MPI–PhT/99–??
BUTP–99/??

Generalized eigenvalue problem for the correlation matrix

Ferenc Niedermayer[1]

Institute for Theoretical Physics
University of Bern
Sidlerstrasse 5, CH-3012 Bern, Switzerland

Peter Weisz

Max-Planck-Institut für Physik
Föhringer Ring 6, D-80805 München, Germany

February 2000

Abstract

We analise the benefits and pitfalls of using the generalised eigenvalue equation to obtain the spectrum of the system from the correlation matrix. We also suggest some improvements. ...???

[1]On leave from the Institute of Theoretical Physics, Eötvös University, Budapest

1 Introduction

In lattice theory information on energy spectra is obtained by measuring the correlation functions of operators separated in Euclidean time. It is well known (and intuitively obvious) that one can determine the spectrum of intermediate states more precisely by measuring the correlation matrix $C(t)$ of several operators (with the same quantum numbers) instead of just the correlation function of a single operator.

An efficient way to extract the spectral information encoded in the correlation matrix is to consider the generalized eigenvalue equation,

$$C(t)v_n = \lambda_n(t, y_0)C(t_0)v_n\,.$$

This equation for the correlation matrix has been proposed first by Michael [1] who obtained it as a result of a variational approach, and applied it to determine the static potential in lattice gauge theory. Lüscher and Wolff [2] arrived at the same equation from another starting point. They applied this method to evaluate the 2-particle phase shifts in the 2d O(3) non-linear sigma-model. This required a very precise determination of the 2-particle energy levels in a periodic box to obtain a reasonable error for the phase shifts.

The generalized eigenvalue equation is especially valuable when there are nearly degenerate excited states with the same quantum numbers. This situation is met not only in the determination of the 2-particle energies mentioned above but also in many phenomena of physical interest e.g. the static potential in pure gauge theory, heavy $Q\overline{Q}$ mesons, etc. The method is now used widely: for the glueball spectrum [3], to observe the string breaking by adjoint sources [4], and in many other works. Nevertheless, a systematic study of the method seems to be lacking.

The purpose of this paper is to discuss the generalized eigenvalue equation in more detail, and correct some misconceptions concerning its properties which have appeared in the literature. Furthermore, we show how to check self consistency of the assumptions made, and suggest further improvements.

We shall illustrate the method on the same physical problem considered in ref. [2] and [5], measuring the energy levels of the 2-particle states in the 2d O(3) sigma model. (In the latter work some of the suggestions presented here were used.) Our aim here is only to illustrate different modifications of the method, investigate possible systematic errors in the procedure and compare it with other methods. For this purpose we shall restrict the investigation to the $I = 0$ isospin channel, one single coupling and one lattice size.

2 The generalized eigenvalue problem

Consider the general case of a correlation matrix of some operators $O_i,\ i = 1, \ldots, N$ on an infinite lattice in the Euclidean time direction

$$\langle O_i(0) O_j(t) \rangle = C_{ij}(t) = \sum_{n=1}^{\infty} \mathrm{e}^{-E_n t} \psi_{ni} \psi_{nj} \,, \tag{1}$$

where

$$(\psi_n)_i \equiv \psi_{ni} = \langle \mathrm{vac} | O_i | n \rangle \,, \tag{2}$$

are the 'wave functions' of the state $|n\rangle$. For notational simplicity we consider only real observables, hence $C(t)$ is real symmetric. The states are labeled with increasing energy, $E_1 \leq E_2 \leq \ldots$.

In eq. (1) the sum extends over all eigenstates of the transfer matrix. However, contribution of states with $n > N$ will be small for $t \geq t_0$ when t_0 is sufficiently large for the given set of N operators.[2] Following ref. [2] consider first the contribution of the first N terms only and include the corrections later. Let

$$C^{(0)}_{ij}(t) = \sum_{n=1}^{N} \mathrm{e}^{-E_n t} \psi_{ni} \psi_{nj} \,, \tag{3}$$

The set of operators O_i should be obviously such that they distinguish between the lowest states of interest. It will thus be assumed

[2] N can be taken to be quite large: in the illustrative example below we took $N = L/2 = 64$.

that the vectors ψ_n, $n = 1, \ldots, N$ are linearly independent. They are, however, not orthogonal to each other in general. As a consequence the eigenvalues and eigenvectors of $C^{(0)}(t)$ are not given by $c_n \exp(-E_n t)$ and ψ_n. However, as shown in the lemma of ref. [2], for large t the eigenvalues approach this value:

$$\lambda_n \left[C^{(0)}(t)\right] = c_n \mathrm{e}^{-E_n t} \left(1 + \mathrm{O}\left(\mathrm{e}^{-\Delta E_n t}\right)\right) , \tag{4}$$

where ΔE_n is the distance of E_n from the nearest energy level,

$$\Delta E_n = \min_k |E_k - E_n| . \tag{5}$$

In the problematic cases where ΔE_n is small this slow decay can cause a systematic error, an apparent shift in E_n. For example, the determination of the static quark potential in pure QCD by measuring a Wilson-loop could be hampered by such systematic errors [6].

Instead of studying the eigenvalues of the correlation matrix for large t it is useful to consider the generalized eigenvalue problem [1, 2]

$$C(t) v_n = \lambda_n(t, t_0) C(t_0) v_n . \tag{6}$$

The 0^{th} approximation to this is

$$C^{(0)}(t) v_n^{(0)} = \lambda_n^{(0)}(t, t_0) C^{(0)}(t_0) v_n^{(0)} . \tag{7}$$

Consider the set $\{v_n^{(0)}\}$ $(n = 1, \ldots, N)$ dual to the set of the first N wave functions $\{\psi_n\}$, i.e.

$$\left(v_n^{(0)}, \psi_{n'}\right) = \delta_{nn'} , \quad \text{for } n, n' \leq N . \tag{8}$$

One has then from eq. (3)

$$C^{(0)}(t) v_n^{(0)} = \mathrm{e}^{-E_n t} \psi_n . \tag{9}$$

From here it follows that $v_n^{(0)}$ satisfies eq. (7) with

$$\lambda_n^{(0)} = \mathrm{e}^{-E_n (t - t_0)} . \tag{10}$$

The eigenvectors $v_n^{(0)}$ are orthogonal to each other with the weight $C^{(0)}(t)$ (for any t):

$$(v_{n'}^{(0)}, C^{(0)}(t) v_n^{(0)}) = \delta_{nn'} \mathrm{e}^{-E_n t} . \tag{11}$$

Note that the generalized eigenvalue problem gives both the energy spectrum and the wave functions ψ_n of the intermediate states.

