Marshall's sign rule and density-matrix renormalization-group acceleration

Ulrich Schollwöck
Sektion Physik, Ludwig-Maximilians-Universität München, Theresienstrasse 37, 80333 Munich, Germany
(Received 20 April 1998; revised manuscript received 9 June 1998)

In applications of White's density-matrix renormalization-group (DMRG) algorithm, computation time is dominated by the diagonalization of large sparse Hamiltonians by iterative diagonalization algorithms, whose convergence can be decisively accelerated by the usage of good start vectors. In this paper I show how, using the Marshall sign rule, in a wide class of antiferromagnetic models the number of diagonalization iterations can be reduced below 10, sometimes down to 2, accelerating the DMRG by an order of magnitude. This acceleration, applicable during the growth of long chains, complements the acceleration procedure proposed by White. To illustrate the feasibility of the approach, I show how it performs if applied to the calculation of the Haldane gap for $S=2$. [S0163-1829(98)02037-2]

I. INTRODUCTION

In recent years, the density-matrix renormalization-group algorithm proposed by White[1] in 1992, has become together with exact diagonalization and quantum Monte Carlo methods the algorithm of choice in low-dimensional quantum models, including magnetic, fermionic and bosonic systems.

White has proposed two different DMRG algorithms, the so-called finite-size and infinite-size algorithm. The infinite-size algorithm precedes the application of the finite-size algorithm: a chain is grown symmetrically to its final size, whereas the finite-size algorithm treats the fully-sized system splitting it asymmetrically. This increases the precision of the results. Sometimes, it is however advantageous to stick to the infinite-size algorithm and increase its precision by increasing the number of block states M. For example, when the Hamiltonian is invariant under reflection and parity thus a good quantum number, it is of great advantage to retain this quantum number, for easier classification of states and for thinning out the Hilbert space by splitting it into more sectors invariant under the operations of the Hamiltonian. This gives access to more states and speeds up the algorithm.

In either case, DMRG precision is dominated by the number of block states M, and time consumption scales as M^3. The most time-consuming part of the algorithm is the determination of low lying eigenstates of the approximate Hamiltonians that are construed during the iterative application of the decimation process of the DMRG. To do this, iterative diagonalization algorithms such as Lanczos[2,3] are used. They can be accelerated if a good initial guess for the targeted state is available, as the number of (Lanczos) iterations drops.

During the application of the finite size algorithm, a previous ground state for full-length system under consideration as well as the (incomplete) DMRG basis transformations are available. White has used this information to make a very good guess for the initial state of the diagonalization algorithm,[4] basically carrying out basis transformations on the previous ground state. During the application of the infinite size algorithm, previous ground state(s) were all obtained on shorter chains, such that a basis transformation is not feasible. What I want to show in this paper is how in the infinite size algorithm the old information can be used to also make a good initial guess.

What the DMRG essentially does is to find a fixpoint in density matrix space, i.e., for large systems, essentially the same (incomplete) transformation from a basis with MN states (M block states and N spin states) to a new (decimated) basis with M states is carried out.[5] A first guess would be that the target state converges to a fixpoint in Hilbert space also, implying that simply using the target state found in the last iteration as the initial state vector should be a good guess.

This is not so: the overlap between the initial guess and the result is typically far from 1, and $\langle v_1|\mathcal{H}|v_1\rangle$, the energy expectation value of the initial guess is much larger than the true target state energy. One finds that for longer chains, the absolute values of wave function coefficients (if labeled suitably, see below) converge fast to fixed values, while the signs vary randomly. If one could predict the sign changes, the old wave function would indeed provide an excellent starting point. The randomness in signs has two origins. First, real eigenvectors are only determined up to a global sign, which is attributed unpredictably by the density-matrix diagonalization algorithm. Second, there are deterministic sign changes with chain length. We will see how the latter can be used to fix the global signs of eigenvectors such that almost all signs of the new wave function are correctly reproduced in the old wave function, which then can serve as an excellent prediction for the new wave function.

