{8}$$

Figure 1: HMC configurations (dots) for the 2-site model on a single time slice ($N_t = 1$) for $U = 18$, $\beta = 1$ and $\kappa = 1$. The configurations are compared to the exact unnormalized distribution, shown as a contour plot. The HMC simulations were conducted with $T = \frac{\pi}{2}\sqrt{U\beta/N_t}$ and $N_{\mathrm{MD}} = 60$, achieving an acceptance rate > 99%. The left panel shows the simulation without radial updates, while the right panel includes radial updates. In the latter, a single radial update was performed before each HMC step with a proposal standard deviation of $\sigma_R = 1.84$. Each panel shows 10^5 of the $N_{\mathrm{conf}} = 2 \times 10^5$ recorded configurations, with measurements taken after each HMC step. The plots at the margins compare the exact marginal distribution to the histograms obtained from the visualized trajectories, respectively.

which eliminates autocorrelations originating from the harmonic part of the action [15]. Furthermore, throughout this work, the number of molecular dynamics steps is always tuned to obtain a fine integrator with acceptance rate > 99%, leading to $N_{\mathrm{md}} = 60$ in the present simulation. We record $N_{\mathrm{config}} = 2 \times 10^5$ configurations, saving the configuration after each HMC step. In the simulation with radial updates, we employ one radial update per HMC step with standard deviation $\sigma_R = 1.84$.

The first 10^5 configurations in the Markov chain are visualized in Figure 1. Without radial updates, shown in the left panel, the HMC becomes trapped within the middle diagonal band in configuration space, which is separated from the adjacent bands by the aforementioned infinite potential barriers. As a result, the simulation fails to sample regions of high probability in the adjacent bands, leading to a severe ergodicity problem. However, when radial updates are activated, as shown in the right panel of Figure 1, the ergodicity violations are resolved entirely. This qualitative observation is further supported when histograms of the recorded configurations are compared to the exact marginal distributions, as depicted at the margins of the two plots in Figure 1. Without radial updates, the middle peak at $\phi_i = 0$ is heavily oversampled in this example. However, when radial updates are employed, the simulation closely aligns with the exact distribution.

Next, we increase the dimensionality d of the simple two-site model by increasing the number of time slices N_t, while keeping the parameters U, β and κ fixed. In this scenario, the potential
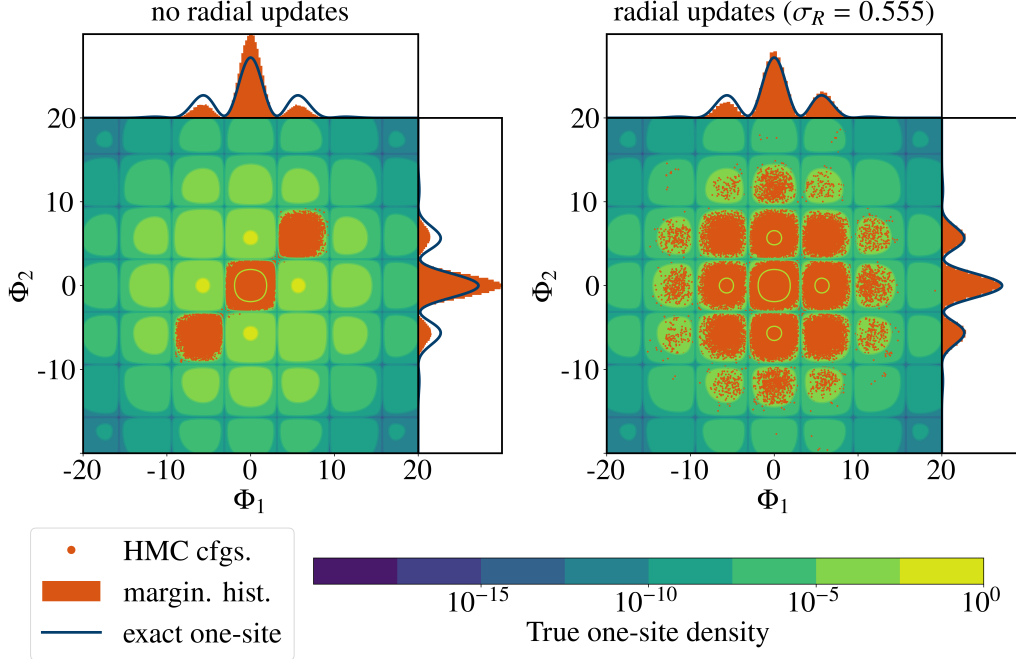
Figure 2: HMC configurations (dots) for the 2-site model with $N_t = 8$ for $U = 18$, $\beta = 1$ and $\kappa = 1$. The configurations $\Phi = (\Phi_1, \Phi_2)$ with $\Phi_x = \sum_t \phi_{tx}$ are compared to the exact one-site distribution, recovered in the strong-coupling limit $U/\kappa \gg 1$ and shown as a contour plot. The HMC simulations were conducted with $T = \frac{\pi}{2}\sqrt{U\beta/N_t}$ and $N_{\text{MD}} = 60$, achieving an acceptance rate > 99%. The left panel shows the simulation without radial updates, while the right panel includes radial updates. In the latter, a single radial update was performed before each HMC step with a proposal standard deviation of $\sigma_R = 1.84$. Each panel shows 10^5 of the $N_{\text{conf}} = 3 \times 10^5$ recorded configurations, with measurements taken after each HMC step. The plots on the margins compare the marginal one-site distribution to the histograms obtained from the visualized trajectories, respectively.

