Time-evolution methods for matrix-product states

Sebastian Paeckel [a], Thomas Köhler [a,b], Andreas Swoboda [c],
Salvatore R. Manmana [a], Ulrich Schollwöck [c,d],
Claudius Hubig [e,d,*]

[a] *Institut für Theoretische Physik, Universität Göttingen, 37077 Göttingen, Germany*
[b] *Department of Physics and Astronomy, Uppsala University, Box 516, S-75120 Uppsala, Sweden*
[c] *Department of Physics, Arnold Sommerfeld Center for Theoretical Physics (ASC), Fakultät für Physik, Ludwig-Maximilians-Universität München, 80333 München, Germany*
[d] *Munich Center for Quantum Science and Technology (MCQST), 80799 München, Germany*
[e] *Max-Planck-Institut für Quantenoptik, 85748 Garching, Germany*

A R T I C L E   I N F O

A B S T R A C T

Matrix-product states have become the de facto standard for the representation of one-dimensional quantum many body states. During the last few years, numerous new methods have been introduced to evaluate the time evolution of a matrix-product state. Here, we will review and summarize the recent work on this topic as applied to finite quantum systems. We will explain and compare the different methods available to construct a time-evolved matrix-product state, namely the time-evolving block decimation, the MPO $W^{I,II}$ method, the global Krylov method, the local Krylov method and the one- and two-site time-dependent variational principle. We will also apply these methods to four different representative examples of current problem settings in condensed matter physics.

1. Introduction

The dynamics of correlated quantum systems is a rich playground for non-equilibrium physics. The actual study, however, is a challenge for experiments as well as for theory. In the context

---

* Corresponding author at: Max-Planck-Institut für Quantenoptik, 85748 Garching, Germany.
E-mail address: claudius.hubig@mpq.mpg.de (C. Hubig).

of tensor networks, numerical methods were introduced around 2004 and opened the path to study the dynamical behavior of low-dimensional systems in a controlled fashion. In particular, the Trotterized time evolution of matrix-product states (MPS) as produced by the density-matrix renormalization group (TEBD, tDMRG, tMPS) [1–8] made a breakthrough as it combined powerful time-evolution schemes with the efficient truncation of the exponentially-large Hilbert space intrinsic to quantum many-body systems. Nowadays, DMRG methods formulated in the language of matrix-product states [9–14] (MPS) are the de facto standard for the investigation of one-dimensional systems. Here, we will focus on the time evolution of finite-dimensional quantum states along the real and the imaginary time axis. That is, we restrict ourselves to those methods which produce a time-evolved MPS representation of a quantum state as opposed to evaluating the time-dependence of a specific observable in some subspace (as done by the Chebyshev [15–19] and other similar approaches) or methods which directly work in frequency space (such as, for example, the correction vector [20–22] approach). Specifically, we will review and compare in some detail the various MPS-based approaches which have been developed in the past few years with sometimes large improvements over the first generation of methods.

The main goal of these schemes is to combine the efficient truncation of the Hilbert space with an accurate time-evolution method. Naturally, time-evolution methods are dealt with regularly in the mathematical literature (cf. the review [23]), but the applicability to MPS-based problem settings is often limited due to the peculiarity of the MPS approach in approximating the Hilbert space. Since we have quantum systems in mind, the underlying differential equation is the time-dependent Schrödinger equation (TDSE), whose solution can formally be recast to the task of applying the time-evolution operator

$$\hat{U}(\delta) = e^{-i\delta\hat{H}} \tag{1}$$

with the time-independent Hamiltonian $\hat{H}$, time step size δ and $\hbar \equiv 1$. It is therefore natural to ask for efficient MPS methods for the time evolution of quantum states by combining suitable approaches to matrix exponentials with the MPS formulation of quantum (lattice) systems. Indeed, most of this review article is concerned precisely with this question and explores under which circumstances which ansatz will be the most successful one in terms of accuracy and efficiency.

In the context of quantum systems, first attempts to numerically integrate the time-dependent Schrödinger equation beyond the reach of exact diagonalization using sparse matrix exponentials [24] are reported in quantum chemistry [25,26]. When attempting to do the same for MPS, two main approaches have to be differentiated: the first evaluates or constructs the time-evolved state $|\psi(t + \delta)\rangle = \hat{U}(\delta)|\psi(t)\rangle$ directly by approximating the action of $\hat{U}(\delta)$ in a sufficiently small subspace. In this approach, $\hat{U}(\delta)$ itself is never calculated or approximated. The global Krylov method [27–29] presented in Section 5 simply translates the Lanczos formalism [30] for unitary time evolution to matrix-product states. The local Krylov approach [29,31–35] (Section 6.1) and the time-dependent variational principle [36,37] (Section 6.2) further adapt the Krylov method to the special MPS structure and re-derive the TDSE in suitable local subspaces.

Alternatively, one may ask for an approximation of $\hat{U}(\delta)$ which can be obtained efficiently and also can be applied efficiently to the quantum state. One way to deal with such exponentials is well-known in quantum field theory where it permits to set up path integrals [38]. It decomposes the time-evolution operator for "sufficiently small" time steps using the Suzuki–Trotter [39] decomposition. This decomposition leads to conveniently smaller matrix exponentials. The method can be directly applied in the context of matrix-product states [4–8], as explained in Section 4.1. Extending on it, we may ask for efficient matrix-product operator (MPO) approximations of $\hat{U}(\delta)$ which exploit the MPO structure directly [40] to allow efficient exponentiation, cf. Section 4.2.

A direct comparison of widely used approaches, including the recently developed methods, is missing so far. Since those methods have different strengths and weaknesses, it is difficult to say a priori which one will be the most efficient for a given problem setting. We will attempt to fill this gap by providing an overview and comparison. In particular, we discuss in detail the Suzuki–Trotter decomposition and MPO $W^{I,II}$ methods for constructing $\hat{U}(\delta)$ directly, and TDVP- and Krylov-subspace-based approaches evaluating $\hat{U}(\delta)|\psi\rangle$. All methods are tested and compared by treating four representative problems of dynamics in correlated systems. This includes the
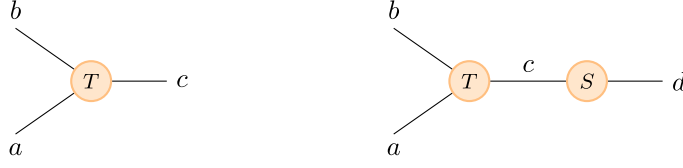
Fig. 1. Left: Graphical representation of the example tensor $T_{a,b,c}$ with three indices (legs) a, b and c. Right: Graphical representation of the tensor contraction $\sum_c T_{a,b,c} S_{c,d} = T \cdot S$.

real and the imaginary time evolution of chains, time-dependent correlators, critical and gapped systems, and attempts at two-dimensional systems. To cross-check our results, we use two independent implementations of the time-evolution methods within the SciPal-SymMPS [41,42] and SyTen[43,44] toolkits.

2. Matrix-product states and operators

The main problem in the numerical treatment of quantum many-body systems is the large Hilbert space which grows exponentially with system size [14]. This strongly restricts exact diagonalization approaches to small system sizes; for spin systems one can reach $\sim$ 40–50 lattice sites [45]. A variety of numerical methods (e.g., quantum Monte Carlo [46], dynamical mean-field theory [47], …,) have been developed to overcome this restriction. For one-dimensional systems, matrix-product state (MPS) methods are extremely successful because they can efficiently represent weakly-entangled states that obey the area law [14,48,49]. Here, we start by recapitulating general properties of the MPS ansatz class and its efficient numerical use, with the scaling costs of each application summarized in Table 1. As in the rest of the review, we focus on finite matrix-product states, representing quantum states on finite-dimensional Hilbert spaces. While a few of the methods tested here translate also to infinite matrix-product states and a careful choice of boundary conditions [50,51] alleviates finite-size effects, the review of time-evolution methods treating the thermodynamic limit is a topic on its own and is left for future review articles.

2.1. Tensor notation

The fundamental objects of MPS algorithms are tensors. In our context, tensors are multi-dimensional collections of complex or real numbers. Each index of a tensor corresponds to one of its dimensions; graphically, tensors are represented as shapes (circles, triangles, squares) with one leg per tensor index (cf. Fig. 1). A tensor T on e.g. three associated vector spaces A, B and C has three indices $a = [1, \ldots, \dim(A)]$, $b = [1, \ldots, \dim(B)]$ and $c = [1, \ldots, \dim(C)]$. The tensor then has scalar entries $T_{a,b,c}$. We explicitly do not consider global symmetries[1] here and hence do not associate a direction to our tensor legs (or, conversely, differentiate between vector and dual spaces). As such, the upstairs/downstairs location of tensor indices is meaningless and $T_{a,b,c} = T^a_{b,c} = T^{a,b,c}$. By complex-conjugating every element of a tensor T, we obtain a new tensor $\bar{T}$ with elements $\bar{T}_{\bar{a},\bar{b},\bar{c}} = T^\star_{a,b,c}$. We will use $\bar{a}$ to denote indices of conjugated tensors. Contracting $\bar{T}$ and T over the indices a and b, we write:

$$\sum_{a,b} \bar{T}_{a,b,\bar{c}} T_{a,b,c} = X_{\bar{c},c} \tag{2}$$

which is equivalent to

$$\sum_{\bar{a},a,\bar{b},b} \delta_{\bar{a},a} \delta_{\bar{b},b} \bar{T}_{\bar{a},\bar{b},\bar{c}} T_{a,b,c} \ . \tag{3}$$

---

[1] The use of global symmetries is certainly crucial for efficient implementations, but at the same time orthogonal to the time-evolution methods discussed here. For introductions to the topic of global symmetries in MPS and more generically tensor networks, cf. [11,44,52–58].
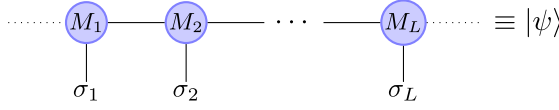
Fig. 2. Schematic of the tensor network of a matrix-product state (MPS). Horizontal lines denote the internal indices with bond dimension m, whereas the vertical lines denote physical indices with dimension σ. Dotted lines to the left and right indicate the dummy indices m_0 and m_L.

Given two tensors $A_{a,b,c}$ and $B_{b,d,c}$, the shorthand $A{\cdot}B$ denotes the contraction over all shared indices:

$$A \cdot B = \sum_{b,c} A_{a,b,c} B_{b,d,c} = Y_{a,d} . \tag{4}$$

Tensor contractions can also be represented graphically by drawing tensors with connected legs, cf. Fig. 1. Note that we will, where possible without confusion, also use a, b, c, …, to refer to the dimension dim(A) etc. respectively.

2.2. Matrix-product states (MPS)

Matrix-product states are efficient representations for one-dimensional weakly-entangled quantum states. The main idea is to represent the coefficient tensor $c_{\sigma_1 \ldots \sigma_L}$ of a general quantum state on a discrete lattice

$$|\psi\rangle = \sum_{\sigma_1,\ldots,\sigma_L} c_{\sigma_1 \ldots \sigma_L} |\sigma_1 \cdots \sigma_L\rangle \tag{5}$$
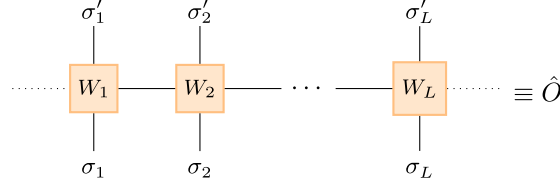
as a product of L rank-3 tensors M_i (cf. Fig. 2)

$$|\psi\rangle = \sum_{\substack{\sigma_1,\ldots,\sigma_L, \\ m_0,\ldots,m_L}} M^{\sigma_1}_{1;m_0,m_1} \cdots M^{\sigma_L}_{L;m_{L-1},m_L} |\sigma_1 \cdots \sigma_L\rangle , \tag{6}$$

where m_0 and m_L are 1-dimensional dummy indices introduced for consistency. For a specific set of local states $\{\sigma_1 \ldots \sigma_L\}$, we use a single matrix $M^{\sigma_j}_j$ per site. We hence evaluate a matrix product to give a single entry $c_{\sigma_1 \ldots \sigma_L}$, resulting in the name "matrix-product" states. We will use σ_j to index the local physical states on site j. Depending on the bond dimensions m_j of the tensors M_j (also called virtual or auxiliary dimension), different quantum states can be represented exactly. If we let m_j grow exponentially by a factor of σ per site towards the center of the system, any quantum state can be written as an MPS. The entanglement between the part of an MPS to the left of bond $(j, j+1)$ and to the right of this bond is bounded by $\log(m_j)$. The required bond dimension $m = \max_j(m_j)$ to represent a quantum state exactly hence grows exponentially with its entanglement along left/right partitions. On the other hand, lowly-entangled states require only a small bond dimension, leading to an efficient representation of the quantum state [14,48,49].

2.3. Matrix-product operators (MPO)

By analogy to quantum states, we can express every operator as a matrix-product operator (MPO), i.e., a contraction of L rank-4 tensors W_j:

$$\hat{O} = \sum_{\substack{\sigma_1,\ldots,\sigma_L, \\ \sigma'_1,\ldots,\sigma'_L}} c_{\sigma_1 \ldots \sigma_L, \sigma'_1 \ldots \sigma'_L} |\sigma_1 \cdots \sigma_L\rangle \langle \sigma'_1 \cdots \sigma'_L| \tag{7}$$

Fig. 3. Schematic of the tensor network of a matrix-product operator (MPO). Horizontal lines denote the internal indices with bond dimension w, whereas the vertical lines denote physical indices with dimension σ.

$$= \sum_{\substack{\sigma_1,...,\sigma_L, \\ \sigma'_1,...,\sigma'_L, \\ w_0,...,w_L}} W^{\sigma_1,\sigma'_1}_{1;w_0,w_1} \ldots W^{\sigma_L,\sigma'_L}_{L;w_{L-1},w_L} |\sigma_1 \cdots \sigma_L\rangle\langle\sigma'_1 \cdots \sigma'_L| . \tag{8}$$

The only difference is that the tensor components W_j are now rank-4 tensors to account for the domain and image Hilbert spaces (see Fig. 3). There are different avenues to constructing matrix-product operators [59–61]. When used to represent the Hamiltonian $\hat{H}$ or other operators which are sums of local terms, the construction can be understood by splitting the system at bond j (connecting sites j and $j + 1$). We then separate terms of the operator that act only within their partition $\hat{H}^L_{j-1}$, $\hat{H}^R_{j+1}$ and those that connect the partitions $\hat{h}^L_{j;a_j}$, $\hat{h}^R_{j;a_j}$ [2]

$$\hat{H} = \hat{H}^L_{j-1} \otimes \hat{\mathbf{1}}^R_j + \hat{\mathbf{1}}^L_j \otimes \hat{H}^R_{j+1} + \sum_{a_j=1}^{N_j} \hat{h}^L_{j;a_j} \otimes \hat{h}^R_{j;a_j} . \tag{9}$$

Based on the tensor product structure there is an operator-valued matrix $\hat{W}_j$ which relates the partitioned representations between the bonds $j - 1$ and j; e.g., for the right partition we have

$$\begin{pmatrix} \hat{H}^R_{j-1} \\ \hat{h}^R_{j-1} \\ \hat{\mathbf{1}}^R_{j-1} \end{pmatrix} = \underbrace{\begin{pmatrix} \hat{\mathbf{1}}_j & \hat{C}_j & \hat{D}_j \\ & \hat{A}_j & \hat{B}_j \\ & & \hat{\mathbf{1}}_j \end{pmatrix}}_{\hat{W}_j} \otimes \begin{pmatrix} \hat{H}^R_j \\ \hat{h}^R_j \\ \hat{\mathbf{1}}^R_j \end{pmatrix} . \tag{10}$$

The operator-valued matrices $\hat{A}$, $\hat{B}$, $\hat{C}$ and $\hat{D}$ then define the recursion relations to iteratively build the complete operator $\hat{H}$. This picture directly leads to a construction of MPOs based on finite-state machines (FSM) [61,63].

In analogy to matrix-product states with bonds m_j and a maximal bond dimension m, the bonds of matrix-product operators are labeled by w_j with the maximal bond dimension denoted by w.

2.4. Canonical form

Considering Eq. (6), we can clearly insert resolutions of the identity XX^{-1} in between any two MPS tensors M_j and M_{j+1}. Multiplying X into M_j and X^{-1} into M_{j+1} changes the numerical content of

---

[2] This is actually what is called *superblock Hamiltonian* in the standard DMRG [62].
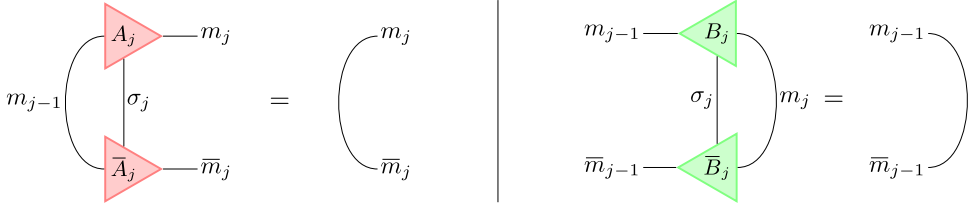
Fig. 4. Left (Right) normalized tensor A_j (red, right-pointing triangle) (B_j (green, left-pointing triangle)) contracted with its adjoint resulting in an identity.

each tensor while keeping the state invariant. This gauge freedom can be exploited to increase the numerical stability of the algorithm and simplify many tensor contractions. Two possible choices to fix the gauge are enforcing the *left* or *right* normalization of the tensors M_j. We write A_j for a left-normalized tensor and B_j for a right-normalized tensor with the defining property that

$$\sum_{\sigma_j, m_{j-1}} \bar{A}^{\sigma_j}_{j; m_{j-1}, \bar{m}_j} A^{\sigma_j}_{j; m_{j-1}, m_j} = \mathbf{1}_{\bar{m}_j, m_j} \tag{11}$$

$$\sum_{\sigma_j, m_j} B^{\sigma_j}_{j; m_{j-1}, m_j} \bar{B}^{\sigma_j}_{j; \bar{m}_{j-1}, m_j} = \mathbf{1}_{m_{j-1}, \bar{m}_{j-1}}. \tag{12}$$

Graphically they are represented by red (left-normalized) or green (right-normalized) triangles (see Fig. 4) where the orientation of the triangles also indicates the normalization. Left-/Right-canonical MPS are now defined by requiring that they consist of left-/right-normalized tensors only. Furthermore, a mixed-canonical MPS is defined by fixing a site j with unnormalized tensor M_j and demanding all site tensors to the left/right to be left-/right-normalized tensors (see Fig. 5). The unnormalized site is often called active site or orthogonality center.

2.5. Normalizing an MPS

Given an unnormalized MPS with site tensors M_j, the first step is often to bring it into a left or right-canonical form. This can be done by a series of QR decompositions: The tensor M_j is reshaped into a matrix $\tilde{M}$ with the left and physical tensor legs forming the rows and the right tensor leg the columns of the matrix, that is $\tilde{M}_{(\sigma_j m_{j-1}), m_j} = M^{\sigma_j}_{j; m_{j-1}, m_j}$. Applying a QR decomposition to this matrix, one obtains two new matrices Q_j and R_j (we label tensors living on bonds $(j, j+1)$ with an underlined index $\underline{j}$). The reshaping operation on M_j is done in reverse on Q to give the new left-normalized site tensor A_j while the transfer tensor $R_{\underline{j}}$ is multiplied into the next site tensor M_{j+1}. Likewise reshaping the M_j with right and physical legs as rows and left tensor leg as columns results in a right-normalized tensor B_j (with the transfer tensor multiplied into the previous site tensor M_{j-1}).

Starting on the left edge of the system and subsequently performing left-normalizations on each MPS tensor results in a complete left-normalized state. Equivalently, starting on the right edge of the system and moving to the left results in a right-normalized state.

2.6. Truncating an MPS

Operations on matrix-product states typically increase the bond dimension of the state (e.g. MPO–MPS applications or the addition of two MPS). Finding an optimal approximation to such a quantum state for a smaller bond dimension is the purpose of this section. This is of particular relevance to time-evolution methods, as entanglement generically grows during real-time evolution
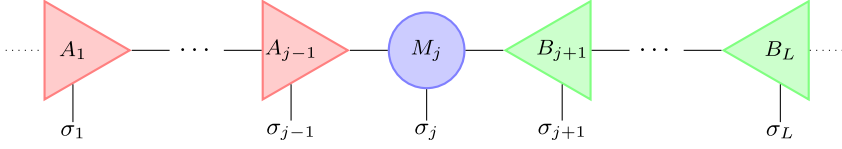
Fig. 5. MPS with active site j and consequently left (right) normalized site tensor left (right) of site j as defined in Eq. (11).

and time-evolved states hence per se already need a larger bond dimension than e.g. ground states. Hence finding good approximation methods is crucial.

Let us consider a state $|\psi\rangle$ which is represented by an MPS with an initial large bond dimension m. We wish to find another state $|\psi'\rangle$ with smaller bond dimension m' which approximates $|\psi\rangle$ well in the sense that it minimizes the Hilbert space distance

$$\| |\psi\rangle - |\psi'\rangle \| . \tag{13}$$

The most direct way to proceed is to use a series of singular value decompositions to successively truncate each bond of the MPS. On each individual bond, the optimal choice is made, but this does not have to result in the globally optimal state $|\psi'\rangle$. It is also possible to optimize each site tensor of $|\psi'\rangle$ sequentially to maximize the overlap between $|\psi\rangle$ and $|\psi'\rangle$. Multiple sweeps of this variational optimization can be done to approach the global optimum as well as possible.

2.6.1. Direct truncation via SVD

Consider a cut of the MPS on one bond $(j, j+1)$ into a left and right half such that the state $|\psi\rangle$ can be represented in effective left and right basis sets as

$$|\psi\rangle = \sum_{l,r=1}^{m_j} \Psi_{l,r} |l\rangle_L \otimes |r\rangle_R . \tag{14}$$

The coefficient tensor $\Psi_{l,r}$ then occupies the bond between sites j and $j+1$, such rank-2 tensors are called bond tensors in the following. For orthonormal left and right basis sets as realized in an MPS with orthogonality center on bond $(j, j+1)$, we can use a singular value decomposition (SVD, cf. Ref. [14] Sec. 4.5) of the Ψ tensor to obtain the approximation which is optimal under the condition that all other site tensors are kept fixed. That is, we decompose

$$\Psi_{l,r} = \sum_{s=1}^{m} U_{l,s} S_{s,s} V_{s,r}^\dagger \tag{15}$$

with U and $V^\dagger$ left-/right-unitary and S diagonal and non-negative. An approximation is then given by

$$\Psi'_{l,r} = \sum_{s=1}^{m' \leq m} U'_{l,s} S'_{s,s} V'^\dagger_{s,r} \tag{16}$$

where we only keep the m' largest singular values and also correspondingly truncate the rows and columns of U and $V^\dagger$. The error of this approximation of $|\psi\rangle$ is given by

$$\epsilon \approx \sqrt{\sum_{k=m'+1}^{m} S_{k,k}^2} \tag{17}$$

where the argument to the square root is also known as the discarded weight. In practice, one selects m' such that some target discarded weight (e.g. 10^{-10}, corresponding to error $\approx 10^{-5} \cdot L$) is obtained. Instead of working on the bond tensor $\Psi_{l,r}$ we can also apply the decomposition directly to the MPS tensor M_j by a suitable reshaping (cf. Figs. 6a and 6b). Starting e.g. on the left end of a

(a)

(b)
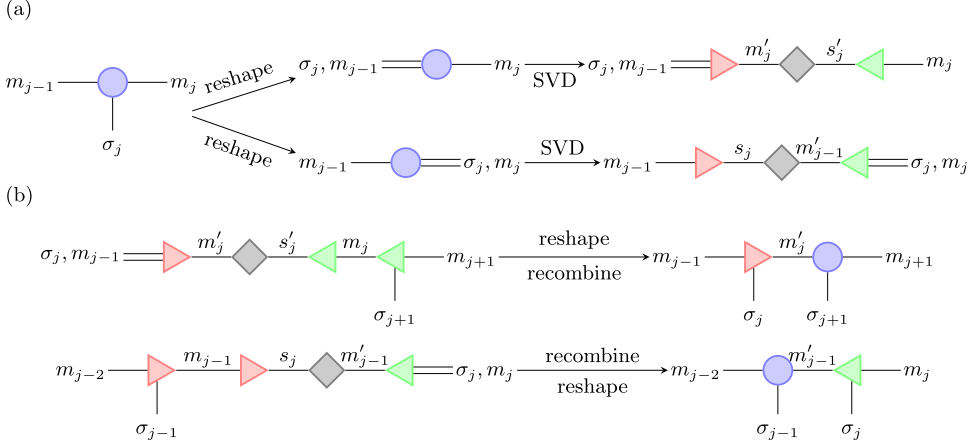
Fig. 6. (a) The singular value decomposition of a rank-three tensor M into USV within a truncation sweep to the right (top) or to the left (bottom). (b) Assignment of the result of the singular value decomposition (a) into the new left (right) normalized rank three tensor A_j (B_j) and the new active site M_{j+1} (M_{j-1}) at the top (bottom).

right-normalized MPS, we can sweep left-to-right and sequentially truncate every bond, each time obtaining the locally optimal state.

Due to this sweeping through the system, the truncation of site $L-1$ becomes dependent on the truncation of site 1 but not vice versa. In the case of small truncations the error resulting from this asymmetry is small and can be ignored. If truncation errors are large, however, this asymmetry leads to a series of optimal approximations which together do not constitute a globally optimal approximation, as each of the singular value decompositions always only produces the optimal approximation on a particular bond. To increase accuracy in this case, a subsequent variational optimization of the state may be necessary.
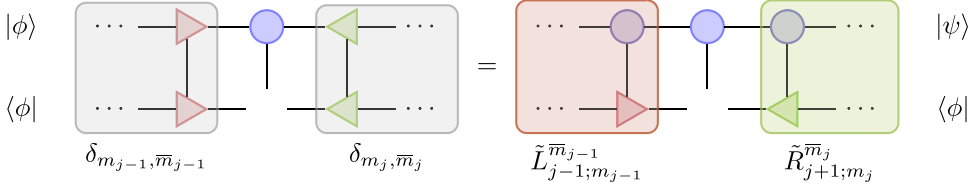
2.6.2. Variational truncation

To overcome the problems of the direct truncation by SVD, an iterative sweeping mechanism is often employed. By sweeping multiple times through the system and finding on each site the locally optimal tensor, it is more likely that one obtains the globally optimal MPS. We start from an initial guess state $|\phi\rangle$ with tensors M'_j and a chosen bond dimension m' and variationally minimize the distance to the un-truncated state $|\psi\rangle$ with tensors M_j. We must stress that the convergence of the variational optimization algorithm strongly depends on the initial guess state. Being unlucky and starting from an unsuitable initial state, the variational optimization may take a long time to converge or even may get completely stuck in a locally optimal but globally suboptimal state. Typically, a good choice for the initial state is produced by the direct truncation using the SVD. Once we have an initial state, the distance to minimize is given by

$$\| |\psi\rangle - |\phi\rangle \|^2 = \langle\psi|\psi\rangle - \langle\psi|\phi\rangle - \langle\phi|\psi\rangle + \langle\phi|\phi\rangle . \tag{18}$$

We now keep all tensors but M'_j fixed and hence only minimize the single site tensor M'_j, differentiating with respect to $\overline{M}'_j$. Because this tensor only occurs in the second half of Eq. (18), the new (optimized) M'_j can be obtained via

$$\frac{\partial}{\partial \overline{M}'_j} \left(\langle\phi|\phi\rangle - \langle\phi|\psi\rangle \right) = 0 . \tag{19}$$

Fig. 7. Iterative variational truncation considering the truncated state to be in a mixed canonical form. The left hand side can then be reduced to the new optimal site tensor M'_j. The right hand side, which needs to be considered completely, can nevertheless be calculated iteratively via the bond tensors $\tilde{L}$ and $\tilde{R}$.

Let us now consider that the truncated state is in a mixed canonical form with the active site at position j. Then the new tensor is given by

$$M'^{\sigma_j}_{j;m_{j-1},m_j} = \sum_{\bar{m}_{j-1}} \tilde{L}^{\bar{m}_{j-1}}_{j-1;m_{j-1}} \left(\sum_{\bar{m}_j} \tilde{R}^{\bar{m}_j}_{j+1;m_j} M^{\sigma_j}_{j;m_{j-1},m_j} \right) , \tag{20}$$

in which the left (right) parts of the tensor network (Fig. 7, right hand side) are contracted into $\tilde{L}$ $(\tilde{R})$,

$$\tilde{L}^{\bar{m}_{j-1}}_{j-1;m_{j-1}} = \sum_{\substack{\sigma_{j-1}, \\ \bar{m}_{j-2}, \\ m_{j-2}}} \bar{A}'^{\sigma_{j-1}}_{j-1;\bar{m}_{j-2},\bar{m}_{j-1}} \left(\cdots \left(\sum_{\substack{\sigma_1, \\ \bar{m}_0, \\ m_0}} \bar{A}'^{\sigma_1}_{1;\bar{m}_0,\bar{m}_1} M^{\sigma_1}_{1;m_0,m_1} \right) \cdots \right) M^{\sigma_{j-1}}_{j-1;m_{j-2},m_{j-1}} \tag{21}$$

$$\tilde{R}^{\bar{m}_j}_{j+1;m_j} = \sum_{\substack{\sigma_{j+1}, \\ \bar{m}_{j+1}, \\ m_{j+1}}} \bar{B}'^{\sigma_{j+1}}_{j+1;\bar{m}_j,\bar{m}_{j+1}} \left(\cdots \left(\sum_{\substack{\sigma_L, \\ \bar{m}_L, \\ m_L}} \bar{B}'^{\sigma_L}_{L;\bar{m}_{L-1},\bar{m}_L} M^{\sigma_L}_{L;m_{L-1},m_L} \right) \cdots \right) M^{\sigma_{j+1}}_{j+1;m_j,m_{j+1}} . \tag{22}$$

It is never necessary to calculate the complete contraction of the boundary tensors Eq. (21), because the next tensor in sweep direction was already calculated in the sweep before and the other tensor is obtained by reusing the one from the previous sweep step.

In practice, one should consider a two-site variational optimization in which neighboring site tensors M'_j, M'_{j+1} are optimized at the same time. This allows for flexibility in the bond dimension and distribution of quantum number sectors on the bond m_j. For convenience we will demonstrate the necessity of a two-site update to permit for an increased bond dimension. Consider the reshaped updated tensor

$$\Theta^{\sigma_j m_{j-1}}_{\sigma_{j+1} m_{j+1}} = M'^{\sigma_j}_{j;m_{j-1},m_j} \cdot M'^{\sigma_{j+1}}_{j+1;m_j,m_{j+1}} . \tag{23}$$

Any matrix factorization $\Theta = X \cdot Y$ may then create an index with intermediate bond dimension which is bounded by $\min\{m_{j-1}\sigma_j, \sigma_{j+1}m_{j+1}\}$ and potentially larger than m_j, which is possible iff at least two MPS sites are considered at the same time.

2.7. Finite temperatures

MPS by default can only represent pure quantum states. To describe the mixed states encountered at finite temperatures, one therefore requires an additional ingredient. Two possible choices exist within the MPS framework: minimally entangled typical thermal states [64] and thermofield

doublet states [8,65,66]. In the following, we will concentrate on thermofield doublet states or "purifications" of a mixed state; a detailed comparison between both methods can be found in [51]. The purification works by doubling the Hilbert space $\mathcal{H} \rightarrow \mathcal{H}_p \otimes \mathcal{H}_a$ where a denotes the auxiliary or ancilla degrees of freedom. The density matrix we wish to represent is just the trace over the auxiliary space of our purified quantum state $|\psi\rangle \in \mathcal{H}_p \otimes \mathcal{H}_a$:

$$\hat{\rho} = \text{Tr}_a |\psi\rangle\langle\psi| \quad . \tag{24}$$

At infinite temperature ($\beta = 0$), $\hat{\rho}$ is the identity matrix. Hence, an infinite-temperature quantum state in the grand-canonical ensemble can be constructed simply as a product state of maximally entangled states of physical and auxiliary sites. To construct a *canonical* infinite-temperature state, it is best to start with a complete vacuum state. In the Hubbard model, for instance, with creators on physical (auxiliary) sites denoted by $\hat{c}_{j,\sigma}^\dagger$ ($\hat{c}_{a(j),\sigma}^\dagger$), one applies the operator

$$\hat{C}_{\text{tot}}^\dagger = \sum_{j=1}^{L} \hat{c}_{j,\uparrow}^\dagger \hat{c}_{a(j),\uparrow}^\dagger + \hat{c}_{j,\downarrow}^\dagger \hat{c}_{a(j),\downarrow}^\dagger \tag{25}$$

repeatedly to create equal superpositions of particles until the desired filling is reached [67]. To obtain a finite-temperature state at β, we time-evolve the infinite-temperature initial state along the imaginary axis over a range $\beta/2$. The density matrix is then

$$\hat{\rho} \propto \text{Tr}_a\{e^{-(\beta/2)\hat{H}}|\psi\rangle\langle\psi|e^{-(\beta/2)\hat{H}}\} \propto e^{-\beta\hat{H}} \tag{26}$$

up to normalization which is taken care of by keeping the purified state $|\psi\rangle$ normalized at all times. For a more detailed discussion, cf. Ref. [14] Section 7.2.