Provided the perturbations to $C^{(0)}(t)$ are sufficiently small the quantity

$$E_n^{\rm eff}(t,t_0) = \log \frac{\lambda_n(t,t_0)}{\lambda_n(t+1,t_0)} \tag{12}$$

suggested in ref. [2] should converge rapidly to E_n with increasing t. The plateau of this 'effective mass plot' serves as a good estimator for E_n.

3 Perturbative expansion

Here we calculate the perturbative expansion for the generalized eigenvalue equation. For the actual case there are two different perturbations to $C^{(0)}(t)$:

1. the contributions from the $n > N$ states,
2. statistical noise.

First we consider the general case of the equation

$$Av_n = \lambda_n B v_n \,, \tag{13}$$

with perturbations (considered to be first order)

$$A = A^{(0)} + A^{(1)} \,, \qquad B = B^{(0)} + B^{(1)} \,. \tag{14}$$

The corresponding expansions for λ_n and v_n are:

$$\lambda_n = \lambda_n^{(0)} + \lambda_n^{(1)} + \lambda_n^{(2)} + \dots \,, \qquad v_n = v_n^{(0)} + v_n^{(1)} + \dots \,. \tag{15}$$

The eigenvectors of the $0^{\rm th}$ order equation are assumed to be normalized by the condition

$$\left(v_n^{(0)}, B^{(0)} v_{n'}^{(0)}\right) = \delta_{nn'} \,, \tag{16}$$

and without loss of generality $v_n^{(1)}$ is such that $(v_n^{(1)}, B^{(0)} v_n^{(0)}) = 0$. Introducing the notation

$$\Delta_n^{(1)} \equiv A^{(1)} - \lambda_n^{(0)} B^{(1)} \tag{17}$$

one has in first order

$$\lambda_n^{(1)} = \left(v_n^{(0)}, \Delta_n^{(1)} v_n^{(0)}\right) , \tag{18}$$

and

$$v_n^{(1)} = \sum_{k \neq n} v_k^{(0)} \frac{\left(v_k^{(0)}, \Delta_n^{(1)} v_n^{(0)}\right)}{\lambda_n^{(0)} - \lambda_k^{(0)}} . \tag{19}$$

The 2^{nd} order correction to the eigenvalue is given by

$$\lambda_n^{(2)} = \sum_{k \neq n} \frac{\left(v_k^{(0)}, \Delta_n^{(1)} v_n^{(0)}\right)^2}{\lambda_n^{(0)} - \lambda_k^{(0)}} - \left(v_n^{(0)}, \Delta_n^{(1)} v_n^{(0)}\right) \left(v_n^{(0)}, B^{(1)} v_n^{(0)}\right) . \tag{20}$$

Now we apply these expressions to the case of interest. As the leading term we take $A^{(0)} = C^{(0)}(t)$, $B^{(0)} = C^{(0)}(t_0)$. In this case $\lambda_n^{(0)}$ and $v_n^{(0)}$ are the exact quantities, while the actual quantities λ_n and v_n obtained solving eq. (6) are distorted by the two types of unwanted perturbations. Instead of (16) it is more convenient to use the normalization condition (8) which is independent on t_0. This involves a trivial rescaling of $v_n^{(0)}$ in the expressions above.

The final results for the perturbation due to states $l > N$ is given by

$$\lambda_n^{(0)} = \mathrm{e}^{-E_n (t - t_0)} , \tag{21}$$

$$\frac{\lambda_n^{(1)}}{\lambda_n^{(0)}} = - \sum_{l > N} \mathrm{e}^{-(E_l - E_n) t_0} \left(1 - \mathrm{e}^{-(E_l - E_n)(t - t_0)}\right) (v_n^{(0)}, \psi_l)^2 , \tag{22}$$

and

$$\frac{\lambda_n^{(2)}}{\lambda_n^{(0)}} = \sum_{k\neq n} \frac{1}{1-\mathrm{e}^{-(E_k-E_n)(t-t_0)}} \left\{ \sum_{l>N} \mathrm{e}^{-(E_l-\frac{1}{2}E_n-\frac{1}{2}E_k)t_0} \right.$$
$$\left. \times \left(1-\mathrm{e}^{-(E_l-E_n)(t-t_0)}\right)(v_n^{(0)},\psi_l)(\psi_l,v_k^{(0)}) \right\}^2$$
$$+ \sum_{l>N} \mathrm{e}^{-(E_l-E_n)t_0} \left(1-\mathrm{e}^{-(E_l-E_n)(t-t_0)}\right)(v_n^{(0)},\psi_l)^2$$
$$\times \sum_{l'>N} \mathrm{e}^{-(E_{l'}-E_n)t_0}(v_n^{(0)},\psi_{l'})^2 . \quad (23)$$

(The summation over k is restricted of course to $k \leq N$.) Note that $\lambda_n^{(1)} < 0$ for all n and $\lambda_1^{(2)} > 0$ for these perturbations.

To first order we have

$$\lambda_n^{(0)} + \lambda_n^{(1)} = \mathrm{e}^{-E_n(t-t_0)} c_n \left[1 + \mathrm{O}\left(\mathrm{e}^{-(E_{N+1}-E_n)(t-t_0)}\right)\right] , \quad (24)$$

where c_n depends on t_0 (but not on t):

$$c_n = 1 + \mathrm{O}\left(\mathrm{e}^{-(E_{N+1}-E_n)t_0}\right) . \quad (25)$$

Hence both corrections are suppressed by exponential factors containing the large energy difference $E_{N+1} - E_n$ (as opposed to ΔE_n). However, this is true only in the first order approximation. In the second order the factor $1/[1-\exp(-(E_k-E_n)(t-t_0))]$ in eq. (23) introduces an unwanted slow dependence. For the ground state, e.g. this correction gives

$$\frac{\lambda_1^{(2)}(t,t_0)}{\lambda_1^{(0)}(t,t_0)} = c_1(t_0) + c_2(t_0)\mathrm{e}^{-(E_2-E_1)(t-t_0)} + \ldots \quad (26)$$

The behaviour (24) has, in some papers (e.g. [4]), been claimed as being generally valid for perturbations $l > N$. As we see, unfortunately, this statement is valid only in the first order.

The slowly decreasing terms come from the corrections to $C^{(0)}(t_0)$: states with $l > N$ in $C(t)$ produce fast decaying corrections only.

From this observation it is quite simple to see that slowly decaying corrections should be present in $\lambda_n(t,t_0)$ in the general case. Suppose one neglects deformations of $C(t)$ (due to $l > N$ states) but keeps them in $C(t_0)$. The latter could be arbitrary, in particular $C(t_0)$ could be deformed to a constant times the unit matrix. In this case $\lambda_n(t,t_0)$ becomes an ordinary eigenvalue of $C(t)$, hence according to lemma of ref. [2], eq. (4), such slowly decreasing corrections should be present for a general deformation of $C(t_0)$.