barriers in configuration space can be visualized by considering $\Phi_x = \sum_t \phi_{tx}$, because in the strong-coupling limit $U/\kappa \gg 1$, the probability weights of a configuration $\Phi = (\Phi_1, \Phi_2)$ are well-approximated by the exact one-site distribution given in [3]. We first examine the exemplary case of $N_t = 8$ by visualizing the first 10^5 of a total of $N_{\text{conf}} = 3 \times 10^5$ recorded configurations, both without and with radial updates. In the simulation with radial updates we employ one radial update with standard deviation $\sigma_R = 0.555$ per HMC step. The results are shown in Fig. 2. Similarly to the $N_t = 1$ case, we observe that without radial updates, the trajectory remains trapped within the middle diagonal band in the two-dimensional Φ-plane, whereas, with radial updates, the algorithm efficiently explores the entire configuration space, thereby restoring ergodicity.

4.2 Autocorrelations, parameter tuning and scaling

In the next phase of our analysis, we examine the influence of the radial updates on autocorrelations and utilize the integrated autocorrelation time to optimize the proposal standard deviation of the radial updates, while also exploring their scaling properties. We begin by computing the integrated autocorrelation time τ_{int} for the 2-site model with $N_t = 1$ as a function of the standard deviation σ_R. For a comprehensive introduction to the estimation of errors and autocorrelation
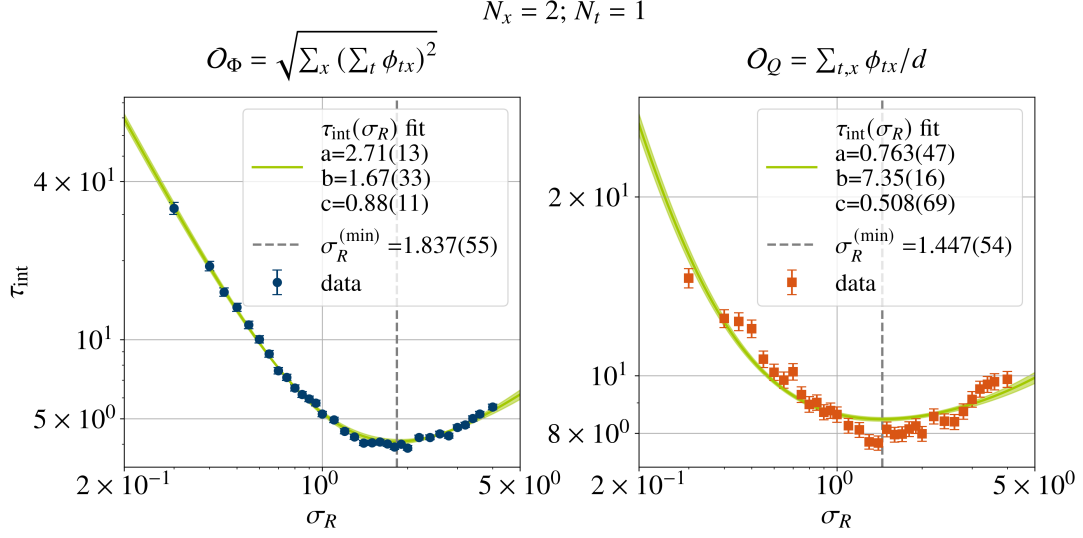
Figure 3: Integrated autocorrelation time τ_{int} for the observables O_Φ (left) and O_Q (right), defined in (9), as a function of the proposal standard deviation σ_R. The underlying model is the 2-site model on a single time slice ($N_t = 1$) for $U = 18$, $\beta = 1$ and $\kappa = 1$. The HMC simulations were conducted with $T = \frac{\pi}{2}\sqrt{U\beta/N_t}$ and $N_{\text{MD}} = 60$, achieving an acceptance rate > 99%. Each HMC step was preceded by a single radial update and a total of $N_{\text{conf}} = 2 \times 10^5$ configurations were recorded, with measurements taken after each HMC step.