The dynamics of this density operator is described by the von-Neumann equation. In the case of the purification, this is identical to the Liouville equation which has two copies of the Hamiltonian acting on both the physical and the auxiliary space [68], but with a negative sign in the auxiliary degrees of freedom. Formally, this corresponds to a time evolution, which is forward in time on the physical degrees of freedom, but backward in time on the auxiliary space. Despite this formal resemblance, as one traces out the auxiliary degrees of freedom, a gauge freedom exists, which allows to apply *any* unitary operator $\hat{V}_a$ acting only on the auxiliary space to the state $|\psi\rangle$. It is clear that this leaves the density matrix $\hat{\rho} = \text{Tr}_a|\psi\rangle\langle\psi| = \text{Tr}_a \hat{V}_a |\psi\rangle\langle\psi| \hat{V}_a^\dagger$ invariant. The simplest choice then is to obtain the real-time evolution by time-evolving the purified state with the Hamiltonian on the physical space, and leave the auxiliary degrees untouched.

Even more, the question arises, if there are unitaries acting on the auxiliary space, which may reduce the entanglement of the purified state $|\psi\rangle$ − finding such an optimized scheme would be extremely important in practice. Multiple choices for such unitary operators have been worked out, whose effect on computational efficiency depends on the system at hand. For example, during a real-time evolution of the physical system under the Hamiltonian $\hat{H}$, an easy and often helpful [69,70] choice is to evolve the auxiliary system under the Hamiltonian $-\hat{H}$, which formally resembles to the solution of the Liouville equation for the purified density operator as mentioned above. Finding further schemes is a topic of present research, and for details and more elaborate approaches we refer to the literature, e.g. Refs. [71,72]. Furthermore, also cf. Section 7.5 later for a detailed discussion on how specifically one should purify a density matrix when aiming to evaluate a series of time-dependent observables.

2.8. Application of an MPO to an MPS

One of the most important operations within the framework of matrix-product states is the application of an MPO to an MPS. In theory this can be done by a straightforward tensor multiplication of the corresponding site tensors of the MPS and MPO. In practice this is not the method of choice as in most applications the resulting target state $\hat{O}|\psi\rangle$ has a much higher bond dimension $m' = m \cdot w$, which, however, is mostly not required to represent the target state efficiently. It is therefore helpful to look at different approaches − nevertheless, for pedagogical reasons we will begin with the direct application before we turn to more elaborate application schemes.
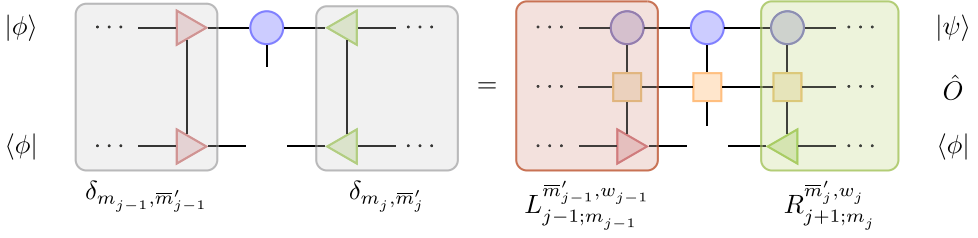
Fig. 8. The variational application of an MPO to an MPS considering the truncated guess state $|\phi\rangle$ to be in a mixed canonical form. The left hand side can then be reduced to the active side j that we want to optimize. The right hand side, which needs to be evaluated completely, can nevertheless be calculated iteratively via the bond tensors L_{j-1} and R_{j+1}.

2.8.1. Direct application

The direct application of an MPO to an MPS is obtained by regrouping the contractions such that the target MPS tensor can be obtained as a tensor product of the individual site tensors

$$\hat{O}|\psi\rangle = \sum_{\substack{\sigma_1,\ldots,\sigma_L, \\ \sigma'_1,\ldots,\sigma'_L, \\ \sigma''_1,\ldots,\sigma''_L}} \sum_{\substack{m_0,\ldots,m_L, \\ w_0,\ldots,w_L}} \left(W^{\sigma_1\sigma'_1}_{1;w_0,w_1} \cdots W^{\sigma_L\sigma'_L}_{L;w_{L-1},w_L} \right) |\sigma_1\ldots\sigma_L\rangle\langle\sigma'_1\ldots\sigma'_L|$$

$$\times \left(M^{\sigma''_1}_{1;m_0,m_1} \cdots M^{\sigma''_L}_{L;m_{L-1},m_L} \right) |\sigma''_1\ldots\sigma''_L\rangle \tag{27}$$

$$= \sum_{\substack{\sigma_1\cdots\sigma_L, \\ m_0,\ldots,m_L, \\ w_0,\ldots,w_L}} M'^{\sigma_1}_{1;(w_0 m_0),(w_1 m_1)} \cdots M'^{\sigma_L}_{L;(w_{L-1}m_{L-1}),(w_L m_L)} |\sigma_1\ldots\sigma_L\rangle = |\phi\rangle, \tag{28}$$

with the tensors M'_j given by

$$M'^{\sigma_j}_{j;(w_{j-1}m_{j-1}),(w_j m_j)} = \sum_{\sigma'_j} W^{\sigma_j\sigma'_j}_{j;w_{j-1},w_j} M^{\sigma'_j}_{j;m_{j-1},m_j}. \tag{29}$$
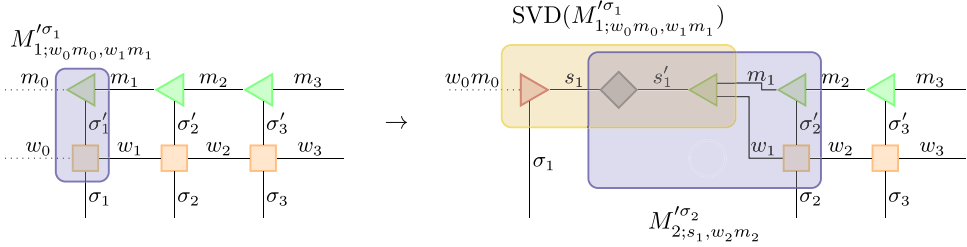
The resulting state $|\phi\rangle$ is therefore again an MPS, but with a larger dimension $m' = m \cdot w$ with w being the matrix dimension of the MPO site tensors W_j. Repeated application of an operator onto a state in this manner hence quickly increases the dimension of the state and truncation becomes necessary with computational costs scaling as $m^3 w^3 d$ per site if we do an SVD truncation or $m'^2 m w d$ per site for the variational compression.

2.8.2. Variational application

In the spirit of variationally compressing a state towards a target state we can try to apply the same considerations to the application of an MPO to a source state $|\psi\rangle$ with the subsequent compression of the state in one optimization step. Therefore we seek to minimize the distance between a guess state $|\phi\rangle$ with tensors M'_j and bond dimension m' and the source state with the MPO applied to it, which we denote as $|\hat{O}\psi\rangle$,

$$|\phi\rangle = \min_{|\phi\rangle} \left\| |\phi\rangle - |\hat{O}\psi\rangle \right\|^2 = \min_{|\phi\rangle} \left(\langle\phi|\phi\rangle - \langle\phi|\hat{O}\psi\rangle - \langle\hat{O}\psi|\phi\rangle + \langle\hat{O}\psi|\hat{O}\psi\rangle \right) \tag{30}$$

$$\Rightarrow \frac{\partial}{\partial\overline{M}'_j} \left(\langle\phi|\phi\rangle - \langle\phi|\hat{O}\psi\rangle \right) \overset{!}{=} 0 \tag{31}$$

Fig. 9. The essential steps of the zip-up method proposed in Ref. [64]. (left) Initial tensor network that shall be contracted, consisting of a right-canonical MPS and an MPO that only slightly destroy the canonical form. The first step is to interpret the combination of the first MPS tensor and the first MPO tensor as a new tensor $M'^{\sigma_1}_{1;w_0 m_0,w_1 m_1}$, which is slightly non-normalized and therefore framed in blue. (right) The next step is to apply an SVD (including a relaxed truncation) on the tensor M'_1 and finally build the next tensor M'_2 leaving the contracted, left-normalized MPS tensor $A'^{\sigma_1}_{1;m_0 w_0,s_1}$ on the left side.

for all guess site tensors $\overline{M}'_j$. If we keep the current guess state in a mixed canonical form, the above set of coupled equations again reduces to a local update scheme for the target tensors

$$M'^{\sigma_j}_{j;\overline{m}'_{j-1},\overline{m}'_j} = \sum_{m_{j-1},w_{j-1}} L^{\overline{m}'_{j-1},w_{j-1}}_{j-1;m_{j-1}} \left(\sum_{\sigma'_j,m_j,w_j} W^{\sigma_j\sigma'_j}_{j;w_{j-1},w_j} M^{\sigma'_j}_{j;m_{j-1},m_j} R^{\overline{m}'_j,w_j}_{j+1;m_j} \right), \tag{32}$$

where the boundary tensors L_{j-1}, R_{j+1} can be built recursively by sweeping through the system and evaluating the contractions

$$L^{\overline{m}'_{j-1},w_{j-1}}_{j-1;m_{j-1}} = \sum_{\substack{\sigma_{j-1},\sigma'_{j-1}, \\ \overline{m}'_{j-2},m_{j-2},w_{j-2}}} \overline{A}'^{\sigma_{j-1}}_{j-1;\overline{m}'_{j-2},\overline{m}'_{j-1}} L^{\overline{m}'_{j-2},w_{j-2}}_{j-2;m_{j-2}} W^{\sigma_{j-1}\sigma'_{j-1}}_{j-1;w_{j-2},w_{j-1}} M^{\sigma'_{j-1}}_{j-1;m_{j-2},m_{j-1}} \tag{33}$$

$$R^{\overline{m}'_j,w_j}_{j+1;m_j} = \sum_{\substack{\sigma_{j+1},\sigma'_{j+1}, \\ w_{j+1},\overline{m}'_{j+1},m_{j+1}}} W^{\sigma_{j+1}\sigma'_{j+1}}_{j+1;w_j,w_{j+1}} M^{\sigma'_{j+1}}_{j+1;m_j,m_{j+1}} R^{\overline{m}'_{j+1},w_{j+1}}_{j+2;m_{j+1}} \overline{B}'^{\sigma_{j+1}}_{j+1;\overline{m}'_j,\overline{m}'_{j+1}}. \tag{34}$$

The overall onsite contractions are depicted in Fig. 8. Care must be taken to use the (typically) optimal contraction order $\left((L_{j-1} \cdot A_j) \cdot W_j \right) \cdot \overline{A}_j$.

2.8.3. The zip-up method

An alternative to the direct application of an MPO to an MPS is the zip-up method described in Ref. [64]. The central assumption is that the operator, such as a time-evolution operator, only slightly destroys the canonical form of the MPS. Hence, a modest truncation is already possible during the contraction process without too much loss of information because there is only a small loss of orthogonality in the used left and right basis sets.

The first step is similar to Eq. (29), but we work with a right-normalized initial tensor B,

$$M'^{\sigma_1}_{1;w_0 m_0,w_1 m_1} = \sum_{\sigma'_1} W^{\sigma_1\sigma'_1}_{1;w_0,w_1} B^{\sigma'_1}_{1;m_0,m_1}. \tag{35}$$

Note that m_0 and w_0 are dummy indices and can therefore easily be fused into a new dummy index. Applying the SVD with a relaxed truncation criterion we obtain the left-normalized tensor A_1, with a single right index m'_1

$$\text{SVD}(M'^{\sigma_1}_{1;w_0 m_0,w_1 m_1}) \approx \sum_{s_1,s'_1} A'^{\sigma_1}_{1;m_0 w_0,s_1} S'_{\underline{1};s_1,s'_1} V'_{\underline{1};s'_1,w_1 m_1}. \tag{36}$$
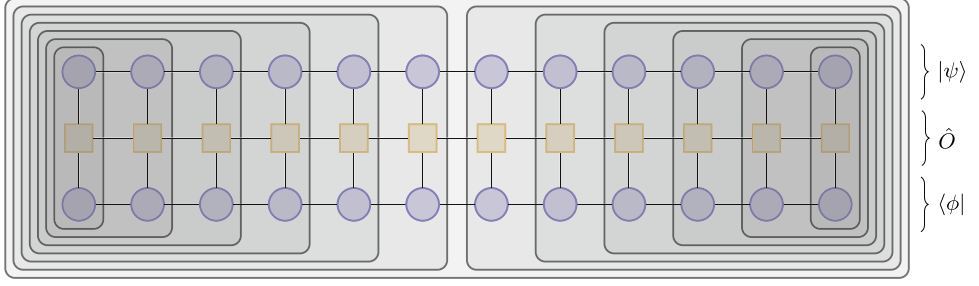
Fig. 10. The expectation value $\langle\phi|\hat{O}|\psi\rangle$ of an operator between two (possibly different) states represented as MPO and MPS respectively. The optimal contraction order is sideways, e.g. from left to right, as indicated by the shading. To add an additional column, the optimal evaluation order is $L_{j-1} \cdot M_j \cdot W_j \cdot M_j^\dagger$. An easy option for two-fold parallelization is the concurrent evaluation of L_{j-1} and R_j from left and right respectively. Here and in all following diagrams, we leave off the dummy left/right indices indicated by dotted lines earlier.

In the next step, the remaining parts of the result of the SVD are incorporated as before, now also including the MPO tensor, to obtain the next slightly unnormalized tensor

$$M'^{\sigma_2}_{2;s_1,w_2m_2} = \sum_{\sigma'_2,s'_1,m_1,w_1} S'_{1;s_1,s'_1} V'_{1;s'_1,w_1m_1} W^{\sigma_2\sigma'_2}_{2;w_1,w_2} B^{\sigma'_2}_{2;m_1,m_2} . \tag{37}$$

This procedure is depicted in Fig. 9. It is repeated until the right end of the system is reached and hence the complete operator is applied. For the relaxed truncation scheme a maximal growth factor of the MPS bond dimension $m' = 2m$ and a truncated weight of $1/10$ of the target weight turns out to be a suitable choice. The contraction then has complexity $2m^3w\sigma + 2m^2w^2\sigma^2$. The main cost is in the SVD of a $2m\sigma \times mw$ matrix at cost $\mathcal{O}(m^3\sigma w)$, i.e. linear in w. The MPS is in left canonical form now. A subsequent compression as described in Section 2.6 should be applied to obtain the resulting MPS with the target bond dimension.

Our experience. The zip-up method (Section 2.8.3) is typically sufficiently accurate and fast. In some cases, it is worthwhile to follow up on it with some additional sweeps of the variational optimization (Section 2.8.2) to increase accuracy. In particular if the bond dimension of the state is already large and the operator close to the identity (as would be the case for a small time step), variational optimization will converge quickly and accurately. It hence may be useful to also consider the MPS bond dimension when dynamically selecting the best application method. Conversely, when using the zip-up method, repeated truncation sweeps according to a defined singular value threshold should be used to obtain the most efficient representation of the state. Further interesting possibilities may be the truncation based on the left/right density matrix [73] or by optimizing the local density matrix [74] which aims at preserving local observables.

2.9. Expectation values

In standard dense numerical linear algebra, to evaluate the expectation value $\langle\phi|\hat{O}|\psi\rangle$, it is necessary to evaluate $\hat{O}|\psi\rangle$ and subsequently the overlap between $\langle\phi|$ and $\hat{O}|\psi\rangle$. As we have seen in the previous section, MPO–MPS products are relatively costly to evaluate. Luckily, the tensor network representing $\langle\phi|\hat{O}|\psi\rangle$ (cf. Fig. 10) allows many different contraction orders. Contracting it from left-to-right (using iteratively updated tensors L_j as defined in Eq. (33)) or right-to-left (using R_j as defined in Eq. (34) respectively) leads to an asymptotic contraction cost of $O\left(L\left(m^3w\sigma + m^2w^2\sigma^2\right)\right)$. The optimal contraction sequence is also indicated by the shading in Fig. 10.

Table 1

Approximate cost of various MPS and MPO operations. MPS bond dimensions are m and – if two MPS are used – n, MPO bond dimensions are w and the local dimension is σ. Different operations (matrix multiplication, QR and singular value decompositions) are given separately due to the very different cost of each. The overall factor L of the system size is not given explicitly, each method scales linearly in L. Particularly expensive terms are highlighted in bold.

Matrix multiplications	QR	SVD
Overlap of two MPS $mn^2\sigma + m^2n\sigma$	None	None
Expectation value of MPO between two MPS $mn^2\sigma w + mn\sigma^2 w^2 + m^2n\sigma w$	None	None
Variational truncation of MPS from m to n; single-site; one optimization sweep $2mn^2\sigma + \mathbf{2m^2n\sigma} + n^3\sigma$	$n^3\sigma$	None
Variational truncation of MPS from m to n; two-site; one optimization sweep $2mn^2\sigma + \mathbf{2m^2n\sigma} + mn\sigma^2 + n^3\sigma$	None	$\mathbf{n^3\sigma^3}$
Direct truncation of MPS from m to n with SVD (incl. initial normalization) $m^3 + \mathbf{m^3\sigma} + mn^2 + m^2n\sigma$	$\mathbf{m^3\sigma}$	$\min\{\mathbf{m^2n\sigma}, \mathbf{mn^2\sigma^2}\}$
Direct MPO–MPS application and subsequent direct truncation with SVD to bond dimension n $m^2\sigma^2 w^2 + m^3 w^3 + \mathbf{m^3\sigma w^3} + mn^2 w + m^2n\sigma w^2$	$\mathbf{m^3\sigma w^3}$	$\mathbf{mn^2\sigma^2 w}$
Zip-up MPO–MPS application with target bond dimension n and intermediate bond dimension $2n$ (incl. initial normalization; second line is the final truncation sweep from $2n$ to n) $m^3 + m^3\sigma + \mathbf{2m^2n\sigma w} + 2mn\sigma^2 w^2 + 4mn^2 w$ $+4n^3\sigma + 2n^3$	$m^3\sigma$	$\mathbf{4\sigma^2 n^2 mw}$ $+4n^3\sigma$
Variational MPO–MPS application with target bond dimension n; single-site; one optimization sweep $2mn^2\sigma w + 2mn\sigma^2 w^2 + \mathbf{2m^2n\sigma w} + n^3\sigma$	$n^3\sigma$	None
Variational MPO–MPS application with target bond dimension n; two-site; one optimization sweep $mn^2\sigma w + 3mn\sigma^2 w^2 + \mathbf{3m^2n\sigma w} + mn^2\sigma^2 w + n^3 + n^3\sigma$	None	$\mathbf{n^3\sigma^3}$

3. Overview of time-evolution methods

In the following sections, we will discuss in detail five different time-evolution methods for MPS which are currently in use to solve the time-dependent Schrödinger equation (TDSE). Each of them has different strengths and weaknesses, requiring a careful consideration of each individual problem to determine the most promising approach. In Section 8 we then examine some prototypical examples in an attempt to provide some guidance.

The first two methods (Section 4) approximate the time-evolution operator $\hat{U}(\delta) = e^{-i\delta\hat{H}}$, which is then repeatedly applied to the MPS $|\psi(t)\rangle$ to yield time-evolved states $|\psi(t + \delta)\rangle$, $|\psi(t + 2\delta)\rangle$ etc. The *time-evolving block decimation,* also known as the Trotter method and abbreviated here as TEBD, was developed around 2004 both for MPS [4,7,8] and the original classical DMRG [5,6]. It uses a Trotter–Suzuki decomposition of $\hat{U}(\delta)$, which is in particular applicable to short-ranged Hamiltonians. The time evolution is unitary up to the inherent truncation error, but the energy is typically not conserved. The main variants are using a second-order decomposition (TEBD2 in the following) and a fourth-order decomposition (TEBD4) to minimize the error due to the finite time step. While TEBD2 is essentially always useful, as it produces only two- to three times as many exponential terms as a first-order decomposition, TEBD4 produces four to five times as many exponentials as TEBD2. Depending on the desired error, this may or may not be worthwhile. In contrast, the *MPO $W^{I,II}$* method was introduced only recently [40] and relies on a different decomposition scheme for the time-evolution operator $\hat{U}(\delta)$ particularly suited to construct an efficient representation as a matrix-product operator. It can directly deal with long-range interactions and generally generates smaller MPOs than TEBD. Its primary downside is that the evolution is not exactly unitary. The time step error of both TEBD and the MPO $W^{I,II}$ method is larger than in the other methods described below.

Compared to these methods, algorithms based on Krylov subspaces [75,76] directly approximate the action of $\hat{U}(\delta)$ onto $|\psi\rangle$, i.e. produce a state $|\psi(t + \delta)\rangle$ without explicitly constructing $\hat{U}(\delta)$ in

the full Hilbert space while preserving the unitarity of the time evolution. The main advantage lies in a very precise evaluation of $|\psi(t+\delta)\rangle$ with a very small inherent finite time-step error [77].

The *global Krylov* algorithm (Section 5) is related to the time-dependent Lanczos method in exact diagonalization approaches to time evolution [24–26] and has only partially [27–29] been adapted to the MPS setting. Interestingly, evaluation of observables on a very fine time grid (e.g. $\delta = 10^{-5}$) is possible, which would be prohibitively expensive using any of the time-stepping methods. The downside of this global Krylov method is its need to represent potentially highly entangled Krylov vectors as matrix-product states.

These highly-entangled global Krylov vectors do not need to be represented if instead of working globally, one works in the local reduced basis (Section 6). From a DMRG perspective, this can be seen as a variant of the time-step targeting method [32–35]. Its primary objective is to solve the TDSE locally on each pair of lattice sites to construct an optimal MPS for the time-evolved state. Conversely, this local method can no longer evaluate observables precisely at arbitrarily-small time steps $\delta' \ll \delta$ (as no complete MPS is available at such intermediate times) but works much like TEBD2 as a time-stepper, producing a single final state $|\psi(t+\delta)\rangle$. In contrast to TEBD and the MPO W^{II} it allows for the treatment of arbitrary Hamiltonians. An uncontrolled projection error may, however, lead to incorrect results as the MPS-projected Hamiltonian cannot represent the actual physical Hamiltonian well if the MPS used for the projection is very small. A further development of this approach is the *time-dependent variational principle*[36,37] (TDVP). The TDVP can be considered an approach to remedy the dominant errors in the local Krylov approach by a thorough theoretical analysis leading to an optimized evolution method. Its implementation is nearly identical to the local Krylov method, but the different derivation of the local equations leads to smaller errors because it arrives at a closed solution of a series of coupled equations. In particular, using the two-site variant of TDVP (2TDVP), we know that nearest-neighbor Hamiltonians do not incur a projection error which is often the primary error source in the local methods. The single-site variant (1TDVP) has a larger projection error and also always works at constant MPS bond dimension but violates neither unitarity nor energy conservation during the time evolution.

We complement this part of the review by presenting a subjective selection of useful tricks which are applicable to most of the time-evolution methods and which can help in the treatment of complicated systems. We will discuss in some detail: (i) how to combine the time evolution in the Heisenberg and the Schrödinger picture, respectively, to reach longer times; (ii) how to select time steps to increase the accuracy of the integrator; (iii) how removing low-lying eigenstates and the application of linear prediction helps in calculating Green's functions; (iv) how to specifically select the optimal choice of purification schemes for finite-temperature calculations; and (v) briefly summarize the local basis optimization technique to treat systems with many bosonic degrees of freedom.

4. Approximations to $\hat{U}(\delta)$

The following two subsections will summarize two popular choices to approximate the time-evolution operator $\hat{U}(\delta)$, which, when applied to an MPS $|\psi(t)\rangle$ gives a new MPS $|\psi(t+\delta)\rangle$ at time $t+\delta$. Future developments to construct such $\hat{U}(\delta)$ in an easier and more generic manner are conceivable.

4.1. Time-evolving block decimation (TEBD) or Trotter decomposition

At its heart, this method relies on a Trotter–Suzuki [39] decomposition and subsequent approximation of the time-evolution operator $\hat{U}^{\text{exact}}(\delta)$. It was developed around 2004 both in the (then new) context of MPS [4,8] and the classical system-environment DMRG [5,6] and remained the most popular time-evolution method for both settings in the following years.[3]

---

[3] In these papers, TEBD is introduced without relying on MPO/MPS arithmetic but instead by applying quantum gates to individual state tensors. Here, we will employ MPOs to simplify notation and implementation.

As an illustrative example let us first consider the nearest-neighbor Heisenberg chain. Its Hamiltonian is given by

$$\hat{H} = \sum_j \hat{h}_{j,j+1} \tag{38}$$

$$\hat{h}_{j,j+1} = \hat{s}_j^x \hat{s}_{j+1}^x + \hat{s}_j^y \hat{s}_{j+1}^y + \hat{s}_j^z \hat{s}_{j+1}^z \ . \tag{39}$$

The exact evolution is then given by

$$\hat{U}^{\text{exact}}(\delta) = e^{-i\delta\hat{H}} \ . \tag{40}$$

If we split $\hat{H}$ into two summands $\hat{H}_{\text{even}}$ and $\hat{H}_{\text{odd}}$ as

$$\hat{H}_{\text{even}} = \sum_{j \text{ even}} \hat{h}_{j,j+1} \tag{41}$$

$$\hat{H}_{\text{odd}} = \sum_{j \text{ odd}} \hat{h}_{j,j+1} \tag{42}$$

$$\hat{H} = \hat{H}_{\text{even}} + \hat{H}_{\text{odd}} \ , \tag{43}$$

we can use the Baker–Campbell–Hausdorff formula to approximate

$$\hat{U}^{\text{exact}}(\delta) = e^{-i\delta\hat{H}} \tag{44}$$

$$= e^{-i\delta\hat{H}_{\text{even}}} e^{-i\delta\hat{H}_{\text{odd}}} e^{-i\delta^2 \left[\hat{H}_{\text{even}}, \hat{H}_{\text{odd}}\right]} \tag{45}$$

$$\approx e^{-i\delta\hat{H}_{\text{even}}} e^{-i\delta\hat{H}_{\text{odd}}} \tag{46}$$

$$\equiv \hat{U}^{\text{TEBD1}}(\delta) \ . \tag{47}$$

Now, both $e^{-i\delta\hat{H}_{\text{even}}}$ and $e^{-i\delta\hat{H}_{\text{odd}}}$ are easy to evaluate: All the summands $\hat{h}_{j,j+1}$ commute with each other so it suffices to exponentiate them individually:

$$e^{-i\delta\hat{H}_{\text{even}}} = e^{-i\delta \sum_{j \text{ even}} \hat{h}_{j,j+1}} = \prod_{j \text{ even}} e^{-i\delta\hat{h}_{j,j+1}} \ . \tag{48}$$

Writing $e^{-i\delta\hat{h}_{j,j+1}}$ into an MPO, we have identity tensors with bond dimension $w = 1$ everywhere except on sites j and $j+1$. Splitting the two-site gate into two individual tensors (e.g. with an SVD) results in a bond between those sites which has at most dimension σ^2. Hence writing $e^{-i\delta\hat{H}_{\text{even}}}$ as an MPO is also efficient, it has alternating bond dimensions σ^2 and 1. Accordingly, $e^{-i\delta\hat{H}_{\text{odd}}}$ written as an MPO has alternating bond dimensions 1 and σ^2 and their product $e^{-i\delta\hat{H}_{\text{even}}} e^{-i\delta\hat{H}_{\text{odd}}}$ can be written with bond dimensions σ^2 throughout.

If we keep the decomposition of $\hat{H}$ into $\hat{H}_{\text{even}}$ and $\hat{H}_{\text{odd}}$ fixed and hence the commutator $\left[\hat{H}_{\text{even}}, \hat{H}_{\text{odd}}\right]$ constant, the error per time step δ between $\hat{U}^{\text{TEBD1}}(\delta)$ and $\hat{U}^{\text{exact}}(\delta)$ is of second order:

$$\hat{U}^{\text{exact}}(\delta) = \hat{U}^{\text{TEBD1}}(\delta) + O(\delta^2) \tag{49}$$

because we can approximate

$$e^{-i\delta^2 \left[\hat{H}_1, \hat{H}_2\right]} \approx \hat{\mathbf{1}} - i\delta^2 \left[\hat{H}_1, \hat{H}_2\right] \ . \tag{50}$$

Considering a longer interval T, which we divide into T/δ smaller intervals, at which we apply $\hat{U}^{\text{TEBD1}}(\delta)$, our errors accumulate to $T/\delta \, O(\delta^2) = O(\delta)$ per time interval T. The error is hence of first order in δ which gives $\hat{U}^{\text{TEBD1}}(\delta)$ its name as first-order TEBD. By symmetrizing the decomposition, it is straightforward to construct

$$\hat{U}^{\text{TEBD2}}(\delta) \equiv e^{-i\delta/2\hat{H}_{\text{even}}} e^{-i\delta\hat{H}_{\text{odd}}} e^{-i\delta/2\hat{H}_{\text{even}}} \ , \tag{51}$$

which has a third-order error per step

$$\hat{U}^{\text{exact}}(\delta) = \hat{U}^{\text{TEBD2}}(\delta) + O(\delta^3) \tag{52}$$

and hence a second-order error per time interval. Similarly, we can construct a fourth-order TEBD evolution operator as

$$\hat{U}^{\text{TEBD4}}(\delta) \equiv \hat{U}^{\text{TEBD2}}(\delta_1)\hat{U}^{\text{TEBD2}}(\delta_1)\hat{U}^{\text{TEBD2}}(\delta_2)\hat{U}^{\text{TEBD2}}(\delta_1)\hat{U}^{\text{TEBD2}}(\delta_1) \tag{53}$$

with

$$\delta_1 \equiv \frac{1}{4 - 4^{1/3}}\delta \tag{54}$$

$$\delta_2 \equiv (1 - 4\delta_1)\,\delta\ . \tag{55}$$

We then have

$$\hat{U}^{\text{exact}}(\delta) = \hat{U}^{\text{TEBD4}}(\delta) + O(\delta^5) \tag{56}$$

and hence a fourth-order error per time step.