We want to stress, however, that in spite of this negative result the method of generalized eigenvalue problem is a significant improvement over considering the ordinary eigenvalues. The slowly decaying corrections $\exp(-\Delta E_n t)$ appear already in the eigenvalues of $C^{(0)}(t)$ due to the non-orthogonality of the wave functions ψ_n. The generalised eigenvalue problem for $C^{(0)}(t)$, $C^{(0)}(t_0)$ avoids this – the eigenvalues $\lambda_n^{(0)}(t,t_0)$ give the exact result, eq. (10). Moreover, perturbations to $C^{(0)}(t_0)$ in first order produce fast decaying corrections – the slowly decaying ones appear only in second order. Furthermore, these corrections can be controlled by choosing appropriate values for t_0 and N (although this is not without additional problems, as will be discussed later).

For the perturbations coming from the statistical noise one gets similar effects. To first order they do not shift the expectation values, but in second order there is a systematic shift and also the same slowly decaying factor. Under some natural assumptions about the correlations one expects that for the ground state $\lambda_1^{(2)} > 0$. One has to be careful also here: the presence of small eigenvalues in $C(t_0)$ enhances the effect of statistical noise and it is advisable to study whether one is still in the linear regime. This could be done e.g. by an artificial rescaling of the fluctuations.

NEW:

The presence of small eigenvalues of $C(t_0)$ leads to different kinds of instabilities. It can happen that the method finds a wave function with very small (and fake) wave function ψ_n with some arbitrary energy which accidentally can be the smallest one. Because the contribution of such a term is suppressed by $|\psi_n|^2$ it can easily arise

from small statistical fluctuations. It can appear even when the lowest energies are still stable, but an automatic ordering of the energy levels will misinterpret them. Another type of instability is the split of the levels: One intermediate state can be split into two states with nearly equal wave functions and energies. We stress here again that it can happen only if $C(t_0)$ has nearly zero eigenvalues. To avoid such artifacts it is advisable to monitor the eigenvalues of $C(t_0)$, the wave functions ψ_n and the dual vectors v_n.

TOY EXAMPLE SHORTLY ???

4 Improvements

A proper measure of how well a set of energies E_n and the wave functions ψ_{ni} describe the correlation matrix at a given t is

$$\chi^2(t) = \sum_{ij} w_{ij}(t) \left(C_{ij}(t) - \sum_{n=1}^{N} \mathrm{e}^{-E_n t} \psi_{ni} \psi_{nj} \right)^2 , \qquad (27)$$

where $C_{ij}(t)$ is the correlation matrix measured in the Monte Carlo experiment and $1/w_{ij}(t)$ is its dispersion. An overall measure is then $\chi^2 = \sum_t \chi^2(t)$.

It is desirable to check the calculations by using several reference values t_0 and number of operators N. By increasing t_0 or N one meets a difficulty, however. The solutions to the generalized eigenvalue problem, eq. (6) are very sensitive to the small eigenvalues of $C(t_0)$, since the procedure of solving it involves inversion of $C(t_0)$. (Observe that eq. (7) can be transformed to an ordinary eigenvalue problem for the matrix $C(t_0)^{-1/2} C(t) C(t_0)^{-1/2}$.) If the small eigenvalues are unstable against the statistical noise the extracted energy values will be distorted and unstable. Increasing t_0 (for a given N) could lead to such instability because the signal decreases with t faster than the noise.[3] Of course, whether a given t_0 and N is satisfactory depends on the choice of the observables O_i.

[3] In ref. [2] only $t_0 \leq 1$ has been considered to avoid this instability.

How can it happen that by increasing N, i.e. by increasing the information obtained about the system, the results deteriorate? The reason can obviously only lie in the improper way to extract the results. We propose the following procedure to avoid the possible instability. For a given t_0 calculate the eigenvalues of $C(t_0)$, preferably together with their statistical errors.[4] Take then the M largest eigenvalues ($M < N$) and the corresponding eigenvectors φ_k:

$$C(t_0)\varphi_k = \gamma_k \varphi_k \,, \;\; k = 1 \ldots M \,. \tag{28}$$

(Measuring the statistical errors of the eigenvalues could give a hint how large M can be chosen, but this is not really necessary.) We restrict the correlation matrix to this subspace:

$$\overline{C}_{kl}(t) = (\varphi_k, C(t)\varphi_l) \equiv \sum_{i,j=1}^{N} \varphi_{ki} C_{ij}(t) \varphi_{lj}, \;\; k, l = 1 \ldots M \,. \tag{29}$$

Repeating the previous considerations for the corresponding $M \times M$ matrices one obtains the wave functions $\overline{\psi}_k$ and the dual vectors $\overline{v}_k$ of the truncated problem. Since the smallest eigenvalues of $C(t_0)$ (i.e. $M < k \leq N$) are excluded one avoids the instability.

It is also useful to introduce $\overline{\chi}^2(t)$ analogously to eq. (27) in the truncated basis. This is expected to be a more sensitive characteristic of the fit than $\chi^2(t)$.

Finally (if needed) one obtains the first M wave functions in the original basis

$$(\psi_k)_i = \sum_{l=1}^{M} \overline{\psi}_{kl} \varphi_{li} \,, \quad i = 1 \ldots N \,, \tag{30}$$

and similarly the M dual vectors v_k. From the orthogonality of vectors φ_l it follows that $(v_k, \psi_{k'}) \propto \delta_{kk'}$ for $k, k' = 1, \ldots, M$, so the orthogonality still holds for the first M vectors.

Analysing MC data it is useful to consider several values of t_0, N and M, calculating the spectrum by different methods and monitoring the χ^2 values.

[4] Alternatively, one can use the eigenvectors of $C(t_1)$ for some $t_1 < t_0$.

For larger correlation length it is possible to increase the signal/noise ratio by a further modification of the generalized eigenvalue problem. We propose adding several equations (6) for different pairs of t, t_0 with appropriate weights having a factorization property. We replace eq. (6) by

$$K(\tau)v_n = \alpha_n(\tau)K_0 v_n\,, \tag{31}$$

where

$$K_0 = \sum_t w_0(t)C(t)\,, \tag{32}$$

$$K(\tau) = \sum_{t,\Delta} w_0(t)w_1(\Delta)C(t+\Delta+\tau)\,, \tag{33}$$

and

$$\alpha_n(\tau) = \sum_{\Delta} w_1(\Delta)\mathrm{e}^{-E_n(\Delta+\tau)}\,. \tag{34}$$

Here $w_0(t)$ and $w_1(\Delta)$ are some positive weights. E_n can be obtained easily from $\alpha_n(\tau)$. Alternatively, one can define again an 'effective mass',

$$\log\frac{\alpha_n(\tau)}{\alpha_n(\tau+1)}\,. \tag{35}$$

With increasing τ this effective mass should approach E_n as well .

