hep-lat 0910.1465

arXiv:0910.1465v2 [hep-lat] 16 Oct 2009

Investigating the critical slowing down of QCD simulations

Stefan Schaefer*
Humboldt Universität zu Berlin, Institut für Physik, Newtonstr. 15, 12489 Berlin, Germany
E-mail: sschaef@physik.hu-berlin.de

Rainer Sommer and Francesco Virotta
NIC/DESY, Platanenallee 6, 15738 Zeuthen, Germany
E-mail: Rainer.Sommer@desy.de, Francesco.Virotta@desy.de

Simulations of QCD are known to suffer from serious critical slowing down towards the continuum limit. This is particularly prominent in the topological charge. We investigate the severeness of the problem in the range of lattice spacings used in contemporary simulations and propose a method to give more reliable error estimates.

The XXVII International Symposium on Lattice Field Theory
July 26-31, 2009
Peking University, Beijing, China

*Speaker.

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike Licence.

1. Introduction

Simulations of QCD on the lattice seem to have reached the stage at which all systematic errors can be controlled in a well defined way. This includes control over the continuum limit $a \to 0$ for which calculations at a series of decreasing lattice spacings a have to be performed. Approaching the critical point at $a = 0$ using contemporary algorithms, however, one encounters severe critical slowing down. As we will show in the following, it is most prominent in the topological charge, whose auto-correlation time increases dramatically as the lattice spacing drops below $a \approx 0.08$fm. Critical slowing down is a very general problem in Markov chain Monte Carlo simulations and also the slowing down of topological modes has been observed before. Recently, Del Debbio, Manca and Vicari[1] studied the issue in the $\mathrm{CP}^{(N-1)}$ model and found evidence for the auto-correlation times rising exponentially with the correlation length. Also in Yang-Mills theory[2] and QCD[3] large auto-correlations associated with topology have been observed.

Our interest in the subject was spurred by the observation of a significant slowing down of the topological charge in our simulations with two flavors of dynamical Wilson fermions in the context of the CLS effort[1]. We are particularly interested in lattice spacings which are fine enough to reliably simulate a relativistic charm quark for which we estimate $a \approx 0.04$fm to be necessary[4].

2. Dynamical simulations

To illustrate the problem, let us start with the dynamical simulations of $N_f = 2$ degenerate flavors of non-perturbatively improved Wilson quarks and Wilson gauge action. We use three values of the coupling constant, $\beta = 5.3$, 5.5 and 5.7, which correspond to roughly $a \approx 0.08$fm, 0.06fm and 0.04fm respectively [5, 6, 7]. For each coupling constant, several values of the sea quark mass are simulated with the DD-HMC algorithm [8], which uses a block decomposition of the lattice to separate the infrared from the short distance modes. As a consequence, only gauge links within the blocks are updated ("are active"). We will show in the pure gauge study that this has no noticeable effect on the problems discussed in this section beyond rescaling by the fraction of active links.

Each of the ensembles was generated with trajectories of length $\tau = 0.5$ and after each trajectory, the lattice is shifted randomly to change the active links. On the saved configurations we measure the topological charge using three times HYP smeared gauge fields [9] from which we constructed the clover field strength tensor $F_{\mu\nu}$. Then the charge Q is given by

$$Q = \frac{1}{16\pi^2} a^4 \sum_x \mathrm{tr}\left[F(x)\tilde{F}(x)\right] \ . \tag{2.1}$$

Although this definition of the charge does in general not give an integer and is therefore not particularly useful for the computation of physical observables, it is known to be well correlated with, e.g. the definition of the charge via the number of zero modes of a chiral Dirac operator. In the end, the precise interpretation does not matter, since a slow evolution shows a problem of the algorithm. Also note that since Q in Eq. 2.1 is parity odd, its expectation value vanishes.