   This approach can be generalized to more complicated divisions of the full Hamiltonian $\hat{H}$ into N_H internally-commuting parts $\hat{H}_1, \hat{H}_2, \ldots$:

$$\hat{H} = \sum_{\alpha=1}^{N_H} \hat{H}_\alpha\ , \tag{57}$$

where each $\hat{H}_\alpha$ is a sum

$$\hat{H}_\alpha = \sum_{k=1}^{N_\alpha} \hat{h}_\alpha^k \tag{58}$$

such that $\hat{h}_\alpha^k$ can be diagonalized efficiently and $[\hat{h}_\alpha^k, \hat{h}_\alpha^l] = 0$. The local terms $\hat{h}_\alpha^k$ do not need to be limited to two sites, but it must still be possible to evaluate their exponential efficiently. A first-order TEBD time evolution operator $\hat{U}^{\text{TEBD1}}(\delta)$ can then be written as

$$\hat{U}^{\text{TEBD1}}(\delta) = e^{-i\delta\hat{H}_1}e^{-i\delta\hat{H}_2}e^{-i\delta\hat{H}_3}\cdots e^{-i\delta\hat{H}_{N_H}} \tag{59}$$

$$= \prod_{\alpha=1}^{N_H} e^{-i\delta\hat{H}_\alpha}\ . \tag{60}$$

The error per time interval is still of first order as before. The second-order decomposition is likewise obtained as a symmetrization of the $\hat{U}^{\text{TEBD1}}(\delta)$:

$$\hat{U}^{\text{TEBD2}}(\delta) = e^{-i\delta/2\hat{H}_1}e^{-i\delta/2\hat{H}_2}\cdots e^{-i\delta/2\hat{H}_{N_H-1}}e^{-i\delta\hat{H}_{N_H}}e^{-i\delta/2\hat{H}_{N_H-1}}\cdots e^{-i\delta/2\hat{H}_2}e^{-i\delta/2\hat{H}_1} \tag{61}$$

$$= \prod_{\alpha=1}^{N_H} e^{-i\delta/2\hat{H}_\alpha}\prod_{\alpha=N_H}^{1} e^{-i\delta/2\hat{H}_\alpha}\ . \tag{62}$$

It has the same second-order error per time step as above, an exemplary tensor network for the case of a three-site interaction is given in Fig. 11. A fourth-order decomposition could be constructed, but the number of terms required grows very quickly with the number of summands N_H in the Hamiltonian, making the fourth-order decomposition less attractive. Recently, Ref. [78] investigated optimized decomposition schemes beyond the naive approaches presented here. These schemes have the potential to further reduce errors and the number of exponentials required, but still need to be tested in practice.
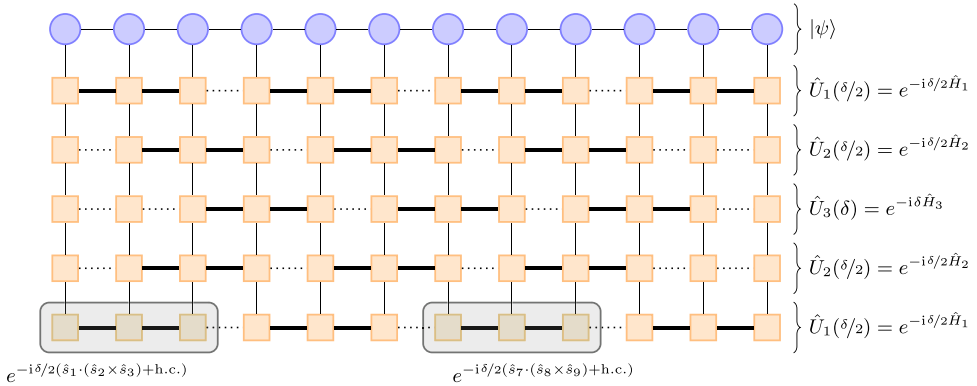
Fig. 11. Exemplary structure of a tensor network representing a time-evolved state $\hat{U}^{\text{TEBD2}}(\delta)|\psi\rangle$ of 12 sites. The state is written as an MPS $|\psi\rangle$ at the top and evolved using TEBD2 under the Hamiltonian $\hat{H} = \sum_{j=1}^{10} \hat{s}_j \cdot (\hat{s}_{j+1} \times \hat{s}_{j+2}) + \text{h.c.}$ The three-site action of the Hamiltonian necessitates the decomposition into three sums $\hat{H}_{1,2,3}$. To achieve a second-order error, the evolution is symmetrized by applying the operators $\hat{U}_{1,2,3}$ in reverse order after each half-step. The individual time-evolution operators written as MPOs have two thick bonds alternating with a 1-dimensional dummy bond.

4.1.1. Errors

TEBD suffers from two errors, both of which are controllable and can be estimated fairly straightforwardly. The first is the time step error of order $O(\delta^2)$ per time step for first-order TEBD and $O(\delta^3)$ per time step for second-order TEBD. With a fixed total interval T divided into $N = T/\delta$ steps, the error over the entire interval is $O(\delta)$ and $O(\delta^2)$ respectively for first and second order TEBD. This time step-inherent error does not violate unitarity of a real-time evolution as each of the constructed operators $e^{-i\delta\hat{H}_\alpha}$ is unitary. However, the energy and other conserved quantities are not necessarily constant if the time-step error is large. The primary option to reduce this discretization error is to chose a smaller time-step size δ or a higher-order decomposition like TEBD4. It is not always optimal to use TEBD4, as higher-order decompositions typically result in more MPOs to apply to the state sequentially to do a single time step. For example, if a Trotter error of 10^{-2} per unit time is desired, 1TEBD needs a step size $\delta = 0.01$ and accordingly 100 steps to reach time $t = 1$. For the same error, TEBD2 needs a step size $\delta = 0.1$ and hence 10 steps to reach $t = 1$ while TEBD4 could work with a step size of $\sqrt{0.1} \approx 0.3$ and hence approximately three steps. However, where TEBD1 (e.g.) needs three exponentials $e^{-i\delta H_1}e^{-i\delta H_2}e^{-i\delta H_3}$, TEBD2 has to apply five exponentials $e^{-i\delta/2H_1}e^{-i\delta/2H_2}e^{-i\delta H_3}e^{-i\delta/2H_2}e^{-i\delta/2H_1}$ (or four, if we combine $e^{-i\delta/2H_1}$ of neighboring steps) and TEBD4 has to work with 21 (20 when combining neighboring steps) exponentials. That is, while TEBD4 only needs three times fewer steps than TEBD2 for a final error of 10^{-2}, each individual step needs approximately five times as much work. In comparison, if we target an error of 10^{-8}, TEBD1 needs 10^8 steps, TEBD2 needs 10^4 steps and TEBD4 only needs 10^2 steps − for doing five times as much work per step, we reduce the number of steps by a factor of 100 when going from TEBD2 to TEBD4.

As an alternative to such higher-order decompositions, one may combine a large-scale Trotter decomposition with a small-scale Krylov-based approach to make the *commutator* of the Hamiltonian summands $\hat{H}_\alpha$ smaller [79].

The second source of errors is due to the mandatory truncation of the time-evolved state to a realistic bond dimension. This truncation error can also be controlled and measured by the discarded weight during the truncation as usual. It affects both the unitarity of the evolution and the conserved quantities. By gradually increasing the MPS bond dimension, it is straightforward to estimate convergence of the calculation with respect to this error.

4.1.2. Long-range interactions

Long-range terms in the Hamiltonian can be treated by introducing swap gates [64] in order to exponentiate each individual contribution. Each swap gate exchanges the physical indices of two nearest neighbors, hence moving the interacting sites next to each other. The evolution term $e^{-i\delta \hat{h}_\alpha^k}$ is then sandwiched between two series of swap gates which first bring the sites onto which $\hat{h}_\alpha^k$ acts next to each other and then, after the evolution, move them back to their original places. A smart ordering of the individual evolution terms often allows to cancel two series of swap gates from different $\hat{h}_\alpha^k$ and $\hat{h}_\alpha^l$ against each other. Nearly the entire overhead of such swap gates can be removed in this way.

4.1.3. Algorithm

The two steps of the TEBD method (constructing the evolution operator and repeatedly applying it to the state) are described in Alg. 1 at the example of the second-order TEBD2 method.

To construct the time-evolution operator, one first needs to split the Hamiltonian into internally-commuting parts $\hat{H}_\alpha$ (cf. Eq. (58)). Each summand $\hat{h}_\alpha^k$ of each part is then individually exponentiated as $\exp\left(-i\delta \hat{h}_\alpha^k\right)$. If the summand is non-local, swap gates are introduced before and after the exponentiation such that the exponentiation works on nearest-neighbor sites. One then obtains a number of MPO component tensors (one for each site between the leftmost and rightmost sites on which $\hat{h}_\alpha^k$ acts). These tensors are placed within an MPO spanning the entire system to eventually yield $\hat{U}_\alpha$. Once all $\hat{U}_\alpha$ are constructed (in a specific, selected ordering), one has a representation of the first-order decomposition of $\hat{U}$. Going from this first-order decomposition to a second-order decomposition is very straightforward, one only has to first replace δ by $\delta/2$ in the initial exponentiation and second append the reversed list of MPOs to itself, i.e. store a list $\{\hat{U}_1, \hat{U}_2, \hat{U}_3, \hat{U}_3, \hat{U}_2, \hat{U}_1\}$. This new $\hat{U}$ is then symmetric, each half moves time forward by $\delta/2$ for a total step of size δ.

To then time-evolve the state, one only has to apply $\hat{U}$ repeatedly to it, each time advancing the state by a step δ in time. In practice, we found it useful to write the time-evolved state $|\psi(t)\rangle$ to disk after every step. In this way, the time evolution can proceed as quickly as possible and evaluations (e.g. for expectation values, overlaps with other states etc.) may proceed in parallel by using the generated state files, possibly even on separate computers.

An interesting avenue to parallelization [80] in TEBD can be used if instead of working in the MPO framework all n-site gates are kept separate. For very small time steps, gates should not affect singular value spectra on far-away bonds. It should hence be possible to apply many gates independently and also truncate many bonds independently (using a slightly different gauge not discussed here). While not extensively utilized in the literature yet, such real-space parallelization promises much larger speed-ups than local (e.g. BLAS) parallelization options if the associated errors are kept sufficiently small.

Splitting or combining $\hat{U}(\delta)$. It may provide an additional computational advantage to combine multiple MPOs representing $\hat{U}_\alpha(\delta)$ into one larger MPO representing $\hat{U}(\delta)$. For example in the case of the nearest-neighbor Heisenberg model, one arrives naively at two MPOs $\hat{U}_{\text{even}}$ and $\hat{U}_{\text{odd}}$, each exponential of $\hat{H}_{\text{even}}$ and $\hat{H}_{\text{odd}}$ respectively. These MPOs have alternating bond dimensions $1-4-1-4-\cdots$ and $4-1-4-1-\cdots$. Multiplying both MPOs together, one obtains a single new MPO with bond dimension 4 throughout the system. Applying the new MPO only requires a single sweep over the system, whereas applying the two original MPOs requires two sequential sweeps. While the computational effort to leading order is identical, the former option is much more cache friendly, requiring access of each individual site tensor from RAM (or even a hard disk drive) only once instead of twice.

Similarly, if the Hamiltonian needs to be split into multiple terms $\hat{H}_\alpha$, it is potentially helpful to first multiply the MPOs for $\hat{U}_\alpha$ and then apply the resulting $\hat{U}$ to the MPS. In general, applying many $\hat{U}_\alpha$ with very small bond dimensions leads to a large overhead, whereas $\hat{U}_\alpha$ with very large bond dimensions are more costly to multiply with states. The optimal trade-off depends on the problem at hand. As a rule of thumb, one should multiply two time-evolution operators if this does not increase

the maximal bond dimension (as in the example above). If the bond dimension of the time-evolution operators is very small, it might also make sense to combine them until a comparatively large bond dimension (e.g. $w \approx 10$) is obtained which reduces the overhead associated to many small matrix multiplications.

Algorithm 1 The second-order TEBD2 method. The main input is an analytic expression for the Hamiltonian $\hat{H}$ necessary to split it into its constituent internally-commuting parts. Additionally, one needs to select an application method *apply* to evaluate MPO–MPS products (including truncation as required). To run, one first calls PREPARE-$U(\delta)$ and then repeatedly TIMESTEP to generate the time-evolved states.

1: **procedure** PREPARE-$U_\alpha(\delta)$(Analytic expression for $\hat{H}_\alpha$, step size δ)
2:     **for each** summand $\hat{h}_\alpha^k$ of $\hat{H}_\alpha$ **do**
3:        $\hat{u}_\alpha^k \leftarrow$ EXP$(-\mathrm{i}\delta\hat{h}_\alpha^k)$ ▷ If $\hat{h}_\alpha^k$ is long-range, swap gates are used around this line
4:        Split $\hat{u}_\alpha^k$ into MPO component tensors by SVD or QR
5:        Place those tensors on appropriate sites of MPO $\hat{U}_\alpha$
6:     **end for**
7:     **return** $\hat{U}_\alpha$
8: **end procedure**
9: **procedure** PREPARE-$U(\delta)$(Analytic expression for $\hat{H}$, step size δ)
10:     Split $\hat{H}$ into internally-commuting parts $\{\hat{H}_1, \ldots, \hat{H}_{N_H}\}$ ▷ typically done by hand
11:     $\hat{U} \leftarrow \{$PREPARE-$U_\alpha(\delta/2)(\hat{H}_1), \ldots,$ PREPARE-$U_\alpha(\delta/2)(\hat{H}_{N_H})\}$
12:     $\hat{U} \leftarrow \hat{U} +$ REVERSE$(\hat{U})$ ▷ Construct second-order Trotter decomposition, hence $\delta/2$
13:     Multiply pairs of constituents $\hat{U}_\alpha, \hat{U}_{\alpha+1}$ of $\hat{U}$ if possible within bond dimensions
14:     **return** $\hat{U}$
15: **end procedure**
16: **procedure** TIMESTEP(List of MPOs $\hat{U}$, initial state $|\psi\rangle$, application method *apply*)
17:     **for each** term $\hat{U}_\alpha$ in $\hat{U}$ **do**
18:        $|\psi\rangle \leftarrow$ APPLY$(\hat{U}_\alpha, |\psi\rangle)$
19:     **end for**
20:     **return** $|\psi\rangle$
21: **end procedure**

4.2. The MPO $W^{\mathrm{I,II}}$ method

In Ref. [40] Zaletel et al. proposed a generalization of the Euler approximation of the operator exponential $e^{-\mathrm{i}\delta\hat{H}}$, which can be implemented efficiently using MPOs. In this scheme, the error per site is independent of the system size. Furthermore, the construction is capable of dealing with long-ranged interaction terms making it suitable for two-dimensional systems, too. In this section, we will sketch the derivation of the MPO representations $W^{\mathrm{I,II}}$ and discuss how to construct these operators as well as the numerical errors and the stability of the overall method.

4.2.1. Motivation and construction

The general idea is to exploit the intrinsic factorization of MPO representations of operators. We consider operators which have a local structure in the sense that they are given by a sum of terms $\hat{H}_j$ acting on a subset of the lattice, starting at site j: $\hat{H} = \sum_j \hat{H}_j$. The Euler approximation of the operator exponential is $e^{-\mathrm{i}\delta\hat{H}} = 1 - \mathrm{i}\delta\sum_j \hat{H}_j + \mathcal{O}(\delta^2)$ and since there are $\sim L^2$ contributions from all possible combinations of local terms with finite support the error *per site* is $\sim L\delta^2$; hence the approximation becomes more and more unstable with increasing system size. Ref. [40] introduced

the following local version of the Euler stepper

$$e^{-i\delta\hat{H}} = 1 - i\delta \sum_j \hat{H}_j - \delta^2 \sum_{j<k}{}' \hat{H}_j\hat{H}_k + \mathcal{O}(L\delta^2) + \mathcal{O}(\delta^3) \tag{63}$$

$$\approx \prod_j (1 - i\delta\hat{H}_j) \equiv \hat{U}^{\mathrm{I}}(\delta) , \tag{64}$$

where the primed sum indicates that the local operator terms $\hat{H}_j$, $\hat{H}_k$ do not act on a common subset of the lattice, i.e., they do not overlap. Even though the error of $\hat{U}^{\mathrm{I}}$ is still of order δ^2 in the step size there are only $\mathcal{O}(L)$ contributions which are missed, namely those combinations of local terms with overlapping support. Hence, the overall error is bounded by $\mathcal{O}(L\delta^2)$, and thus the error per site is constant in the system size.

Recall now the decomposition of an MPO into a left, right and local part

$$\hat{H} = \hat{H}^L_{j-1} \otimes \hat{\mathbf{1}}^R_j + \hat{\mathbf{1}}^L_j \otimes \hat{H}^R_{j+1} + \sum_{a_j=1}^{N_j} \hat{h}^L_{j;a_j} \otimes \hat{h}^R_{j;a_j} . \tag{65}$$

with N_j interaction terms crossing bond j, or equivalently:

$$\begin{pmatrix} \hat{H}^R_{j-1} \\ \hat{h}^R_{j-1} \\ \hat{\mathbf{1}}^R_{j-1} \end{pmatrix} = \underbrace{\begin{array}{c} {}^1 \\ {}^{N_{j-1}} \\ {}^1 \end{array} \begin{pmatrix} \hat{\mathbf{1}}_j & \hat{C}_j & \hat{D}_j \\ & \hat{A}_j & \hat{B}_j \\ & & \hat{\mathbf{1}}_j \end{pmatrix}}_{\hat{W}_j} \otimes \begin{pmatrix} \hat{H}^R_j \\ \hat{h}^R_j \\ \hat{\mathbf{1}}^R_j \end{pmatrix} . \tag{66}$$

The operator-valued matrices $\hat{A}_j$, $\hat{B}_j$, $\hat{C}_j$, $\hat{D}_j$ specify the local structure of the interactions described by $\hat{H}$ and the set of all matrices $\hat{W}_1 \ldots \hat{W}_L$ define the MPO representation of $\hat{H}$. The MPO representation W^{I} of $\hat{U}^{\mathrm{I}}$ is given by

$$\hat{W}^{\mathrm{I}}_j = \begin{pmatrix} \hat{\mathbf{1}}_j + \delta\hat{D}_j & \sqrt{\delta}\hat{C}_j \\ & \\ \sqrt{\delta}\hat{B}_j & \hat{A}_j \end{pmatrix} . \tag{67}$$

In this case the representation of $\hat{H}$ as finite state machine is particularly useful since it permits to directly deduce the MPO representation for W^{I}. The MPO bond dimension of W^{I} is $w - 1$ with w the bond dimension of $\hat{H}$. Hence, this MPO can be applied numerically very efficiently.

However, the restriction of W^{I} to treat only non-overlapping local operator terms $\hat{H}_j$ is very strong and fails to even reproduce the correct time evolution of a purely on-site Hamiltonian. For example, W^{I} for $\hat{H} = \sum_j \hat{s}^z_j$ generates only operator strings $\hat{s}^z_k \cdots \hat{s}^z_{k+n}$ but not $\hat{s}^z_k \hat{s}^z_k$. An improvement

is to permit operator strings that overlap on one site:

$$\hat{U}^{\mathrm{II}}(\delta) = 1 - \mathrm{i}\delta \sum_j \hat{H}_j - \frac{\delta^2}{2} {\sum_{j,k}}'' \hat{H}_j \hat{H}_k + \cdots, \tag{68}$$

where the double-primed sum only excludes terms $\hat{H}_j, \hat{H}_k$ overlapping at more than one site, sharing a bond. For instance, consider the expansion of the time-evolution operator for the $S = 1/2$-Heisenberg chain: expressions of the form $\left(\hat{s}_j^+ \hat{s}_{j+1}^-\right)\left(\hat{s}_{j+1}^z\right)$ are kept, while those of the form $\left(\hat{s}_j^+ \hat{s}_{j+1}^-\right)\left(\hat{s}_j^z \hat{s}_{j+1}^z\right)$ are discarded. Hence, the error is again of order δ^2 but contributions with arbitrary powers of single-site terms are treated exactly. There is no closed general MPO representation for $\hat{U}^{\mathrm{II}}$ but we can give an approximation which has an error $\mathcal{O}(\delta^3)$ and hence does not affect the second-order approximation of $\hat{U}^{\mathrm{II}}$. In the following we will first explicitly demonstrate how to numerically construct the MPO representation W^{II} from the block-triangular structure of the MPO representation of $\hat{H}$ and subsequently motivate the used formalism. As for $\hat{U}^{\mathrm{I}}$, the MPO representation of $\hat{U}^{\mathrm{II}}$ is of the form

$$\hat{W}_j^{\mathrm{II}} = \begin{pmatrix} \hat{W}_{D_j}^{\mathrm{II}} & \hat{W}_{C_j}^{\mathrm{II}} \\ \hat{W}_{B_j}^{\mathrm{II}} & \hat{W}_{A_j}^{\mathrm{II}} \end{pmatrix}. \tag{69}$$

In order to construct the matrices $W_{\{A_j, B_j, C_j, D_j\}}^{\mathrm{II}}$ we employ transition amplitudes between hard-core bosonic states (cf. Section 4.2.2) following the notation in Ref. [40]. Let $\mathcal{H}_{2,a_j}$ denote the a_jth hard-core bosonic Hilbert space. N_j of these spaces form $\mathcal{H}_c = \bigotimes_{a_j=1}^{N_j} \mathcal{H}_{2,a_j}$ and we work in the joint Hilbert space $\mathcal{H}_c \otimes \mathcal{H}_{\bar{c}}$ which spans $2N_j$ individual hard-core bosonic Hilbert spaces. The ladder operators acting on $\mathcal{H}_c$ and $\mathcal{H}_{\bar{c}}$ are $\hat{c}_{a_j}^\dagger$, $\hat{c}_{a_j}$, $\hat{\bar{c}}_{\bar{a}_j}^\dagger$ and $\hat{\bar{c}}_{\bar{a}_j}$ respectively.

The generator of the matrix elements is a map on the joint bosonic and physical Hilbert spaces $\mathcal{H}_c \otimes \mathcal{H}_{\bar{c}} \otimes \mathcal{H}_{\mathrm{phys}}$ and given by the operator-valued exponentials

$$\hat{\Phi}_{j;a_j,\bar{a}_j} = e^{\hat{F}_{j;a_j,\bar{a}_j}} = e^{\hat{c}_{a_j}^\dagger \hat{\bar{c}}_{\bar{a}_j}^\dagger \hat{A}_{j;a_j,\bar{a}_j} + \sqrt{\delta}(\hat{c}_{a_j}^\dagger \hat{B}_{j;a_j} + \hat{\bar{c}}_{\bar{a}_j}^\dagger \hat{C}_{j;\bar{a}_j}) + \delta \hat{D}_j}. \tag{70}$$

Denoting the combined bosonic vacuum state by $|0\rangle \otimes |\bar{0}\rangle \equiv |0, \bar{0}\rangle$ the following transition amplitudes determine the operator-valued entries of the MPO representation $\hat{W}_j^{\mathrm{II}}$:

$$\hat{W}_{A_j;a_j,\bar{a}_j}^{\mathrm{II}} = \langle 0, \bar{0}|\hat{c}_{a_j} \hat{\bar{c}}_{\bar{a}_j} \hat{\Phi}_{j;a_j,\bar{a}_j}|0, \bar{0}\rangle = \langle 0|\hat{c}_{a_j} \hat{\bar{c}}_{\bar{a}_j} e^{\hat{c}_{a_j}^\dagger \hat{\bar{c}}_{\bar{a}_j}^\dagger \hat{A}_{j;a_j,\bar{a}_j} + \sqrt{\delta}(\hat{c}_{a_j}^\dagger \hat{B}_{j;a_j} + \hat{\bar{c}}_{\bar{a}_j}^\dagger \hat{C}_{j;\bar{a}_j}) + \delta \hat{D}_j}|0\rangle \tag{71}$$

$$\hat{W}_{B_j;a_j}^{\mathrm{II}} = \langle 0, \bar{0}|\hat{c}_{a_j} \hat{\Phi}_{j;a_j,\bar{a}_j}|0, \bar{0}\rangle = \langle 0|\hat{c}_{a_j} e^{\sqrt{\delta}\hat{c}_{a_j}^\dagger \hat{B}_{j;a_j} + \delta \hat{D}_j}|0\rangle \tag{72}$$

$$\hat{W}_{C_j;\bar{a}_j}^{\mathrm{II}} = \langle 0, \bar{0}|\hat{\bar{c}}_{\bar{a}_j} \hat{\Phi}_{j;a_j,\bar{a}_j}|0, \bar{0}\rangle = \langle \bar{0}|\hat{\bar{c}}_{\bar{a}_j} e^{\sqrt{\delta}\hat{\bar{c}}_{\bar{a}_j}^\dagger \hat{C}_{j;\bar{a}_j} + \delta \hat{D}_j}|\bar{0}\rangle \tag{73}$$

$$\hat{W}_{D_j}^{\mathrm{II}} = \langle 0, \bar{0}|\hat{\Phi}_{j;a_j,\bar{a}_j}|0, \bar{0}\rangle = e^{\delta \hat{D}_j}. \tag{74}$$

For clarity, in the following we will explicitly demonstrate the calculation of the elements of $W_{A_j;a_j,\bar{a}_j}^{\mathrm{II}}$. Explicitly, we use

$$|0\rangle_{a_j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathcal{H}_{2,a_j}, \quad \hat{c}_{a_j} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \hat{c}_{a_j}^\dagger = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \hat{\mathbf{1}}_{a_j} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{75}$$

to represent the hard-core bosonic states and operators. Since $\hat{F}_{j;a_j,\bar{a}_j}$ only contains creation operators for the modes $a_j, \bar{a}_j$ the vacuum expectation value over all $2N_j$ bosonic modes can be simplified to the relevant modes, i.e., we can replace $|0, \bar{0}\rangle \to |0_{a_j}, \bar{0}_{\bar{a}_j}\rangle$. Therefore, the generator is obtained by exponentiating the matrix

$$\hat{F}_{j;a_j,\bar{a}_j} = \hat{c}^\dagger_{a_j} \otimes \hat{\bar{c}}^\dagger_{\bar{a}_j} \hat{A}_{j;a_j,\bar{a}_j} + \sqrt{\delta}\left(\hat{c}^\dagger_{a_j} \otimes \hat{\mathbf{1}}_{\bar{a}_j} \hat{B}_{j;a_j} + \hat{\mathbf{1}}_{a_j} \otimes \hat{\bar{c}}^\dagger_{\bar{a}_j} \hat{C}_{j;\bar{a}_j}\right) + \hat{\mathbf{1}}_{a_j} \otimes \hat{\mathbf{1}}_{\bar{a}_j} \delta\hat{D}_j \tag{76}$$

$$= \begin{pmatrix} \delta\hat{D}_j & 0 & 0 & 0 \\ \sqrt{\delta}\hat{C}_{j;\bar{a}_j} & \delta\hat{D}_j & 0 & 0 \\ \sqrt{\delta}\hat{B}_{j;a_j} & 0 & \delta\hat{D}_j & 0 \\ \hat{A}_{j;a_j,\bar{a}_j} & \sqrt{\delta}\hat{B}_{j;a_j} & \sqrt{\delta}\hat{C}_{j;\bar{a}_j} & \delta\hat{D}_j \end{pmatrix}, \tag{77}$$

which has been truncated to the relevant bosonic Hilbert spaces $\mathcal{H}_{2,a_j}$, $\mathcal{H}_{2,\bar{a}_j}$. The entries of $\hat{W}^{\mathrm{II}}_{A_j;a_j,\bar{a}_j}$ are then obtained by evaluating the power series of the matrix exponential and calculating the vacuum expectation value

$$\hat{W}^{\mathrm{II}}_{A_j;a_j,\bar{a}_j} = \langle 0, \bar{0} | \hat{c}_{a_j} \hat{\bar{c}}_{\bar{a}_j} \hat{\Phi}_{j;a_j,\bar{a}_j} | 0, \bar{0} \rangle = \langle 1_{a_j}, \bar{1}_{\bar{a}_j} | \hat{\Phi}_{j;a_j,\bar{a}_j} | 0, \bar{0} \rangle$$

$$= \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} \exp\left\{ \begin{pmatrix} \delta\hat{D}_j & 0 & 0 & 0 \\ \sqrt{\delta}\hat{C}_{j;\bar{a}_j} & \delta\hat{D}_j & 0 & 0 \\ \sqrt{\delta}\hat{B}_{j;a_j} & 0 & \delta\hat{D}_j & 0 \\ \hat{A}_{j;a_j,\bar{a}_j} & \sqrt{\delta}\hat{B}_{j;a_j} & \sqrt{\delta}\hat{C}_{j;\bar{a}_j} & \delta\hat{D}_j \end{pmatrix} \right\} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{78}$$

A compact notation for all matrix elements of W^{II} can be obtained if we let the annihilation operators $\hat{c}_{a_j}$, $\hat{c}_{\bar{a}_j}$ in Eqs. (71) to (74) act on the bra $\langle 0, \bar{0} |$ and define the formal symbol $\mathcal{S}_j \in \{A_j, B_j, C_j, D_j\}$

$$\hat{W}^{\mathrm{II}}_{\mathcal{S}_j;a_j,\bar{a}_j} = \begin{pmatrix} \delta_{\mathcal{S}_j,D_j} & \delta_{\mathcal{S}_j,C_j} & \delta_{\mathcal{S}_j,B_j} & \delta_{\mathcal{S}_j,A_j} \end{pmatrix}$$

$$\times \exp\left\{ \begin{pmatrix} \delta\hat{D}_j & 0 & 0 & 0 \\ \sqrt{\delta}\hat{C}_{j;\bar{a}_j} & \delta\hat{D}_j & 0 & 0 \\ \sqrt{\delta}\hat{B}_{j;a_j} & 0 & \delta\hat{D}_j & 0 \\ \hat{A}_{j;a_j,\bar{a}_j} & \sqrt{\delta}\hat{B}_{j;a_j} & \sqrt{\delta}\hat{C}_{j;\bar{a}_j} & \delta\hat{D}_j \end{pmatrix} \right\} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{79}$$

Hence, the N_j^2 exponentials $\exp\left\{\hat{F}_{j;a_j,\bar{a}_j}\right\}$ already contain all relevant information to construct the stepper W^{II} and in particular there is no need to calculate different exponentials as suggested by Eqs. (71)–(74).

4.2.2. Detailed derivation of the W^{II} representation

We will present the construction of W^{II} in more detail, following Ref. [40], in this section. Before digging into the details, let us briefly sketch the derivation: The goal is to find an MPO representation of the time stepper $\hat{U}^{\mathrm{II}}(\delta)$ capturing interaction terms in the series expansion as described above. The corresponding MPO site tensors W^{II} are obtained by making use of the MPO recursion Eq. (66) to factorize the exponential $e^{-\mathrm{i}\delta\hat{H}}$. The factorization itself is performed exploiting complex Gaussian integrals via the introduction of auxiliary fields ϕ_j, $\bar{\phi}_j$ on each bond which are integrated over. The resulting MPO is of the form

$$W^{\mathrm{II}} = \int D[\phi_1, \bar{\phi}_1] U_{\bar{\phi}_1} e^{-\bar{\phi}_1\phi_1} \int D[\phi_2, \bar{\phi}_2] U_{\phi_1, \bar{\phi}_2} e^{-\bar{\phi}_2\phi_2} U_{\phi_2, \bar{\phi}_3} \cdots, \tag{80}$$

that is, the bond indices ϕ_j, $\bar{\phi}_j$ are continuous degrees of freedom. In a final step, these bond indices are discretized using bosonic coherent state path integrals yielding the desired expression for the MPO site tensors.