II. PREDICTION MECHANISM

Let us consider, to simplify the description, an isotropic Heisenberg antiferromagnetic spin chain with integer spin length S. Let us assume that we have reached a certain block length L, such that the total chain under consideration has length $2L+2$. Let us call the block states of the block of length L $|m'_L\rangle$, those of the block of length $L+1$ $|m_{L+1}\rangle$, with total magnetizations $S^z_{m'_L}$ and $S^z_{m_{L+1}}$. If we call σ the spin state on site $L+1$, and the magnetization S^z_σ, the DMRG decimation procedure yields

$$\langle m'_L\sigma|m_{L+1}\rangle \neq 0 \tag{1}$$

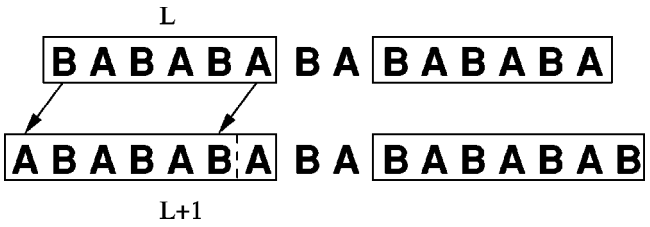

FIG. 1. DMRG chains for different block lengths L and $L+1$ with sublattice structure. Sublattices are labeled such that central spins remain on the same sublattice. Blocks are represented by rectangles. Note the shift in sublattice position of block spins during system growth.

with

$$S^z_{m'_L}+S^z_\sigma=S^z_{m_{L+1}} \tag{2}$$

as a necessary condition. Both the states of the block with length L and the block with length $L+1$ are now *ordered and numbered in the respective magnetization sectors according to their importance, i.e., the associated eigenvalue of the density matrix when they were formed.* I assume now that the respective magnetization sectors for the blocks of length L and $L+1$ have equal numbers of states. This is normally the case as soon as chains exceed certain, rather small lengths; if not, typically one or two states of extremely small weight are redistributed. One can easily modify the proposed procedure to take this into account, which for reasons of clarity I will not discuss; all results shown below were obtained by switching off the prediction mechanism in these rare circumstances. Otherwise the condition could be ensured by adding or dropping a state which has very little weight, at minor loss of precision.

Fundamental rule. If this assumption holds, the global sign of the eigenvector of the density matrix giving $|m_{L+1}\rangle$ is chosen such that

$$(-1)^{S^z_{m_{L+1}}}\langle m_L\sigma|m_{L+1}\rangle>0 \tag{3}$$

for the case where $m_L\equiv m_{L+1}$, meaning that the states have *equal label number* (I will call them ''equivalent'').

How can this rule be motivated? Let us consider Marshall's sign rule.[6] In its original form it states that for a bipartite connected (i.e., that only sites on different sublattices A and B interact and that each site can be linked to each other site by a chain of bonds) isotropic antiferromagnet the lowest energy state $|\psi^M\rangle$ in a sector with total magnetization M can be written in the Ising basis

$$|\psi^M\rangle=\sum_n c_n(-1)^{\Sigma_{i\in B}-S+S^z_i}|n\rangle, \tag{4}$$

such that $c_n>0$. The $|n\rangle$ are the Ising basis states, the sum in the exponent effectively yields the total magnetization of sublattice B. Marshall's sign rule applies to the two most interesting states for the DMRG, the ground and first excited states. Let us compare the two chains arising in the DMRG procedure shown in Fig. 1, for block length L and block length $L+1$. If we assign the same sublattice positions to the center spins, the positions of the spins in the L-block shift their position from one sublattice to the other. Let us now take the two (say, ground state) wave functions

$$|\psi_L\rangle=\sum \psi_{m^l_L\sigma^l\sigma^r m^r_L}|m^l_L\sigma^l\sigma^r m^r_L\rangle, \tag{5}$$

$$|\psi_{L+1}\rangle=\sum \psi_{m^l_{L+1}\sigma^l\sigma^r m^r_{L+1}}|m^l_{L+1}\sigma^l\sigma^r m^r_{L+1}\rangle. \tag{6}$$