[1] https://twiki.cern.ch/twiki/bin/view/CLS

Figure 1: Examples of the Monte Carlo time history of the topological charge. The data shown in the left plot is from a run on a 64×32^3 lattice at $\beta = 5.3$, $m_\pi \approx 360$MeV. The central plot shows a 64×32^3 lattice at $\beta = 5.5$, $m_\pi \approx 460$MeV run, the right plot a 128×64^3 lattice at $\beta = 5.7$, $m_\pi \approx 480$MeV.

Examples of the Monte Carlo time history for the three values of the coupling constant are given in Fig. 1. At the coarsest lattice spacing $a \approx 0.08$fm, the run with roughly 5000 trajectories shows quite long correlations, but a quite decent sampling of all topological sectors. At the intermediate lattice spacing $a \approx 0.06$fm, the auto-correlations get worse and for the finest lattice spacing, even 1000 trajectories of length 0.5 do not manage to move the charge considerably. The qualitative picture does not depend on the value of the sea quark mass. What matters most is the value of the gauge coupling constant. However, due to the high cost of the dynamical simulations our time histories are too short to make a definitive statement about the auto-correlation time of the topological charge.

3. Pure gauge study

Since the sea quark mass seems to have little effect, we now study the problem in pure gauge theory. We measure the auto-correlations of the topological charge and of Wilson loops in simulations with the same DD-HMC algorithm as in the dynamical runs. As a comparison serve some runs with the standard HMC algorithm, to check for effects of the block-decomposition, and with hybrid over-relaxation (HOR) in order to see whether this is a particular problem of the molecular dynamics based algorithms.

We use Wilson gauge action and start on a 24^4 lattice at $\beta = 6.0$ and match lattices at constant physical volume using r_0 as a scale. With $r_0 = 0.49$fm, the lattices of size $L/a = 16$, 24, 32 and 48 have therefore a lattice spacing of about 0.14fm, 0.09fm, 0.07fm and 0.046fm, respectively [10].

With the DD-HMC algorithm and a fixed trajectory length of $\tau = 4$, we measure the auto-correlation time of Q^2 as a function of the lattice spacing. In the next section we discuss the method used to determine τ_{int} and why we use Q^2 and not Q. Here, and in all other plots, we multiply the auto-correlation times by the fraction of active links. In pure gauge theory, also the cost of a trajectory scales with this ratio. The result is shown on the left of Fig. 2. We observe a steep rise towards the continuum limit, which is roughly compatible with $\tau_{\text{int}} \propto a^{-5}$, but also the exponential hypothesis of Ref. [1] can describe the data. Our precision and our range in a are not enough to distinguish between the two scenarios.

The problem is not limited to molecular dynamics based algorithms only. On the right of Fig. 2 we show data by M. Lüscher and F. Palombi using sweeps combined of heatbath and over-

Figure 2: Left: Auto-correlation time of Q^2 and the $(0.5\text{fm}) \times (0.5\text{fm})$ square Wilson loop as a function of the lattice spacing using the DD-HMC algorithm. The two curves are possible functional forms which describe the data. At $a \approx 0.09$fm we show values of $\tau_{\text{int}}(Q^2)$ for several domain decompositions. **Right:** Same plot, but using the HOR algorithm on lattices of size $(4.8\text{fm}) \times (2.4\text{fm})^3$. The scales of the y-axis of the two plots are unrelated, the line in the right plot is just to guide the eye.

Figure 3: Left: Dependence of the AC time of Q^2 on the trajectory length for the HMC and DD-HMC algorithm on the 24^4 lattices. The units of τ_{int} are such that the differences in cost due to the inactive links has been scaled out. The values for several block decompositions at the same τ slightly shifted for better visibility, 6^4, $6^2 \times 12^2$ and 12^4 from left to right. **Right:** Estimator $\varphi^{\text{eff}}_{O,1} = \Gamma_O(t)/\Gamma_O(0)[\Gamma_{11}(t/2)/\Gamma_{11}(t)]^2$ for $\varphi_{O,1}$ of the $(0.5\text{fm})^2$ Wilson loop.

relaxation steps. The rise of the auto-correlation time is about as steep as for the DD-HMC algorithm. This indicates a universal problem, because the two algorithms are rather different.