We begin by employing the bond representation of $\hat{H}$ (compare Eq. (65)) and write the operator as a formal scalar product over auxiliary degrees of freedom $a_j = 1 \ldots N_j$, where N_j again is the

number of interaction terms crossing the bond j,

$$\hat{H} = \hat{H}^L_{j-1} \otimes \hat{\mathbf{1}}^R_j + \hat{\mathbf{1}}^L_j \otimes \hat{H}^R_{j+1} + \left(\begin{array}{ccc} \hat{h}^L_{j;1} & \cdots & \hat{h}^L_{j;N_j} \end{array} \right) \left(\begin{array}{c} \hat{h}^R_{j;1} \\ \vdots \\ \hat{h}^R_{j;N_j} \end{array} \right) \tag{81}$$

$$\equiv \hat{H}^L_{j-1} \otimes \hat{\mathbf{1}}^R_j + \hat{\mathbf{1}}^L_j \otimes \hat{H}^R_{j+1} + \hat{J}^t_j \hat{\hat{J}}_j \tag{82}$$

and the product between the entries of the operator-valued "vectors" $\hat{J}_j$ and $\hat{\hat{J}}_j$ are tensor products between operators acting on partitioned Hilbert spaces to the left and right of the bond j, respectively. Next, we introduce complex vector fields $\phi_{j;a_j}$ and their complex conjugate $\bar{\phi}_{j;a_j}$ and define a mapping from the auxiliary indices into the full Hilbert space

$$\hat{J}^t_j \cdot \bar{\phi}_j : \quad \left(\mathcal{H}^L_j \to \mathcal{H}^L_j \right)^{\otimes N_j} \times \mathbb{C}^{N_j} \longrightarrow (\mathcal{H} \to \mathcal{H})$$

$$\left(\hat{J}_j, \bar{\phi}_j \right) \longmapsto \hat{J}^t_j \cdot \bar{\phi}_j := \hat{h}^L_{j;1} \bar{\phi}_{j;1} \hat{\mathbf{1}}^R_j + \cdots + \hat{h}^L_{j;N_j} \bar{\phi}_{j;N_j} \hat{\mathbf{1}}^R_j \tag{83}$$

$$\phi_j \cdot \hat{\hat{J}}_j : \quad \mathbb{C}^{N_j} \times \left(\mathcal{H}^R_j \to \mathcal{H}^R_j \right)^{\otimes N_j} \longrightarrow (\mathcal{H} \to \mathcal{H})$$

$$\left(\phi_j, \hat{\hat{J}}_j \right) \longmapsto \phi_j \cdot \hat{\hat{J}}_j := \hat{\mathbf{1}}^L_j \phi_{j;1} \hat{h}^R_{j;1} + \cdots + \hat{\mathbf{1}}^L_j \phi_{j;N_j} \hat{h}^R_{j;N_j} , \tag{84}$$

where we will suppress the identities acting on the left and right partitions of the Hilbert space in the following. We then use complex Gaussian integrals to decouple each left and right interaction term

$$e^{\hat{J}_{j;a_j} \hat{\hat{J}}_{j;a_j}} = \frac{1}{\pi} \int d\phi_{j;a_j} d\bar{\phi}_{j;a_j} e^{-\bar{\phi}_{j;a_j} \cdot \phi_{j;a_j} + \hat{J}_{j;a_j} \cdot \bar{\phi}_{j;a_j} + \phi_{j;a_j} \cdot \hat{\hat{J}}_{j;a_j}} . \tag{85}$$

Furthermore, we factorize the operator exponential $e^{-i\delta \hat{H}}$ (setting $\tau = -i\delta$)

$$e^{\tau \hat{H}} = e^{\tau \left(\hat{H}^L_{j-1} \otimes \hat{\mathbf{1}}^R_j + \hat{\mathbf{1}}^L_j \otimes \hat{H}^R_{j+1} + \hat{J}^t_j \hat{\hat{J}}_j \right)}$$

$$= e^{\tau \hat{H}^L_{j-1} \otimes \hat{\mathbf{1}}^R_j} e^{\tau \hat{\mathbf{1}}^L_j \otimes \hat{H}^R_{j+1}} e^{\tau \hat{J}^t_j \hat{\hat{J}}_j} + \mathcal{O}(\tau^2) \tag{86}$$

where the error occurs at second order in τ if the operators $\hat{J}_j, \hat{\hat{J}}_j, \hat{H}^L_{j-1}, \hat{H}^R_{j+1}$ do not commute.[4] Exploiting N_j complex Gaussian integrals and defining $D\left[\phi_j, \bar{\phi}_j \right] = \prod_{a_j} \frac{d\phi_{j;a_j}}{\sqrt{\pi}} \frac{d\bar{\phi}_{j;a_j}}{\sqrt{\pi}}$ we obtain

$$e^{\tau \hat{H}} = \int D\left[\phi_j, \bar{\phi}_j \right] e^{\tau \hat{H}^L_{j-1} \otimes \hat{\mathbf{1}}^R_j} e^{\tau \hat{\mathbf{1}}^L_j \otimes \hat{H}^R_{j+1}} e^{-\bar{\phi}_j \cdot \phi_j + \sqrt{\tau} \hat{J}_j \cdot \bar{\phi}_j + \sqrt{\tau} \phi_j \cdot \hat{\hat{J}}_j} + \mathcal{O}(\tau^2)$$

$$= \int D\left[\phi_j, \bar{\phi}_j \right] e^{\tau \hat{H}^L_{j-1} \cdot \hat{\mathbf{1}}^R_j + \sqrt{\tau} \hat{J}_j \cdot \bar{\phi}_j} e^{-\bar{\phi}_j \cdot \phi_j} e^{\tau \hat{\mathbf{1}}^L_j \cdot \hat{H}^R_{j+1} + \sqrt{\tau} \phi_j \cdot \hat{\hat{J}}_j} + \mathcal{O}(\tau^2). \tag{87}$$

Here we used that the identities $\hat{\mathbf{1}}^L_j$ and $\hat{\mathbf{1}}^R_j$ can also be interpreted as maps from an auxiliary index into the full Hilbert space.

Now the MPO recursion Eq. (66) can be applied to, e.g., shift the active bond $j \to j+1$ in the third exponential

$$e^{\tau \hat{\mathbf{1}}^L_j \cdot \hat{H}^R_{j+1} + \sqrt{\tau} \phi_j \hat{\hat{J}}_j} = e^{\tau \hat{\mathbf{1}}^L_j \cdot \hat{H}^R_{j+1} + \sqrt{\tau} \phi_j \cdot \hat{A}_j \hat{\hat{J}}_{j+1} + \sqrt{\tau} \phi_j \cdot \hat{B}_j \hat{\mathbf{1}}^R_{j+1} + \tau \hat{C}_j \hat{\hat{J}}_{j+1} + \tau \hat{D}_j \cdot \hat{\mathbf{1}}^R_{j+1}}. \tag{88}$$

---

[4] The commutators need to be evaluated on the full Hilbert space by completing the partitioned operators $\hat{J}_j, \hat{\hat{J}}_j, \hat{H}^L_{j-1}, \hat{H}^R_{j+1}$ which is achieved by taking the appropriate tensor products with $\hat{\mathbf{1}}^L_j$ and $\hat{\mathbf{1}}^R_j$.

From now on we replace the formal dot product $\hat{D}_j \cdot \hat{\mathbf{1}}_{j+1}^R$ by its action on the full Hilbert space which reduces for the case of $\hat{D}_j$ to on-site terms only. The same holds for the dot product $\phi_j \cdot \hat{B}_j \cdot \hat{\mathbf{1}}_{j+1}^R \equiv \phi_j \cdot \hat{B}_j$ which only connects sites to the left of j. In the exponent the summands $\sqrt{\tau} \phi_j \cdot \hat{B}_j \cdot \hat{\mathbf{1}}_{j+1}^R$, $\tau \hat{D}_j \cdot \hat{\mathbf{1}}_{j+1}^R$ and $\tau \hat{\mathbf{1}}_j \cdot \hat{H}_{j+1}^R$ already act separately on site j and the right partition of the Hilbert space $\mathcal{H}_{j+1}^R$. In order to separate the remaining summands, too, we introduce another set of auxiliary fields $\phi_{j+1}, \bar{\phi}_{j+1}$ by insertion of more complex Gaussian integrals

$$e^{\sqrt{\tau}\phi_j \cdot \hat{A}_j \hat{J}_{j+1} + \tau \hat{C}_j \cdot \hat{J}_{j+1}} = \int D\left[\phi_{j+1}, \bar{\phi}_{j+1}\right] e^{-\bar{\phi}_{j+1}\phi_{j+1}} e^{\phi_j \cdot \hat{A}_j \cdot \bar{\phi}_{j+1} + \sqrt{\tau}\hat{C}_j \cdot \bar{\phi}_{j+1} + \sqrt{\tau}\phi_{j+1} \cdot \hat{J}_{j+1}}$$

$$\Rightarrow e^{\tau \hat{\mathbf{1}}_j^L \cdot \hat{H}_{j+1}^R + \sqrt{\tau}\phi_j \cdot \hat{J}_j} = \int D\left[\phi_{j+1}, \bar{\phi}_{j+1}\right] U_{\phi_j, \bar{\phi}_{j+1}} e^{-\bar{\phi}_{j+1}\phi_{j+1}} e^{\tau \hat{\mathbf{1}}_{j+1}^L \cdot \hat{H}_{j+2}^R + \sqrt{\tau}\phi_{j+1} \cdot \hat{J}_{j+1}} \tag{89}$$

with $U_{\phi_j, \bar{\phi}_{j+1}} = e^{\phi_j \cdot \hat{A}_j \cdot \bar{\phi}_{j+1} + \sqrt{\tau}\phi_j \cdot \hat{B}_j + \sqrt{\tau}\hat{C}_j \cdot \bar{\phi}_{j+1} + \tau \hat{D}_j}$. The last equation defines a recursion

$$e^{\tau \hat{\mathbf{1}}_j^L \cdot \hat{H}_{j+1}^R + \sqrt{\tau}\phi_j \cdot \hat{J}_j} = \mathcal{F}_{j, j+1}\left[e^{\tau \hat{\mathbf{1}}_{j+1}^L \cdot \hat{H}_{j+2}^R + \sqrt{\tau}\phi_{j+1} \cdot \hat{J}_{j+1}}\right] \tag{90}$$

generated by the integration of $U_{\phi_j, \bar{\phi}_{j+1}} e^{-\bar{\phi}_{j+1}\phi_{j+1}}$ over the continuous bond degrees of freedom $\phi_j, \bar{\phi}_{j+1}$. Iterating through the whole system we finally obtain

$$e^{\tau \hat{H}} = \int D[\phi_1, \bar{\phi}_1] U_{\bar{\phi}_1} e^{-\bar{\phi}_1 \phi_1} \int D[\phi_2, \bar{\phi}_2] U_{\phi_1, \bar{\phi}_2} e^{-\bar{\phi}_2 \phi_2} U_{\phi_2, \bar{\phi}_3} \cdots + \mathcal{O}(\tau^2), \tag{91}$$

which is a first-order approximation in τ of $e^{\tau \hat{H}}$ in terms of an MPO with continuous bond labels $\phi_j = (\phi_1, \ldots, \phi_{N_j})$.

In a final step the continuous bond labels ϕ_{a_j} are transformed into discrete ones. We note that for a set of N_j bosonic ladder operators $\hat{c}_{a_j}, \hat{c}_{a_j}^\dagger$ we can rewrite the tensors at site j as expectation values for coherent states

$$|\phi_{a_j}\rangle = \sum_n \frac{\phi_{a_j}^n}{n!} \left(\hat{c}_{a_j}^\dagger\right)^n |0\rangle, \tag{92}$$

which have the properties

$$\hat{c}_{a_j}|\phi_{a_j}\rangle = \phi_{a_j}|\phi_{a_j}\rangle, \quad \langle\phi_{a_j}|\hat{c}_{a_j}^\dagger = \langle\phi_{a_j}|\bar{\phi}_{a_j}, \quad \langle\phi_{a_j}|\phi_{a_j}\rangle = e^{\bar{\phi}_{a_j}\phi_{a_j}}. \tag{93}$$

For instance, we may consider the case when $N_j \equiv 1$. Then a single bond label $\phi_{j;a_j} \equiv z_j$ is sufficient and (abbreviating $\hat{c}_{a_j} \equiv \hat{c}_j$, $\hat{c}_{a_j}^\dagger \equiv \hat{c}_j^\dagger$) we have at the bond $(j, j+1)$

$$e^{z_j \hat{A}_j \bar{z}_{j+1} + \sqrt{\tau} z_j \hat{B}_j + \sqrt{\tau} \hat{C}_j \bar{z}_{j+1} + \tau \hat{D}_j} = \langle \bar{z}_j, z_{j+1} | e^{\hat{c}_j^\dagger \hat{A}_j \hat{c}_{j+1}^\dagger + \sqrt{\tau} \hat{c}_j^\dagger \hat{B}_j + \sqrt{\tau} \hat{C}_j \hat{c}_{j+1}^\dagger + \tau \hat{D}_j} |0, 0\rangle$$

$$= \sum_{n_j, n_{j+1}} \underbrace{\langle n_j, \bar{n}_{j+1} | e^{\hat{c}_j^\dagger \hat{A}_j \hat{c}_{j+1}^\dagger + \sqrt{\tau} \hat{c}_j^\dagger \hat{B}_j + \sqrt{\tau} \hat{C}_j \hat{c}_{j+1}^\dagger + \tau \hat{D}_j} |0, 0\rangle}_{\equiv \hat{U}_{n_j, \bar{n}_{j+1}}} \frac{z_j^{n_j} \bar{z}_{j+1}^{\bar{n}_{j+1}}}{\sqrt{n_j! \bar{n}_{j+1}!}}. \tag{94}$$

In the last equality we expanded the coherent states $|z_j\rangle$, $|z_{j+1}\rangle$ in the bosonic occupations number basis $|n_j\rangle$, $|\bar{n}_{j+1}\rangle$. Now, we turn to the integral over one pair of bond labels $z_j, \bar{z}_j$, considering only those factors $\frac{z_j^{k_j}}{\sqrt{k_j!}}$ which contribute to the integration,

$$\int \frac{dz_j d\bar{z}_j}{\pi} \hat{U}_{z_{j-1}, \bar{z}_j} e^{-\bar{z}_j z_j} \hat{U}_{z_j, \bar{z}_{j+1}} = \sum_{n_j, \bar{n}_j} \int \frac{dz_j d\bar{z}_j}{\pi} \hat{U}_{n_{j-1}, \bar{n}_j} e^{-\bar{z}_j z_j} \hat{U}_{n_j, \bar{n}_{j+1}} \frac{\bar{z}_j^{\bar{n}_j} z_j^{n_j}}{\sqrt{\bar{n}_j! n_j!}}$$

$$
= \sum_{n_j, \bar{n}_j} \hat{U}_{n_{j-1}, \bar{n}_j} \hat{U}_{n_j, \bar{n}_{j+1}} \int_0^\infty d\rho_j \frac{\rho_j^{\bar{n}_j + n_j + 1}}{\sqrt{\bar{n}_j! \, n_j!}} e^{-\rho_j^2} \int_0^{2\pi} \frac{d\varphi_j}{\pi} e^{i\varphi_j(n_j - \bar{n}_j)}
$$

$$
= \sum_{n_j} \hat{U}_{n_{j-1}, n_j} \hat{U}_{n_j, \bar{n}_{j+1}} \frac{2}{n_j!} \underbrace{\int_0^\infty d\rho \, \rho^{2n_j+1} e^{-\rho^2}}_{= \frac{1}{2}\Gamma(n_j)}
$$

$$
= \sum_{n_j} \hat{U}_{n_{j-1}, n_j} \hat{U}_{n_j, \bar{n}_{j+1}}. \tag{95}
$$

We parametrized the complex integration in polar coordinates in the second identity and exploited the representation of the Γ function in the last line. For the general case $N_j > 1$ we define the boson occupation number $\mathbf{n}_j = (n_1 \cdots n_{N_j})$ for each bond. Since we can always reorder the ϕ_{a_j}, we can perform the same integrals for each continuous bond label so that we finally find

$$
\hat{U}(\delta) = \sum_{\mathbf{n}_1 \cdots \mathbf{n}_L} \hat{U}_{\bar{\mathbf{n}}_1} \hat{U}_{\mathbf{n}_1, \bar{\mathbf{n}}_2} \cdots + \mathcal{O}(\delta^2), \tag{96}
$$

yielding the desired MPO representation.

A pedestrian way to obtain a compact MPO representation from this very formal derivation is to analyze $\hat{U}_{\mathbf{n}_{j-1}, \bar{\mathbf{n}}_j}$ when truncating the bosonic Hilbert spaces to a maximal boson occupation b, i.e., $n_{j;a_j} \in \{0, \ldots, b\}$. We define bosonic field operators $\hat{\varphi}_j^\dagger = \left(\hat{c}_1^\dagger \cdots \hat{c}_L^\dagger \right)$ and denote the maximal occupation number by an upper index (b). Then, for each b we have operator-valued bosonic matrix elements

$$
\hat{U}^{(b)}_{j; \mathbf{n}_j, \bar{\mathbf{n}}_{j+1}} = \langle \mathbf{n}_j, \bar{\mathbf{n}}_{j+1} | \underbrace{e^{\hat{\varphi}_j^\dagger \cdot \hat{A}_j \cdot \hat{\bar{\varphi}}_{j+1}^\dagger + \sqrt{\tau}\hat{\varphi}_j^\dagger \cdot \hat{B}_j + \sqrt{\tau}\hat{C}_j \cdot \hat{\bar{\varphi}}_{j+1}^\dagger + \tau\hat{D}_j}}_{\equiv \hat{\Phi}_j} |\mathbf{0}, \bar{\mathbf{0}}\rangle \tag{97}
$$

of maps $\hat{\Phi}_j$ on the joint Hilbert spaces of b bosons $\mathcal{H}_{b;a_j}$ for every $a_j \in \{1, \ldots, N_j\}$ at each bond and the physical Hilbert space $\mathcal{H}_{\text{phys}}$ at each site

$$
\hat{\Phi}_j : \bigotimes_{a_j} \mathcal{H}_{b;a_j} \otimes \bigotimes_{\bar{a}_j} \mathcal{H}_{\bar{b};\bar{a}_j} \otimes \mathcal{H}_{\text{phys}} \longrightarrow \bigotimes_{a_j} \mathcal{H}_{b;a_j} \otimes \bigotimes_{\bar{a}_j} \mathcal{H}_{\bar{b};\bar{a}_j} \otimes \mathcal{H}_{\text{phys}}. \tag{98}
$$

Let us consider $b \equiv 0$, then the matrix elements collapse to pure on-site terms $\hat{U}^{(0)}_{j;\mathbf{0}, \bar{\mathbf{0}}} = \langle \mathbf{0}, \bar{\mathbf{0}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle = e^{\tau\hat{D}_j}$. The next order $b \equiv 1$ additionally collects contributions from all interactions crossing either the bond $(j-1, j)$ or $(j, j+1)$

$$
\hat{U}^{(1)}_{j;\mathbf{0}, \bar{\mathbf{0}}} = \langle \mathbf{0}, \bar{\mathbf{0}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle, \quad \hat{U}^{(1)}_{j;\mathbf{0}, \bar{\mathbf{1}}} = \langle \mathbf{0}, \bar{\mathbf{1}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle,
$$
$$
\hat{U}^{(1)}_{j;\mathbf{1}, \bar{\mathbf{0}}} = \langle \mathbf{1}, \bar{\mathbf{0}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle, \quad \hat{U}^{(1)}_{j;\mathbf{1}, \bar{\mathbf{1}}} = \langle \mathbf{1}, \bar{\mathbf{1}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle, \tag{99}
$$

which can be arranged as an operator-valued matrix

$$
\hat{U}^{(1)}_j = \begin{pmatrix} \langle \mathbf{0}, \bar{\mathbf{0}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle & \langle \mathbf{0}, \bar{\mathbf{1}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle \\ \langle \mathbf{1}, \bar{\mathbf{0}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle & \langle \mathbf{1}, \bar{\mathbf{1}} | \hat{\Phi}_j | \mathbf{0}, \bar{\mathbf{0}}\rangle \end{pmatrix} \equiv \begin{pmatrix} \hat{W}^{\text{II}}_{D_j} & \hat{W}^{\text{II}}_{C_j} \\ \hat{W}^{\text{II}}_{B_j} & \hat{W}^{\text{II}}_{A_j} \end{pmatrix} \equiv \hat{W}^{\text{II}}_j \tag{100}
$$

and yields the suggested form of W^{II}. The lower right matrix element $\hat{W}^{\text{II}}_{A_j}$ contains two bosons. When formally contracting $\hat{U}^{(1)}_j$ with the neighboring matrix $\hat{U}^{(1)}_{j+1}$, these bosons connect, for instance, all local operators from $\hat{A}_j$ to local operators from $\hat{B}_{j+1}$. Following this procedure, we find that in this way truncating the MPO approximation to $b = 1$ we keep all local operator strings that overlap at most at one site and hence the error is $\mathcal{O}(\tau^2)$.

Algorithm 2 The Krylov method. The main input is the Hamiltonian, the initial state, and the (possibly complex) time step. Additionally, a procedure APPLYANDORTHONORMALIZE is needed, which in turn requires the operator-state product and the orthogonalization of states. For details on a variational approach to this, see Section 5.3. The function COMPUTEEFFECTIVEH only needs to update the new elements of T_{j+1} compared to T_j.

1: **procedure** ORTHONORMALIZE($|w\rangle$, $\{|v_0\rangle, \cdots, |v_k\rangle\}$)
2:     $|w'\rangle \leftarrow |w\rangle - \sum_{\alpha=0}^{k} \langle w|v_\alpha\rangle |v_\alpha\rangle$ ▷ or variational orthonormalization for MPS
3:     **if** $\|\,|w'\rangle\,\| < \varepsilon$ **then**
4:         **return** invariant subspace found   ▷ evolution exact using just $\{|v_0\rangle, \cdots, |v_k\rangle\}$
5:     **end if**
6:     **return** $\frac{|w'\rangle}{\|\,|w'\rangle\,\|}$
7: **end procedure**
8: **procedure** APPLYANDORTHONORMALIZE($\hat{H}$, $\{|v_k\rangle, \cdots |v_0\rangle\}$)
9:     $|w'_k\rangle \leftarrow \hat{H}|v_k\rangle$
10:    **return** ORTHONORMALIZE($|w'_k\rangle$, $\{|v_0\rangle, \cdots |v_k\rangle\}$)
11: **end procedure**
12: **procedure** TIMESTEP($\hat{H}$, $|\psi(t)\rangle$, δ)
13:    $|v_0\rangle \leftarrow |\psi(t)\rangle / \|\,|\psi(t)\rangle\,\|$
14:    **for** $j \leftarrow 1 \ldots$ **do**
15:       $|v_j\rangle \leftarrow$ APPLYANDORTHONORMALIZE($\hat{H}$, $\{|v_0\rangle, \cdots |v_{j-1}\rangle\}$)
16:       $T_{j+1} \leftarrow$ COMPUTEEFFECTIVEH(T_j, $\hat{H}$, $\{|v_0\rangle, \cdots |v_j\rangle\}$)   ▷ $\left(T_{j+1}\right)_{k,l} = \langle v_k|\hat{H}|v_l\rangle$
17:       $c_{j+1} \leftarrow e^{-i\delta T_{j+1}} e^1_{j+1}$
18:       **if** c_{j+1} converged **then**
19:          **break**
20:       **end if**
21:    **end for**
22:    **return** $\|\,|\psi(t)\rangle\,\| \sum_{i=0}^{j} c^i_{j+1} |v_i\rangle$
23: **end procedure**

4.2.3. Errors

The MPO W^{II} approximation to the time-evolution operator $\hat{U}(\delta)$ primarily exhibits an error $\mathcal{O}(\delta^2)$ due to the truncation of the auxiliary degrees to hard-core bosons with maximal occupation $b \equiv 1$. The Trotter error created by Eq. (86) was shown in Ref. [40] to be $\mathcal{O}(\tau^3)$ so that it is subleading compared to the auxiliary boson field truncation error. However, the W^{II} MPO representation is not invariant under the particular choice of the decomposition of the Hamiltonian into local terms $\hat{H} = \sum_j \hat{H}_j$. This degree of freedom can be used to reduce the truncation error and is discussed in detail in Ref. [40].

We want to point out that the method of complex time steps discussed in Section 7.2 can transform a first-order stepper with error $\mathcal{O}(\delta^2)$ into a second-order stepper with error $\mathcal{O}(\delta^3)$. This is particularly suitable for the W^{II} MPO representation since the W^{II} bond dimension $w = N_j + 1$ is comparably small also for long-ranged Hamiltonians. An improved stepper is hence available without too much numerical effort and should be used (it is used in this form in the experiments run later). In turn, this construction in general destroys the unitarity of the stepper. Furthermore, as already discussed in Section 4.1.1 for TEBD, the choice of the MPO application method can significantly improve (or reduce) the accuracy of the time stepper.

5. The global Krylov method

Krylov subspace methods [24–29] (e.g. the Lanczos method [30,76]) are well-known iterative techniques from the field of numerical linear algebra. In their application to time-dependent problems, one approximates the action of $\hat{U}^{\text{exact}}(\delta)$ onto the quantum state $|\psi(t)\rangle$ directly, resulting

in a time-evolved state $|\psi(t+\delta)\rangle$. It does *not* provide access to the exponential $e^{-i\delta\hat{H}}$ in the standard physical basis. The most straightforward approach is to ignore the special structure of the MPS/MPO representation and directly implement the iterative procedure, as detailed below. This is what we call the *global Krylov method*. In contrast, a variant exploiting the structure of the MPS ansatz will be discussed in Section 6.1. We first introduce the Krylov method independent of the specific representation (dense vectors as in exact diagonalization, MPS, tree tensor networks etc.) used. This algorithm is also used as the local integrator for the local Krylov method of Section 6.1 and the TDVP in Section 6.2. Subsequently, we will discuss particular caveats when applying the method globally to matrix-product states.

The Krylov subspace $\mathcal{K}_N$ of a Hamiltonian $\hat{H}$ and initial state $|\psi\rangle$ is defined as the span of vectors $\{|\psi\rangle, \hat{H}|\psi\rangle, \ldots, \hat{H}^{N-1}|\psi\rangle\}$. This space is spanned by the Krylov vectors $|v_0\rangle, |v_1\rangle, \ldots, |v_{N-1}\rangle$ such that the first Krylov vector $|v_0\rangle$ is set to $|\psi\rangle$ normalized to have norm 1, and the subsequent Krylov vectors $|v_i\rangle$ are constructed by applying $\hat{H}$ to $|v_{i-1}\rangle$ and orthonormalizing against all previous Krylov vectors equivalently to the Lanczos algorithm. In exact arithmetics with Hermitian $\hat{H}$, this way to construct a Krylov subspace reduces to orthogonalizing against the previous two vectors $|v_{i-1}\rangle$ and $|v_{i-2}\rangle$, which is equivalent to the Lanczos algorithm [76]. However, due to round-off errors intrinsic to a numerical implementation, orthogonality of the Krylov vectors is usually lost. If the precision required of each solution is low, one may abstain from avoiding this problem and simply work in a very small subspace. However, due to the accumulation of errors during a time evolution and the calculation of spectral or time-dependent properties, it is necessary to cure this problem. Hence, one typically needs to explicitly orthogonalize each new Krylov vector against *all* previous Krylov vectors.[5] The method then proceeds by searching for the element of $\mathcal{K}_N$ which approximates the result of the exact evolution most closely:

$$\hat{U}^{\text{exact}}(\delta)|\psi(t)\rangle \approx \arg\min_{|u\rangle\in\mathcal{K}_N} \||u\rangle - \hat{U}^{\text{exact}}(\delta)|\psi(t)\rangle\| \equiv |\psi_N(t+\delta)\rangle . \tag{101}$$

To do so, we define the projector onto $\mathcal{K}_N$

$$\hat{P}_N = \sum_{i=0}^{N-1} |v_i\rangle\langle v_i| \tag{102}$$

$$= \left(\begin{pmatrix} \vdots \\ |v_0\rangle \\ \vdots \end{pmatrix} \begin{pmatrix} \vdots \\ |v_1\rangle \\ \vdots \end{pmatrix} \cdots \begin{pmatrix} \vdots \\ |v_{N-1}\rangle \\ \vdots \end{pmatrix} \right) \cdot \begin{pmatrix} \langle\cdots v_0 \quad \cdots| \\ \langle\cdots v_1 \quad \cdots| \\ \vdots \\ \langle\cdots v_{N-1}\cdots| \end{pmatrix} \equiv V_N^\dagger V_N \tag{103}$$

where we have introduced matrices V_N and $V_N^\dagger$ to represent the maps from the Hilbert space onto the Krylov space and vice versa. The solution to the above minimization problem is given by

$$|\psi_N(t+\delta)\rangle = \hat{P}_N^\dagger \hat{U}^{\text{exact}}(\delta)\hat{P}_N|\psi(t)\rangle . \tag{104}$$

Note that for $N = \dim\mathcal{H} \equiv N_\mathcal{H}$ this is exact. Expanding the projectors and writing down the formal Taylor series for $\hat{U}^{\text{exact}}(\delta)$, we find:

$$|\psi_{N_\mathcal{H}}(t+\delta)\rangle = \sum_{i=0}^{N_\mathcal{H}-1} |v_i\rangle\langle v_i|e^{-i\delta\hat{H}} \sum_{i'=0}^{N_\mathcal{H}-1} |v_{i'}\rangle\langle v_{i'}|\psi(t)\rangle \tag{105}$$

$$= \sum_{i=0}^{N_\mathcal{H}-1} |v_i\rangle\langle v_i| \sum_{n=0}^{\infty} \frac{(-i\delta)^n}{n!}\hat{H}^n \sum_{i'=0}^{N_\mathcal{H}-1} |v_{i'}\rangle\langle v_{i'}|\psi(t)\rangle \tag{106}$$

---

[5] Alternatively, one may only orthogonalize against the previous two vectors and achieve complete orthogonality by a subsequent basis transformation as detailed in Ref. [28].

$$= V_{N_{\mathcal{H}}}^{\dagger} \sum_{n=0}^{\infty} \frac{(-\mathrm{i}\delta)^n}{n!} \underbrace{V_{N_{\mathcal{H}}} \hat{H}^n V_{N_{\mathcal{H}}}^{\dagger}}_{\equiv (T_{N_{\mathcal{H}}})^n} V_{N_{\mathcal{H}}} |\psi(t)\rangle \tag{107}$$

$$\approx V_N^{\dagger} \sum_{n=0}^{\infty} \frac{(-\mathrm{i}\delta)^n}{n!} (T_N)^n V_N |\psi(t)\rangle = V_N^{\dagger} e^{-\mathrm{i}\delta T_N} V_N |\psi(t)\rangle , \tag{108}$$

where $N \ll N_{\mathcal{H}}$ and T_N is the Krylov-space representation of $\hat{H}$ with coefficients

$$(T_N)_{i,i'} = \langle v_i | \hat{H} | v_{i'} \rangle . \tag{109}$$

The Krylov approximation is introduced in Eq. (108). Note that for $n > N - 1$

$$V_{N_{\mathcal{H}}} \hat{H}^n V_{N_{\mathcal{H}}}^{\dagger} V_{N_{\mathcal{H}}} |\psi(t)\rangle \neq (T_N)^n V_N |\psi(t)\rangle . \tag{110}$$

This implies that the error in the Taylor-expansion of the time- evolution operator is of order $\delta^n/n!$, indicating that already a few iterations suffice to give a very small error in the integrator. If $V_N^{\dagger} V_N$ were a proper identity, we could insert it in between any power of $\hat{H}$ and obtain exactly $V_N \hat{H}^n V_N^{\dagger} = T_N^n$. However, our Krylov subspace is much smaller than the true Hilbert space and the projection induced by $V_N^{\dagger} V_N$ is hence very large, $V_N^{\dagger} V_N \neq \mathbf{1}$. But due to the special structure of this Krylov space, $V_N^{\dagger} T_N^n V_N |\psi(t)\rangle$ converges much more quickly to $\hat{H}^n |\psi(t)\rangle$ than expected if we, e.g., selected random vectors in the full Hilbert space to construct our subspace.

In exact arithmetic and with Hermitian $\hat{H}$, T_N would be tridiagonal, i.e., $(T_N)_{i,i'} = 0$ if $|i - i'| > 1$. While in practice this is not necessarily true, we found that it improves numerical stability and accuracy of the results to *assume* T_N to be tridiagonal and only evaluate those elements while forcibly setting all other elements to zero. Returning to our equation Eq. (108), $V_N |\psi(t)\rangle$ is the Krylov-space vector

$$(\| |\psi(t)\rangle \|, 0, 0, \ldots)^T \tag{111}$$

as all other Krylov vectors are orthogonal to $|v_0\rangle$ and $|v_0\rangle$ is the normalized version of $|\psi(t)\rangle$. T_N can be exponentiated efficiently using standard diagonalization routines from LAPACK, as it is only of size $N \times N$. With $T_N = Q_N^{\dagger} D_N Q_N$ this yields

$$e^{-\mathrm{i}\delta T_N} = Q_N^{\dagger} e^{-\mathrm{i}\delta D_N} Q_N . \tag{112}$$

For a given number of Krylov vectors and step size δ, we hence obtain

$$|\psi_N(t + \delta)\rangle = \| |\psi(t)\rangle \| V_N^{\dagger} Q_N^{\dagger} e^{-\mathrm{i}\delta D_N} Q_N e_N^1 \tag{113}$$

$$= \| |\psi(t)\rangle \| V_N^{\dagger} c_N \tag{114}$$

with the coefficient vector $c_N = Q_N^{\dagger} e^{-\mathrm{i}\delta D_N} Q_N e_N^1$ and e_N^1 the N-dimensional unit vector $(1, 0, 0, \ldots)^T$. For typical problems as presented in the example section, the number of Krylov vectors used by us was between 3 and 10. The algorithmic procedure is summarized in Alg. 2.