If we make now the assumption that the DMRG is close enough to its fix point that we may identify both systems, the wave function coefficients will be (almost) identical *but for the sign*, as sublattice positions of spins have changed. This is confirmed by numerical evidence, and motivates calling states equivalent; while physically distinct, they make (almost) the same contribution to (physically also distinct) wave functions as also supported by the arguments of Ref. 5. It remains to determine the sign. If we consider two equivalent states $|m_L\rangle$ and $|m_{L+1}\rangle$, the matrix element $\langle m_L\sigma|m_{L+1}\rangle$ is nonzero only if $S^z_\sigma=0$. Expanding $|m_L\rangle$ in the Ising basis, the signs of the wave function coefficients will be, up to a global sign, be given by $(-1)^{[S^z]^B_n}$, where $[S^z]^B_n$ is the sum over the magnetization on sublattice B of Ising state n. If we identify now $|m_L\rangle$ in $|\psi_L\rangle$ with $|m_{L+1}\rangle$ in $|\psi_{L+1}\rangle$, one sees that the added spin (sitting on sublattice A) makes no sign contribution, while the spins of the L-block have changed sublattice, such that the sign will be $(-1)^{[S^z]^A_n}$. The relative sign change will be

$$(-1)^{[S^z]^A_n-[S^z]^B_n}=(-1)^{[S^z]^A_n+[S^z]^B_n}=(-1)^{S^z_{m_L}}=(-1)^{S^z_{m_{L+1}}}. \tag{7}$$

The last expression is independent of the underlying Ising states and leads immediately to rule (3).

Majority rule. Retracing the above argument, its central weakness is the assumption that block states can be identified for different lengths. Calculating the overlap between predicted wave functions and calculated wave functions, it is found to have increased to typically well above 0.9, but this does not lead yet to a large increase in performance of iterative diagonalizations. Thus, rule (3) seems to catch an important point, but is also oversimplifying. Block identification implied that the added spin had zero magnetization, which is physically not true; there will be other magnetizations contributing, in particular in excited states. Rule (3) may therefore not apply.

To go beyond, magnetizations $S^z_\sigma\neq 0$ have to be considered also. It seems there is no strict rule to do this—also because the overlaps between states may be so small that the sign is up to numerical arbitrariness—but we have found the following procedure to work extremely well.

The expression

$$(m_{L+1},m'_{L+1}):=\text{sign}\,(-1)^{S^z_\sigma}\langle m'_L\sigma|m_{L+1}\rangle \times\langle m_L(-\sigma)|m'_{L+1}\rangle, \tag{8}$$

where $|m_L\rangle$ and $|m_{L+1}\rangle$, and $|m'_L\rangle$ and $|m'_{L+1}\rangle$ are equivalent states respectively, is evaluated. If it is negative, the sign is considered ''wrong.'' One finds, that almost all states $|m_{L+1}\rangle$ have all signs (m_{L+1},m'_{L+1}) right, and that some states (typically less than 10 out of several hundred) have

almost all signs wrong. The wrong signs of the ''almost correct'' states $|m_{L+1}\rangle$ are typically those (m_{L+1},m'_{L+1}), where $|m'_{L+1}\rangle$ is an ''almost wrong'' state. By globally flipping the sign of the few ''almost wrong'' states, nearly all signs can be made right; the majority character of the rule makes it extremely stable against numerical imprecisions. The fact that only few flips are necessary justifies the assumptions made for Eq. (3): As we will show, the majority rule added to Eq. (3) ensures that the old wave function has an almost perfect overlap with the desired new wave function.

One of several ways to see how this rule goes beyond Eq. (3) is to imagine that we are adding two sites to a block, assuming that m_L and m_{L+2} can be identified, with spin $-\sigma$ on the first and spin σ on the second site. There is no shift in sublattice positions, one of the new spins makes a sign contribution $(-1)^{S^z_\sigma}$. Therefore,

$$\text{sign}\,[(-1)^{S^z_\sigma}\langle m_L(-\sigma)\sigma|m_{L+2}\rangle]=1, \tag{9}$$

for identified states. Inserting $\Sigma_{m'_{L+1}}|m'_{L+1}\rangle\langle m'_{L+1}|$, we find

$$\text{sign}\left[(-1)^{S^z_\sigma}\sum_{m'}\langle m_L(-\sigma)|m'_{L+1}\rangle\langle m'_{L+1}\sigma|m_{L+2}\rangle\right]=1.$$

Now we are not making a statement about $\langle m_L(-\sigma)|m'_{L+1}\rangle$ [this would not go beyond Eq. (3)], but about a product. Identifying the decimations $L\rightarrow L+1$ and $L+1\rightarrow L+2$ in the above for large L,[5] we obtain the majority rule at least for the dominant overlaps.