functions, we refer the reader to [16]. The observables that we consider are the net charge of the system O_Q and the heuristically motivated Φ-plane radius O_Φ, defined by

$$O_Q = \frac{1}{d}\sum_{t,x}\phi_{tx} \qquad \text{and} \qquad O_\Phi = \sqrt{\sum_x\left(\sum_t\phi_{tx}\right)^2}. \qquad (9)$$

The results are depicted in Fig. 3, where we observe that τ_{int} initially decreases as σ_R increases, reaches a minimum, and then starts to rise. To describe the dependence of the integrated autocorrelation time on the proposal standard deviation σ_R, we adopt the fitting ansatz

$$\tau_{\text{int}}(\sigma_R) = a\sigma_R^{-2} + b + c\sigma_R. \qquad (10)$$

In this expression, the first term accounts for the expected random walk behavior at small σ_R, resulting in a diffusive regime with $\tau_{\text{int}} \propto \sigma_R^{-2}$. The third term quantifies the large σ_R regime where proposed steps are large and the autocorrelation time increases with decreasing acceptance rate, such that $\tau_{\text{int}} \propto \sigma_R$. The extrapolation is displayed in Fig. 3 and throughout this work, fit results are obtained by fitting the respective ansatz to $N_{\text{boot}} = 10^3$ bootstrap samples of the measured data. The results clearly indicate that the chosen ansatz effectively captures the behavior of the data and, additionally, it enables the sensible estimation of the position of the minimum $\sigma_R^{(\text{min})}$. To examine the scaling behavior of the radial updates, we conduct simulations across several values of N_t and repeat the previous analysis. The respective estimates for $\sigma_R^{(\text{min})}$ and the values for the integrated autocorrelation time at the minimum, denoted by $\tau_{\text{int}}^{(\text{min})}$, are shown in Fig. 4. Theoretical considerations suggest that the position of the minimum should scale as $\sigma_R^{(\text{min})}(d) \propto d^{-0.5} + O(d^{-1})$ at leading order [10], motivating the fit-model $\sigma_R^{(\text{min})}(d) = \alpha d^\beta$. We find that the extrapolation of
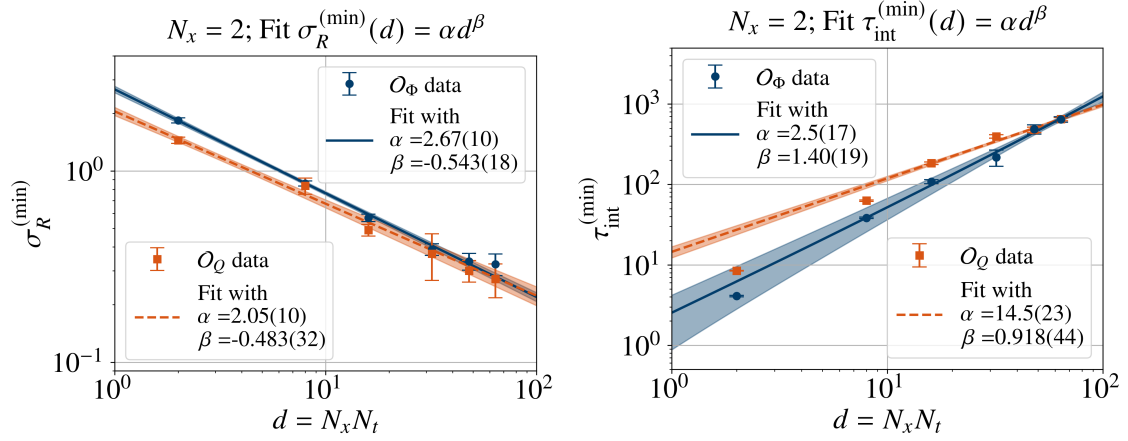
Figure 4: The left panel shows the position of the minimum $\sigma_R^{(\mathrm{min})}$ of the fit ansatz (10) as a function of dimensionality $d = N_x N_t$, along with bootstrapped fits using the model $\sigma_R^{(\mathrm{min})} = \alpha d^\beta$. The right panel illustrates the integrated autocorrelation time at this minimum, denoted by $\tau_{\mathrm{int}}^{(\mathrm{min})}$, also as a function of dimensionality d, with corresponding bootstrapped fits using the model $\tau_{\mathrm{int}}^{(\mathrm{min})} = \alpha d^\beta$. Both panels display results for the 2-site model with varying N_t for $U = 18$, $\beta = 1$ and $\kappa = 1$.

our numerical results closely matches the predicted theoretical $d^{-0.5}$-scaling for both observables. To determine the leading order scaling of the minimal integrated autocorrelation time, we apply the same fit-model $\tau_{\mathrm{int}}^{(\mathrm{min})}(d) = \alpha d^\beta$, resulting in an almost linear leading order scaling of the minimal integrated autocorrelation time with the dimension d.