Although it is not denoted here, it is understood that truncation to the M-dimensional subspace of the stable eigenvectors of K_0 (or of $C(t_0)$) is performed, if necessary.

The optimal choice of the weights $w_0(t)$ and $w_1(\Delta)$ depends on many things. We did not investigate this choice even in the illustrative example given later, but a few general points could be suggested, nevertheless. One should perhaps avoid the appearance of the same correlation matrix in K and K_0. The optimal number of terms in K or K_0 is a result of a compromise – including more terms is useful only if this increases the signal/noise ratio. Hence too large t's should not be included.

The results for the energy spectrum should be independent of the various choices in some range. In the regions where the results are inconsistent, one should see this also by checking the corresponding χ^2 values.

A POSSIBLE STRATEGY:

Define the dual vectors in some optimal way, by the procedure described above. Using these *fixed* dual vectors determine $(v_n, C(t)v_{n'})$. For $n \neq n'$ it should be consistent with zero (apart from the noise and the contribution of states with $l > N$), and the diagonal terms are used to build an effective mass plot to obtain E_n. This procedure will avoid the uncertainty of using different set of dual vectors for different pairs of (t, t_0) and expected to be more robust.

AN OTHER ONE:

Use the χ^2 in eq. (27) (summed over some range in t) in fitting $C(t)$ with the $N(N+1)$ parameters E_n, ψ_{ni}. The number of independent terms is $n_t N(N+1)/2$ where n_t is the number of time slices used in χ^2. For $n_t = 2$ this minimization problem is equivalent to the generalized eigenvalue problem and gives exactly $\chi^2 = 0$ (provided $C(t)$ is positive definite for both values of t considered). By taking $n_t > 2$ the minimization problem of χ^2 cannot be worse than for the $n_t = 2$ case, i.e. the generalized eigenvalue problem. Obviously, too large values of t, dominated by the noise have to be excluded. Similarly, for too small values of t (especially for $t = 0$) the sates with $n > N$ (or $n > M$ for the truncated case) could have a significant contribution, and have to be excluded as well. Note that this method has the same instability as the generalized eigenvalue equation, so it must be truncated first to the stable subspace of $C(t_0)$ (or by some other way) and used with care. Of course, the nonlinear minimization is very time consuming, but taking the solution of the generalized eigenvalue equation as a starting value considerably simplifies the procedure. ONE SHOULD TRY OUT THIS METHOD AGAIN!

VISUALISATION: plot the whole set of energies E_n for each the jack-knife sample, and for different values of the parameters used (M, t_0, t, etc.), together with the value of χ^2. An instability (fake

eigenvalue, splitting, etc.) could be easily seen this way.

5 An illustrative example

Here we apply the method of the generalized eigenvalue problem, together with some of its proposed modifications to the case of the 2d O(3) non-linear sigma-model. We calculate the 2-particle energies in the $I = 0$ isospin channel in a periodic box of size L. If L is large enough the energy difference between the 2-particle states with different momenta is small, $\mathrm{O}(1/(mL^2))$ compared to the lowest energy $E_1 = \mathrm{O}(2m)$. With a single operator it would be very difficult to distinguish these states, and large systematic errors are expected. For the given problem one has to determine the energy levels E_n to a high precision since the value of the phase shift is very sensitive to E_n. To see this we remind that the procedure suggested in ref. [2] is to determine the momentum p_n from $E_n = 2\sqrt{m^2 + p_n^2}$ and then the phase shift $\delta(p_n)$ from the periodicity condition $p_n L + 2\delta(p_n) = 2\pi n$. It is clear that the error on E_n is strongly enhanced in $\delta(p_n)$.

Note, however, that this example has also a property which is not typical in MC simulations: the wave functions ψ_n in this problem (with the definitions of the operators O_i given below) are nearly orthogonal to each other. Therefore a simple diagonalization of $C(t)$ leads to eigenvalues close to const $\times \exp(-E_n t)$. We shall also investigate this naive method.

For the illustration of the procedure discussed here we used MC results for the O(3) model with the standard action at $\beta = 1.40$ on a 128×256 lattice (the longer size being the 'time' direction) from ref. [5]. The correlation length, $\xi = 1/m$ is 6.883(3) at this coupling. We made 120 runs with 10k measurements in each run.

The observables used here are

$$O(x,t) = \frac{1}{L} \sum_{z=0}^{L-1} S^a(z,t) S^a(z+x,t)\,, \qquad x = 0..L/2\,. \tag{36}$$

These operators connect the vacuum with states containing even

number of particles in a total $I = 0$ isospin channel and with zero total momentum. The lowest lying such states are the 2-particle scattering states characterized by momenta $p_n, -p_n$. Actually, we consider only the connected part, i.e.

$$C_{xy}(t) = \langle O(x,0)O(y,t)\rangle - \langle O(x,0)\rangle\langle O(y,0)\rangle\,. \qquad (37)$$

Since $O(0,t) = 1$ the vacuum contribution can be subtracted easily from the full correlation matrix measured in MC. In the connected correlation matrix we have therefore $N = L/2$ observables, for $x = 1, \ldots, L/2$.

In table 1 the prediction for the first three energy levels of the corresponding 2-particle states are given. They are calculated from the analytic result of the S-matrix [7] (for the $I = 0$ channel)

$$\mathrm{e}^{2i\delta_0(p)} = \frac{\theta + 2i\pi}{\theta - 2i\pi}\ , \qquad (38)$$

where θ is the rapidity, $p = m\sinh\theta$, with the measured 1-particle mass. The lattice results which came out from the analysis [5] are also given. The fact that they are in agreement with the analytic results within the errors is not important in this study – we provide these numbers only for comparison, to be able to see the possible systematic errors in the methods considered below.

k	exact	lattice
1	0.294970	0.2956(5)
2	0.328017	0.3281(6)
3	0.385342	0.3859(9)

Table 1: The analytic result for the three lowest energy levels of 2-particle states in the $I = 0$ channel for $L = 128$ and $1/m = 6.883(3)$ ($\beta = 1.40$). These and the best lattice results are given to be able to judge the presence of systematic errors of results obtained by different methods.

