Published by Institute of Physics Publishing for SISSA/ISAS

Received: *November 12, 2002*
Revised: *January 11, 2003*
Accepted: *January 22, 2003*

Locality and statistical error reduction on correlation functions

Harvey B. Meyer

Theoretical Physics, University of Oxford
1 Keble Road, Oxford, OX1 3NP, UK
Email: meyer@thphys.ox.ac.uk

Abstract: We propose a multilevel Monte-Carlo scheme, applicable to local actions, which is expected to reduce statistical errors on correlation functions. We give general arguments to show how the efficiency and parameters of the algorithm are determined by the low-energy spectrum. As an application, we measure the euclidean-time correlation of pairs of Wilson loops in SU(3) pure gauge theory with constant relative errors. In this case the ratio of the new method's efficiency to the standard one increases as $e^{m_0 t/2}$, where m_0 is the mass gap and t the time separation.

Keywords: Lattice Quantum Field Theory, Lattice Gauge Field Theories, Lattice QCD.

Contents

1. Introduction

In equilibrium statistical mechanics and quantum field theory, the physical information on the theory is encoded in n-point functions. The short-range nature of interactions in the former and the causality requirement in the latter case lead to the property of locality of the hamiltonian (resp. action). While certain lower-dimensional systems have been solved analytically, Monte-Carlo simulations have become an invaluable tool in the study of interacting theories such as non-abelian gauge theories. In particular, the properties of the spectrum are extracted from numerically calculated 2-point functions in the euclidean formulation:

$$C(t) \equiv \langle \mathcal{O}^\dagger(t)\mathcal{O}(t=0)\rangle = \sum_n |\langle 0|\mathcal{O}|n\rangle|^2\, e^{-E_n t}\,.$$

This formula follows directly from the insertion of the complete set of energy eigenstates $H|n\rangle = E_n|n\rangle$. In theories with a mass gap, the exponential decay of each term singles out the lightest state compatible with the symmetry of the operator, thus enabling us to extract the low-lying spectrum of the theory. However, it is precisely this decay that makes the 2-point function numerically difficult to compute at large t. Indeed, standard algorithms keep the absolute error roughly constant, so that the error on the local effective mass

$$am_{\rm eff}\left(t+\frac{a}{2}\right) = \log\frac{C(t)}{C(t+a)}$$

increases exponentially with the time separation. For that reason, it would be highly desirable to have a more efficient method to compute correlations functions at large time separation t. The task amounts to reduce uncorrelated fluctuations between the two time slices separated by euclidean time t.

In the context of non-abelian gauge theories, two types of techniques have proved useful: the first aims essentially at increasing the "signal". The widely used "fuzzing" algorithms, such as smearing [1] and blocking [2], increase the overlap onto the lightest states. As the lattice spacing a is made much smaller than the correlation length ξ,[1] the number of smearing/blocking steps as well as their weights have to be increased in order to maintain the quality of the overlap.

The other type of technique aims at reducing the "noise", that is, the variance of the quantity being evaluated. For instance, when computing a Polyakov loop with the multihit technique [3], a link is replaced by its thermal average under fixed neighbouring links. Last year, Lüscher and Weisz [4] demonstrated how to make even better use of the locality of the theory to exponentially improve the efficiency of the algorithm that computes Polyakov loop correlators. In that method, pairs of time-like links are averaged over in time slices of increasing width with fixed boundary conditions before they are multiplied together to form the loops. In this paper, we present a "noise reduction" method that exploits the locality property in a similar way. It has the advantage of being compatible with the popular link-smearing and -blocking: one can cumulate the advantages of both types of techniques. Indeed, while the fuzzing algorithms also help reduce short wavelength fluctuations inside a time-slice, till now no attempt has been made to tame the noise induced by fluctuations appearing in neighbouring time-slices. Our method aims at averaging out the latter by performing additional sweeps with fixed time-slices as boundary conditions; as the continuum is approached, the volume over which the average is performed ought to be kept fixed in physical units to maintain the efficiency of the algorithm. Finally, we note that the idea is very general and is expected to be applicable in other types of theories.