Since the auto-correlation times are much larger than 0.5, the minimum value, a longer trajectory length might have a beneficial effect on the performance of the algorithm. The random walk is then composed of longer, more directed pieces of evolution. An example for this has already been demonstrated in the Schrödinger functional [12]. In Fig. 3 on the left we show the result for $\beta = 6$. We observe an improvement—in total cost of the simulation—which is roughly compatible with the expected $1/\sqrt{\tau}$ behavior. This gain, however, is far too small to solve the problem. In the same plot we put numbers from the DD-HMC algorithm with several block decompositions (6^4, $6^2 \times 12^2$ and 12^4) and the HMC algorithm without domain decomposition. All these values are compatible with another. The simple rescaling by the fraction of active links accounts for all the differences in the auto-correlation times within errors.

The question arises in how far other observables are affected. This can be only answered on a case by case basis. Here we study rectangular Wilson loops, both, constructed from thin links and from once HYP smeared links, in particular the one with the longest auto-correlation time, the smeared (0.5fm)×(0.5fm) square loop. Fig. 3 shows the result. On coarse lattices $\tau_{\rm int}$ of the loop and the charge are comparable, however, the loops are much less affected by critical slowing down as one approaches the continuum. This can be interpreted as a dynamical decoupling of the topological modes from the Wilson loops as the continuum limit is approached.

4. Stabilized error analysis and dynamical (de)coupling

The previous sections present the description of a grave problem for lattice simulations. The topological charge moves very slowly in the simulations. We have already seen that other observables may be affected less, because they are only weakly coupled to this "mode". Here we want to discuss what such a decoupling means and how error estimates can be stabilized in this difficult situation.

We consider an observable $O = F(\bar{A})$, which is a function of the mean $\bar{A}_\alpha = \langle A_\alpha \rangle$ of primary observables $A_\alpha = A_\alpha^*$ averaged over N subsequent measurements in the Markov chain. The error, δO is given by $(\delta O)^2 = \mathrm{var}(O)\frac{2\tau_{\rm int}(O)}{N}$, where the variance is a property of the theory and the auto-correlation time a property of the algorithm. It is given [11] by the summed auto-correlation function $\Gamma(t)$

$$\begin{aligned} \tau_{\rm int}(O) &= \frac{1}{2} + \sum_{t=1}^{\infty} \frac{\Gamma_O(t)}{\Gamma_O(0)}\,, \\ \Gamma_O(t) &= \sum_{\alpha\beta} F_\alpha F_\beta \Gamma_{\alpha\beta}(t)\,, \qquad \Gamma_{\alpha\beta}(t) = \langle (A_\alpha(t) - \bar{A}_\alpha)(A_\beta(0) - \bar{A}_\beta) \rangle\,, \end{aligned} \tag{4.1}$$

where t is the separation in Monte Carlo time and $F_\alpha = \partial F / \partial \bar{A}_\alpha$.

The function Γ can be expressed in terms of the transition matrix of the Markov chain, $M(q',q)$, which gives the transition probability for a change from a state q to a state q'. Denoting the equilibrium probability distribution by $W(q)$, and assuming the algorithm to satisfy detailed balance means that the matrix $T(q',q) = (W(q'))^{-1/2} M(q',q) (W(q))^{1/2}$ is symmetric. Its eigenfunctions are labeled $\chi_n(q)$ and its eigenvalues λ_n. The autocorrelation function can then be written as

$$\Gamma_{\alpha\beta}(t) = \sum_n (\lambda_n)^t \eta^*_{n,\alpha} \eta_{n,\beta} = \sum_n [\mathrm{sign}(\lambda_n)]^t \, \exp(-t/\tau_n) \; \eta^*_{n,\alpha} \eta_{n,\beta}\,, \tag{4.2}$$

where $\lambda_n < 1$ and $\tau_{n+1} \le \tau_n = -1/\ln|\lambda_n|$ and

$$\eta_{n,\alpha} = (\chi_n, W^{1/2} A_\alpha) = \int \mathrm{d}q \, \chi_n^*(q) \, W^{1/2}(q) A_\alpha(q)\,. \tag{4.3}$$