As the simplest method to evaluate the lowest energy value E_1 consider the correlation function $C_{AA}(t) = \langle A(0)A(t)\rangle$ of a single operator $A(t) = \sum_x O(x,t)$. Fig. 1 shows the 'effective mass plot', i.e. $\log(C_{AA}(t)/C_{AA}(t+1))$. The plot shows that the result obtained in this way is significantly above the 'true' result shown here (and in other plots) for comparison. Of course, this systematic error is

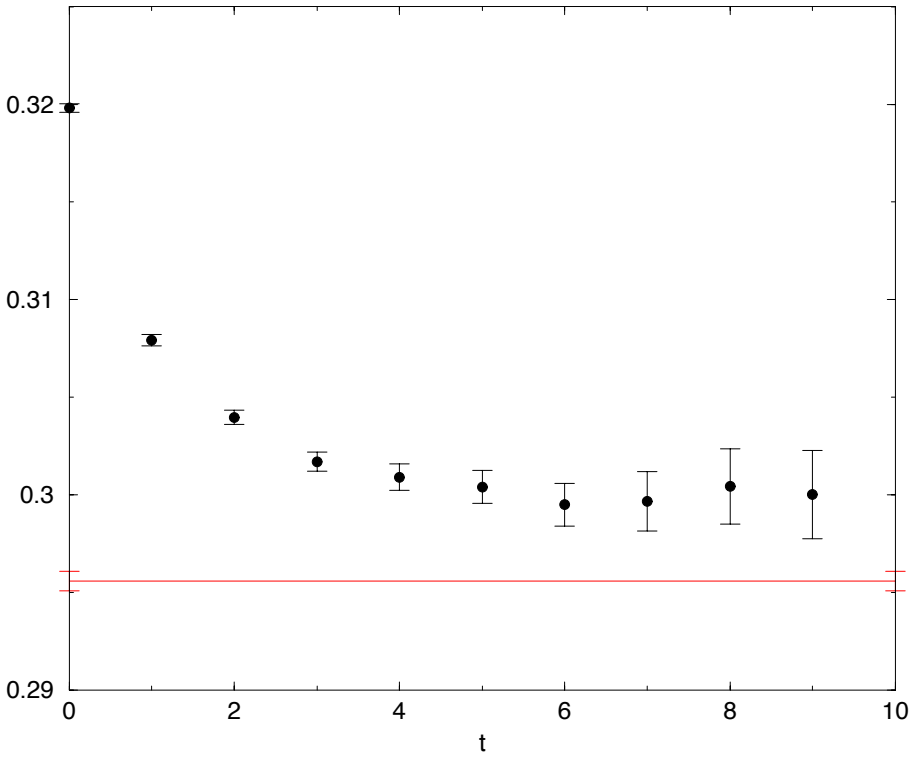

Figure 1: The effective mass plot, $\log(C_{AA}(t)/C_{AA}(t+1))$ vs. t for a single operator $A(t) = \sum_x O(x,t)$. The solid line shows the 'true' lattice result, obtained by the generalized eigenvalue method, shown here for comparison. one-op-em

due to the fact that the gap $\Delta E = E_2 - E_1 \approx 0.033$ is too small, $1/\Delta E \approx 30$, and the higher excitations do not die out at these t values.

A more sophisticated method is to measure the eigenvalues of the correlation matrix $C(t)$ and build the effective mass plot from its lowest eigenvalue. This would be exact if the wave functions were orthogonal to each other (and there was no statistical noise). Even if the wave functions are far from being orthogonal (which is the general case) for large enough t the value of the effective mass converges to E_1, the true value. But again, the rate of convergence is controlled by $\exp(-\Delta E t)$. Fig. 2 shows the effective mass for the largest eigenvalue of $C(t)$ (small filled circles). For $t \geq 3$ it agrees

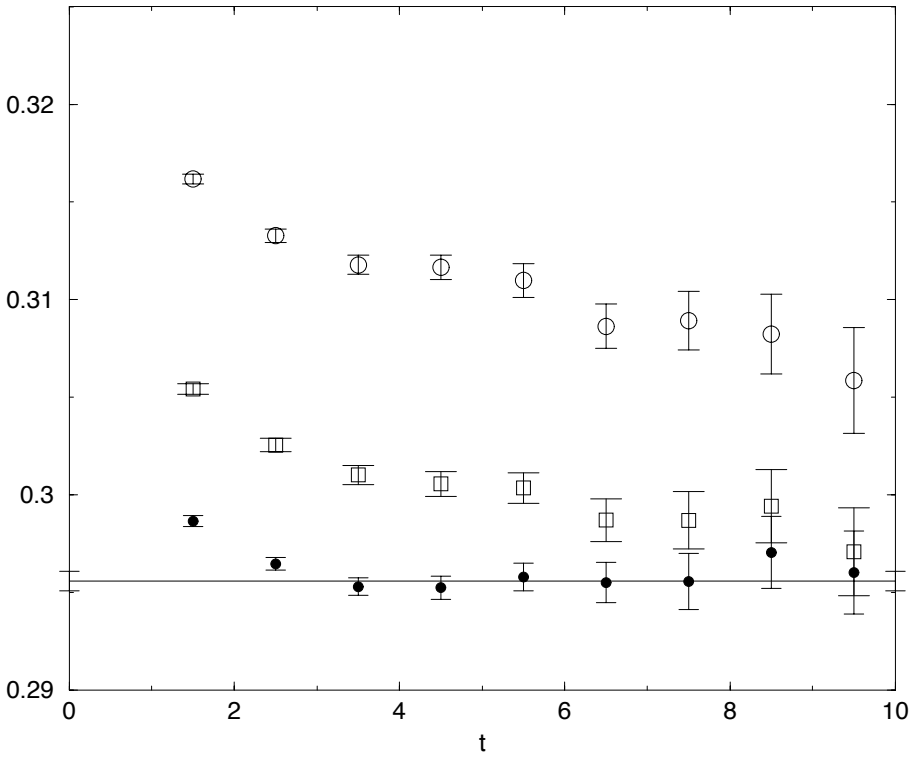

Figure 2: The filled circles show the 'effective mass' for the largest eigenvalue of $C(t)$, $\log(\lambda_1[C(t)]/\lambda_1[C(t+1)])$ vs. t for the full matrix $C_{xy}(t)$, $x, y \leq L/2 = 64$. The solid line shows the 'true' lattice result, as before. The open boxes show the same for the matrix restricted to $x, y \leq 60$, while the open circles for $x, y \leq 48$. The good agreement for the $x, y \leq 64$ is due to the fact that the wave functions are nearly orthogonal for this case. c0-ev-em

well with the 'best' value. However, this is just the consequence of the fact that in this special case (as mentioned above) the wave functions are nearly orthogonal to each other. Restricting the range of the matrix $C_{xy}(t)$ from $x, y \leq L/2 = 64$ to $x, y \leq 60$ (open boxes) or to $x, y \leq 48$ (open circles) no convergence to the true value is seen from the figure. (Of course, for $t \geq 30$ all these should approach the true result.)