5. Summary and Outlook

In this work, we demonstrated that the augmentation of HMC with radial updates successfully resolves ergodicity violations of a standalone HMC in the Hubbard model. Furthermore, the radial updates reduce autocorrelation times. We also discussed the scaling and tuning of additional parameters. Due to the global multiplicative nature of the radial updates, their utilization is computationally inexpensive compared to the HMC itself.

In future research, we aim to investigate the scaling properties of radial updates further by extending our analysis to simulations of larger, more realistic systems, such as the simulation of Perylene [17]. Additionally, performing a realistic simulation also involves utilizing a coarser MD integrator, which leads to energy violations in the integration that also enable tunnelling through potential barriers. The employment of radial updates also has great implications for the geometrical convergence of HMC on non-compact manifolds, which will be subject to future work.

Acknowledgments

References

[1] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.

[2] M.V. Ulybyshev and S.N. Valgushev, *Path integral representation for the Hubbard model with reduced number of Lefschetz thimbles*, 2017, 1712.02188.

[3] J.-L. Wynen, E. Berkowitz, C. Körber, T.A. Lähde and T. Luu, *Avoiding ergodicity problems in lattice discretizations of the Hubbard model*, *Phys. Rev. B* **100** (2019) 075141 [1812.09268].

[4] D. Smith and L. von Smekal, *Monte Carlo simulation of the tight-binding model of graphene with partially screened Coulomb interactions*, *Phys. Rev. B* **89** (2014) 195429 [1403.3620].

[5] S. Beyl, F. Goth and F.F. Assaad, *Revisiting the hybrid quantum Monte Carlo method for Hubbard and electron-phonon models*, *Phys. Rev. B* **97** (2018) 085144 [1708.03661].

[6] P. Buividovich, D. Smith, M. Ulybyshev and L. von Smekal, *Hybrid Monte Carlo study of competing order in the extended fermionic Hubbard model on the hexagonal lattice*, *Phys. Rev. B* **98** (2018) 235129 [1807.07025].

[7] S.R. White, R.L. Sugar and R.T. Scalettar, *Algorithm for the simulation of many-electron systems at low temperatures*, *Phys. Rev. B* **38** (1988) 11665.

[8] D. Schuh, J. Kreit, E. Berkowitz, L. Funcke, T. Luu, K.A. Nicoli, M. Rodekamp, *Simulating the Hubbard Model with Equivariant Normalizing Flows*, *PoS* **LATTICE2024** 069.

[9] A. Kennedy and X. Yu, *The convergence of HMC on non-compact Riemannian Manifolds*, forthcoming.

[10] J. Ostmeyer, *Radial updates in non-compact spaces from continuous distributions*, forthcoming.

[11] S. Dutta, *Multiplicative random walk Metropolis-Hastings on the real line*, *Sankhya B* **74** (2012) 315 [1008.5227].

[12] S. Dutta and S. Bhattacharya, *Markov chain Monte Carlo based on deterministic transformations*, *Stat. Methodol.* **16** (2014) 100 [1106.5850].

[13] M. Rodekamp et al., *Nanosystem Simulation Library (NSL)*, URL: https://github.com/Marcel-Rodekamp/NSL.

[14] J. Ostmeyer, *comp-avg: Compare Averages of time series and more*, Zenodo, March 2024, version v2.1.1, 10.5281/zenodo.10794620. URL: https://github.com/j-ostmeyer/comp-avg.

[15] J. Ostmeyer and P. Buividovich, *Minimal Autocorrelation in Hybrid Monte Carlo simulations using Exact Fourier Acceleration*, 2024, 2404.09723.

[16] U. Wolff, *Monte Carlo errors with less errors*, *Comput. Phys. Commun.* **156** (2004) 143 [hep-lat/0306017].

[17] M. Rodekamp, E. Berkowitz, C. Gäntgen, S. Krieg, T. Luu, J. Ostmeyer and G. Pederiva, *Single Particle Spectrum of Doped* $C_{20}H_{12}$*-Perylene*, 2024, `2406.06711`.

[18] Jülich Supercomputing Centre, *JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre*, *Journal of large-scale research facilities* **7** (2021) .

PoS(LATTICE2024)068