Outline. We first present the idea in its full generality, without reference to the specific form of the action or to the quantity being computed. We next formulate a multilevel scheme for the case of a 2-pt function and point out how the efficiency and parameters of the algorithm are determined by the low-lying spectrum of the theory. We finally apply this algorithm to the case of (3+1) SU(3) gauge theory.

2. Locality, hierarchical formula and multilevel algorithm

2.1 Locality

The locality property of most studied actions allows us to derive an interesting way of computing correlation functions. First we give a general, 'topological' definition of locality in continuum field theories. For simplicity we shall use a symbolic notation; if $\mathcal{C}$ denotes a configuration, let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{A}$ be mutually disjoint subsets of $\mathcal{C}$. If Ω_X, Ω_Y and Ω_A are their

[1] e.g. defined by $1/\sqrt{\sigma}$, where σ is the fundamental string tension

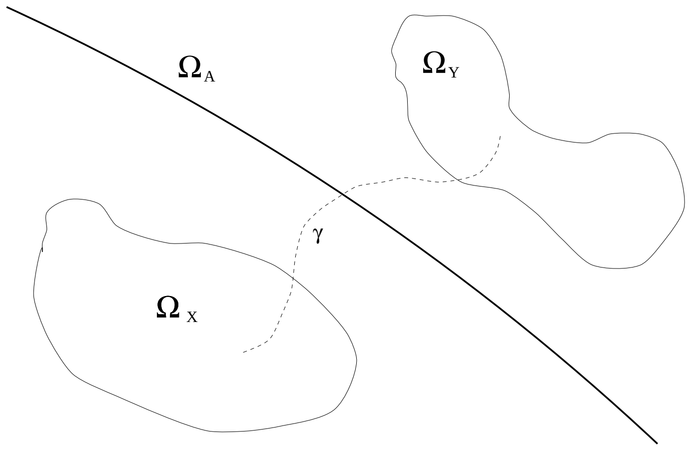

Figure 1: The sets Ω_X, Ω_Y and Ω_A on the space-time manifold.

(disjoint) supports on the (continuous) space-time manifold $\mathcal{M}$, suppose furthermore that any continuous path $\gamma : I \to \mathcal{M}$ joining Ω_X and Ω_Y (i.e. $\gamma(I) \cap \Omega_X \neq \emptyset$ and $\gamma(I) \cap \Omega_Y \neq \emptyset$) passes through Ω_A (i.e. $\gamma(I) \cap \Omega_A \neq \emptyset$). See figure 1. The theory with probability distribution $p(\mathcal{C})$ is *local* if there exists functionals $p_{\mathcal{A}}$ and $\tilde{p}_{\mathcal{A}}$ such that, for any setup with this topology,

$$p(\mathcal{X}, \mathcal{Y}) = \sum_{\mathcal{A}} p(\mathcal{A})\ p_{\mathcal{A}}(\mathcal{X})\ \tilde{p}_{\mathcal{A}}(\mathcal{Y})\,. \tag{2.1}$$

That is, "$\mathcal{X}$ and $\mathcal{Y}$ influence each other only through $\mathcal{A}$". This condition is obviously satisfied by continuum euclidean field theories whose lagrangian density contains a finite number of derivatives. With a suitable notion of "continuity" of the path γ, one can extend this definition to lattice actions. For instance, the Wilson action is also local in this sense, but note that $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ must be separated by more than one lattice spacing in order to realise the setup in the first place.

2.2 Hierarchical formula

As a consequence, for two operators $\mathcal{O}_x$ and $\mathcal{O}_y$, functionals of $\mathcal{X}$ and $\mathcal{Y}$ respectively, we have

$$\langle \mathcal{O}_x \mathcal{O}_y \rangle \equiv \sum_{\mathcal{C}} \mathcal{O}_x(\mathcal{C}) \mathcal{O}_y(\mathcal{C}) p(\mathcal{C}) = \sum_{\mathcal{A}} p(\mathcal{A})\ \langle \mathcal{O}_x \rangle_A\ \langle \mathcal{O}_y \rangle_A\,, \tag{2.2}$$

where

$$\begin{aligned} \langle \mathcal{O}_x \rangle_A &= \sum_{\mathcal{X}} p_A(\mathcal{X})\ \mathcal{O}_x(\mathcal{X}) \\ \langle \mathcal{O}_y \rangle_A &= \sum_{\mathcal{Y}} \tilde{p}_A(\mathcal{Y})\ \mathcal{O}_y(\mathcal{Y}) \end{aligned} \tag{2.3}$$