In ref. [2] the matrix $C(t)$ has been considered in Fourier space in the relative momentum (instead of x as we do here), taking the lowest $N = L/4$ momenta. The filled circles in fig. 3 show the result

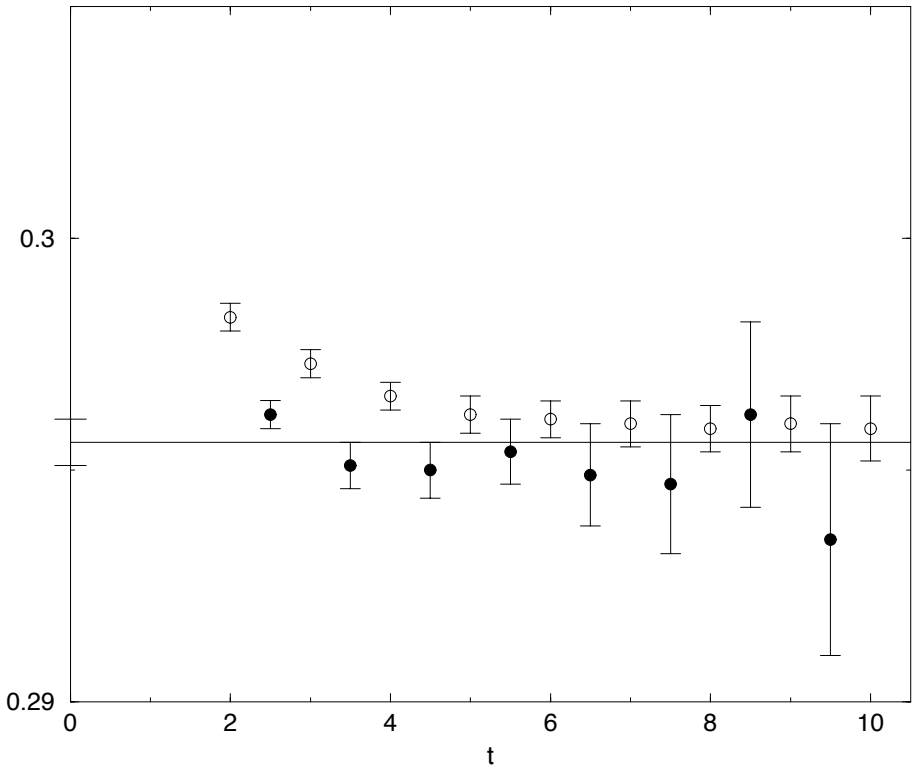

Figure 3: Results obtained from $\lambda_1(t,t_0)$ with $t_0 = 1$ and different values of t using the matrix $C(t)$ in the Fourier space with $N = L/4 = 32$ states, as in ref. [2]. The filled circles are the 'effective mass' values of eq. (12) plotted at $t + 1/2$, while the open circles show $-\log\lambda_1(t,t_0)/(t-t_0)$. The solid line is as in fig. 1. gen-four-em

for the 'effective mass', eq. (12) (as used in [2]) for $t_0 = 1$, plotted at $t+1/2$. The values $-\log\lambda_1(t,t_0)/(t-t_0)$ are also shown by open circles. This figure indicates that at $t_0 = 1$ and $N = 32$ (in the given basis) the $n > N$ states are not completely negligible. It would be desirable to choose t_0 and N so that the sum in eq. (1) could be restricted to $n \leq N$ within the statistical errors for all $t \geq t_0$.

For completeness (and to underline the contrast with the standard eigenvalue problem, the results of which were shown in fig. 2) we did the same calculation also with $C_{xy}(t)$ restricted to $x, y \leq N = L/4 = 32$, taking $t_0 = 1$ again. The results coincide with those

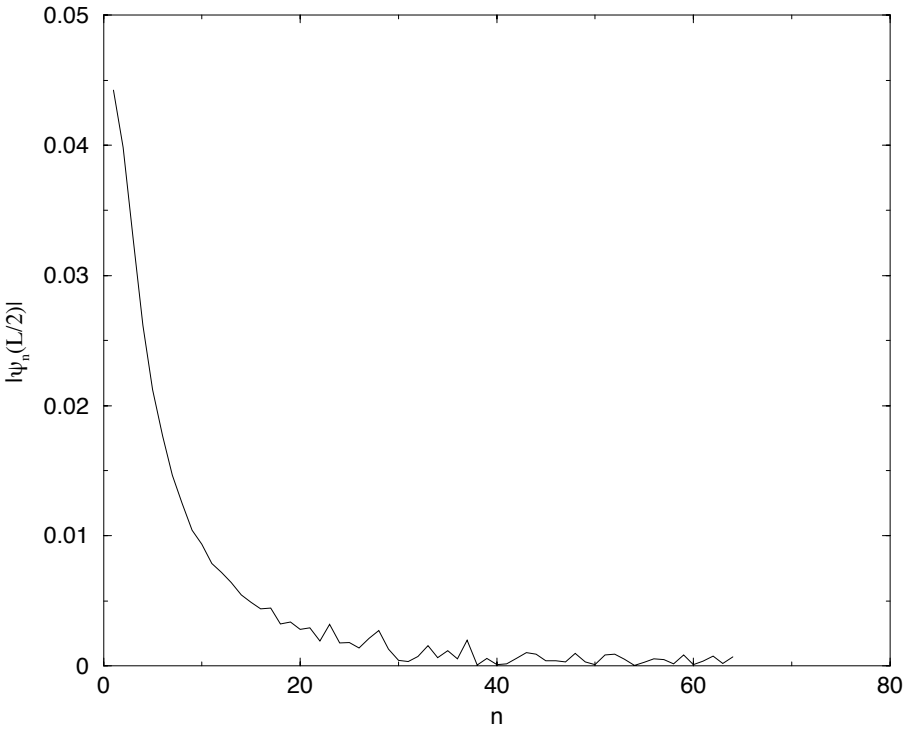

Figure 4: Values of $|\psi_n(L/2)|$ vs. n, as obtained from the generalized eigenvalue problem with $t_0 = 1$, $t = 2$. gen-psi

without truncation to at least 4 digits (for the first 3 levels checked).

It is instructive to consider the dependence of the wave functions on n. In figure 4 we plotted $|\psi_n(L/2)|$ vs. n as obtained from eq. (6) with $t_0 = 1$, $t = 2$. The fast decrease of ψ_n with n is also a specific feature of the problem. When using e.g. in gauge theories operators given by different loops, with different smearing prescriptions one does not expect such sharp decrease in n, i.e. higher excited states are less suppressed in general. This affects, in particular the smallness of the coefficient of the slowly decreasing corrections.