Evaluation of expectation values $\langle \hat{O}(t + \delta) \rangle$. It is not actually necessary to construct the time-evolved state $|\psi_N(t + \delta)\rangle$ to evaluate $\langle \hat{O}(t + \delta) \rangle$ for arbitrary δ. Instead, evaluating $\langle v_i | \hat{O} | v_{i'} \rangle$ for all $i, i' = [0, \ldots, N-1]$ and constructing only the coefficient vector c_N is sufficient to evaluate the observable. One can hence potentially avoid adding up the Krylov vectors and constructing the time-evolved state $|\psi(t + \delta)\rangle$. Indeed (cf. Section 5.4 later), by evaluating the expectation values $\langle v_i | \hat{O} | v_{i'} \rangle$, it becomes possible to calculate the value of the observable at *any* time $\delta' < \delta$ by only operating on small $N \times N$ matrices. Hence very small time steps unattainable by the other time-stepping methods (e.g. $\delta' = 10^{-5}$) become available.

5.1. Errors

Elaborate bounds are available to estimate the error incurred during the Krylov approximation [77]. Unfortunately, these bounds are proven under the assumption of exact arithmetic and

hence do not necessarily apply in the context of matrix-product states. The main take-away, which is confirmed in practice, is that the Krylov error is in $O(\delta^N)$ as long as $\sqrt{W\delta} \leq N$ where W is the spectral range, which, in turn, is roughly of the same order of magnitude as the system size. For a typical system of $L = 100$ sites with $\delta = 0.1$, this condition is fulfilled as soon as $N \approx 3$; more precise bounds are available in exact arithmetic.[6]

Hence, there are two approaches to measure the convergence of the Krylov space: (i) The bottom-rightmost entry of the effective matrix T_n measures the weight scattered out of the Krylov space by application of the Hamiltonian and typically decays exponentially; (ii) the full Hilbert-space 2-norm distance between two sequential iterations is cheaply available through the coefficients of Krylov vectors produced in the two iterations. In our experience, this second measure makes for an excellent convergence criterion.

In addition to the inherent Krylov error which can often be made extremely small ($O(10^{-10})$ or smaller), the Krylov method of course also suffers from the standard MPS truncation error — this error, too, can be measured precisely (via the discarded weight) and be made very small. As such, both errors of the global Krylov method can be made extremely small at *finite time-step size*, albeit at relatively large numerical cost. The method hence in particular excels if used to evaluate states very precisely, e.g., to measure the errors of other methods on short time scales.

5.2. Application to matrix-product states

Up to this point, there was no need to narrow the description down to a specific representation, which serves as a proof for the versatility of the Krylov method. In our practical calculations, however, we wish to use MPS to represent the time-evolved quantum states and intermediate Krylov vectors and an MPO to represent the Hamiltonian $\hat{H}$, which requires a few minor adaptations for efficiency and accuracy. Note that in stark contrast to the TEBD and MPO $W^{\mathrm{I,II}}$ method, *only* an MPO representation of $\hat{H}$ and no analytical or other decomposition is required.

First and foremost, the most obvious improvement is in the calculation of the last entry of the effective Krylov matrix T_N. In exact or dense arithmetic, the evaluation of $\langle v_{N-1}|\hat{H}|v_{N-1}\rangle$ requires computing the matrix–vector product $\hat{H}|v_{N-1}\rangle$. This is not the case in the MPS approach: Indeed, evaluating the expectation value $\langle v_{N-1}|\hat{H}|v_{N-1}\rangle$ is much cheaper than calculating the MPS representing $\hat{H}|v_{N-1}\rangle$, cf. Section 2.9. As such, to generate a $N \times N$-dimensional effective Krylov matrix T_N, one only needs to evaluate $N - 1$ MPO–MPS products and avoids the MPO–MPS product for the most costly application on the last Krylov vector. In our experience, the bond dimension of every additional Krylov vector grows superlinearly, making this optimization very worthwhile.

To construct the time-evolved state $|\psi(t + \delta)\rangle$, it is necessary to sum N Krylov vectors together. Various methods to do so exist, in our implementation, we sequentially add two vectors (resulting in a new MPS with bond dimension $2m$) which are then truncated back down to the target bond dimension m. In $N - 1$ steps, all N Krylov vectors are summed together at cost $O(N(2m)^3)$. One could follow-up on this procedure with some sweeps of variational optimization or alternatively directly variationally optimize, but this does not appear to be necessary for our application.

5.3. Loss of orthogonality

A typical problem of Krylov-subspace methods is the possible loss of orthogonality of the basis vectors due to the finite-precision arithmetic of floating point operations. This matter becomes substantially more pressing in the matrix-product algebra, as truncation is crucial to keeping computations feasible. If many Krylov vectors are desired, truncation errors affecting the orthogonality of the basis vectors do not simply add to the overall error (see above), but may quickly degrade the overall quality of the Krylov space, leading to a poor result. In this case, it is necessary to check for orthogonality in the basis and eventually re-orthogonalize the basis vectors successively. However,

---

[6] In exact arithmetic for a spectral width W of $\hat{H}$, the error is smaller than $10e^{-N^2/(5W\delta)}$ if N is between $\sqrt{4W\delta}$ and $2W\delta$ and the error is smaller than $(^{10}/w_\delta)\, e^{-W\delta}\, (^{eW\delta}/N)^N$ if $N > 2W\delta$ [77]. It is unknown how this translates to the case of inexact arithmetic.

if one uses a simple Gram–Schmidt procedure to orthogonalize vectors by successive additions of MPS, new truncation errors are introduced during this procedure, which will quite often entail the same problem.

In our experience, it has been proven fruitful to orthogonalize the new Krylov states variationally against all other Krylov states. This is essentially variational compression of a state under the additional constraints of having zero overlap with all previous Krylov states, see Section 2.6. The additional constraints can be incorporated with the method of Lagrange multipliers: For each constraint (orthogonal vector $|v_i\rangle$), introduce the respective Lagrange multiplier β_i. To minimize $\|\hat{H}|v_i\rangle - |v_{i+1}\rangle\|^2$ under the constraints $\{\langle v_{i+1}|v_{i'}\rangle = 0\}_{0 \leq i' \leq i}$, we actually minimize

$$\left\|\hat{H}|v_i\rangle - |v_{i+1}\rangle\right\|^2 + \sum_{i'} \beta_{i'} \langle v_{i+1}|v_{i'}\rangle \tag{115}$$

with regard to $|v_i\rangle$ and the $\beta_{i'}$. As with variational compression, this can also be solved by iteratively solving the local one- or two-site problems (without explicitly evaluating $\langle \hat{H}^2 \rangle$). Care should be taken to ensure local orthogonality by using the pseudo-inverse of the Gram matrix as explained in Ref. [29]. Using a two-site approach entails an additional truncation step after each local optimization step and implies again a loss of orthogonality. However, the two-site approach converges much better than the single-site approach towards the global optimum. In practice, we hence first do a few sweeps using the two-site optimization (or, similarly, a single-site optimization with subspace expansion [81]) and follow up with a few sweeps of fully single-site optimization without expansion and hence also without truncation. The resulting state is then exactly orthogonal to all previous states. Note that when initially starting the optimization Eq. (115), the available vector space on the first few sites is likely to be very small (e.g., $\sigma^2 \cdot (m_2 = \sigma^2)$) and the orthogonalization hence overconstrained. To avoid this problem, one should add the constraints one-by-one during subsequent sweeps.

This variational orthogonalization can either be used as a separate orthogonalization step after the MPO–MPS application (using any of the algorithms presented in Section 2.8) or it can be combined with the variational operator application (cf. Section 2.8.2). Whether it is better to first do the MPO–MPS application using the zip-up method and then variationally orthogonalize the result or to do both steps at once depends on the system at hand: in particular with long-range interactions, the variational approach may require more sweeps to converge whereas short-range interactions are dealt with very efficiently there.

5.4. Dynamic step sizing

Dynamic step sizing is one of the most interesting and powerful features of this method and can be used in several ways. The idea is that a Krylov subspace, which was computed for some time step δ, can be recycled for some other step length. It is possible to distinguish two cases: interpolation and extrapolation.

5.4.1. Interpolation

In some applications, the time evolution needs to be performed on a very fine grid in time. The time-stepping methods would involve a single step for each point of the grid, which can quickly turn cumbersome or even impossible. On the other hand, if we have a Krylov subspace that we used to perform a large time step, it can be re-used to compute any intermediate smaller time step at the same or higher accuracy. This immediately follows from the construction of the Krylov space and the convergence criteria/assumptions made above. As the diagonalization of the effective Hamiltonian is already known, all we need to do is exponentiate the diagonal times the new time step, map back into the Krylov basis to get the coefficient vector, and compute the new MPS as a superposition of Krylov vectors. If one is only interested in expectation values of an observable $\hat{O}$, it is advantageous to compute its projection into the Krylov space via $(O_N)_{i,i'} = \langle \phi_i|\hat{O}|\phi_{i'}\rangle$ with complexity $\sim \mathcal{O}(n^2)$. With the coefficient vector c_N, the desired expectation value can be computed as $c_N^\dagger O_N c_N$, entirely skipping the more expensive superposition of Krylov states.

5.4.2. Extrapolation

Albeit trickier to implement, extrapolation can significantly improve performance when used as a kind of automatic adaptive step sizing scheme. The idea is as follows: Given a Krylov space, it is also often possible to recycle it for larger step sizes, by only adding a small number of additional Krylov vectors (or none at all). It follows that the optimal Krylov subspace dimension minimizes the ratio of the time needed to compute its basis and the number of steps that it can be used for. As crude approximations of these quantities, we assume that the cost of any new Krylov vector grows exponentially, i.e. the ratio of the costs of successive vectors is fixed. Furthermore, we also assume that any new Krylov vector allows us as many additional time steps as the previous Krylov vector. We then continuously monitor the time needed to construct a new Krylov vector and the number of steps we are able to take with it. Once a decision has to be taken whether to extend the Krylov space or rebuild it from scratch, we use those values as estimates for our decision. In practice, this heuristic has proven to be quite reliable.

6. MPS-local methods

The global Krylov method of the previous section uses generic features of Krylov subspaces to compute the time evolution of an MPS. However, working "globally" and only using MPS to represent the Krylov vectors comes with the major drawback that those vectors are typically much more entangled than the actual time-evolved state $|\psi(t + \delta)\rangle$. To circumvent this problem, it may be desirable to work in the local basis of reduced sites.[7] There are two different ways to obtain such a local basis, one being via an attempt to directly Lie–Trotter decompose the Hamiltonian into local subspaces [27,32–35], the other by enforcing a constraint that the evolution lies within the MPS manifold [36,37]. Both have been used in the literature, but their precise derivation requires some work. In the following, we will first present a very heuristic understanding of this approach, followed by a detailed derivation of the first approach, which leads to the local Krylov method presented in Section 6.1, and subsequently the second approach, which results in the time-dependent variational principle, TDVP, presented in Section 6.2.

We start with an MPS $|\psi\rangle$, a Hamiltonian $\hat{H}$ and a site j. The tensors $\psi^L_{j-1} \equiv (M_1, \ldots, M_{j-1})$ and $\psi^R_{j+1} \equiv (M_{j+1}, \ldots, M_L)$ then constitute a left and a right map, respectively, from the joint Hilbert space on sites 1 through $j - 1$ onto the bond space m_{j-1} and from the joint Hilbert space on sites $j + 1$ through L onto the bond space m_j.

By sandwiching the Hamiltonian between two copies of ψ^L_{j-1} and ψ^R_{j+1}, we can transform it to act on the effective single-site space on site j (of dimension $\sigma_j \times m_{j-1} \times m_j$). Let us call this effective single-site Hamiltonian $\hat{H}^{\text{eff}}_j$, cf. Fig. 13. Likewise, we can take the MPS $|\psi\rangle$ and act on it with a single copy of $\bar{\psi}^L_{j-1}$ and $\bar{\psi}^R_{j+1}$ to obtain the effective single-site tensor $\tilde{M}_j$ representing the effective state $|\psi^{\text{eff}}_j\rangle$, cf. Fig. 12. If we move the orthogonality center of the MPS to site j first, the transformation of the state is an identity operation, as the $\bar{A}_1$ of the transformation operator cancels with the A_1 in the state etc. This is desirable for many reasons, primarily numerical stability and avoiding a generalized eigenvalue problem in favor of a standard eigenvalue problem.

Instead of mapping onto the space of a single site, we can also map onto the space of two sites. This results in an effective state $|\psi^{\text{eff}}_{(j,j+1)}\rangle$ given by the product of two site tensors $M_j \cdot M_{j+1}$ over their shared bond and likewise an effective Hamiltonian $\hat{H}^{\text{eff}}_{(j,j+1)}$. On the other hand, if we map onto the space of a single bond between sites j and $j + 1$ (without a physical index), we obtain the center matrix[8] C_j to represent the effective state $|\psi^{\text{eff}}_j\rangle$ and the effective center site Hamiltonian $\hat{H}^{\text{eff}}_j$.

In any of these smaller spaces – typically in the basis of left- and right-orthogonal environments as produced naturally during left-to-right and right-to-left sweeps – it is very easy to solve the time-dependent Schrödinger equation exactly from time t to time $t + \delta$ using a local Krylov

---

[7] From a practical point of view, this is also motivated by the DMRG formulation which *always* works in local spaces and hence makes it very difficult to formulate a global Krylov method.

[8] As before, we will use underlined indices $_j$ to refer to bond tensors between sites j and $j+1$ whereas $_j$ (without the underline) refers to the site tensor or effective Hamiltonian tensors on site j.
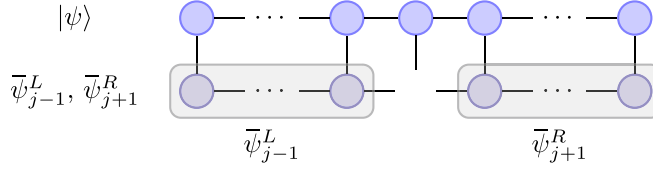
Fig. 12. The effective state $|\psi_j^{\text{eff}}\rangle$ is obtained by projecting the MPS with itself. In case of a mixed-orthogonal MPS with orthogonality center on site j, $|\psi_j^{\text{eff}}\rangle$ is simply the local site tensor M_j and the left and right projectors are identity matrices.
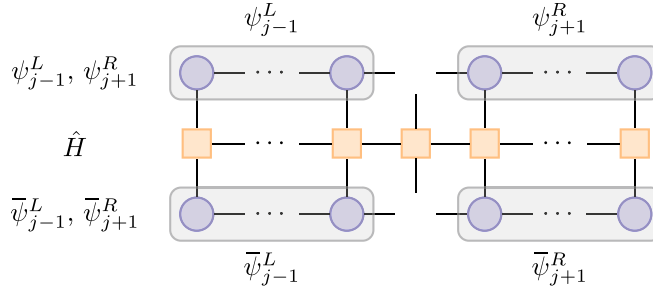
Fig. 13. The effective Hamiltonian $\hat{H}_j^{\text{eff}}$ obtained by projecting the MPO for $\hat{H}$ using ψ_{j-1}^L, ψ_{j+1}^R, $\overline{\psi}_{j-1}^L$, $\overline{\psi}_{j+1}^R$. Note that this tensor is never explicitly constructed! Only its action on the state tensor M_j is evaluated.

procedure or any other exponential solver [23]. This results then in a new effective state $|\psi_j^{\text{eff}}(t+\delta)\rangle$. Unfortunately, the new state will be represented in a relatively bad basis of the environments ψ^L and ψ^R which are optimized to represent the original state at time t. To alleviate this issue, after evolving the site tensor M_j forwards by δ, one now wishes to obtain a global state $|\psi(t)\rangle$ at the original time but in the new left auxiliary basis. If this is achieved, we may iteratively update the basis transformation tensors to be suitable for the new state $|\psi(t+\delta)\rangle$.

This is where the local Krylov method and the TDVP diverge and proceed differently.[9] However, regardless of how one obtains this old state in the new basis, by sweeping once through the system, all site tensors are incrementally optimized for a representation of the next time step while the global state is kept at the original time. Upon reaching the far end of the system, we keep the local state tensor at the new time and hence obtain a global state at the new time $t + \delta$ written using near-optimal basis transformation tensors.

Instead of solving the forward problem a single site at a time, it is also possible to consider the effective Hamiltonian and state of two neighboring sites. In this two-site scheme, one obtains a forward-evolved tensor $M_{(j,j+1)}$ which needs to be split up using a SVD, with then the single-site tensor on site $j + 1$ backwards-evolved to keep the state at the original time. The advantage is the variable bond dimension which allows a successive adaptation of the MPS bond dimension to the entanglement growth in the system. The downside is that an additional truncation error is introduced. In practice, it appears that a hybrid scheme which first uses the two-site method to increase the bond dimension to the usable maximum and then switches to the single-site variant may fare best, at least for some observables [84].

Depending on the particular way in which we obtain the local problems, solving the system of local Schrödinger equations exactly (using a small direct solver each time) is possible. The result is that both the energy and the norm (where applicable) of the state are conserved exactly. This

---

[9] An interesting intermediate between the local Krylov method based on time-step targeting and the TDVP as introduced in Refs. [36,37] was presented in Refs. [82,83] where the time-step targeting was combined with a variational principle to solve the local TDSE.

conservation may extend to other global observables [84,85]. In practice, the TDVP achieves this goal.

Finally, one may turn the first-order integrator of sweeping once through the system into a second-order integrator by sweeping once left-to-right and then right-to-left through the system, each time with a reduced time step $\delta/2$.

6.1. The local Krylov method

Deriving a Lie–Trotter integration scheme while working in the local reduced spaces is in fact equivalent to translating the time-step targeting DMRG [27,32–35] into the MPS framework. Crucially, the MPS framework makes it possible to precisely analyze the errors made, something which would be very difficult – and to our knowledge has not been done – in the standard environment-system DMRG picture. To integrate the local time-dependent Schrödinger equations resulting from the Lie–Trotter decomposition, we will use a Krylov-based approach [27,33,34]. This approach has the advantage of a very precise solution of the local equations and a large degree of similarity with both ground-state search DMRG and the time-dependent variational principle (cf. Section 6.2). Alternatively, Runge–Kutta integrators have also been used extensively with only minor changes to the overall method [32,35]. In particular, the error analysis presented here is also valid for the Runge–Kutta integrator, though one of course also has to include the additional time-step error of this integrator.

We begin by looking at the Lie–Trotter decomposition of the time-dependent Schrödinger equation

$$-\mathrm{i}\frac{d}{dt}|\psi\rangle = \hat{H}|\psi\rangle \equiv \sum_{\nu} \hat{H}_{\nu}|\psi\rangle. \tag{116}$$
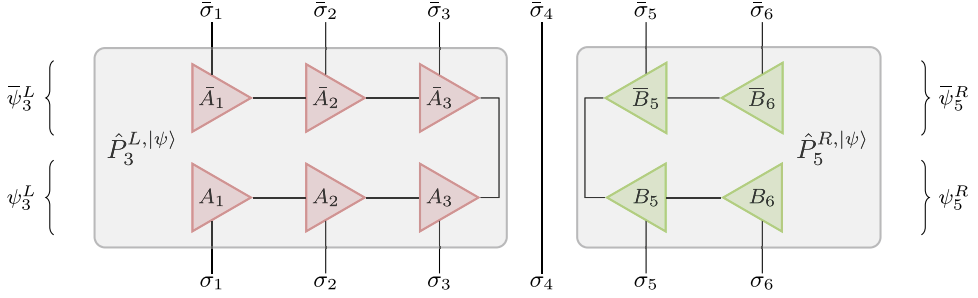
The goal is to find a decomposition scheme $\hat{H} = \sum_{\nu}\hat{H}_{\nu}$ such that we can integrate each summand separately by taking advantage of the MPS representation of the state vector $|\psi\rangle$. Therefore we define orthogonal projectors $\hat{P}_j^{L,|\psi\rangle}$ and $\hat{P}_j^{R,|\psi\rangle}$ acting on the physical degrees of freedom in a partition of the Hilbert space. For that purpose we introduce bipartitions $\mathcal{H} = \mathcal{H}_j^L \otimes \mathcal{H}_{j+1}^R$ where $\mathcal{H}_j^L = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_j$ and $\mathcal{H}_{j+1}^R = \mathcal{H}_{j+1} \otimes \cdots \otimes \mathcal{H}_L$ and declare

$$\hat{P}_j^{L,|\psi\rangle} : \mathcal{H}_j^L \otimes \mathcal{H}_{j+1}^R \longrightarrow \mathcal{H}_j^L \otimes \mathcal{H}_{j+1}^R$$
$$\hat{P}_j^{L,|\psi\rangle} = \sum_{\substack{\sigma_1,\ldots,\sigma_j, \\ \bar{\sigma}_1,\ldots,\bar{\sigma}_j, \\ m_j}} \underbrace{A_1\cdots A_j}_{\equiv \psi_{j;m_j}^L} \underbrace{\bar{A}_j \cdots \bar{A}_1}_{\equiv \bar{\psi}_{j;m_j}^L} |\sigma_1 \cdots \sigma_j\rangle\langle\bar{\sigma}_1 \cdots \bar{\sigma}_j| \otimes \hat{\mathbf{1}}_{j+1}^R \tag{117}$$

$$\hat{P}_j^{R,|\psi\rangle} : \mathcal{H}_{j-1}^L \otimes \mathcal{H}_j^R \longrightarrow \mathcal{H}_{j-1}^L \otimes \mathcal{H}_j^R$$
$$\hat{P}_j^{R,|\psi\rangle} = \hat{\mathbf{1}}_{j-1}^L \otimes \sum_{\substack{\sigma_j,\ldots,\sigma_L, \\ \bar{\sigma}_j,\ldots,\bar{\sigma}_L, \\ m_{j-1}}} \underbrace{\bar{B}_L \cdots \bar{B}_j}_{\equiv \bar{\psi}_{j;m_{j-1}}^R} \underbrace{B_j \cdots B_L}_{\equiv \psi_{j;m_{j-1}}^R} |\sigma_j \cdots \sigma_L\rangle\langle\bar{\sigma}_j \cdots \bar{\sigma}_L| \tag{118}$$

with mappings $\psi_{j;m_j}^L$ from a part of the physical Hilbert space into the bond space m_j. By construction, these operators fulfill $\left(\hat{P}_j^{L/R,|\psi\rangle}\right)^2 = \hat{P}_j^{L/R,|\psi\rangle}$ and $\left(\hat{P}_j^{L/R,|\psi\rangle}\right)^{\dagger} = \hat{P}_j^{L/R,|\psi\rangle}$, i.e., they are projectors. They are explicitly constructed from left-/right orthogonalized MPS site tensors. The action of such projectors onto an MPS representation of $|\psi\rangle$ in canonical form with orthogonality center at site $j+1$ $(j-1)$ is then given via

$$\hat{P}_j^{L,|\psi\rangle}|\psi\rangle = \sum_{\substack{\sigma_1,\ldots,\sigma_j, \\ \sigma_1',\ldots,\sigma_j', \\ \bar{\sigma}_1',\ldots,\bar{\sigma}_j', \\ m_1,\ldots,m_j,\bar{m}_j'}} \psi_{j;m_j'}^{L;\sigma_1',\ldots,\sigma_j'} \underbrace{\bar{\psi}_{j;\bar{m}_j'}^{L;\bar{\sigma}_1',\ldots,\bar{\sigma}_j'} A_{1;m_1}^{\sigma_1}\cdots A_{j;m_{j-1},m_j}^{\sigma_j}}_{=\delta_{\bar{m}_j',m_j}\delta_{\sigma_1',\sigma_1}\cdots\delta_{\bar{\sigma}_j',\sigma_j}} \sum_{\sigma_{j+1},\ldots,\sigma_L} M_{j+1}B_{j+2}\cdots B_L |\sigma_1\cdots\sigma_L\rangle$$

Fig. 14. Projector $\hat{P}_3^{L,|\psi\rangle} \otimes \hat{1}_4 \otimes \hat{P}_5^{R,|\psi\rangle}$ as defined in Eq. (121) and also used in the TDVP projector Eq. (158), here at the example of a six-site system. The other terms in Eq. (158) are constructed correspondingly.

$$= \sum_{\sigma_1,\dots,\sigma_L} A_1 \cdots A_j M_{j+1} B_{j+2} \cdots B_L |\sigma_1 \cdots \sigma_L\rangle = |\psi\rangle \tag{119}$$

$$\hat{P}_j^{R,|\psi\rangle}|\psi\rangle = \sum_{\sigma_1,\dots,\sigma_{j-1}} A_1 \cdots A_{j-2} M_{j-1} \sum_{\substack{\sigma_j,\dots,\sigma_L, \\ \sigma_j',\dots,\sigma_L', \\ \bar{\sigma}_j',\dots,\bar{\sigma}_L', \\ m_j,\dots,m_L,\bar{m}_j'}} \underbrace{B_{j;m_{j-1},m_j}^{\sigma_j} \cdots B_{L;m_{L-1}}^{\sigma_L} \bar{\psi}_{j;\bar{m}_{j-1}'}^{R;\bar{\sigma}_j',\dots,\bar{\sigma}_L'} \psi_{j;m_{j-1}'}^{R;\sigma_j',\dots,\sigma_L'}}_{\delta_{m_j,\bar{m}_j'}\delta_{\bar{\sigma}_j',\sigma_j}\cdots\delta_{\bar{\sigma}_L',\sigma_L}} |\sigma_1 \cdots \sigma_L\rangle$$
$$= \sum_{\sigma_1 \cdots \sigma_L} A_1 \cdots A_{j-2} M_{j-1} B_j \cdots B_L |\sigma_1 \cdots \sigma_L\rangle = |\psi\rangle . \tag{120}$$

That is, if the state $|\psi\rangle$ has orthogonality center to the right (left) of the target index j, the projectors constructed from it act as an identity on their Hilbert space partition. Next we define the projector on the reduced site-space at site j (cf. Fig. 14) via

$$\hat{\Pi}_j^{|\psi\rangle} \equiv \hat{P}_{j-1}^{L,|\psi\rangle} \otimes \hat{1}_j \otimes \hat{P}_{j+1}^{R,|\psi\rangle} . \tag{121}$$

The action of such a projector $\hat{\Pi}_j^{|\psi\rangle}$ on a state $|\psi\rangle$ is to shift the orthogonality center of $|\psi\rangle$ to the site j which can be shown by applying the manipulation depicted in Fig. 15 on to $|\psi\rangle$ repeatedly. Therein gauge invariance is employed so that the action of the projector is trivial on the site tensors A_k/B_k, $k \neq j$ rendering j the center of orthogonality. Thus, the quantum state $|\psi\rangle$ remains unchanged under the action of $\hat{\Pi}_j^{|\psi\rangle}$ and therefore

$$\langle\phi|\hat{\Pi}_j^{|\psi\rangle}|\psi\rangle = \langle\phi|\psi\rangle \tag{122}$$

$$\Rightarrow \hat{\Pi}_j^{|\psi\rangle}|\phi\rangle = \left(\frac{|\psi^\perp\rangle\langle\psi^\perp|}{\langle\psi^\perp|\psi^\perp\rangle} + \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} \right) \hat{\Pi}_j^{|\psi\rangle}|\phi\rangle$$

$$= \frac{\langle\psi^\perp|\hat{\Pi}_j^{|\psi\rangle}|\phi\rangle}{\langle\psi^\perp|\psi^\perp\rangle}|\psi^\perp\rangle + \frac{\langle\psi|\hat{\Pi}_j^{|\psi\rangle}|\phi\rangle}{\langle\psi|\psi\rangle}|\psi\rangle = \frac{\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle}|\psi\rangle \tag{123}$$

It is instructive to think of $\hat{\Pi}_j^{|\psi\rangle}$ as an operator acting on both the physical and gauge degrees of freedom of the MPS representation. In the physical system it acts as a projector on the physical indices σ_j of the source state $|\psi\rangle$. In the gauge degrees of freedom of the MPS representation, $\hat{\Pi}_j^{|\psi\rangle}|\phi\rangle$ fixes the orthogonality center to site j. As the physical content of the state is independent of the location of its orthogonality center, we must have

$$\langle\psi|\psi\rangle = \left(\langle\psi|\hat{\Pi}_j^{|\psi\rangle} \right) \left(\hat{\Pi}_i^{|\psi\rangle}|\psi\rangle \right) \tag{124}$$
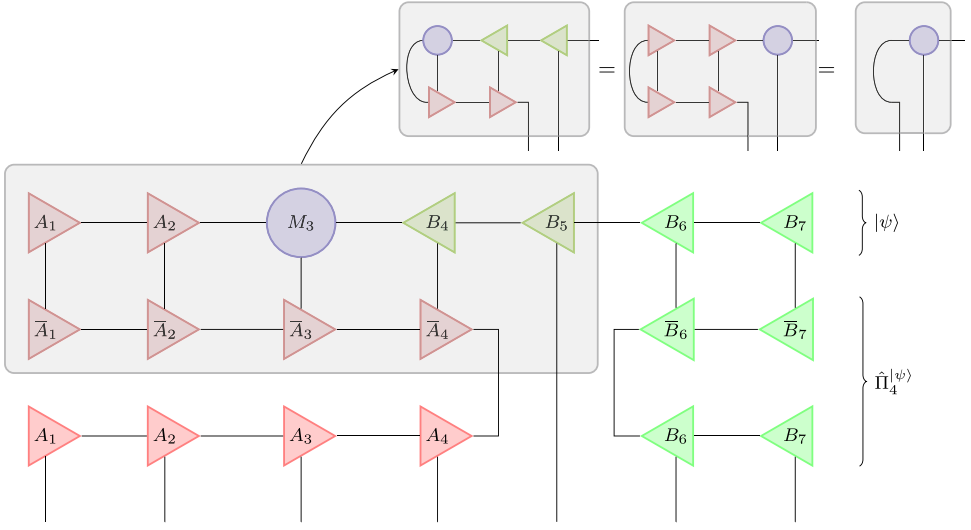
Fig. 15. Shifting the center of orthogonality from $3 \to 5$ under the action of $\hat{\Pi}_5^{|\psi\rangle}|\psi\rangle = \hat{P}_4^{L,|\psi\rangle} \otimes \hat{\mathbf{1}}_5 \otimes \hat{P}_6^{R,|\psi\rangle}|\psi\rangle$. The orthogonality center is implicitly shifted to site 5 (without changing the state content, cf. gray boxes) which gives identities on sites 1 through 4 and sites 6 and 7 and a new orthogonality center tensor M_5. The completely contracted upper two rows then define the new tensor M_5.

as also immediately follows from Eq. (122). Now, we can reformulate the action of the Hamiltonian by decomposing it into representations $\hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}$ acting only onto reduced site-spaces:

$$\hat{H} \approx \frac{1}{L\||\psi\rangle\|} \sum_j \hat{\Pi}_j^{|\psi\rangle} \hat{H} \hat{\Pi}_j^{|\psi\rangle} \equiv \frac{1}{L\||\psi\rangle\|} \sum_j \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}} \tag{125}$$

$$\hat{H}|\psi\rangle \approx \frac{1}{L\||\psi\rangle\|} \sum_j \hat{\Pi}_j^{|\psi\rangle} \hat{H} \hat{\Pi}_j^{|\psi\rangle} |\psi\rangle \tag{126}$$

$$= \frac{1}{L\||\psi\rangle\|} \sum_j \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}} |\psi\rangle \tag{127}$$

which indeed yields a Lie–Trotter decomposition of the time-dependent Schrödinger equation

$$-\mathrm{i}\frac{d}{dt}|\psi(t)\rangle = \frac{1}{L\||\psi\rangle\|} \sum_j \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}} |\psi(t)\rangle \; . \tag{128}$$

Typically, this decomposition is not exact, it depends for instance on the size of the chosen basis, but given a sufficiently large MPS bond dimension, the error made will be small.