are the average values of the operators at a fixed value of $\mathcal{A}$. Thus the averaging process factorises into an average *at* fixed "boundary conditions" and an average *over* these boundary conditions. There are several ways in which this factorisation can be iterated: first, if the operator $\mathcal{O}_x \equiv \mathcal{O}_{x_1}\ \mathcal{O}_{x_2}$ itself factorises, the decomposition can be carried out also at this level, where p_A now plays the role of p. This means that the decomposition (2.2) allows us to treat the general n-point functions in the same way as the $n = 2$ case that we shall investigate in more detail: each factor can be averaged over separately.

There is another way the decomposition can be iterated: we can in turn write $\langle \mathcal{O}_x \rangle_A$ and $\langle \mathcal{O}_y \rangle_A$ as factorised averages over yet smaller subspaces, thus obtaining a nested expression for the correlation function. A three-level version of (2.2) would be

$$\begin{aligned}\langle \mathcal{O}_x \mathcal{O}_y \rangle = \sum_{\mathcal{A}} p(\mathcal{A}) \; \times \; &\sum_{\mathcal{A}_1} p_A(\mathcal{A}_1) \; \sum_{\mathcal{A}_2} p_{A_1}(\mathcal{A}_2) \langle \mathcal{O}_x \rangle_{\mathcal{A}_2} \; \times \\ \times &\sum_{\tilde{\mathcal{A}}_1} \tilde{p}_A(\tilde{\mathcal{A}}_1) \; \sum_{\tilde{\mathcal{A}}_2} \tilde{p}_{\tilde{A}_1}(\tilde{\mathcal{A}}_2) \langle \mathcal{O}_y \rangle_{\tilde{\mathcal{A}}_2} \, . \end{aligned} \tag{2.4}$$

This type of formula is the basis of our multilevel algorithm for the 2-point function.

2.3 Multilevel algorithm for the 2-point function

The hierarchical formula (2.4) is completely analogous to the expression derived in [4] in the case of the Polyakov loop, where it was also proved that it can be realised in a Monte-Carlo simulation by generating configurations in the usual way, then keeping the subset $\mathcal{A}$ fixed and updating the regions $\mathcal{X}$ and $\mathcal{Y}$. Suppose we do N updates of the boundary and n measurements of the operators for each of these updates. We are thus performing $N \cdot n$ updates.[2] But because the two sums in (2.2) are factorised, this in effect achieves Nn^2 measurements. As long as

- the latter are independent;
- that the fluctuations on the boundary $\Omega_{\mathcal{A}}$ have a much smaller influence than those occurring in $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$;
- that no phase transition occurs [4] due to the small volume and the boundary conditions;

error bars reduce with the computer time τ like $1/\tau$ rather than $1/\sqrt{\tau}$: to divide the error by two, we double n. The fluctuations of the boundary are only reduced in the usual $1/\sqrt{\tau}$ regime. Thus, for a fixed overall computer time, one should tune parameters of the multilevel algorithm so that the fluctuations in $\mathcal{X}$ and $\mathcal{Y}$ are reduced down to the level of those coming from $\mathcal{A}$.

The one-level setup we shall use in practise is illustrated in figure 2: $\Omega_{\mathcal{X}}$, $\Omega_{\mathcal{Y}}$ and $\Omega_{\mathcal{A}}$ are time-slices. An indication of how many "submeasurements" should be chosen at each level is given by the following consideration: if we measure an operator located in the middle of a time-block (of width Δt) bounded by $\mathcal{A}$, then any fluctuation occurring on $\mathcal{A}$ can be decomposed on the basis which diagonalises the hamiltonian of the theory. If the lowest-lying state compatible with the symmetry of the operator being measured has mass m_0 and $\Delta t > 1/m_0$, that state will act as the main 'carrier' of the fluctuation, so that it will induce a fluctuation of relative magnitude $e^{-m_0 \Delta t/2}$ on the time-slice where

[2] An update in general consists of several sweeps.