TAKE $t_0 = 0$ and $N =$?? TO ILLUSTRATE THE SLOWLY DECREASING CORRECTIONS. PERHAPS ONE COULD RESCALE $\psi_n \propto 1/\sqrt{n}$(???) TO SEE THE EFFECT BETTER.

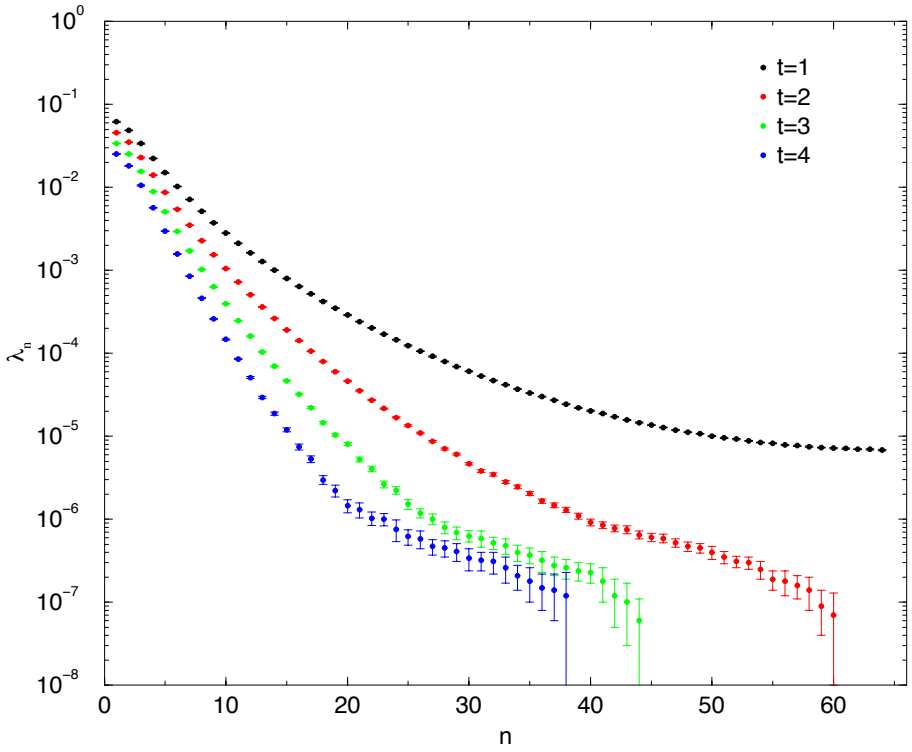

Figure 5: Eigenvalues $\lambda_n[C(t)]$ for $t = 1, 2, 3, 4$ vs. n. It shows that for $t = 1$ all eigenvalues are well measured, while for $t > 1$ some small eigenvalues are not, and one needs a truncation in the number of the eigenvectors kept ($M < N$). c0-ev

Note that in the present problem the x-dependence of the wave functions $\psi_n(x)$ (not shown here) contains additional information on the relative momentum of the two particles, and has been used in ref. [5].

By trying to take $t_0 > 1$ one meets the problem discussed in section 2 and one has to truncate to a subspace of dimension $M < N$. Fig. 5 shows the eigenvalues of $C(t)$ for $t = 1, \ldots, 4$. It is seen that for $t \geq 2$ the smallest eigenvalues are unstable.

For $t_0 = 2$ the result is practically insensitive to $M \leq 50$ but it becomes completely unstable for $M = 52$ (with our statistics, of course) so the change occurs abruptly. The results are shown in fig.

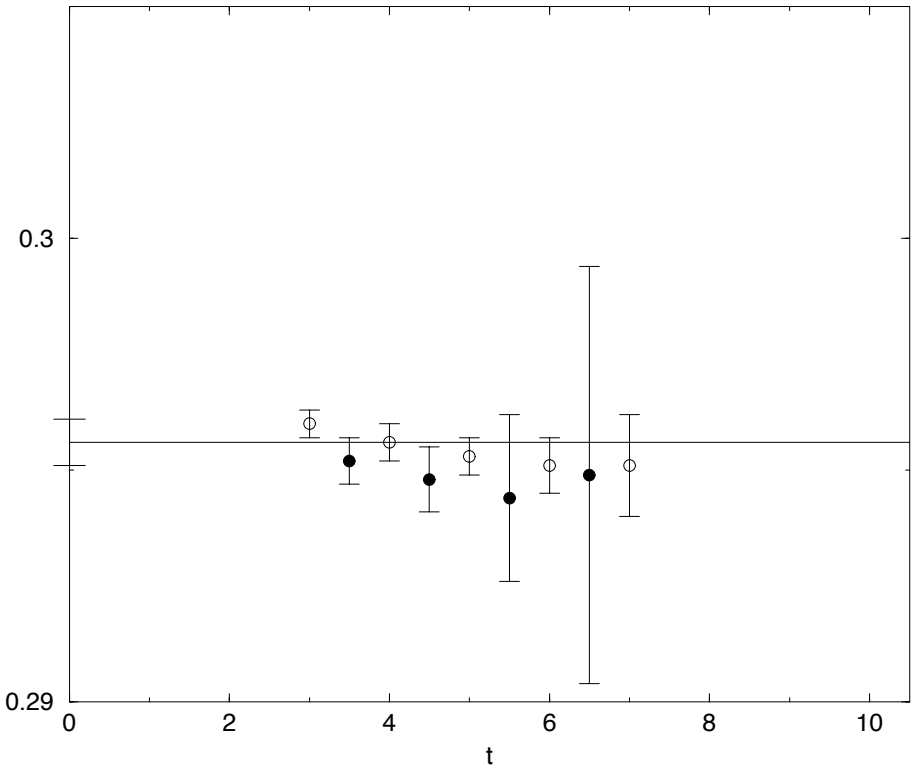

Figure 6: Results obtained from $\lambda_1(t,t_0)$ with $t_0 = 2$, with truncation to first $M = 40$ eigenstates of $C(t_0)$. The filled circles are the 'effective mass' values of eq. (12) plotted at $t + 1/2$, while the open circles show $-\log\lambda_1(t,t_0)/(t-t_0)$. The solid line is as in fig. 1. gen-t0-2

6. Since the decay of the correlation matrix is given by $E_1 \approx 0.3$, it decays perhaps too fast to illustrate the usefulness of taking $t_0 > 1$ or of eq. (31).