Having obtained the Lie–Trotter decomposition, we proceed by formulating a recursive integration scheme which is suitable for the particular structure of the MPS representation. For this purpose note that a first-order approximation to the evolved state can be obtained by solving each problem $-\mathrm{i}\frac{d}{dt}|\psi(t)\rangle = \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}|\psi(t)\rangle$ independently.[10] Let $|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle$ be the solution of the jth problem. An approximation to the overall time-evolved state from $t \to t+\delta$ can then be obtained by sequentially

---

[10] The prefactors $\frac{1}{L\||\psi\rangle\|}$ are absorbed into the normalization of the site tensor that is currently evolved.

solving the initial value problems (setting $|\psi^{\hat{\Pi}_0^{|\psi\rangle}}(t)\rangle \equiv |\psi(t)\rangle$)

$$-\mathrm{i}\frac{d}{dt}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle = \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle$$

$$\text{and} \quad |\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle = |\psi^{\hat{\Pi}_{j-1}^{|\psi\rangle}}(t+\delta)\rangle \tag{129}$$

and identifying $|\psi(t+\delta)\rangle \equiv |\psi^{\hat{\Pi}_L^{|\psi\rangle}}(t+\delta)\rangle$ with the approximated time evolved state. Comparing the formal Taylor expansions of the exactly integrated state $|\psi(t+\delta)\rangle_{\text{exact}}$ with the approximation one readily finds

$$|\psi(t+\delta)\rangle_{\text{exact}} = |\psi(t)\rangle - \mathrm{i}\delta\hat{H}|\psi(t)\rangle - \frac{\delta^2}{2}\hat{H}^2|\psi(t)\rangle + \cdots \tag{130}$$

$$|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t+\delta)\rangle = |\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle + \delta\frac{d}{dt}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle + \frac{\delta^2}{2}\frac{d^2}{dt^2}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle + \cdots \tag{131}$$

$$= |\psi^{\hat{\Pi}_{j-1}^{|\psi\rangle}}(t+\delta)\rangle - \mathrm{i}\delta\hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle - \frac{\delta^2}{2}\hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}\hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle + \cdots$$

$$\Rightarrow |\psi(t+\delta)\rangle = |\psi(t)\rangle - \mathrm{i}\delta\hat{H}|\psi(t)\rangle - \frac{\delta^2}{2}\sum_{i\leq j}\hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}\hat{H}^{\hat{\Pi}_i^{|\psi\rangle}}|\psi(t)\rangle\cdots \tag{132}$$

where we have already applied Eq. (129) to the expansion of the partially time-evolved intermediate state $|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t+\delta)\rangle$ around $\delta = 0$. Replacing $|\psi^{\hat{\Pi}_{j-1}^{|\psi\rangle}}(t+\delta)\rangle$ by its Taylor expansion around $\delta = 0$ generates a recursion for the partially time-evolved states. Applying this recursion L times and using the initial condition $|\psi^{\hat{\Pi}_0^{|\psi\rangle}}(t)\rangle = |\psi(t)\rangle$ yields the last equation. As expected, $|\psi^{\hat{\Pi}_L^{|\psi\rangle}}(t+\delta)\rangle$ coincides with Eq. (128) up to the first order of the expansion. The first error occurs at second-order terms in the expansion and is given by

$$|\psi(t+\delta)\rangle_{\text{exact}} - |\psi(t+\delta)\rangle = \frac{\delta^2}{2}\sum_{i<j}\left[\hat{H}^{\hat{\Pi}_i^{|\psi\rangle}}, \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}\right]|\psi(t)\rangle \tag{133}$$

and further commutators at arbitrary high orders. Due to the dependence of the reduced problem currently solved onto the previous solution in Eq. (129) the projectors $\hat{P}_{j-1}^{L,|\psi\rangle}$ in our integration scheme are in fact time-dependent because they need to be constructed from site tensors $A_{j'<j}(t+\delta)$ that have already been evolved. In contrast, $\hat{P}_{j+1}^{R,|\psi\rangle}$ is constructed from unevolved site tensors $B_{j'>j}(t)$. We hence have to actually solve the L initial value problems

$$-\mathrm{i}\frac{d}{dt}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle = \left[\hat{\Pi}_j^{|\psi\rangle}(t, t+\delta)\hat{H}\hat{\Pi}_j^{|\psi\rangle}(t, t+\delta)\right]|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle$$

$$\text{and} \quad |\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle = |\psi^{\hat{\Pi}_{j-1}^{|\psi\rangle}}(t+\delta)\rangle \tag{134}$$

with $\hat{\Pi}_j^{|\psi\rangle}(t, t+\delta) \equiv \hat{P}_{j-1}^{L,|\psi^{\hat{\Pi}_{j-1}^{|\psi\rangle}}(t+\delta)\rangle} \otimes \hat{\mathbf{1}}_j \otimes \hat{P}_{j+1}^{R,|\psi(t)\rangle}$. Note that while these problems appear to evolve over L time steps with step size δ, they only do so using different local Hamiltonians. Those local problems are then connected via their initial conditions (c.f. Eq. (129)) to generate a global time evolution by δ under (an approximation of) the global Hamiltonian.

Having decoupled the global Schrödinger equation into L local problems, we can take advantage of the local representation of the state by multiplying each local differential equation at site j with the bond maps $\overline{\psi}_{j-1}^L \otimes \hat{\mathbf{1}}_j \otimes \overline{\psi}_{j+1}^R$

$$-\mathrm{i}\frac{d}{dt}\left[\overline{\psi}_{j-1}^L \otimes \hat{\mathbf{1}}_j \otimes \overline{\psi}_{j+1}^R\right]|\psi^{\hat{\Pi}_j^{|\psi\rangle}}\rangle = \left[\overline{\psi}_{j-1}^L \otimes \hat{\mathbf{1}}_j \otimes \overline{\psi}_{j+1}^R\right]\hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}\rangle$$

$$\Rightarrow -\mathrm{i}\frac{d}{dt}M_j = H_j^{\text{eff}}M_j \tag{135}$$

where we defined effective reduced-site state and operator representations M_j and H_j^{eff}, respectively (see also Figs. 12 and 13). It can be checked easily that the projection of the effective problems onto

their reduced site spaces at site j leaves the solution invariant. Carrying out the differentiation on the left-hand site explicitly yields a sum of partial differentiations. However, on the right hand site $\hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}$ acts only on the reduced site spaces and hence all differentials must vanish except the one acting on the tensor in the reduced site space.

Unfortunately, a direct adoption of the recursive solution strategy proposed above is not possible because the current problem at site j requires the projectors to be constructed from left-/right-canonical site-tensors $A_1, \ldots, A_{j-1}, B_{j+1}, \ldots, B_L$. However, the solution M_{j-1} of the previous reduced problem is not in general in a canonical representation so that in order to construct the next projector $\hat{\Pi}_j^{|\psi\rangle}$ we need to perform a basis transformation first. There is no prescription in the derived decomposition scheme which corresponds to such a basis transformation keeping the evolved site tensors in canonical form. On the other hand we can of course rewrite the updated site tensor as $M_{j-1} \to A_{j-1} R_{j-1}$ in a left-canonical representation. But, as already pointed out, we have to discard the transformation matrix R_{j-1} to obey the decomposition scheme. To proceed with the next reduced problem, we need to resolve the situation of having 2 different basis representations between the sites $(j-1, j)$. Hence, we introduce a transformation between the unevolved and evolved site tensors A_{j-1}, M_j. The simplest guess is to consider a direct mapping between the unevolved and evolved site tensor basis sets. This transformation can be readily obtained from the iterative solution strategy and is performed by projecting the evolved onto the unevolved bond space. To see this we write the partially evolved state explicitly in the mixed canonical form

$$|\psi(t, t+\delta)\rangle = \sum_{\substack{\sigma_1, \ldots, \sigma_L, \\ m_{j-1}, m_j}} \underbrace{A_1^{\sigma_1}(t+\delta) \cdots A_{j-1}^{\sigma_{j-1}}(t+\delta)}_{\equiv \psi_{j-1;m_{j-1}}^{L;\sigma_1,\ldots,\sigma_{j-1}}(t+\delta)} M_{j;m_{j-1},m_j}^{\sigma_j}(t) \underbrace{B_{j+1}^{\sigma_{j+1}}(t) \cdots B_L^{\sigma_L}(t)}_{\equiv \psi_{j+1;m_j}^{R;\sigma_{j+1},\ldots,\sigma_L}(t)} |\sigma_1 \cdots \sigma_L\rangle \quad (136)$$

and introduce a compact notation for the left/right partition's basis states

$$|\psi_j^L(t)\rangle_{m_j} = \sum_{\sigma_1,\ldots,\sigma_j} \psi_{j;m_j}^{L;\sigma_1,\ldots,\sigma_j}(t)|\sigma_1 \cdots \sigma_j\rangle \qquad |\psi_j^R(t)\rangle_{m_{j-1}} = \sum_{\sigma_j,\ldots,\sigma_L} \psi_{j;m_{j-1}}^{R;\sigma_j,\ldots,\sigma_L}(t)|\sigma_j \cdots \sigma_L\rangle \ . \quad (137)$$

Without truncation $|\psi_j^L(t)\rangle_{m_j}$ and $|\psi_j^R(t)\rangle_{m_{j-1}}$ would constitute complete maps from the physical degrees of freedom in the left/right partition and we could perform an exact basis transformation mapping the evolved into the unevolved left basis

$$|\psi(t)\rangle = \sum_{m_{j-1}} |\psi_{j-1}^L(t)\rangle_{m_{j-1}} \, _{m_{j-1}}\langle\psi_{j-1}^L(t)|\psi(t, t+\delta)\rangle \quad (138)$$

$$= \sum_{\substack{m_{j-1}, m'_{j-1}, \\ \sigma_j, m_j}} |\psi_{j-1}^L(t)\rangle_{m_{j-1}} \underbrace{_{m_{j-1}}\langle\psi_{j-1}^L(t)|\psi_{j-1}^L(t+\delta)\rangle_{m'_{j-1}}}_{Q_{j-1;m_{j-1},m'_{j-1}}(t,t+\delta)} M_{j;m'_{j-1},m_j}^{\sigma_j}(t)|\sigma_j\rangle|\psi_j^R(t)\rangle_{m_j} \quad (139)$$

$$= \sum_{\substack{\sigma_1,\ldots,\sigma_L, \\ m_{j-1}, m_j \\ m'_{j-1}}} \psi_{j-1;m_{j-1}}^{L;\sigma_1,\ldots,\sigma_{j-1}}(t) Q_{j-1;m_{j-1},m'_{j-1}}(t,t+\delta) M_{j;m'_{j-1},m_j}^{\sigma_j}(t)\psi_{j+1;m_j}^R(t)|\sigma_1\cdots\sigma_L\rangle. \quad (140)$$

and the matrix elements $Q_{j-1;m_{j-1},m'_{j-1}}(t, t+\delta)$ of the basis transformation $Q_{j-1}(t, t+\delta)$ are constructed from

$$Q_{j-1;m_{j-1},m'_{j-1}}(t, t+\delta) = \sum_{\sigma_1,\ldots,\sigma_{j-1}} \overline{\psi}_{j-1;m_{j-1}}^{L;\sigma_1,\ldots,\sigma_{j-1}}(t)\psi_{j-1;m'_{j-1}}^{L;\sigma_1,\ldots,\sigma_{j-1}}(t+\delta) \ . \quad (141)$$

The transformation $Q_j(t, t+\delta)$ maps bond basis states $|\psi_j^L(t+\delta)\rangle_{m_j}$ which are optimized to represent the evolved state $|\psi(t+\delta)\rangle$ into bond basis states $|\psi_j^L(t)\rangle_{m_j}$ which are optimized to represent the unevolved state $|\psi(t)\rangle$. Now, if we would let act $Q_{j-1}(t, t+\delta)$ to the already optimized canonical

site tensor $A_{j-1}^{\sigma_{j-1}}(t+\delta)$ then the effect would be to undo the previous site optimization. Hence, the inverse transformation $\overline{Q}_{j-1}(t, t+\delta)$ is employed to transform the unevolved bond basis labeled by m_{j-1} of the current site tensor $M_j(t)$. Observing that $\overline{Q}_{j-1}$ is constructed from the above introduced bond maps $\overline{\psi}_{j-1}^L(t)$ and $\psi_{j-1}^L(t+\delta)$ which themselves should be build recursively, the transformation can be updated and applied before solving the jth problem via

$$Q_{j-1}(t, t+\delta) = \overline{A}_{j-1}(t)Q_{j-2}(t, t+\delta)A_{j-1}(t+\delta), \quad M_j(t) \longrightarrow \overline{Q}_{j-1}(t, t+\delta)M_j(t). \tag{142}$$

If we allow for truncation the error incurred by this mapping depends on the overlap $\langle\psi(t+\delta)|\psi(t)\rangle$ as well as the discarded weight. This basis transformation is mostly motivated by its straightforward availability during the sweeping procedure. However, to the best of our knowledge there is no mathematical justification and we can only give the physical motivation that for small time steps δ the time-evolved state is expected to be relatively close to the unevolved state (deviation $\propto L\delta^2$ as follows from the consideration in Section 6.1.1).

Instead of mapping onto the space of a single site, in practice we map onto the space of two sites. The two-site local TDSE is solved using the time-dependent Lanczos approach to obtain $A_j(t+\delta)$. The original orthogonality center MPS tensor $M_{j+1}(t)$ is then projected onto the new left basis as described above. This allows for a flexible adaptation of not only the tensor A_j itself but also of the bond basis and – if necessary – MPS bond dimension between sites j and $j+1$.

Historically, only this two-site variant was used; but in analogy to the 2TDVP method presented later, it may well make sense to initially use the two-site local Krylov method until the desired maximal bond dimension has been obtained and then switch to the single-site integrator to save computational effort.
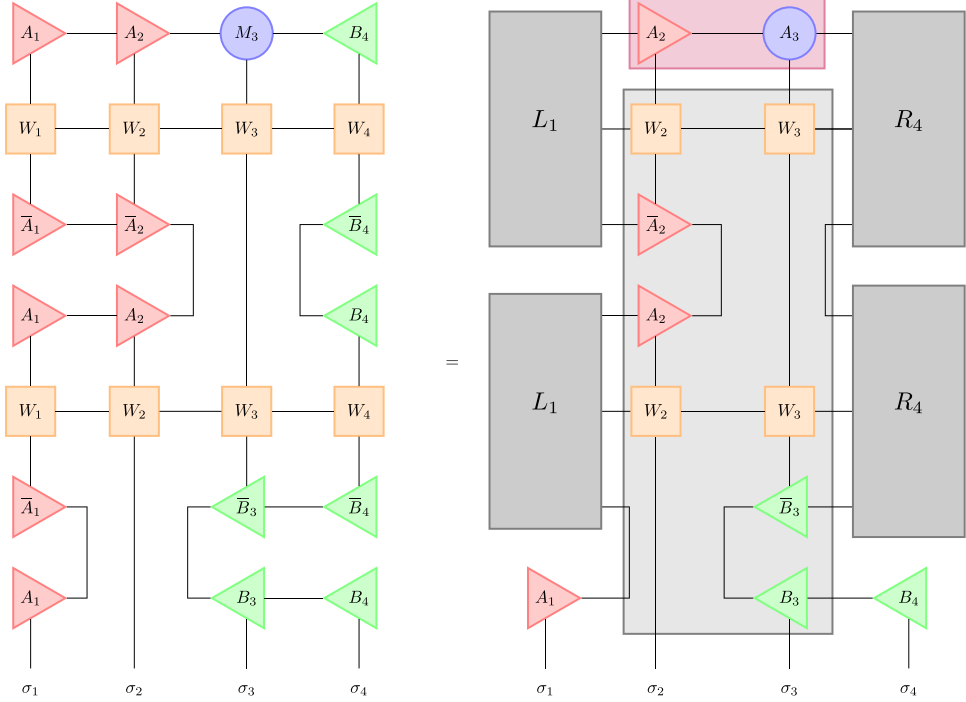
6.1.1. Errors

Four errors are present in the local Krylov method when used in its (standard) two-site variant. The smallest of those stems from the inexact solution of the local TDSE Eq. (135). This error can be made very small using a precise solver; in practice, a Krylov exponential as described in Section 5 with very few (4–5) vectors is sufficient. The second error is the standard truncation error incurred during the SVD to split the merged two-site tensors again while truncating to the desired bond dimension. This error can be measured and observed throughout the calculation and is much the same as in the other methods.

The third error is due to the approximation in Eq. (125). This projection error is difficult to measure and strongly depends on the initial state. If the initial state has a reasonably large bond dimension and the Hamiltonian has reasonably short-range interactions, this error will be very small. The longer the interactions in the Hamiltonian, the larger the state has to be. In the two-site method, nearest-neighbor interactions can be handled at all bond dimensions, in the single-site variant, only on-site interactions are error-free at small bond dimensions. The projection error is in particular problematic when globally quenching from a product state with a very non-local Hamiltonian (e.g. resulting from a 2D → 1D map). When calculating equilibrium Green's functions for short-range Hamiltonians, this error is quite negligible.

Finally, there is an error due to the sequential solution of the local TDSE as resulting from the Lie–Trotter decomposition. This error *can* be quantified, but doing so requires some additional work which will follow now: We continue from the Taylor expansion Eq. (132). We emphasize that the action of the commutators $\left[\hat{H}^{\hat{\Pi}_i^{|\psi\rangle}}, \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}\right]|\psi(t_0)\rangle$ needs to be evaluated with respect to the iteration the commutators are generated from. Consider, for instance, the action of $\hat{H}^{\hat{\Pi}_2^{|\psi\rangle}}\hat{H}^{\hat{\Pi}_3^{|\psi\rangle}}|\psi(t)\rangle$, which is generated from the first-order contribution $\hat{H}^{\hat{\Pi}_3^{|\psi\rangle}}|\psi^{\hat{\Pi}_3^{|\psi\rangle}}\rangle$ and subsequent application of $\hat{H}^{\hat{\Pi}_2^{|\psi\rangle}}$. Thus the commutator also needs to be evaluated considering the partial solution $|\psi^{\hat{\Pi}_j^{|\psi\rangle}}\rangle$, so that in general we have for $i < j$

$$\left[\hat{H}^{\hat{\Pi}_i^{|\psi\rangle}}, \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}\right]|\psi(t)\rangle = \left[\hat{H}^{\hat{\Pi}_i^{|\psi\rangle}}, \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}\right]|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle. \tag{143}$$

Fig. 16. Evaluation of $\hat{H}^{\hat{\Pi}_2^{|\psi\rangle}}\hat{H}^{\hat{\Pi}_3^{|\psi\rangle}}|\psi^{\hat{\Pi}_3^{|\psi\rangle}}(t)\rangle$ at the example of a four-site system. The tensors L_i/R_i correspond to partially contracted MPS–MPO–MPS-networks. The burgundy-shaded rectangular area at the top encloses the reduced site tensor $\psi_{i,\ldots,j;m_{i-1},m_j}^{C;\sigma_i,\ldots,\sigma_j}$. The commutator $\left[\hat{H}^{\hat{\Pi}_2^{|\psi\rangle}}\hat{H}^{\hat{\Pi}_3^{|\psi\rangle}}\right]|\psi_3^{\Pi}(t)\rangle$ requires also the calculation of $\hat{H}^{\hat{\Pi}_3^{|\psi\rangle}}\hat{H}^{\hat{\Pi}_2^{|\psi\rangle}}|\psi_3^{\Pi}(t)\rangle$ which is obtained from vertically flipping the tensors covered by the central rectangular gray area between the sites (i, j), i.e., $(2, 3)$ in the presented example.

In Fig. 16 the action of $\hat{H}^{\hat{\Pi}_i^{|\psi\rangle}}\hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle$ is demonstrated in case of a four-site system with $i = 2, j = 3$ by performing most of the contractions graphically. In order to obtain the matrix element with an arbitrary state $|\phi\rangle$ we will introduce a compact notation for contractions of MPS and MPO site-tensors with the boundary tensors (partially contracted MPS–MPO–MPS-networks L_j/R_j)

$$L_{j-1}E_j \equiv L_{j-1}\overline{M}_j W_j M_j = \sum_{m_{j-1},\sigma_j'}\left(\sum_{w_{j-1},\sigma_j}\left(\sum_{\overline{m}_{j-1}}L_{j-1;m_{j-1}}^{\overline{m}_{j-1},w_{j-1}}\overline{M}_{j;\overline{m}_{j-1},\overline{m}_j}^{\sigma_j}\right)W_{j,w_{j-1},w_j}^{\sigma_j,\sigma_j'}\right)M_{j,m_{j-1},m_j}^{\sigma_j'}$$

(144)

$$E_jR_{j+1} \equiv \overline{M}_j W_j M_j R_{j+1} = \sum_{\overline{m}_j,\sigma_j}\overline{M}_{j;\overline{m}_{j-1},\overline{m}_j}^{\sigma_j}\left(\sum_{w_j,\sigma_j'}W_{j;w_{j-1},w_j}^{\sigma_j,\sigma_j'}\left(\sum_{m_j}M_{j;m_{j-1},m_j}^{\sigma_j'}R_{j+1;m_j}^{\overline{m}_j,w_j}\right)\right)$$

(145)

with the transfer tensors $E_j = \overline{M}_j W_j M_j$. We also need transfer tensors with the target state which we define by $E_j^{\phi} = \overline{M}_j^{\phi} W_j M_j$. Finally, there will be *open* bonds at sites i and j that correspond to the contractions originating from the "brace"-contractions in the projectors $\hat{P}_i^{L,|\psi\rangle}$ and $\hat{P}_i^{R,|\psi\rangle}$ as well as $\hat{P}_{j+1}^{L,|\psi\rangle}$ and $\hat{P}_{j+1}^{R,|\psi\rangle}$, and we will label these bonds explicitly. Considering for instance the first summand

of the commutator at $i = 2, j = 3$, that is $\langle\phi|\hat{H}^{\hat{\Pi}_2^{|\psi\rangle}}\hat{H}^{\hat{\Pi}_3^{|\psi\rangle}}|\psi^{\hat{\Pi}_3^{|\psi\rangle}}(t)\rangle$ (c.f. Fig. 16) the left part of the contractions can be written as

$$\sum_{m_2, \bar{m}_2} \left(L_1 E_2\right)_{\bar{m}_2} \otimes \left(L_1 E_2^\phi\right)_{m_2} \delta_{\bar{m}_2, m_2} \equiv \sum_{m_2, \bar{m}_2} \left(L_1 \otimes L_1\right) \left(E_2 \otimes E_2^\phi\right)_{\bar{m}_2, m_2} \delta_{\bar{m}_2, m_2}. \tag{146}$$

To obtain a compact notation for the right part we introduce the "dangling" transfer tensors $D_j = W_j M_j$ if only the ket site-tensor of the state is included and $\bar{D}_j = \bar{M}_j W_j$ if only the bra site-tensor is considered. In the same manner as for the left part we can now write for the right contractions compactly

$$\sum_{m_3, \bar{m}_3} \left(D_3 R_4\right)_{\bar{m}_3} \otimes \left(\bar{D}_3^\phi R_4\right)_{m_3} \delta_{\bar{m}_3, m_3} \equiv \sum_{m_3, \bar{m}_3} \left(D_3 \otimes \bar{D}_3^\phi\right) \left(R_4 \otimes R_4\right)_{\bar{m}_3, m_3} \delta_{\bar{m}_3, m_3}. \tag{147}$$

In general we obtain for the matrix element of the commutator with the target state $|\phi\rangle$ (suppressing the Kronecker-δ for brevity)

$$\langle\phi|\left[\hat{H}^{\hat{\Pi}_i^{|\psi\rangle}}, \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}\right]|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle$$
$$= \sum_{\substack{m_{j-1}, \bar{m}_{j-1} \\ m_j, \bar{m}_j}} \left(L_{i-1} \otimes L_{i-1}\right) \left(E_i \cdots E_{j-1} \otimes E_i^\phi \cdots E_{j-1}^\phi\right)_{\bar{m}_{j-1}, m_{j-1}} \left(D_j \otimes \bar{D}_j^\phi\right) \left(R_{j+1} \otimes R_{j+1}\right)_{\bar{m}_j, m_j} + \cdots$$
$$- \sum_{\substack{m_{i-1}, \bar{m}_{i-1} \\ m_i, \bar{m}_i}} \left(L_{i-1} \otimes L_{i-1}\right)_{\bar{m}_{i-1}, m_{i-1}} \left(D_i \otimes \bar{D}_i^\phi\right) \left(E_{i+1} \cdots E_j \otimes E_{i+1}^\phi \cdots E_j^\phi\right)_{\bar{m}_i, m_i} \left(R_{j+1} \otimes R_{j+1}\right). \tag{148}$$

It becomes immediately clear that only those terms in the Hamiltonian contribute to the commutator which cross the bond $(j - 1, j)$ or $(i, i + 1)$. For the current purpose it suffices to find a general estimate for the incurred error $\propto \delta^2$ in the Taylor expansion. Thus, we will only consider the contributions from nearest-neighbor interactions and hence set $i = j - 1$ so that

$$\langle\phi|\left[\hat{H}^{\hat{\Pi}_{j-1}^{|\psi\rangle}}, \hat{H}^{\hat{\Pi}_j^{|\psi\rangle}}\right]|\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle$$
$$= \sum_{\substack{m_{j-1}, \bar{m}_{j-1} \\ m_j, \bar{m}_j}} \left(L_{j-2} \otimes L_{j-2}\right) \left(E_{j-1} \otimes E_{j-1}^\phi\right)_{\bar{m}_{j-1}, m_{j-1}} \left(D_j \otimes \bar{D}_j^\phi\right) \left(R_{j+1} \otimes R_{j-1}\right)_{\bar{m}_j, m_j} + \cdots$$
$$- \sum_{\substack{m_{j-2}, \bar{m}_{j-2} \\ m_{j-1}, \bar{m}_{j-1}}} \left(L_{j-2} \otimes L_{j-2}\right)_{\bar{m}_{j-2}, m_{j-2}} \left(D_{j-1} \otimes \bar{D}_{j-1}^\phi\right) \left(E_j \otimes E_j^\phi\right)_{\bar{m}_{j-1}, m_{j-1}} \left(R_{j+1} \otimes R_{j+1}\right). \tag{149}$$

The crucial observation here is that for each "open" bond index pair we can treat the combined MPS–MPS and MPO–MPO tensor-contractions over the bond $(j - 1, j)$ as scalar product between the left and the right part of the system. We can thus write for the first summand for each open index pair

$$\underbrace{\left(L_{j-2} \otimes L_{j-2}\right) \left(E_{j-1} \otimes E_{j-1}^\phi\right)_{\bar{m}_{j-1}, m_{j-1}}}_{\equiv \langle e_{j-1}^L \otimes e_{j-1}^{L,\phi}|} \underbrace{\left(D_j \otimes \bar{D}_j^\phi\right) \left(R_{j+1} \otimes R_{j-1}\right)_{\bar{m}_j, m_j}}_{\equiv |d_j^R \otimes \bar{d}_j^{R,\phi}\rangle}$$
$$= \frac{1}{2} \left(\left||e_{j-1}^L \otimes e_{j-1}^{L,\phi}\rangle + |d_j^R \otimes \bar{d}_j^{R,\phi}\rangle\right|^2 - \langle e_{j-1}^L \otimes e_{j-1}^{L,\phi}|e_{j-1}^L \otimes e_{j-1}^{L,\phi}\rangle - \langle d_j^R \otimes \bar{d}_j^{R,\phi}|d_j^R \otimes \bar{d}_j^{R,\phi}\rangle\right) \tag{150}$$

and for the second summand

$$\underbrace{\left(L_{j-2} \otimes L_{j-2}\right)_{\overline{m}_{j-2}, m_{j-2}} \left(D_{j-1} \otimes \bar{D}_{j-1}^{\phi}\right)}_{\equiv \langle d_{j-1}^L \otimes \bar{d}_{j-1}^{L,\phi}|} \underbrace{\left(E_j \otimes E_j^{\phi}\right)_{\overline{m}_{j-1}, m_{j-1}} \left(R_{j+1} \otimes R_{j+1}\right)}_{\equiv |e_j^R \otimes e_j^{R,\phi}\rangle}$$

$$= -\frac{1}{2}\left(\left||d_{j-1}^L \otimes \bar{d}_{j-1}^{L,\phi}\rangle - |e_j^R \otimes e_j^{R,\phi}\rangle\right|^2 - \langle d_{j-1}^L \otimes \bar{d}_{j-1}^{L,\phi}|d_{j-1}^L \otimes \bar{d}_{j-1}^{L,\phi}\rangle - \langle e_j^R \otimes e_j^{R,\phi}|e_j^R \otimes e_j^{R,\phi}\rangle\right). \tag{151}$$

We can bound the sums over the open bonds for the scalar products in terms of absolute values of expectation values of effective Hamiltonians $\hat{h}_j^{L/R,\text{eff}}$ at site j

$$\sum_{\substack{m_{j-1} \\ \overline{m}_{j-1}}} \langle e_{j-1}^L \otimes e_{j-1}^{L,\phi}|e_{j-1}^L \otimes e_{j-1}^{L,\phi}\rangle \leq \sum_{\substack{m_{j-1} \\ \overline{m}_{j-1}}} \left||e_{j-1}^L \otimes e_{j-1}^{L,\phi}\rangle\right|^2 = \left|\overline{\psi}_{j-1}^{\text{eff}} \hat{h}_{j-1}^{L,\text{eff}} \psi_{j-1}^{\text{eff}} \cdot \overline{\phi}_{j-1}^{\text{eff}} \hat{h}_{j-1}^{L,\text{eff}} \psi_{j-1}^{\text{eff}}\right| \tag{152}$$

$$\sum_{\substack{m_j \\ \overline{m}_j}} \langle d_j^R \otimes \bar{d}_j^{R,\phi}|d_j^R \otimes \bar{d}_j^{R,\phi}\rangle \leq \sum_{\substack{m_j \\ \overline{m}_j}} \left||d_j^R \otimes \bar{d}_j^{R,\phi}\rangle\right|^2 = \left|\overline{\phi}_j^{\text{eff}} \left[\hat{h}_j^{R,\text{eff}}\right]^2 \psi_j^{\text{eff}}\right| \tag{153}$$

$$\sum_{\substack{m_{j-2} \\ \overline{m}_{j-2}}} \langle d_{j-1}^L \otimes \bar{d}_{j-1}^{L,\phi}|d_{j-1}^L \otimes \bar{d}_{j-1}^{L,\phi}\rangle \leq \sum_{\substack{m_{j-2} \\ \overline{m}_{j-2}}} \left||d_{j-1}^L \otimes \bar{d}_{j-1}^{L,\phi}\rangle\right|^2 = \left|\overline{\phi}_{j-1}^{\text{eff}} \left[\hat{h}_{j-1}^{L,\text{eff}}\right]^2 \psi_{j-1}^{\text{eff}}\right| \tag{154}$$

$$\sum_{\substack{m_{j-1} \\ \overline{m}_{j-1}}} \langle e_j^R \otimes e_j^{R,\phi}|e_j^R \otimes e_j^{R,\phi}\rangle \leq \sum_{\substack{m_{j-1} \\ \overline{m}_{j-1}}} \left||e_j^R \otimes e_j^{R,\phi}\rangle\right|^2 = \left|\overline{\psi}_j^{\text{eff}} \hat{h}_j^{R,\text{eff}} \psi_j^{\text{eff}} \cdot \overline{\phi}_j^{\text{eff}} \hat{h}_j^{R,\text{eff}} \psi_j^{\text{eff}}\right| \tag{155}$$

where we defined

$$\hat{h}_j^{L,\text{eff}} = L_{j-1} W_j \quad \text{and} \quad \left[\hat{h}_j^{L,\text{eff}}\right]^2 = \left(L_{j-1} \otimes L_{j-1}\right)\left(W_j \otimes W_j\right) \tag{156}$$

and in a similar way $\hat{h}_j^{R,\text{eff}}$ replacing $L_{j-1} \to R_{j+1}$. The formal absolute values "$|\cdot|$" on the right side can be estimated by replacing the L_j/R_j tensors with fractions of the overall energy expectation value. This is a valid approximation as long as there are no interactions connecting the left/right contracted MPS–MPO–MPS tensor networks with sites to the right/left of them which was exactly the condition for non-vanishing contributions to the commutator. We hence compare only squares of single-site expectation values with either one or two MPO site tensors sandwiched between the effective site tensors at sites $j-1, j$. For a large system with smoothly varying site tensors the differences at neighboring sites are negligible so that the commutator can be estimated to

$$\langle \phi|\left[\hat{H}^{\hat{n}_{j-1}^{|\psi\rangle}}, \hat{H}^{\hat{n}_j^{|\psi\rangle}}\right]|\psi^{\hat{n}_j^{|\psi\rangle}}(t)\rangle \leq \frac{1}{2}\left(\left||e_{j-1}^L \otimes e_{j-1}^{L,\phi}\rangle + |d_j^R \otimes \bar{d}_j^{R,\phi}\rangle\right|^2\right.$$

$$+ \left.\left||d_{j-1}^L \otimes \bar{d}_{j-1}^{L,\phi}\rangle - |e_j^R \otimes e_j^{R,\phi}\rangle\right|^2\right)$$

$$\leq \frac{1}{2}\left(\left||e_{j-1}^L \otimes e_{j-1}^{L,\phi}\rangle\right|^2 + \left||d_j^R \otimes \bar{d}_j^{R,\phi}\rangle\right|^2\right). \tag{157}$$

The last expression can be estimated easily by the coupling strength of the nearest-neighbor interaction term which we denote by Γ. The contributions from the boundary tensors L_j/R_j are bounded by the square of the system size and cancel with the prefactors $1/L$ of each projected Hamiltonian. We hence conclude this analysis by the somewhat straightforward statement that in general the error at second order in δ scales with $L \cdot \Gamma^2$. However, there is one special situation

which simplifies the preceding arguments drastically namely if we set $|\phi\rangle = |\psi^{\hat{\Pi}_j^{|\psi\rangle}}(t)\rangle$. Then the commutator compares only local overlaps between site tensors sandwiched between either one or two MPO tensors on neighboring sites which scales as $1/L$. Therefore equal-time observables are evolved with very high precision $\sim 1/L^2$ and the contribution due to nearest-neighbor interactions is strongly suppressed.