JHEP01(2003)048

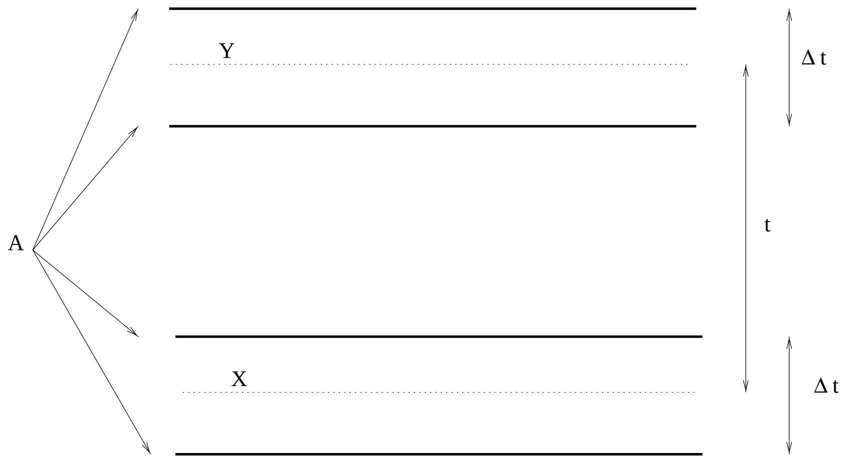

Figure 2: Our choice of $\Omega_{\mathcal{X}}$, $\Omega_{\mathcal{Y}}$ and $\Omega_{\mathcal{A}}$ to implement the hierarchical formula.

the operator is measured (see figure 1); indeed the propagation of fluctuations is damped exponentially in a system that develops a mass gap. Thus it is worth performing roughly $n \sim e^{m_0 \Delta t}$ measurements[3] under fixed boundary $\mathcal{A}$ in order to reduce the fluctuations coming from $\mathcal{X}$ down to the level of those of $\mathcal{A}$. In fact, this estimate is an upper bound, because the vacuum state under the fixed boundary conditions could lie at a higher energy level than the full-lattice vacuum. Finally, if zero modes are present, we expect a power law $n \propto (\Delta t)^\eta$.

At any rate, we need to optimise the parameters of the multilevel Monte-Carlo algorithm "experimentally", but we shall see in a practical example that this order of magnitude estimate is in qualitative agreement with the outcome of the optimisation procedure.

This simple argument also shows that the purpose of the multilevel scheme is to reduce fluctuations occurring at all separations (from the time-slice where the operator is measured) ranging from 0 to $\Delta t/2$ with an appropriate number of updates, in order to reduce their influence down to the level of the out-most boundary, which is then averaged over in the standard way.

2.4 Error reduction

Suppose the theory has a mass gap and that for a given t, the correlation function $C(t) \sim e^{-mt}$ is determined with equal amounts of computer time with the standard algorithm and the multilevel one. If we now want to compute $C(2t)$ with the same *relative* precision with the former, we need to increase the number of measurements by a factor e^{2mt}. With the multilevel algorithm however, we would e.g. introduce an extra level of nesting, so as to multiply the number of submeasurements by e^{mt}, as explained in the preceding section. Thus in this situation, the gain in computer time of this method is a factor e^{mt}; turned the other way, it achieves an error reduction $\propto e^{-mt/2}$ compared to the standard algorithm.[4] Since the quantity determining the error reduction is the product mt, the error reduction

[3]These may be nested, in which case it is the total number of measurements under fixed boundary $\mathcal{A}$ that is meant here.

[4]Note however that the computer effort is still increasing exponentially with t.

is in first approximation independent of β, as long as we measure the correlation function at fixed t in physical units. It would be inefficient to perform sweeps over time-blocks that are much thinner than the physical length scale: none of the three conditions highlighted in the preceding section is likely to be satisfied.

If we compare this to the error reduction analysis in [4], we note a "conservation of difficulty" law: in Wilson loop calculations, the exponent in the error reduction factor is the area of the loop, while here it is the time separation of the two operators.

3. The algorithm in practise

3.1 Preliminaries

We consider pure SU(3) lattice gauge theory in $(3+1)$ dimensions and we use the standard Wilson action

$$S[U] = \beta \sum_{x,\mu<\nu} \left(1 - \frac{1}{3} \operatorname{Re} \operatorname{Tr} P_{\mu\nu}\right), \tag{3.1}$$

where U are the link variables, $P_{\mu\nu}$ is the plaquette and $\beta = 6/g_0^2$.