Extensions. So far, we have only considered isotropic chains. Retracing the derivation of the Marshall sign rule, it becomes evident that (antiferromagnetic) anisotropies and dimerized interactions are compatible with the sign rule; we found the prediction to work.[7] In other cases, such as frustration, where the Marshall sign rule does not apply, Richter, Ivanov, and Retzlaff have shown that for small frustration the weight of the states whose signs violate the sign rule is zero or very small.[8] In such cases, the method will still lead to an efficiency gain, which gradually disappears with increasing frustration. This has been confirmed numerically. We have not considered half-integer spins, as state parity alternates during chain growth; this implies that states that are two DMRG steps apart will be related, making the method more complicated. However, the sign rule still holds, and it should be possible to adapt the prescription to that situation.

III. PERFORMANCE

The above two rules can be implemented very easily (30 or so lines at most), and consume almost no memory and computation time. The gain is however striking. To illustrate that the above procedure is useful, we calculated the $S=2$ Haldane gap using chains with total length $L_{\text{tot}}=600$, while keeping $M=400$ block states, employing magnetization and parity as good quantum numbers. The chain is long enough, that the expected L^{-2} convergence of the finite length gap can be observed, and a gap $\Delta=0.0907(2)$ extrapolated (cf. Ref. 9). The number of Lanczos iterations without prediction varies between ~ 60 for the ground state and ~ 70 for the first excitation; here, we consider the Lanczos algorithm converged once the relative change of the desired eigenvalue is less than 10^{-13} between Lanczos iterations.

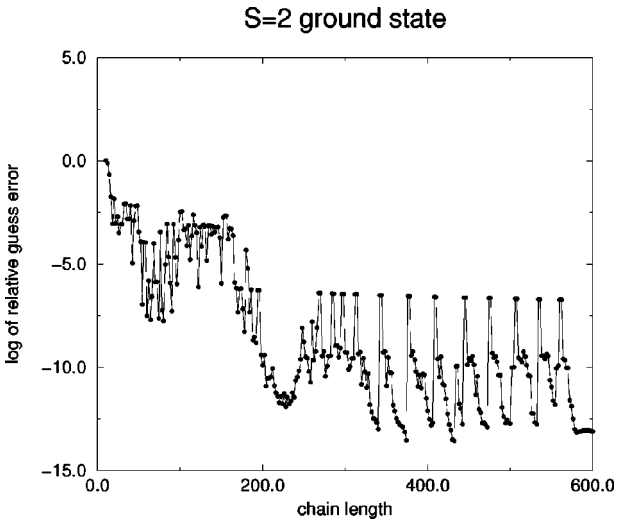

FIG. 2. Decadic logarithm of the relative guess error $(E_{\text{final}}-E_{\text{guess}})/E_{\text{final}}$ for the ground state of a $S=2$ chain vs length.

The prediction algorithm makes the number of Lanczos iterations drop dramatically, once the chain exceeds a certain length. For $L>200$, more than 30 percent of all Lanczos runs finish after 4 or less iterations, where it has to be kept in mind that 2 is the minimum number of iterations to establish that a result has already converged. Less than 15 percent take 10 or more iterations, and the average is 7 iterations only. The intermittent peaks are due to rearrangements of the block states which reflect the growing chain length. In these iterations, prediction works only partially (the same phenomenon can be observed in strong form below $L=200$). The number of Lanczos iterations is closely connected to the guess error (Fig. 2), the difference between the energy expectation value for the reused old ground state and the converged Lanczos result. For $L>200$, the guess error relative to the ground state energy is always better than 10^{-6}, mostly at or below 10^{-10}. This shows that the prediction rules are very powerful indeed. Remaining errors also come from the