In general all modes n contribute to the sum in Eq. (4.2) and slow modes have $\tau_n \gg 1$ producing a long tail in $\Gamma(t)$. This is typically hard to estimate: one quickly reaches the point from which the uncertainty of $\Gamma(t)$ is larger than the signal. Therefore one truncates the sum in Eq. 4.1 at some time w. This, however, introduces a bias which can be sizeable if the tail is very long.

There can be selection rules imposed by common symmetries of the theory (i.e. $W(q)$) and the algorithm (i.e. $T(q',q)$). Parity is such a case relevant for us: Q couples only to parity odd

Figure 4: Left: Extracting the exponential auto-correlation time from an effective mass plot of $\Gamma(t)$ of Q^2 for our 32^4, $\tau = 1$ run with the DD-HMC algorithm. **Right:** Estimate of τ_{int} of the 0.5fm square Wilson loop as a function of the summation window according to Eq. 4.4.

modes, while Q^2 and other parity even observables couple only to even modes. Since in our lattice QCD discretization parity is an exact symmetry, parity odd observables vanish and we only need to consider even ones. Hence only those n, corresponding to parity even $\chi_n(q)$ contribute in Eq. (4.2).

In order to learn about the slowest mode in a simulation, we can then investigate an "effective mass" plot.[2] Indeed we find a long plateau in $1/\tau_{\text{eff}} = \log(\Gamma_{11}(t)/\Gamma_{11}(t-1))$ for $A_1 = Q^2$ in Fig. 4. It provides a good estimate for $1/\tau_{\text{exp}} = -\log(\lambda_1)$.

If λ_1 has been determined well, it can be inserted into the (for even w) exact inequality

$$\tau_{\text{int}} \leq \frac{1}{2} + \sum_{t=1}^{w-1} \frac{\Gamma_O(t)}{\Gamma_O(0)} + \frac{1}{1-\lambda_1}\frac{\Gamma_O(w)}{\Gamma_O(0)} \approx \frac{1}{2} + \sum_{t=1}^{w-1} \frac{\Gamma_O(t)}{\Gamma_O(0)} + \tau_{\text{exp}}\frac{\Gamma_O(w)}{\Gamma_O(0)} \tag{4.4}$$

to obtain an upper bound for the error. [3]

In a practical procedure we first select a set of observables (plaquette and Q^2 after various smearing levels), determine approximate plateaux in their "effective mass" $1/\tau_{\text{eff}}$ and take the largest one as the estimate for τ_{exp} in Eq. (4.4). Note that if the tail of the auto correlation function is dominated by just the slowest mode or by modes with eigenvalues close to it, the upper bound will (almost) be saturated. That such an upper bound is not overly conservative is demonstrated in Fig. 4 on the right, on the 0.07fm lattice where the τ_{eff} is taken from the topological charge and the observable is the 7×7 Wilson loop. We find that generically the upper bound converges faster as a function of the window size w than the more standard estimate $\tau_{\text{int}} = 1/2 + \sum_{t=1}^{w} \frac{\Gamma_O(t)}{\Gamma_O(0)}$. We therefore advocate its use in difficult situations, e.g. in HMC simulations of QCD.

Let us now return to the decoupling of some observables from the critical slowing down. Since in the representation Eq.(4.2) all modes n contribute, a decoupling from the slowest mode means $|\sum_\alpha F_\alpha \eta_{1,\alpha}|^2/\Gamma_O(0) = \varphi_{O,1} \ll 1$. Combining the autocorrelation functions of the Wilson loops and those of Q^2, one can easily form estimators which converge to $\varphi_{O,1}$ for large t. Even if we observed this convergence to be rather slow, we found that the estimators for $\varphi_{O,1}$ become tiny at small lattice spacing, see Fig. 3 (right). This holds for Wilson loops and Creutz ratios.