6.1.2. Algorithm

Algorithm 3 Common helper functions for the local Krylov method and the TDVP algorithms.

1: **procedure** CONTRACT-LEFT(L_{j-1}, W_j, A_j)   ▷ left-contraction as in DMRG

2:    **return** $L_{j;m_j}^{\bar{m}_j,w_j} = \sum_{\sigma_j,\sigma_j',w_{j-1},m_{j-1},\bar{m}_{j-1}} L_{j-1;m_{j-1}}^{\bar{m}_{j-1},w_{j-1}} \bar{A}_{j;\bar{m}_{j-1},\bar{m}_j}^{\sigma_j} W_{j;w_{j-1},w_j}^{\sigma_j,\sigma_j'} A_{j;m_{j-1},m_j}^{\sigma_j'}$

3: **end procedure**

4: **procedure** CONTRACT-RIGHT(R_{j+1}, W_j, B_j)   ▷ right contraction as in DMRG

5:    **return** $R_{j;m_{j-1}}^{\bar{m}_{j-1},w_{j-1}} = \sum_{\sigma_j,\sigma_j',w_j,m_j,\bar{m}_j} \bar{B}_{\bar{m}_{j-1},\bar{m}_j}^{\sigma_j} W_{j;w_{j-1},w_j}^{\sigma_j,\sigma_j'} B_{j;m_{j-1},m_j}^{\sigma_j'} R_{j+1;m_j}^{\bar{m}_j,w_j}$

6: **end procedure**

7: **procedure** INITIALIZE(MPO $\{W_j\}_{j=1}^L$, MPS $\{M_j\}_{j=1}^L$)

8:    $L_{0;m_0}^{\bar{m}_0,w_0} \leftarrow 1$ and $R_{L+1;m_L}^{\bar{m}_L,w_L} \leftarrow 1$

9:    Right-normalize $\{M_j\}_{j=1}^L \rightarrow \{B_j\}_{j=1}^L$ from right to left

10:    **for** $j \in [L, 2]$ **do**

11:        $R_j \leftarrow$ CONTRACT-RIGHT(R_{j+1}, W_j, B_j)

12:    **end for**

13:    **return** L_0, $\{R_j\}_{j=2}^{L+1}$, $\{B_j\}_{j=1}^L$

14: **end procedure**

15: **procedure** TIMESTEP(δ, L_0, $\{R_j\}_{j=2}^{L+1}$, $\{W_j\}$, $\{M_1, B_j\}_{j=2}^L$)

16:    $\{L_j\}_{j=0}^{L-1}$, $\{A_j, M_L\}_{j=1}^{L-1} \leftarrow$ SWEEP-RIGHT($\delta/2$, L_0, $\{R_j\}_{j=2}^L$, $\{W_j\}_{j=1}^L$, $\{M_1, B_j\}_{j=2}^L$)

17:    $\{R_j\}_{j=2}^{L+1}$, $\{M_1, B_j\}_{j=2}^L \leftarrow$ SWEEP-LEFT($\delta/2$, R_{L+1}, $\{L_j\}_{j=0}^{L-1}$, $\{W_j\}_{j=1}^L$, $\{A_j, M_L\}_{j=1}^{L-1}$)

18: **end procedure**

The two-site local Krylov method is described in detail in Alg. 4. It relies on some basic initialization functions familiar from the DMRG algorithm which are summarized in Alg. 3.

6.2. The time-dependent variational principle (TDVP)

There is an alternative to the Lie–Trotter decomposition introduced in Section 6.1 which also results in a series of local problems: the time-dependent variational principle [36,37]. The motivation of this approach is quite different: its primary aim is to constrain the time evolution to a specific manifold of matrix-product states of a given initial bond dimension. To do so, it projects the action of the Hamiltonian into the tangent space to this manifold and then solves the TDSE solely within the manifold. While ideally used in its single-site variant, the two-site variant allows for flexibility in the bond dimension.

6.2.1. Derivation
The main difference between the TDVP and the local Krylov method is in the derivation of the series of these local time-dependent Schrödinger equations and the recovery of the original-time step after each local forward update: Instead of simply projecting the original site tensor onto the new basis as done in the local Krylov approach, the TDVP explicitly solves a backwards-evolution equation. To embark on the derivation for the TDVP, we need to introduce a few additional ingredients: First, we define the *single-site tangent space* $T_{|\psi\rangle}$ of a given MPS $|\psi\rangle$ as the space spanned by variations of single MPS tensors. One may e.g. change the first site tensor of the MPS

Algorithm 4 The local 2-site Krylov method. Also cf. Alg. 3 for definitions of Initialize, Contract-Left, Contract-Right and the needed overall Timestep method.

1: **procedure** Sweep-Right(δ, L_0, $\{R_j\}_{j=3}^{L+1}$, $\{W_j\}_{j=1}^{L}$, $\{M_1, B_j\}_{j=2}^{L}$)
2: $\quad A_1, C_{\underline{1}} \leftarrow M_1$ via QR; $\quad M_2 \leftarrow C_{\underline{1}} \cdot B_2$
3: $\quad P_{\bar{m}_0, m_1}^{\sigma_1} \leftarrow A_{1; m_0, m_1}^{\sigma_1}$ $\qquad\qquad\qquad\qquad$ ▷ entry-for-entry copy, only label of m_0 adapted
4: $\quad$ **for** $j \in [1, L-1]$ **do**
5: $\qquad T_{j,j+1} \leftarrow \sum_{m_j} A_{j; m_{j-1}, m_j}^{\sigma_j} M_{j+1; m_j, m_{j+1}}^{\sigma_{j+1}}$
6: $\qquad T_{j,j+1} \leftarrow \exp(-i\delta \hat{H}_{(j,j+1)}^{\text{eff}}) T_{j,j+1}$ using $\hat{H}_{(j,j+1)}^{\text{eff}} \equiv L_{j-1} \cdot W_j \cdot W_{j+1} \cdot R_{j+2}$
7: $\qquad$ **if** $j \neq L-1$ **then**
8: $\qquad\qquad A_j' \leftarrow T_{j,j+1}$ via SVD and truncation, S and V discarded
9: $\qquad\qquad L_j \leftarrow$ Contract-Left(L_{j-1}, W_j, A_j'); $\quad$ delete R_{j+2}
10: $\qquad\qquad P_{m_j}^{\prime \bar{m}_j} \leftarrow \sum_{\sigma_j \bar{m}_{j-1}} P_{\bar{m}_{j-1}, m_j}^{\sigma_j} \cdot \bar{A}_{j; \bar{m}_{j-1}, \bar{m}_j}^{\prime \sigma_j}$ $\qquad\qquad$ ▷ complete projector
11: $\qquad\qquad M_{j+1; \bar{m}_j, m_{j+1}}^{\sigma_{j+1}} \leftarrow \sum_{m_j} P_{m_j}^{\prime \bar{m}_j} \cdot M_{j+1; m_j, m_{j+1}}^{\sigma_{j+1}}$ $\qquad$ ▷ project M_{j+1} into new left basis
12: $\qquad\qquad A_{j+1}, C_{\underline{j+1}} \leftarrow M_{j+1}$ via QR; $\quad M_{j+2} \leftarrow C_{\underline{j+1}} \cdot B_{j+2}$
13: $\qquad\qquad P_{\bar{m}_j, m_{j+1}}^{\sigma_{j+1}} \leftarrow \sum_{m_j} P_{m_j}^{\prime \bar{m}_j} \cdot A_{j+1; m_j, m_{j+1}}^{\sigma_j}$
14: $\qquad$ **else**
15: $\qquad\qquad A_j', C_{\underline{j}}, B_{j+1} \leftarrow T_{j,j+1}$ via SVD and truncation
16: $\qquad\qquad M_{j+1} \leftarrow C_{\underline{j}} \cdot B_{j+1}$
17: $\qquad$ **end if**
18: $\qquad A_j \leftarrow A_j'$
19: $\quad$ **end for**
20: $\quad$ **return** $\{L_j\}_{j=0}^{L-2}$, $\{A_j, M_L\}_{j=1}^{L-1}$
21: **end procedure**
22: **procedure** Sweep-Left(δ, R_{L+1}, $\{L_j\}_{j=0}^{L-2}$, $\{W_j\}_{j=1}^{L}$, $\{A_j, M_L\}_{j=1}^{L-1}$)
23: $\quad B_L, C_{\underline{L-1}} \leftarrow M_L$ via QR; $\quad M_{L-1} \leftarrow A_{L-1} \cdot C_{\underline{L-1}}$
24: $\quad P_{m_{L-1}, \bar{m}_L}^{\sigma_L} \leftarrow B_{L; m_{L-1}, m_L}^{\sigma_L}$ $\qquad\qquad\qquad\qquad$ ▷ entry-for-entry copy, only label of m_L adapted
25: $\quad$ **for** $j \in [L, 2]$ **do**
26: $\qquad T_{j-1,j} \leftarrow \sum_{m_{j-1}} A_{j-1; m_{j-2}, m_{j-1}}^{\sigma_{j-1}} M_{j; m_{j-1}, m_j}^{\sigma_j}$
27: $\qquad T_{j-1,j} \leftarrow \exp(-i\delta \hat{H}_{(j-1,j)}^{\text{eff}}) T_{j-1,j}$ using $\hat{H}_{(j-1,j)}^{\text{eff}} \equiv L_{j-2} \cdot W_{j-1} \cdot W_j \cdot R_{j+1}$
28: $\qquad$ **if** $j \neq 2$ **then**
29: $\qquad\qquad B_j' \leftarrow T_{j-1,j}$ via SVD and truncation, U and S discarded
30: $\qquad\qquad R_j \leftarrow$ Contract-Right(R_{j+1}, W_j, B_j); $\quad$ delete L_{j-2}
31: $\qquad\qquad P_{m_{j-1}}^{\prime \bar{m}_{j-1}} \leftarrow \sum_{\sigma_j \bar{m}_j} P_{m_{j-1}, \bar{m}_j}^{\sigma_j} \bar{B}_{j; \bar{m}_{j-1}, \bar{m}_j}^{\prime \sigma_j}$ $\qquad\qquad$ ▷ complete projector
32: $\qquad\qquad M_{j-1; m_{j-2}, \bar{m}_{j-1}}^{\sigma_{j-1}} \leftarrow \sum_{m_{j-1}} P_{m_{j-1}}^{\prime \bar{m}_{j-1}} M_{j-1; m_{j-2}, m_{j-1}}^{\sigma_{j-1}}$
33: $\qquad\qquad B_{j-1}, C_{\underline{j-2}} \leftarrow M_{j-1}$ via QR; $\quad M_{j-2} \leftarrow M_{j-2} \cdot C_{\underline{j-2}}$
34: $\qquad\qquad P_{m_{j-2}, \bar{m}_{j-1}}^{\sigma_{j-1}} \leftarrow \sum_{m_{j-1}} P_{m_{j-1}}^{\prime \bar{m}_{j-1}} B_{j-1; m_{j-2}, m_{j-1}}^{\sigma_{j-1}}$
35: $\qquad$ **else**
36: $\qquad\qquad A_{j-1}, C_{\underline{j-1}}, B_j \leftarrow T_{j-1,j}$ via SVD and truncation
37: $\qquad\qquad M_{j-1} \leftarrow A_{j-1} \cdot C_{\underline{j-1}}$
38: $\qquad$ **end if**
39: $\qquad B_j \leftarrow B_j'$
40: $\quad$ **end for**
41: $\quad$ **return** $\{R_j\}_{j=2}^{L+1}$, $\{M_1, B_j\}_{j=2}^{L}$
42: **end procedure**

---

keeping all others fixed to obtain a new state and combine the result with another MPS where only the second site tensor was changed, but one may not change two (or more) site tensors in the same basis state. The projector $\hat{P}_{T_{|\psi\rangle}}$ which projects onto this tangent space is given by [36,37]

$$\hat{P}_{T_{|\psi\rangle}} = \sum_{j=1}^{L} \hat{P}_{j-1}^{L,|\psi\rangle} \otimes \hat{\mathbf{1}}_j \otimes \hat{P}_{j+1}^{R,|\psi\rangle} - \sum_{j=1}^{L-1} \hat{P}_{j}^{L,|\psi\rangle} \otimes \hat{P}_{j+1}^{R,|\psi\rangle} \quad , \tag{158}$$

where $\hat{P}_{j}^{L,|\psi\rangle}$ ($\hat{P}_{j}^{R,|\psi\rangle}$) projects on the sites left (right) of and including site n (cf. Fig. 14) and is exactly the same as the projectors used in the local Krylov method. As before, these projectors use the gauge-fixed left- and right-normalized MPS tensors, i.e. they depend on the MPS $|\psi\rangle$ and can be written as

$$\hat{P}_{j;\bar{\sigma}_1,...,\bar{\sigma}_j,\sigma_1,...,\sigma_j}^{L,|\psi\rangle} = \sum_{m_j} \overline{\psi}_{j;m_j}^{L;\bar{\sigma}_1,...,\bar{\sigma}_j} \otimes \psi_{j;m_j}^{L;\sigma_1,...,\sigma_j} \tag{159}$$

$$\hat{P}_{j;\bar{\sigma}_j,...,\bar{\sigma}_L,\sigma_j,...,\sigma_L}^{R,|\psi\rangle} = \sum_{m_{j-1}} \overline{\psi}_{j;m_{j-1}}^{R;\bar{\sigma}_j,...,\bar{\sigma}_L} \otimes \psi_{j;m_{j-1}}^{R;\sigma_j,...,\sigma_L} \tag{160}$$

where $\psi_j^{L(R)}$ is the collection of left- (right-) normalized MPS tensors on site j and to its left (right) as in the local Krylov approach. The first contributing sum in Eq. (158) filters for all MPS which differ at most on one site from $|\psi\rangle$, whereas the second contributing sum removes all those states which coincide with $|\psi\rangle$. Put differently, individual tangent vectors are constructed by replacing any orthogonality center tensor M_j of the MPS by another tensor N_j which is orthogonal to M_j, i.e., $M_j \cdot \bar{N}_j = 0$. In contrast to the projectors $\hat{\Pi}_j^{|\psi\rangle}$ of the local Krylov method, the total projector $\hat{P}_{T_{|\psi\rangle}}$ projects onto some subspace of the Hilbert space and is only coincidentally written as a sum of local terms.

Second, when inserting the projector $\hat{P}_{T_{|\psi\rangle}}$ into the TDSE, we obtain

$$\frac{\partial}{\partial t}|\psi\rangle = -i\hat{P}_{T_{|\psi\rangle}}\hat{H}|\psi\rangle \tag{161}$$

$$= -i\sum_{j=1}^{L} \hat{P}_{j-1}^{L,|\psi\rangle} \otimes \hat{\mathbf{1}}_j \otimes \hat{P}_{j+1}^{R,|\psi\rangle}\hat{H}|\psi\rangle + i\sum_{j=1}^{L-1} \hat{P}_{j}^{L,|\psi\rangle} \otimes \hat{P}_{j+1}^{R,|\psi\rangle}\hat{H}|\psi\rangle . \tag{162}$$

While an exact solution of Eq. (162) is still not possible, we can approximate it by solving each term individually and sequentially, i.e., solve L forward-evolving equations of the form

$$\frac{\partial}{\partial t}|\psi\rangle = -i\hat{P}_{j-1}^{L,|\psi\rangle} \otimes \hat{\mathbf{1}}_j \otimes \hat{P}_{j+1}^{R,|\psi\rangle}\hat{H}|\psi\rangle \tag{163}$$

and $L-1$ backward-evolving equations of the form

$$\frac{\partial}{\partial t}|\psi\rangle = +i\hat{P}_{j}^{L,|\psi\rangle} \otimes \hat{P}_{j+1}^{R,|\psi\rangle}\hat{H}|\psi\rangle . \tag{164}$$

We then multiply each individual equation above by the single-site map $\overline{\psi}_{j-1}^{L} \otimes \overline{\psi}_{j+1}^{R}$ or the center-bond map $\overline{\psi}_{j}^{L} \otimes \overline{\psi}_{j+1}^{R}$, respectively. As a result, instead of having to work with the full MPS $|\psi\rangle$, we can work with the effective single-site and effective center matrix tensors and associated local Schrödinger equations directly:

$$\frac{\partial}{\partial t}M_j = -i\hat{H}_j^{\text{eff}}M_j \tag{165}$$

$$\frac{\partial}{\partial t}C_j = +i\hat{H}_j^{\text{eff}}C_j \quad . \tag{166}$$

The tensor contraction on the RHS of Eq. (165) is graphically represented in Fig. 17, while the RHS of Eq. (166) is shown in Fig. 18. Each of these equations can be solved with a local application of the Krylov method much as in DMRG or the local Krylov method of the previous section.
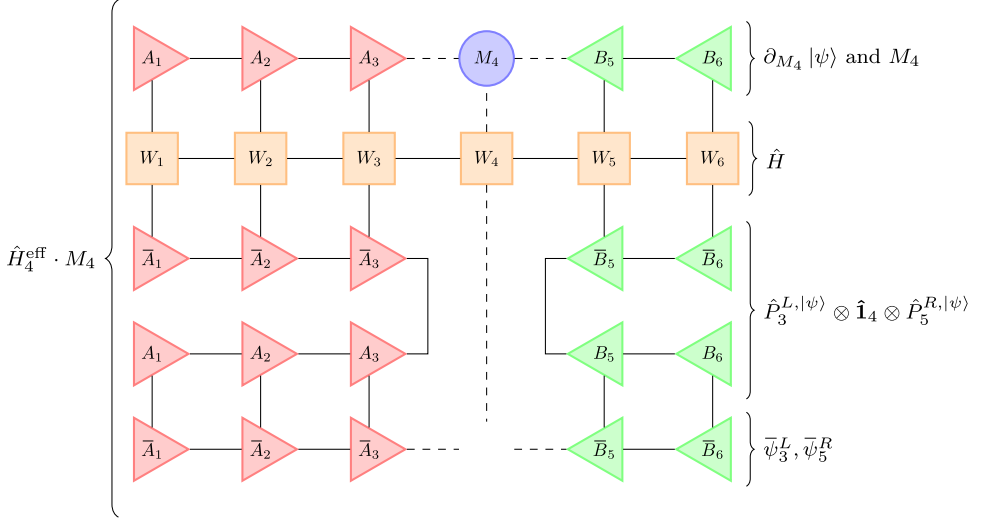
Fig. 17. Right-hand side of the effective single-site forwards-evolving Schrödinger equation (with $j = 4$). The effective Hamiltonian $\hat{H}_j^{\text{eff}}$ is given by the cornered green, orange and red tensors. The effective state is given by the blue circled tensor M_j. During the calculation, the connected dashed lines are contracted, resulting in a new tensor with three legs (the three open dashed lines).
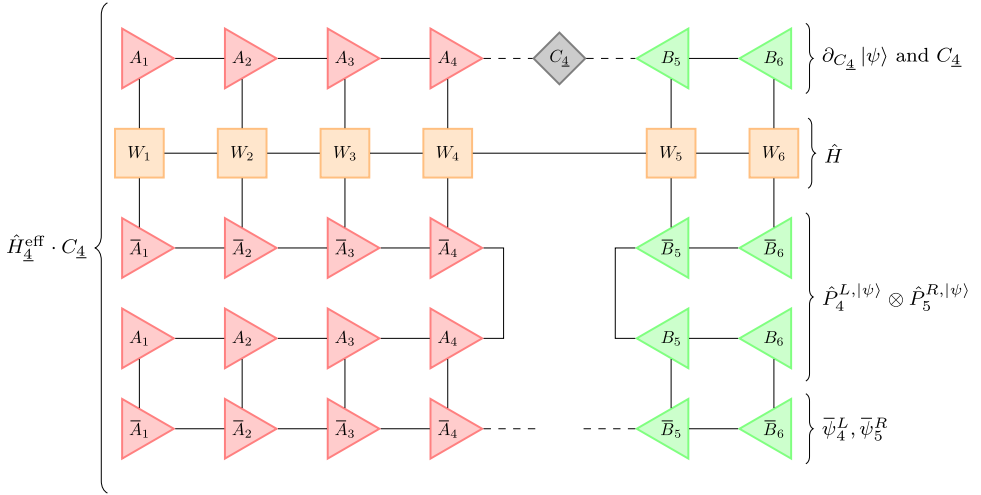
Fig. 18. Right-hand side of the effective center matrix backward-evolving Schrödinger equation with $j = 4$. The effective state over the bond between sites j and $j + 1$ is given by the gray diamond C_j. During the calculation, the connected dashed lines are contracted, resulting in a new tensor with two legs (the two open dashed lines).

Sweeping right-to-left (rather than left-to-right) through the system results in solving the equations in reverse order. This turns the initial first-order integrator into a second-order integrator, reducing the time step error (as described in Section 6.2.2) from $O(\delta)$ to $O(\delta^2)$ if both sweeps are done with halved time steps $\delta/2$.

An interesting property of the single-site TDVP variant (1TDVP) is that the projection of the Hamiltonian onto the MPS manifold occurs before the time evolution, the projection is the *only*

step necessary to obtain the Lie–Trotter decomposition of the Hamiltonian, and no truncation has to happen after the evolution. As such, both the norm and energy of the state are conserved under real-time evolution. This is in contrast to the local Krylov method, where the basis transformation generated by the Q tensors (as described at the end of Section 6.1) is *not* part of the projector and hence introduces an additional error. Alternatively, it is straightforward to extend the mechanism to a two-site variant. This 2TDVP forward-evolves a local tensor $M_{(j,j+1)}$ which needs to be split into two separate site tensors again following the evolution. The advantage is that the bond dimension of the state can be adapted on the fly. However, norm and energy are now no longer conserved exactly if a truncation of the evolved bond is necessary.

6.2.2. Errors

The TDVP has four sources of errors: firstly, there is a *projection error* due to the projection of the full time-dependent Schrödinger equation (TDSE) onto the MPS manifold of limited bond dimension. This error is particularly severe if the MPS in question has a small bond dimension, but it is exactly zero if the MPS has maximal (exponentially growing) bond dimension. However, the projection error occurs during the projection of the TDSE onto the relevant subspace, i.e., *before* the time evolution. As such, it cannot lead to a violation of energy conservation or change the norm of the time-evolved state (during real-time evolution). Using a two- or multi-site variance [86] it is possible to estimate this projection error. If the n-site variance of the state is large, the $(n-1)$TDVP will provide inadequate results. Vice versa, if the up-to-n-site variance of a state is small, the nTDVP will consider this state an eigenstate of the Hamiltonian and the time evolution will only add a global phase to the state. As a corollary, the 2TDVP can evolve Hamiltonians with only nearest-neighbor interactions without incurring a projection error.

Second, the chain of forwards and backwards evolutions can be considered a sequential solution of a series of coupled TDSE (which are the result of the projection above), each describing the evolution of any particular site tensor. Except in the special case that all these evolutions describe exactly the same dynamics (due to the state having maximal bond dimension), there is a finite time-step error of order $O(\delta^3)$ per time step and order $O(\delta^2)$ per unit time. In practice, the prefactor of this error is often much smaller than, e.g., in a TEBD calculation, in particular if the bond dimension of the input state is reasonably large. If the bond dimension is very small, the time-step error will be relatively large.

Third, the 2TDVP contains a SVD to split the evolved two-site tensor into two separate tensors again. During this SVD, a truncation is typically unavoidable, leading to a measurable truncation error. Careful analysis of this truncation error is necessary as always, but also proceeds in much the same way as always. In 1TDVP, this error is exactly zero.

The fourth source of error lies in the inexact solution of the local equations. Using sufficiently many Krylov vectors locally, it is very easy to make this error small. Therefore, one should always use sufficiently many vectors such that the obtained error is at least smaller than the truncation error in the previous step.

Note that changing the time-step size δ in the TDVP affects the four errors differently: the projection and truncation errors affect each time step relatively independently of the size of that step. Hence, increasing the number of time steps during a fixed total time evolution increases the projection and truncation errors. The finite time-step error and the error from the inexact local solution, on the other hand, decrease when increasing the number of time steps and the total time is kept fixed. As such, choosing a smaller δ decreases the time-step error but increases the projection and truncation errors. It is hence typically useful to take some care when choosing, e.g., the truncation threshold and the time-step size such as to approximately balance the induced errors.

Additionally, the energy and norm of the state are conserved exactly within the 1TDVP and only affected by the truncation error in the 2TDVP. This exact conservation may extend to those quantities which are contained within the Hamiltonian [84,85]. While such energy conservation is certainly very helpful to obtain long-time hydrodynamic observables such as diffusion constants, care has to be taken when using only 1TDVP during the calculation as shown in Ref. [87]. Specifically, one has to take great care to ensure that the obtained data is completely converged in the bond dimension of the state at all times.

Algorithm 5 The 1TDVP method. For a more detailed description of the Lanczos solver used in lines 3, 7, 16 and 20 see Section 5. Also cf. Alg. 3 for definitions of INITIALIZE, CONTRACT-LEFT, CONTRACT-RIGHT and the needed overall TIMESTEP method. To run, one first initializes the worker object using INITIALIZE and then does incremental time steps using the TIMESTEP function.

1: **procedure** SWEEP-RIGHT(δ, L_0, $\{R_j\}_{j=2}^{L+1}$, $\{W_j\}_{j=1}^{L}$, $\{M_1, B_j\}_{j=2}^{L}$)
2:      **for** $j \in [1, L]$ **do**
3:          $M_j \leftarrow \exp(-\mathrm{i}\delta/2\hat{H}_j^{\mathrm{eff}})M_j$ using $\hat{H}_j^{\mathrm{eff}} \equiv L_{j-1} \cdot W_j \cdot R_{j+1}$ ▷ e.g. with Lanczos
4:          $A_j, C_{\underline{j}} \leftarrow M_j$ via QR
5:          $L_j \leftarrow$ CONTRACT-LEFT(L_{j-1}, W_j, A_j)
6:          **if** $j \neq L$ **then**
7:              $C_{\underline{j}} \leftarrow \exp(\mathrm{i}\delta/2\hat{H}_{\underline{j}}^{\mathrm{eff}})C_{\underline{j}}$ using $\hat{H}_{\underline{j}}^{\mathrm{eff}} \equiv L_j \cdot R_{j+1}$ ▷ e.g. with Lanczos
8:              $M_{j+1} \leftarrow C_{\underline{j}} \cdot B_{j+1}$
9:              Delete R_{i+1}
10:         **end if**
11:     **end for**
12:     **return** $\{L_j\}_{j=0}^{L-1}$, $\{A_j, M_L\}_{j=1}^{L-1}$
13: **end procedure**
14: **procedure** SWEEP-LEFT(δ, R_{L+1}, $\{L_j\}_{j=0}^{L-1}$, $\{W_j\}_{j=1}^{L}$, $\{A_j, M_L\}_{j=1}^{L-1}$)
15:     **for** $j \in [L, 1]$ **do**
16:         $M_j \leftarrow \exp(-\mathrm{i}\delta/2\hat{H}_j^{\mathrm{eff}})M_j$ using $\hat{H}_j^{\mathrm{eff}} \equiv L_{j-1} \cdot W_j \cdot R_{j+1}$ ▷ e.g. with Lanczos
17:         $B_j, C_{\underline{j-1}} \leftarrow M_j$ via QR
18:         $R_j \leftarrow$ CONTRACT-RIGHT(R_{j+1}, W_j, B_j)
19:         **if** $i \neq 1$ **then**
20:             $C_{\underline{j-1}} \leftarrow \exp(\mathrm{i}\delta/2\hat{H}_{\underline{j-1}}^{\mathrm{eff}})C_{\underline{j-1}}$ using $\hat{H}_{\underline{j-1}}^{\mathrm{eff}} \equiv L_{j-1} \cdot R_j$ ▷ e.g. with Lanczos
21:             $A_{j-1} \leftarrow A_{j-1} \cdot C_{\underline{j-1}}$
22:             Delete L_{j-1}
23:         **end if**
24:     **end for**
25:     **return** $\{R_j\}_{j=2}^{L+1}$, $\{M_1, B_j\}_{j=2}^{L}$
26: **end procedure**

6.2.3. Algorithm

In practice, the 1/2TDVP method is quite similar to the 1/2DMRG method without subspace expansion or density matrix perturbation and nearly identical to the local Krylov method. Compared to the DMRG method, one of course has to replace the local eigensolver by a local exponentiation. Compared to both the local Krylov and the DMRG methods, we also need an additional backwards evolution step either on each bond (1TDVP) or the second site of a two-site evolution (2TDVP). This replacement of the ad-hoc basis transformation done by the local Krylov method with a properly motivated backwards evolution will result in smaller errors in each step.

The 1TDVP algorithm is described in detail in Alg. 5, the 2TDVP algorithm in Alg. 6.

7. Additional tricks

Having reviewed the most commonly used methods in detail, we will now collect and summarize some generic improvements without any claim to completeness. These are essentially tricks which are relatively independent of the actual time-evolution method and can mostly be implemented in all of the methods we just reviewed. They are not strictly necessary to implement in conjunction with any method and as such will not be benchmarked in detail later, but they are useful to keep in mind in case of particularly hard or challenging problems.

Algorithm 6 The 2TDVP method. Also cf. Alg. 3 for definitions of INITIALIZE, CONTRACT-LEFT, CONTRACT-RIGHT and the needed overall TIMESTEP method.

1: **procedure** SWEEP-RIGHT($\delta, L_0, \{R_j\}_{j=3}^{L+1}, \{W_j\}_{j=1}^{L}, \{M_1, B_j\}_{j=2}^{L}$)
2:      **for** $j \in [1, L-1]$ **do**
3:          $T_{j,j+1} \leftarrow \sum_{m_j} M_{j;m_{j-1},m_j}^{\sigma_i} B_{j+1;m_j,m_{j+1}}^{\sigma_{j+1}}$
4:          $T_{j,j+1} \leftarrow \exp(-i\delta/2\hat{H}_{(j,j+1)}^{\text{eff}})T_{j,j+1}$ using $\hat{H}_{(j,j+1)}^{\text{eff}} \equiv L_{j-1} \cdot W_j \cdot W_{j+1} \cdot R_{j+2}$
5:          $A_j, C_j, B_{j+1} \leftarrow T_{j,j+1}$ via singular value decomposition and truncation
6:          $M_{j+1} \leftarrow C_j \cdot B_{j+1}$
7:          **if** $j \neq L-1$ **then**
8:              $L_j \leftarrow$ CONTRACT-LEFT(L_{j-1}, W_j, A_j)
9:              $M_{j+1} \leftarrow \exp(i\delta/2\hat{H}_{j+1}^{\text{eff}})M_{j+1}$ using $\hat{H}_{j+1}^{\text{eff}} \equiv L_j \cdot W_{j+1} \cdot R_{j+2}$
10:              Delete R_{j+2}
11:          **end if**
12:      **end for**
13:      **return** $\{L_j\}_{j=0}^{L-2}, \{A_j, M_L\}_{j=1}^{L-1}$
14: **end procedure**
15: **procedure** SWEEP-LEFT($\delta, R_{L+1}, \{L_j\}_{j=0}^{L-2}, \{W_j\}_{j=1}^{L}, \{A_j, M_L\}_{j=1}^{L-1}$)
16:      **for** $j \in [L, 2]$ **do**
17:          $T_{j-1,j} \leftarrow \sum_{m_{j-1}} A_{j-1;m_{j-2},m_{j-1}}^{\sigma_{j-1}} M_{j;m_{j-1},m_i}^{\sigma_i}$
18:          $T_{j-1,j} \leftarrow \exp(-i\delta/2\hat{H}_{(j-1,j)}^{\text{eff}})T_{j-1,j}$ using $\hat{H}_{(j-1,j)}^{\text{eff}} \equiv L_{j-2} \cdot W_{j-1} \cdot W_j \cdot R_{j+1}$
19:          $A_{j-1}, C_{j-1}, B_j \leftarrow T_{j-1,j}$ via singular value decomposition and truncation
20:          $M_{j-1} \leftarrow A_{j-1} \cdot C_{j-1}$
21:          **if** $j \neq 2$ **then**
22:              $R_j \leftarrow$ CONTRACT-RIGHT(R_{j+1}, W_j, B_j)
23:              $M_{j-1} \leftarrow \exp(i\delta/2\hat{H}_{j-1}^{\text{eff}})M_{j-1}$ using $\hat{H}_{j-1}^{\text{eff}} \equiv L_{j-2} \cdot W_{j-1} \cdot R_j$
24:              Delete L_{j-2}
25:          **end if**
26:      **end for**
27:      **return** $\{R_j\}_{j=2}^{L+1}, \{M_1, B_j\}_{j=2}^{L}$
28: **end procedure**

7.1. Combining Heisenberg and Schrödinger picture time evolution

In the context of MPS methods, combining Schrödinger and Heisenberg picture [88,89] time evolution was first proposed in Ref. [90]. Considering a time-dependent observable

$$\langle \phi | \hat{O}(t) | \psi \rangle \tag{167}$$

between two arbitrary states, in the Schrödinger picture we would evaluate

$$\langle \phi | \hat{O}(t) | \psi \rangle = \left(\langle \phi | e^{it\hat{H}} \right) \hat{O} \left(e^{-it\hat{H}} | \psi \rangle \right) . \tag{168}$$

That is, we apply time-evolution operators to the states $|\phi\rangle$ and $|\psi\rangle$ to obtain time-evolved states $|\phi(t)\rangle$ and $|\psi(t)\rangle$ and then evaluate the time-independent observable between them. The maximum time t obtainable is then typically limited by the entanglement growth in the states, resulting in larger and larger bond dimensions or errors.