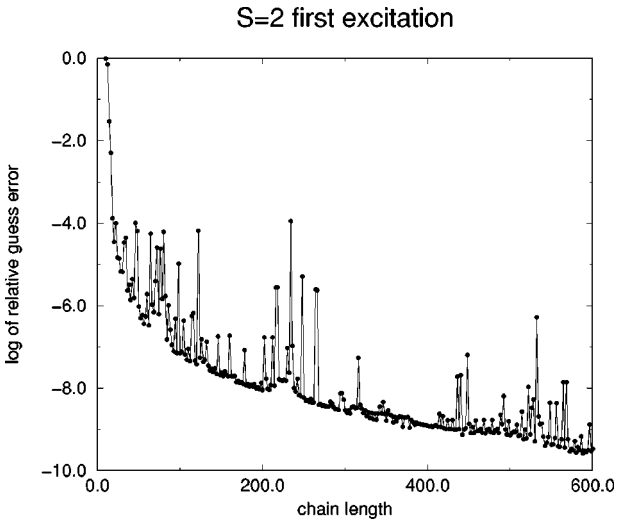

FIG. 3. Decadic logarithm of the relative guess error $(E_{\text{final}}-E_{\text{guess}})/E_{\text{final}}$ for the first excitation of a $S=2$ chain vs length.

fact that the assumption that the absolute weight of states does not with chain length is of course only approximate.

Results for the first excitation are not quite as good, as was to be expected, but still the number of Lanczos iterations drops to about 20, with some peaks up to 40. This is still a saving up to a factor 3. The relative error goes down to 10^{-9}, which shows that the method is also efficient here (Fig. 3).

Let us close with the remark that on an alpha workstation, these calculations could all be done overnight and that $S=2$ is not an ''easy'' case for demonstration purposes: the very long correlation length is damaging for the underlying assumptions of our procedure. All results found here do not change significantly if the Lanczos convergence condition is modified to somewhat higher or lower precision or the convergence of the eigenvector is considered instead. While the number of Lanczos iterations will be globally slightly increased or reduced if the convergence condition is made more or less stringent, the savings remain of the same order and numerical results are only affected in numerically insignificant digits. One may also consider to replace the Lanczos convergence condition by carrying out a fixed number of, say, 5 iterations at each step. However, for short chains, there is a danger not to converge to the right state; for chains significantly longer than the correlation length, this approach is feasible, but precision deteriorates somewhat because of the imperfections of the prediction rule; to control this loss of precision, one should feed results into the speeded-up finite size algorithm[4] afterwards. Then, however, some of the gain in performance is lost once again.

IV. CONCLUSION

We have shown how using information about the nature of antiferromagnetic wave function provided by the Marshall sign rule can be used to strongly reduce the number of diagonalization steps in the infinite system growth phase of the DMRG, allowing the growth of very long high-precision chains in reasonable time. While the proposed procedure is not as completely versatile as White's finite chain procedure,[4] it covers many important scenarios and can be used as complementary to his procedure.

ACKNOWLEDGMENT

I thank J. Richter for drawing my attention to his continuity argument for the Marshall rule in frustrated chains.

[1] S.R. White, Phys. Rev. Lett. **69**, 2863 (1992); Phys. Rev. B **48**, 10 345 (1993).
[2] C. Lanczos, J. Res. Natl. Bur. Stand. **45**, 255 (1950).
[3] C.C. Paige, J. Inst. Math. Appl. **10**, 373 (1972).
[4] S.R. White, Phys. Rev. Lett. **77**, 3633 (1996).
[5] S. Ostlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995).
[6] W. Marshall, Proc. R. Soc. London, Ser. A **232**, 48 (1955).
[7] H. Aschauer and U. Schollwöck, Phys. Rev. B **58**, 359 (1998).
[8] J. Richter, N.B. Ivanov, and K. Retzlaff, Europhys. Lett. **25**, 545 (1994).
[9] U. Schollwöck and Th. Jolicoeur, Europhys. Lett. **30**, 493 (1995); U. Schollwöck, O. Golinelli, and Th. Jolicoeur, Phys. Rev. B **54**, 4038 (1996); Y. Nishiyama, K. Totsuka, N. Hatano, and M. Suzuki, J. Phys. Soc. Jpn. **64**, 414 (1995); S. Yamamoto, Phys. Rev. Lett. **75**, 3348 (1995); Phys. Rev. B **53**, 3364 (1996); Phys. Lett. A **213**, 102 (1996); S. Qin, Y.L. Liu, and L. Yu, Phys. Rev. B **55**, 2721 (1997).