[2] Experience shows that oscillating terms ($\text{sign}(\lambda_n) = -1$) are usually irrelevant at large t.

[3] Actually, the DD-HMC algorithm does not fulfill detailed balance, because of the directed block shifts, and the formulae do not apply. However, we expect this to have little effect on the large time behavior of $\Gamma(t)$.

5. Summary

In this contribution we have studied the critical slowing down of lattice simulations using the HMC and DD-HMC algorithms. We demonstrated a steep increase of the auto-correlation time of the squared topological charge in pure gauge simulations which match our observations with dynamical fermions. Tuning of the parameters of the algorithm can improve the situation a bit, but probably more dramatic changes like the one proposed in [13] will be needed to solve the problem.

In the second part, we studied the impact of such slowly moving modes on the estimation of the errors in Monte Carlo simulations. We proposed a method to give upper bounds on the contribution from the tail of the auto-correlation function, which is normally neglected. The resulting estimates are neither overly conservative nor impractical and therefore are a safer way to determine the errors.

Acknowledgments

Work supported in part by the SFB/TR 9 of the Deutsche Forschungsgemeinschaft and by the European Community through EU Contract No. MRTN-CT-2006-035482, "FLAVIAnet". Our simulations are performed on BlueGene and PC-clusters of the John von Neumann Institute for Computing at FZ Jülich, the HLRN in Berlin, the Universities of Mainz, Rome La Sapienza, Valencia-IFIC and at CERN, as well as on the IBM MareNostrum at the Barcelona Supercomputing Center. We thankfully acknowledge the computer resources and support provided these institutions. We thank M. Lüscher and F. Palombi for very interesting discussions and providing us with unpublished data and C. DeTar for a very useful email exchange.

References

[1] L. Del Debbio, G. M. Manca and E. Vicari, Phys. Lett. B **594** (2004) 315 [arXiv:hep-lat/0403001].

[2] L. Del Debbio, H. Panagopoulos and E. Vicari, JHEP **0208** (2002) 044 [arXiv:hep-th/0204125].

[3] C. Bernard *et al.*, Phys. Rev. D **68** (2003) 114501 [arXiv:hep-lat/0308019].

[4] G. von Hippel, R. Sommer, J. Heitger, S. Schaefer and N. Tantalo, PoS **LATTICE2008** (2008) 227 [arXiv:0810.0214 [hep-lat]].

[5] L. Del Debbio, L. Giusti, M. Lüscher, R. Petronzio and N. Tantalo, JHEP **0702** (2007) 056 [arXiv:hep-lat/0610059].

[6] H. Wittig, PoS **LATTICE2009** (2009) 095

[7] M. Della Morte *et al.*, PoS **LAT2007** (2007) 255 [arXiv:0710.1263 [hep-lat]].

[8] M. Lüscher, Comput. Phys. Commun. **165** (2005) 199 [arXiv:hep-lat/0409106].

[9] A. Hasenfratz and F. Knechtli, Phys. Rev. D **64** (2001) 034504 [arXiv:hep-lat/0103029].

[10] S. Necco and R. Sommer, Nucl. Phys. B **622** (2002) 328 [arXiv:hep-lat/0108008].

[11] U. Wolff [ALPHA collaboration], Comput. Phys. Commun. **156** (2004) 143 [Erratum-ibid. **176** (2007) 383] [arXiv:hep-lat/0306017].

[12] H. B. Meyer, H. Simma, R. Sommer, M. Della Morte, O. Witzel and U. Wolff, Comput. Phys. Commun. **176** (2007) 91 [arXiv:hep-lat/0606004].

[13] M. Lüscher, arXiv:0907.5491 [hep-lat].