We define zero-momentum operators as a sum over a time-slice of Wilson loop traces:

$$\mathcal{O}(t) = \sum_{\vec{x}} \operatorname{Tr} W(\vec{x}, t)\,. \tag{3.2}$$

We are interested in measuring correlation functions of these operators:

$$\langle \mathcal{O}^\dagger(0)\ \mathcal{O}(t)\rangle \equiv \frac{1}{\mathcal{Z}} \int D[U]\ \mathcal{O}^\dagger(0)\mathcal{O}(t)\ e^{-S[U]}, \qquad D[U] = \prod_{x,\mu} dU(x,\mu)\,. \tag{3.3}$$

The main physical information we extract from this numerical measurement is the mass of the lightest state compatible with the symmetry of $\mathcal{O}$, by fitting the exponential decay of the correlation function. Numerically the path integral is evaluated by updating the lattice configuration according to the Boltzmann weight and evaluating the operators at each of these updates. Here we calculate the correlation function of two operators carrying the continuum quantum numbers 0^{++} and 2^{++}, constructed with 4×2 rectangular Wilson loops. The simulation is done at $\beta = 5.70$ on an 8^4 lattice. For the update we use a mixture of heat-bath [6] and over-relaxation [7] in the ratio 1:4, both applied to three SU(2) subgroups [5] . Finally, as was noted in [4], the most elegant way to implement a multilevel Monte-Carlo program is to use a recursive function.

3.2 Averages under fixed boundary conditions

To illustrate our arguments on the way an error reduction is achieved, we measure the change of the average value of an operator carrying quantum numbers 2^{++} — whose expectation value in the full lattice is known to be exactly zero — due to fluctuations on the boundary and the finite number of measurements n performed:

$$G_n(\Delta t) = \sum_{\mathcal{A}} p(\mathcal{A})\ |\langle \mathcal{O}^{2^{++}}\rangle_{\mathcal{A}}^{(n)}|\,.$$