TO DO:

- Construct a $C(t)$ which is exactly $N \times N$ matrix with known energies and wave functions, for checking different things.
- MAKE A FAKE $C(t)$ BUILT FROM N TERMS EXACTLY, WITH KNOWN E_n AND ψ_{ni}, ADDING TO IT A STATISTICAL NOISE SIMILAR TO THE ONE OBSERVED, BY $C_{ij}(t) \to C_{ij}^{\rm exact}(t) + \big(C_{ij}(t) - \overline{C}_{ij}(t)\big)$ for each run separately.

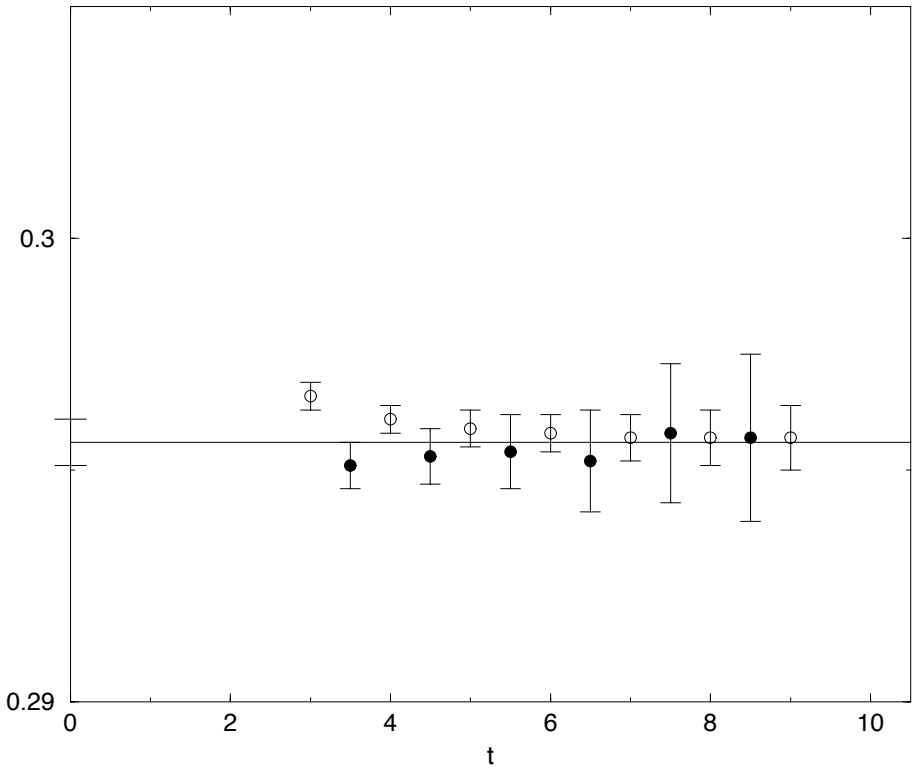

Figure 7: Results obtained from eq. (31) with $t_0 = 1..2$ and $\Delta = 1..1$, with truncation to first $M = 40$ eigenstates of K_0. The filled circles are the 'effective mass' values plotted at $t + 1/2$, while the open circles show ???. The solid line is as in fig. 1. gen-t0-12

$\bar{C}_{ij}(t)$ is the average over the 120 runs. This way one could separate the effect of states with $n > N$ and the effect of non-linearities in the general eigenvalue problem.

References

[1] C. Michael, Nucl. Phys. B259 (1985) 58.

[2] M. Lüscher and U. Wolff, Nucl. Phys. B339 (1990) 222.

[3] Morningstar and Peardon, ...

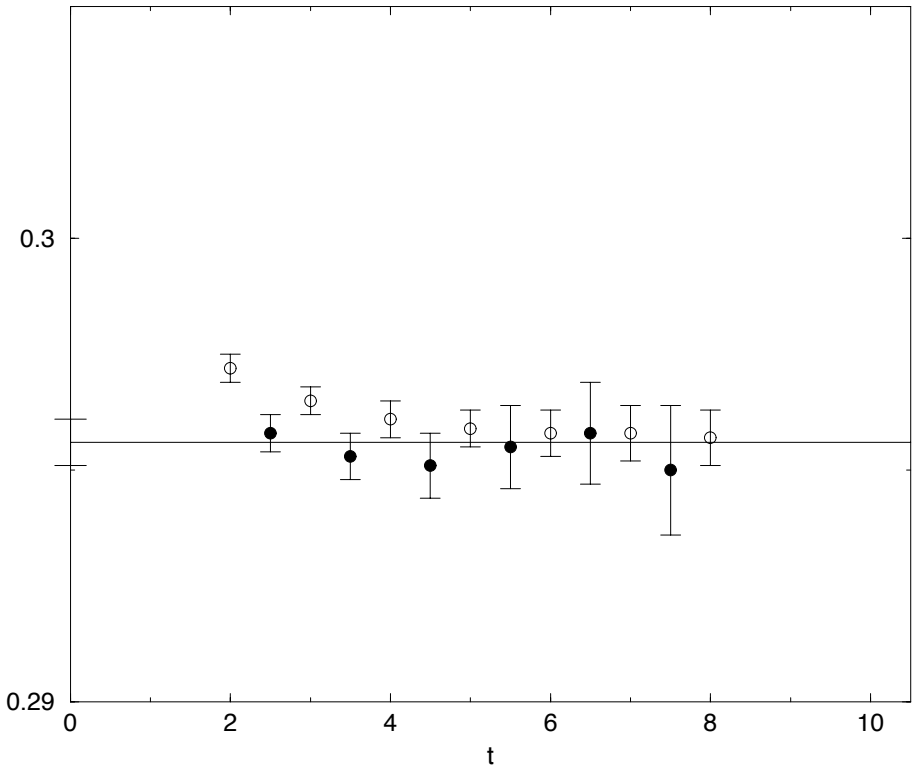

Figure 8: Results obtained from eq. (31) with $t_0 = 1,3$ and $\Delta = 1..1$, with truncation to first $M = 40$ eigenstates of K_0. The filled circles are the 'effective mass' values plotted at $t + 1/2$, while the open circles show ???. The solid line is as in fig. 1. gen-t0-13a

[4] Knechtli and Sommer,...

[5] J. Balog, M. Niedermaier, F. Niedermayer, A. Patrascioiu, E. Seiler and P. Weisz, Comparison of the O(3) bootstrap σ-model with the lattice regularization at low energies. hep-lat/9903???

[6] D. Diakonov and V. Petrov, To what distances do we know the confining potential, hep-lat/9810037.

[7] A.B. and Al.B. Zamolodchikov, Ann. Phys. 120 (1979) 253; Nucl. Phys. B133 (1978) 525.