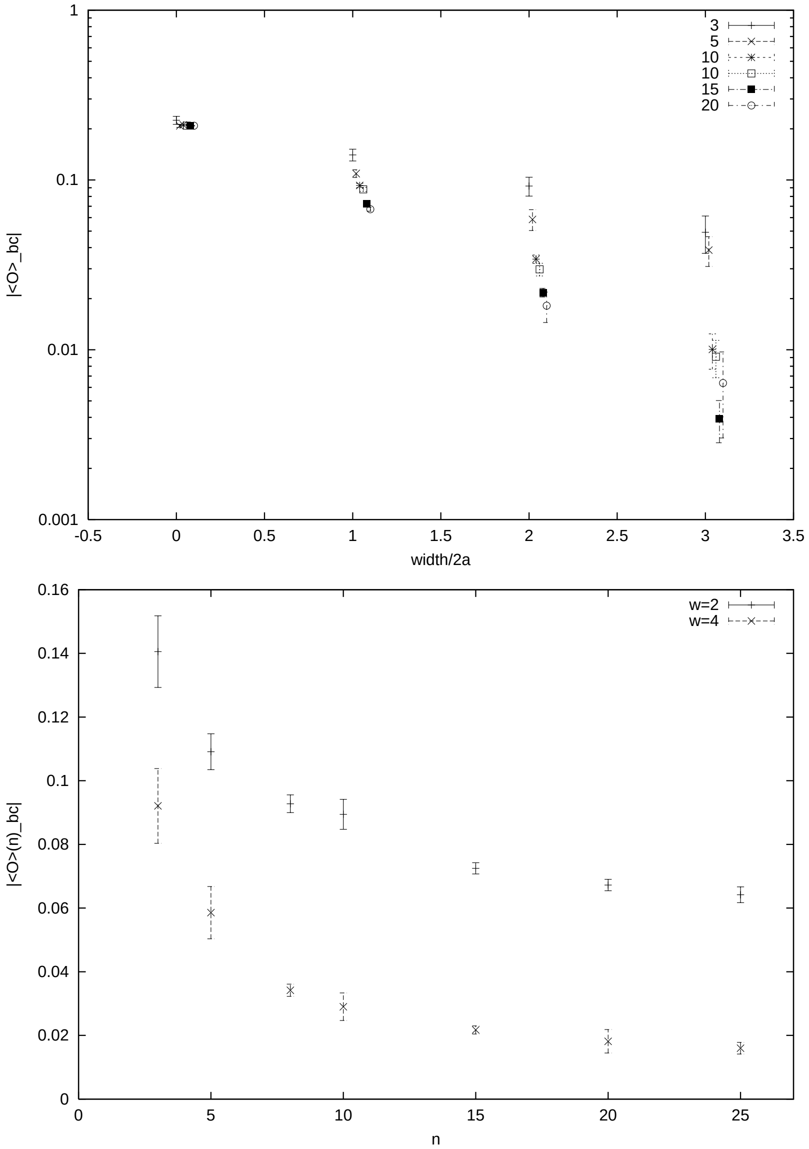

Figure 3: "Error" on the average value of the 2^{++} operator (4×2 rectangles), as a function of the width of the time-block (top) and the number of measurements (bottom).

Figure 3-top shows the dependence on the width Δt of G, for several fixed values of n. The data was obtained by nesting averages on increasing widths of the time-slices. Here, to reduce the number of parameters, we set n to be the same for all time-slices. Clearly, the data confirms that $G(\Delta t)$ vanishes exponentially, as anticipated (this *a posteriori* justifies the choice of keeping n constant). The exponent is a function of n, i.e. of how carefully the average is done: on the bottom plot, we see how fast a good estimate of the average value is obtained as a function of n, for the two fixed widths 2 and 4. We see that at $n = 20$, an accurate estimate of the average value has been reached. For this "true" average $G_{20}(\Delta t) \simeq G_\infty(\Delta t)$, our estimate of the effective mass of the decay $e^{-m_{\text{eff}}\Delta t/2}$ between $\Delta t = 2$ and $\Delta t = 4$ is 1.22(2), and 1.77(20) between $\Delta t = 4$ and $\Delta t = 6$.

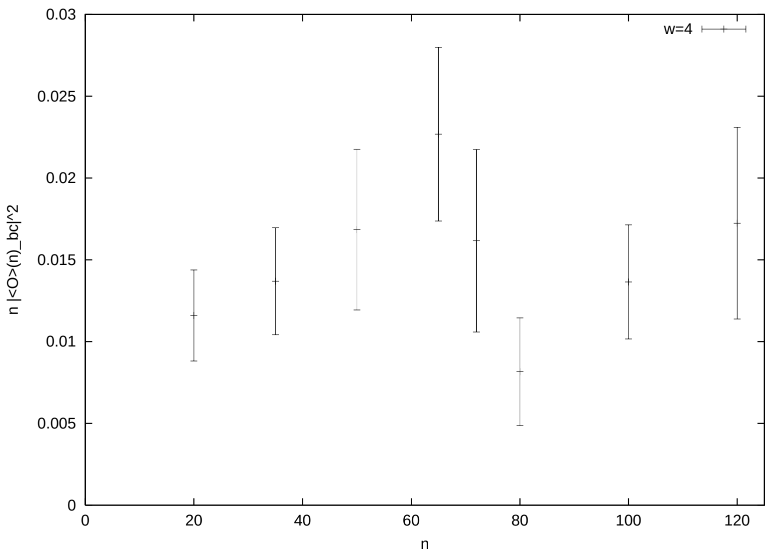

Figure 4: Optimisation curve for the width 4 level.

This is compatible with the mass of the lightest 2^{++} glueball we shall calculate below. Thus it seems that, at large widths, $G_\infty(\Delta t)$ does decay with a mass close to the lightest state sharing the symmetry of our operator. The question now is, when has a good enough estimate of this quantity been achieved so that statistical errors are maximally reduced?

3.3 Optimisation procedure

Since errors are reduced with the number of measurements as $1/\sqrt{n}$ at large n, we plot $(G_n(\Delta t))^2 \times n$ and, at fixed Δt, choose n to minimise this quantity. Eventually this quantity must increase, because the "error" G tends to a constant at large n.

We have already seen (figure 3-right) that for the smallest width, the average is reached very smoothly. As a consequence, the optimal n_2 according to our criterion is small: n_2=5. If we now fix this parameter at that value and move to the width 4 block, figure 4 reveals that the corresponding optimisation curve reaches its minimum between $n_4 = 70$ and 90. Indeed, G reaches its asymptotic value around these values of n_4. The data was obtained as an average over 30 boundary conditions. Between submeasurements, 5 update sweeps are performed. The error reduction achieved with slightly different parameters below shows that the success of the algorithm is not all too dependent on a fine-tuning of parameters.

Finally, we note that to optimise the parameters for a 0^{++} operator, whose expectation value in the full lattice does not vanish, we need to minimise the fluctuation around that value (which has to be known in advance).

3.4 Results

We compute the 0^{++} and 2^{++} correlation functions formed with a 4×2 rectangular Wilson loop at 4 lattice spacings. We use two smearing steps on the operator. Errors are estimated with a standard jackknife analysis using 26 bins. A two-level scheme is implemented: the

t/a	$\langle \mathcal{O}_0^\dagger(0)\ \mathcal{O}_0(t)\rangle$	$am_{\rm eff}^{(0)}(t)$	$\langle \mathcal{O}_2^\dagger(0)\ \mathcal{O}_2(t)\rangle$	$am_{\rm eff}^{(2)}(t)$
0	1.0000(65)		1.0000(14)	
1		1.017(35)		2.151(75)
2	0.1331(99)		$1.36(20) \times 10^{-2}$	
3		0.929(49)		1.794(74)
4	0.0406(39)		$7.49(70) \times 10^{-4}$	

Table 1: The two correlation functions and the corresponding local effective masses.

8 time-slices are split into 2 time-blocks of width 4, which are in turn decomposed into 2 time-blocks of width 2. We restrict ourselves to the measurement of the correlation function at even time-separations.

For the 0^{++} correlation function at 4 lattice spacings, one "measurement" comprises 10 submeasurements at the lower level, 40 at the upper level. When performing the latter, we are free to compute the 0 and 2 lattice spacing correlation in the standard way (thus the error reduction is only applied to the $t = 4a$ correlation). We collected 260 of these compound measurements.

We proceed similarly in the 2^{++} case with following parameters: one "measurement" comprises 8 submeasurements at the lower level, 150 at the upper level, each of these being preceded by 5 sweeps; our program needs about 8.3 minutes on a standard alpha machine to perform one of these compound measurements. We collected 520 of them; because they are time-consuming, we perform $\sim$ 200 sweeps between them to reduce their statistical dependence.

The values in table 1 for the correlation functions, as well as their corresponding local effective masses (extracted from a cosh fit), were obtained. Figure 5 shows a plot of the two correlation functions. The small error bars appear to be roughly constant on the logarithmic scale. The $t = 4a$ correlation of the 2^{++} operator shows that a factor 20 error reduction has been achieved with respect to the $t = 0$ point. It is already at an accuracy that would be inaccessible on current single-processor machines with the standard algorithm. Indeed the latter yields error bars that are roughly independent of t and would have given an error comparable to our error on the $t = 0$ data, where we do not achieve any error reduction. The naive error-reduction estimate of section 2.4 evaluates to $\exp(1.8 \times 4/2) \simeq 36$. Not unexpectedly, the observed reduction factor is somewhat lower than this naive estimate, presumably because the configurations generated at fixed boundary conditions are quite strongly correlated.

4. Conclusion

In this paper we have proposed a general scheme in which the accuracy of numerically computed n-point functions in local field theories could be improved by the use of nested averages. While the procedure is known to be exact, the efficiency of the algorithm must ultimately be tested on a case-by-case basis. However, a simple application to SU(3) Wilson loop correlations showed that, in some cases at least, the multilevel algorithm

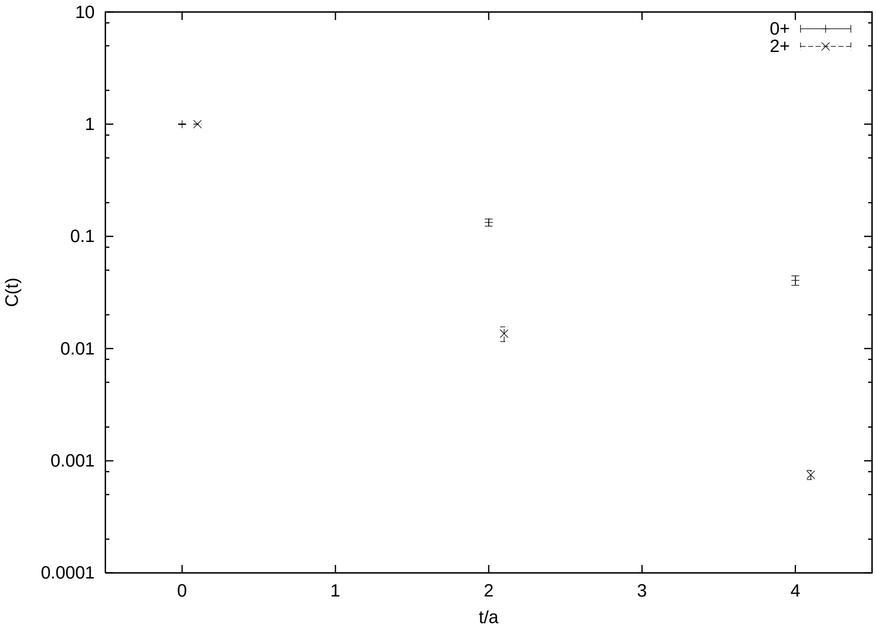

Figure 5: The 0^{++} and 2^{++} correlation functions.

drastically reduces statistical errors. It can straightforwardly be used in conjunction with the smearing and blocking techniques. A further nice feature of the 2-point function case is that previous knowledge of the low-energy spectrum provides useful guidance in the tuning of the algorithm's many parameters. Also, the error reduction achieved was roughly as anticipated. We intend to use multilevel algorithms to extract the higher-spin spectrum [8] of the SU(3) gauge theory, where the higher masses involved and the successful use of the variational method [9] require a high level of accuracy on the correlation functions.

Independently of statistical errors, it is much harder to determine quantities extracted from n-point functions with $n \geq 3$; if these difficulties are overcome, the multilevel scheme could prove a decisive asset in those numerical measurements.

Acknowledgments

The author thanks M. Teper for helpful discussions and M. Lüscher for sharing his experience on nested Monte-Carlo algorithms. He is also grateful to Lincoln College for its hospitality and the Berrow Trust for financial support.

References

[1] APE collaboration, M. Albanese et al., *Glueball masses and string tension in lattice QCD*, *Phys. Lett.* **B 192** (1987) 163.

[2] M. Teper, *An improved method for lattice glueball calculations*, *Phys. Lett.* **B 183** (1987) 345.

[3] G. Parisi, R. Petronzio and F. Rapuano, *A measurement of the string tension near the continuum limit*, *Phys. Lett.* **B 128** (1983) 418.

[4] M. Lüscher and P. Weisz, *Locality and exponential error reduction in numerical lattice gauge theory*, *J. High Energy Phys.* **09** (2001) 010 [hep-lat/0108014].

[5] N. Cabibbo and E. Marinari, *A new method for updating* SU(N) *matrices in computer simulations of gauge theories*, *Phys. Lett.* **B 119** (1982) 387.

[6] K. Fabricius and O. Haan, *Heat bath method for the twisted Eguchi-Kawai model*, *Phys. Lett.* **B 143** (1984) 459;
A.D. Kennedy and B.J. Pendleton, *Improved heat bath method for Monte Carlo calculations in lattice gauge theories*, *Phys. Lett.* **B 156** (1985) 393.

[7] S.L. Adler, *An overrelaxation method for the Monte Carlo evaluation of the partition function for multiquadratic actions*, *Phys. Rev.* **D 23** (1981) 2901.

[8] H.B. Meyer and M.J. Teper, *High spin glueballs from the lattice*, hep-lat/0212026.

[9] M. Lüscher and U. Wolff, *How to calculate the elastic scattering matrix in two-dimensional quantum field theories by numerical simulation*, *Nucl. Phys.* **B 339** (1990) 222.

JHEP01(2003)048