In comparison, the Heisenberg picture would see us time evolve the operator $\hat{O}$ as

$$\langle \phi | \hat{O}(t) | \psi \rangle = \langle \phi | \left(e^{it\hat{H}} \hat{O} e^{-it\hat{H}} \right) | \psi \rangle \tag{169}$$

while keeping the states $|\phi\rangle$, $|\psi\rangle$ static. Again, the maximal obtainable time t is limited by the entanglement growth in the operator $\hat{O}$ and the maximal bond dimension we can use to represent

it.[11] If we now combine the two evolutions as

$$\langle\phi|\hat{O}(t_1+t_2)|\psi\rangle = \left(\langle\phi|e^{it_1\hat{H}}\right)\left(e^{it_2\hat{H}}\hat{O}e^{-it_2\hat{H}}\right)\left(e^{-it_1\hat{H}}|\psi\rangle\right) \tag{170}$$

we can obtain times t_1+t_2 while only requiring MPO and MPS bond dimensions typical of times t_1 and t_2 respectively. Note that in this case, the computationally limiting operation is no longer the time evolution itself but the evaluation of observables given as the tensor network of a large-w MPO between two large-m MPS [90].

7.2. Complex time steps

Let us assume that we have a time-evolution operator $\hat{U}'(\delta) = \hat{\mathbf{1}} - i\delta\hat{H}$ which is exact to first order. Applying this operator to a state will result in an error $O(\delta^2)$ compared to the exact evolution with the operator $\hat{U}(\delta) = e^{-i\delta\hat{H}}$. Repeating the process T/δ times to obtain a state at final time T, we incur an error $O(\delta^2)^{T/\delta} = O(\delta)$. However, if we allow complex intermediate steps δ_1 and δ_2, we can solve

$$\hat{U}'(\delta_1)\hat{U}'(\delta_2) = \hat{\mathbf{1}} - i\delta\hat{H} - \delta/2\hat{H}^2 \tag{171}$$

for δ_1 and δ_2 by expanding the left-hand side:

$$\hat{U}'(\delta_1)\hat{U}'(\delta_2) \tag{172}$$

$$= \left(\hat{\mathbf{1}} - i\delta_1\hat{H}\right)\left(\hat{\mathbf{1}} - i\delta_2\hat{H}\right) \tag{173}$$

$$= \hat{\mathbf{1}} - i(\delta_1+\delta_2)\hat{H} - \delta_1\delta_2\hat{H}^2 \tag{174}$$

$$\Rightarrow \delta_1 + \delta_2 = \delta \quad \wedge \quad \delta_1\delta_2 = \delta/2 \; . \tag{175}$$

Two solutions are possible, one of them is

$$\delta_1 = \frac{1-i}{2}\delta \quad \wedge \quad \delta_2 = \frac{1+i}{2}\delta \quad . \tag{176}$$

Choosing these values for $\delta_{1,2}$ then results in a third-order error per time step and a second-order error overall. This choice of time steps is suggested in particular in combination with the MPO $W^{\mathrm{I,II}}$ method to obtain a better error per time step. The cost of the method only grows linearly with the number of evolution operators and, e.g., four operators $\hat{U}'(\delta_{1,2,3,4})$ are required for a third-order error [40].

The drawback is the loss of unitarity at each individual time step which may be disadvantageous. Furthermore, if the time evolution is purely imaginary (e.g., for finite-temperature calculations) and the Hamiltonian does not contain complex coefficients, one may avoid complex arithmetic entirely and only use real floating-point scalars for 50% less memory usage and an approximately four-fold speed-up on matrix multiplications. Unfortunately, it is then impossible to use this trick to reduce the time-step error.

7.3. Green's functions 1: Removal of low-lying states

This trick was first proposed in Ref. [31] and is relatively straightforward to implement. Assume a ground state $|0\rangle$ as an MPS obtained via DMRG and let us shift the Hamiltonian such that this state has energy $E_0 = 0$, $\hat{H}|0\rangle = 0$. We are then interested in the time-dependent observable

$$x(t) = \langle 0|\hat{A}(t)\hat{B}|0\rangle \tag{177}$$

where $\hat{A}$ and $\hat{B}$ are typically local operators such as creators or annihilators. The evolution of $\hat{B}|0\rangle$ is generically non-trivial and if we want to capture all frequency contributions in $x(t)$, we need to

---

[11] Note that it is difficult to compare errors between MPS and MPO truncations, as the error of the MPO truncation is given by the operator norm whereas during an MPO compression, we only control the 2-norm of the operator [67].

evolve until at least times $t' = 1/E_1$ where E_1 is the energy of the lowest eigenstate with non-zero energy $|1\rangle$ contained in $\hat{B}|0\rangle$. In contrast, to capture contributions of higher-energy states $|n\rangle$ with energies $E_n > E_1$, we only need to evolve to shorter times $t'' = 1/E_n < t'$.

However, a few low-lying eigenstates can often be calculated also with DMRG by orthogonalizing against previously-found eigenstates. Hence if we run DMRG multiple times, we can obtain not just the ground state $|0\rangle$ but also further eigenstates $|1\rangle$, $|2\rangle$ etc. If we use quantum numbers and $\hat{B}$ changes the quantum number of the state, these additional eigenstates should be calculated in the quantum number sector of $\hat{B}|0\rangle$. If we then orthogonalize $\hat{B}|0\rangle$ against $|1\rangle$, $|2\rangle$ etc., we remove the contributions which rotate (in real-time) or decay (in imaginary-time evolutions) the slowest and hence require the longest time evolutions. The evolution of the removed states can then be done exactly as we know both their energy and initial weight in $\hat{B}|0\rangle$. Even missing one of the eigenstates due to convergence problems with DMRG does not introduce an error but merely decreases the effectivity of the method.

7.4. Green's functions 2: Linear prediction

Calculating dynamical structure factors or more generally spectral functions from time-dependent data requires two Fourier transformations: first, one needs to transform from real-space to momentum-space and second from real-time to frequency. The former transformation is typically unproblematic, but the latter transformation suffers either from overdamping or strong spectral leakage if the available maximal time t is insufficient. Linear prediction [91–93] assumes that the real-time momentum-space Green's function $G(k, t)$ is composed of multiple distinct exponentially decaying and oscillating contributions arising from a distinct pole structure of $G(k, \omega)$. If this is true for a time series $x_1, x_2, x_3, \ldots$, an additional data point $\tilde{x}_n$ can be approximated well by the form

$$\tilde{x}_n = -\sum_{i=1}^{p} a_i x_{n-i-1} \tag{178}$$

with suitably chosen coefficients a_i independent of n. We hence first compute a finite time series which is still as long as we can manage with time-dependent MPS methods. Subsequently, we need to find coefficients a_i such that the above holds for the data we computed exactly. Using those a_i, we can then extend the time series to arbitrarily long times to generate a sufficiently long time series that a subsequent Fourier transform only requires minimal damping and hence provides for clear features.

It is useful to divide the available calculated data into three segments: first, one should discard an interval $[x_0, \ldots, x_{D-1}]$ at the beginning which captures short-time physics irrelevant and untypical of the longer-time behavior. Second, a fitting interval $[x_D, \ldots, x_{D+N-1}]$ of N points is selected over which the coefficients a_i are minimized. Third, trust in the prediction is increased if it coincides with additional calculated data $[x_{D+N}, \ldots, x_{\max}]$ outside the fitting interval.

To select the a_i, we want to minimize the error

$$\epsilon = \sum_{k=D}^{D+N-1} |\tilde{x}_k - x_k|^2 \quad . \tag{179}$$

Note that to evaluate $\tilde{x}_k$, D must be larger than the number of coefficients p. The coefficient vector $\underline{a}$ is obtained as

$$\underline{a} = -\underline{R}^{-1}\underline{r} \tag{180}$$

where the matrix $\underline{R}$ and vector $\underline{r}$ have entries

$$R_{i,j} = \sum_{k=D}^{D+N-1} x_{k-i}^{\star} x_{k-j} \tag{181}$$

$$r_i = \sum_{k=D}^{D+N-1} x_{k-i}^{\star} x_k \tag{182}$$

respectively. Once the a_i are obtained, data ideally can be generated initially for the interval $[x_{D+N}, \ldots, x_{max}]$ and, once verified to coincide with the calculated data, extended to arbitrary times.

Several numerical pitfalls need to be considered here: First, the matrix $\underline{R}$ may be singular. Two possible remedies include addition of a small shift ε or reduction of the number of parameters p. Ideally the latter should be considered, but may lead to problems finding the optimal non-singular p. Second, if we construct the vector

$$\underline{x}_n = [x_{n-1}, \ldots, x_{n-p}]^T \tag{183}$$

we can move it forward one step as

$$\tilde{\underline{x}}_{n+1} = \underline{A}\,\underline{x}_n \tag{184}$$

where the matrix $\underline{A}$ is of the form

$$\underline{A} = \begin{pmatrix} -a_1 & -a_2 & -a_3 & \cdots & -a_p \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \tag{185}$$

and its eigenvalues α_i contain the frequencies and dampings of the aforementioned oscillations and exponential decays. As such, $\alpha_i > 1$ are unphysical and need to be dealt with, it appears [93] that setting those contributions to zero works best.

7.5. Purification insertion point (PIP)

Calculating out-of-time-ordered correlators (OTOC) allows us to measure the scrambling of quantum information and finds many interesting and current applications. In general an OTOC of operators $\hat{W}, \hat{V}$ is given as an ensemble average

$$\begin{aligned} C_\beta^{\hat{V},\hat{W}}(t) &= \frac{1}{2}\,\mathrm{Tr}\left\{\hat{\rho}(\beta)\left[\hat{V}(t),\hat{W}\right]^\dagger\left[\hat{V}(t),\hat{W}\right]\right\} \\ &= \underbrace{\mathrm{Re}\left[\mathrm{Tr}\left\{\hat{\rho}(\beta)\hat{V}^\dagger(t)\hat{W}^\dagger\hat{V}(t)\hat{W}\right\}\right]}_{\equiv F_\beta^{\hat{V},\hat{W}}(t)} + \text{time ordered} \end{aligned} \tag{186}$$

wherein we have suppressed the time-ordered terms and define the OTOC as the out-of-time ordered part $F_\beta^{\hat{V},\hat{W}}(t)$. At finite temperature, we have to use a purification to evaluate this quantity. If we would calculate the time evolutions in $F_\beta^{\hat{V},\hat{W}}(t)$ naively by direct evolution only in the physical degrees of freedom we would require $\mathcal{O}(N^2)$ time steps to obtain the OTOC at time $t = N\delta$. Clearly, the growing numerical expenses forbid to reach both large system sizes and long time scales t. The graphical notion (see Fig. 19) immediately suggests to transform the operators in the OTOC in some way as to evenly distribute the required time evolutions leading to only linear scaling of effort in time t. In the following, we will explain how to transform these operators in the purification picture and alter the purification insertion point (PIP). For related work in the framework of matrix-product operators, cf. Ref. [94].

Consider the ensemble average $F_{\hat{X},\hat{Y},\hat{Z},\beta} \equiv \mathrm{Tr}\left\{\hat{\rho}(\beta)\hat{Z}\hat{Y}\hat{X}\right\}$ for some global operators $\hat{X}, \hat{Y}, \hat{Z}$ at inverse temperature β. Using the cyclic property of the trace the ensemble average can now be written as expectation value in the enlarged Hilbert space

$$F_{\hat{X},\hat{Y},\hat{Z},\beta} = \mathrm{Tr}\left\{\hat{\rho}(\beta)\hat{Z}\hat{Y}\hat{X}\right\} = \langle 0|\hat{\rho}(\beta/2)\hat{Z}\hat{Y}\hat{X}\hat{\rho}(\beta/2)|0\rangle \equiv \langle \beta/2|\hat{Z}\hat{Y}\hat{X}|\beta/2\rangle, \tag{187}$$

where we have introduced the purified finite temperature state $|\beta/2\rangle \equiv \hat{\rho}(\beta/2)|0\rangle$ based on the infinite temperature state $|0\rangle$ (cf. Section 2.7). A graphical representation of recasting the
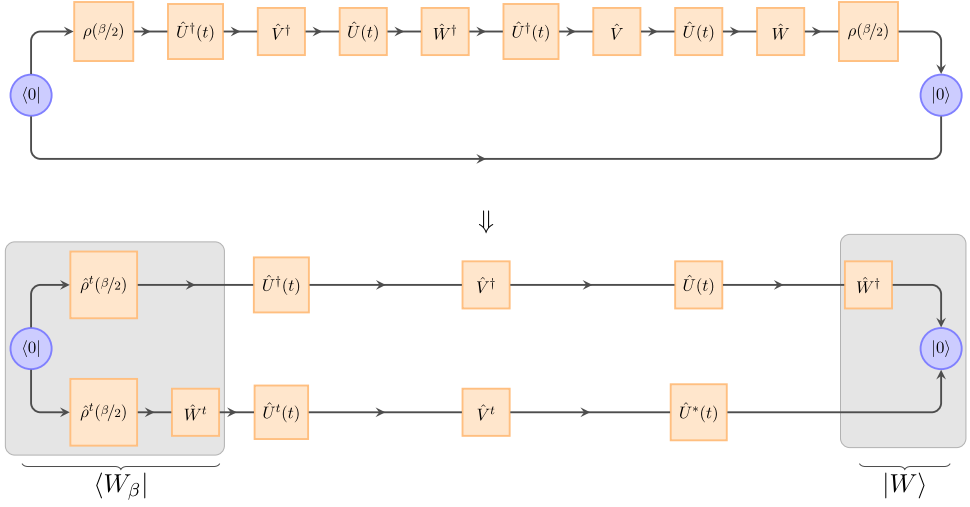
Fig. 19. Shifting the purification insertion point to reduce computational complexity of the time evolution in OTOCs.
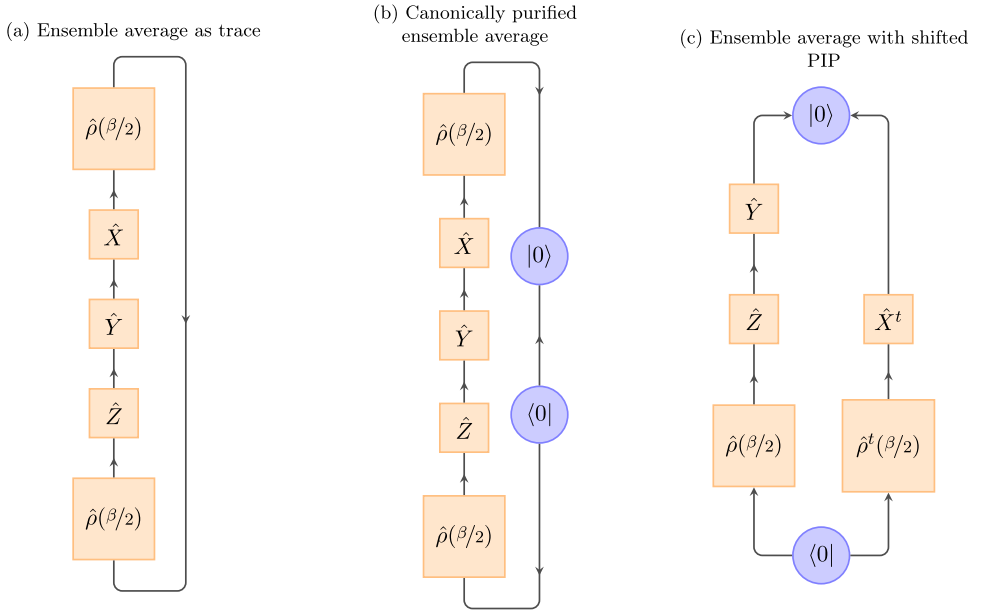
Fig. 20. Different choices of operator purifications for ensemble average $F_{\hat{X},\hat{Y},\hat{Z},\beta} = \mathrm{Tr}\left\{\hat{\rho}(\beta)\hat{Z}\hat{Y}\hat{X}\right\}$.

trace into an expectation value is given by the two networks Figs. 20a and 20b with out-going indices representing row vectors and in-going indices column vectors. From the pictographical representation we motivate the infinite temperature state $|0\rangle$ to be represented by a rank $(2, 0)$ tensor $|0\rangle \equiv \sum_{a,\bar{b}} D^{a,\bar{b}}|a\rangle|\bar{b}\rangle$ and correspondingly $\langle 0|$ by a rank $(0, 2)$ tensor $\langle 0| \equiv \sum_{a,\bar{b}} D_{a,\bar{b}}\langle a|\langle\bar{b}|$, where we have placed a bar over those indices labeling ancilla degrees of freedom.

These tensors have to fulfill the orthogonality conditions

$$\sum_{\bar{b}} D^{a,\bar{b}} D_{c,\bar{b}} = \delta_c^a, \quad \sum_a D^{a,\bar{b}} D_{a,\bar{c}} = \delta_{\bar{c}}^{\bar{b}} \tag{188}$$

so that the tensor elements can be chosen to be $D^{a,\bar{b}} \equiv D^{a,\bar{b}} \delta_{\bar{b}}^{\bar{a}}$ and $D_{a,\bar{b}} \equiv D_{a,\bar{b}} \delta_{\bar{a}}^{\bar{b}}$. When contracted over physical degrees of freedom, the action of these tensors is to convert row vectors into column vectors and vice versa

$$D\hat{O}D^{\dagger} = \sum_{a,c} \sum_{\bar{b},\bar{d}} D^{a,\bar{b}} \hat{O}_a^c D_{c,\bar{d}} |\bar{b}\rangle \langle \bar{d}| = \sum_{\bar{b},\bar{d}} \hat{O}_{\bar{c}}^{\bar{a}} |\bar{a}\rangle \langle \bar{c}| = \hat{O}^t . \tag{189}$$

If we now interpret indices carrying a bar as maps between ancilla degrees of freedom we can reformulate the purification in terms of the D tensors

$$F_{\hat{X},\hat{Y},\hat{Z},\beta} = \sum_{a,c,\dots,g,\bar{b}} D_{a,\bar{b}} \hat{\rho}_c^a(\beta/2) \hat{Z}_d^c \hat{Y}_e^d \hat{X}_f^e \hat{\rho}_g^f(\beta/2) D^{g,\bar{b}} . \tag{190}$$

Inserting identities on the physical Hilbert space between $\hat{\rho}$ and $\hat{X}$ as well as $\hat{X}$ and $\hat{Y}$ and making explicit use of the representation of $\hat{D}$ we obtain

$$F_{\hat{X},\hat{Y},\hat{Z},\beta} = \sum_{\substack{a,c,d,g,\bar{b}, \\ \bar{e},\bar{f},e_l,e_r,f_l,f_r}} D_{a,\bar{b}} \rho_c^a(\beta/2) \hat{Z}_d^c \hat{Y}_{e_l}^d \underbrace{D^{e_l,\bar{e}} D_{e_r,\bar{e}}}_{\delta_{e_r}^{e_l}} \hat{X}_{f_l}^{e_r} \underbrace{D^{f_l,\bar{f}} D_{f_r,\bar{f}}}_{\delta_{f_r}^{f_l}} \rho_g^{f_r}(\beta/2) D^{g,\bar{b}}$$

$$= \sum_{\substack{a,c,d,\bar{b}, \\ \bar{e},\bar{f},e_l}} D_{a,\bar{b}} \rho_c^a(\beta/2) \hat{Z}_d^c \hat{Y}_{e_l}^d D^{e_l,\bar{e}} \underbrace{\hat{X}_{\bar{e}}^{\bar{f}} \rho_{\bar{f}}^{\bar{b}}(\beta/2)}_{\text{act on } \mathcal{H}_A} \tag{191}$$

$$= \langle 0| \left(\hat{\rho}^t(\beta/2) \hat{X}^t \right)_A \otimes \left(\hat{\rho}(\beta/2) \hat{Z} \hat{Y} \right)_P |0\rangle \tag{192}$$

so that now $\sum_{\bar{f}} \hat{X}_{\bar{e}}^{\bar{f}} \rho_{\bar{f}}^{\bar{b}}(\beta/2) \equiv \hat{\rho}^t(\beta/2) \hat{X}^t$ are acting on the ancilla space $\mathcal{H}_A$, i.e., we have shifted the purification insertion point. Again these manipulations can be represented efficiently in a graphical notation and are given in Fig. 20c.

Using this procedure, we can rewrite the OTOC $F_\beta^{\hat{V},\hat{W}}(t)$ as

$$F_\beta^{\hat{V},\hat{W}}(t) = \text{Re} \left[\text{Tr} \left\{ \hat{\rho}(\beta/2) \hat{U}^{\dagger}(t) \hat{V}^{\dagger} \hat{U}(t) \hat{W}^{\dagger} \hat{U}^{\dagger}(t) \hat{V} \hat{U}(t) \hat{W} \hat{\rho}(\beta/2) \right\} \right]$$

$$= \text{Re} \left[\langle 0| \left(\hat{U}^{\dagger}(t) \hat{V} \hat{U}(t) \hat{W} \hat{\rho}(\beta/2) \right)_A^t \otimes \left(\hat{\rho}(\beta/2) \hat{U}^{\dagger}(t) \hat{V}^{\dagger} \hat{U}(t) \hat{W}^{\dagger} \right)_P |0\rangle \right] . \tag{193}$$

Defining the initial states

$$|W\rangle \equiv \hat{W}_P^{\dagger} \otimes \hat{\mathbf{1}}_A |0\rangle , \tag{194}$$

$$|W_\beta\rangle \equiv \hat{\rho}_P(\beta/2) \otimes \left(\hat{W}^* \hat{\rho}_A(\beta/2) \right) |0\rangle , \tag{195}$$

and their purified time evolutions

$$|W(t)\rangle \equiv \hat{U}_P(t) \otimes \hat{U}_A^*(t) |W\rangle , \tag{196}$$

$$|W_\beta(t)\rangle \equiv \hat{U}_P(t) \otimes \hat{U}_A^*(t) |W_\beta\rangle \tag{197}$$

the OTOC can be obtained by calculating the expectation value

$$F_\beta^{\hat{V},\hat{W}}(t) = \text{Re} \left[\langle W_\beta(t)| \hat{V}_P^{\dagger} \otimes \hat{V}_A^t |W(t)\rangle \right] . \tag{198}$$

We hence only need N steps to evaluate all expectation values for times $t = N\delta$.

From a more general point of view shifting the purification insertion point in the OTOCs reformulates the multiple Schrödinger time evolutions of the physical system in the canonical
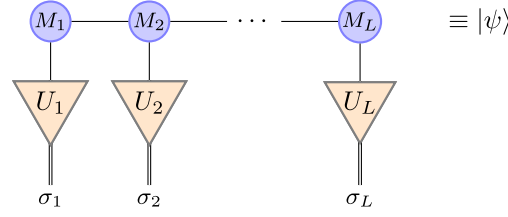
Fig. 21. Local basis optimization matrices U_j are inserted on the physical legs of the MPS to transform a large physical basis (indicated by doubled lines) into a smaller effective basis of the MPS tensors M_j.

choice of the PIP into a Heisenberg time evolution on both the physical and ancilla system of a generalized observable $\hat{V}_P^\dagger \otimes \hat{V}_A^t$.

7.6. Local basis optimization

While the dimension of local Hilbert spaces is typically very limited in spin and electronic systems, bosonic systems potentially require a large local dimension $\sigma = \mathcal{O}(100)$. As this local dimension typically enters at least quadratically in some operations on matrix-product states, some way to dynamically select the most relevant subspace of the local Hilbert space is potentially extremely helpful. The local basis transformation [95–97] method provides for just this: by inserting an additional matrix U_j on each physical leg of the MPS (cf. Fig. 21), the large physical dimension is transformed into a smaller effective basis. The rank-3 MPS tensor then only has to work with the smaller effective basis. The method was adapted for TEBD time evolution in Ref. [97] but it is also straightforward to use in the other time-evolution methods presented here. For the MPO $W^{I,II}$ and global Krylov methods, only the MPO–MPS product has to be adapted to generate additionally a new optimal local basis after each step. The DMRG, TDVP and local Krylov methods translate [44,98,99] directly in much the same way as DMRG.

8. Examples

The following four subsections serve as exemplary applications of the different time-evolution methods to demonstrate, justify and verify the theoretical remarks of the earlier reviews of each method. To some extent it is our hope that the examples here are sufficiently general to be used as benchmarks for *future* time-evolution methods as a way to increase the comparability with previous work. A pairwise comparison of a new method with one of the methods tested here would hence also serve as a comparison with all other methods tested here. To this end, we focus on a clear description of the problem setting and analysis of the runtime behavior of the methods instead of analyzing the physical results in great detail.

In Section 8.1, we will calculate the dynamical structure factor of a XXZ spin chain in a staggered magnetic field at zero temperature. This is a standard application of one-dimensional time-dependent MPS techniques which is straightforward to reproduce in any implementation. Section 8.2 studies the cooling of a Hubbard chain in the canonical ensemble as an application of imaginary-time evolution and a study of the stability of each evolution method against the build-up of errors. Section 8.3 considers the melting of Néel order on a two-dimensional lattice in real-time, a very challenging problem due to very long-range interactions coming into play when projecting the system onto a chain geometry, which is a necessary step for treating higher-dimensional systems with MPS. Finally, in Section 8.4 we simulate the evolution of out-of-time-order correlators on an interacting spin chain at infinite temperature. This problem setting combines the need for a purification ansatz with real-time evolution and the calculation of different-times correlators, which, like the dynamical structure factor, requires the correct treatment of phases during the evolution.

8.1. Dynamical spin structure factor (DSF)

In this section we examine the longitudinal dynamical spin structure factor (DSF) $S^{zz}(q, \omega)$. This quantity can be measured directly in e.g. neutron scattering experiments. Our system of choice here is an anisotropic $S = 1/2$ Heisenberg chain in a staggered magnetic field described by the Hamiltonian

$$\hat{H} = J \sum_{j=1}^{L} \left[\hat{s}_j^x \hat{s}_{j+1}^x + \hat{s}_j^y \hat{s}_{j+1}^y + \Delta \hat{s}_j^z \hat{s}_{j+1}^z - h_j^s \hat{s}_j^z \right]$$

$$= J \sum_{j=1}^{L} \left[\frac{1}{2} \left(\hat{s}_j^+ \hat{s}_{j+1}^- + \hat{s}_j^- \hat{s}_{j+1}^+ \right) + \Delta \hat{s}_j^z \hat{s}_{j+1}^z - h_j^s \hat{s}_j^z \right], \quad h_j^s = (-1)^j h . \tag{199}$$

We set $\Delta \equiv J^{-1}$ and vary $J = 0.1, 0.2, 0.3$ and $J \cdot h = 0.01, 0.05, 0.1$. The calculation is carried out on systems of sizes $L = 100, 150, 200$ with open boundary conditions. $S^{zz}(q, w)$ can then be obtained from the double Fourier transformation of the time-dependent spin–spin correlation function

$$\langle \hat{s}_j^z(t) \hat{s}_{L/2}^z(0) \rangle_{cc} = \langle 0 | \hat{U}^\dagger(t) \hat{s}_j^z \hat{U}(t) \hat{s}_{L/2}^z | 0 \rangle - \langle 0 | \hat{s}_j^z \hat{s}_{L/2}^z | 0 \rangle \tag{200}$$

with $|0\rangle$ the ground state of Eq. (199). We then have

$$S^{zz}(q, \omega) = \frac{1}{L} \sum_{j=1}^{L} e^{-iq(j-L/2)} \int_{-\infty}^{\infty} dt \, e^{i\omega t} \langle \hat{s}_j^z(t) \hat{s}_{L/2}^z(0) \rangle_{cc} \tag{201}$$

$$\simeq \frac{2\pi}{LT} \delta \sum_{j=1}^{L} e^{-iq(j-L/2)} \sum_{n=0}^{N} e^{i(\omega+i\eta)t_n} 2 \operatorname{Re} \langle \hat{s}_j^z(t_n) \hat{s}_{L/2}^z(0) \rangle_{cc} \tag{202}$$

where we discretize the time coordinate $t_n = n\delta$ and truncate the integral at a finite, maximum evolution time $T = N\delta$. It is of course important to chose the system size L sufficiently large that the initial central excitation does not reach the edges of the open chain during the finite simulation time T. Due to this finite simulation time T, it is also necessary to introduce a damping factor $\eta > 0$. Without the damping factor, we would observe large spectral leakage due to the finite interval. With $\eta > 0$, we obtain an artificial broadening of spectral lines instead.

The overall simulation procedure was to calculate the ground state $|0\rangle$ of Eq. (199) with DMRG at a maximum discarded weight of 10^{-14} and energy convergence better than 10^{-12}. The excited state $|1\rangle = \hat{s}_{L/2}^z |0\rangle$ was then evolved in time until the final time $T = 200$ in units of Δ was reached. During the evolution, we need to evaluate the time-dependent correlator

$$S^{zz}(j, t_n) = \langle 0 | \hat{s}_j^z(t_n) \hat{s}_{L/2}^z(0) | 0 \rangle - \langle 0 | \hat{s}_j^z \hat{s}_{L/2}^z | 0 \rangle \tag{203}$$

$$= \langle 0 | \hat{U}^\dagger(t_n) \hat{s}_j^z \hat{U}(t_n) \hat{s}_{L/2}^z | 0 \rangle - \langle 0 | \hat{s}_j^z \hat{s}_{L/2}^z | 0 \rangle \tag{204}$$

$$= e^{iE_0 t_n} \langle 0 | \hat{s}_j^z | 1(t) \rangle - \langle 0 | \hat{s}_j^z \hat{s}_{L/2}^z | 0 \rangle \tag{205}$$

where the term $\langle 0 | \hat{s}_j^z \hat{s}_{L/2}^z | 0 \rangle$ is of course only evaluated once. The last equality only holds for representations of the time stepper $\hat{U}(\delta)$ which act on the ground state by multiplying a phase, i.e. $\hat{U}(\delta) | 0 \rangle = e^{-iE_0 \delta} | 0 \rangle$. This should of course be the case ideally, but in practice is not achieved by all methods. In general, $S^{zz}(q, \omega)$ is very sensitive to the build-up of global phases by the particular time stepper rendering this quantity an important testcase.

8.1.1. Real-space evolution

In Fig. 22, $S^{zz}(j, t_n)$ as obtained from SYTEN 2TDVP and TEBD2 is shown in the top panels. Here we already see the main error source when using the raw data to calculate the spin structure factor: a global phase shift in the correlator which can be identified by a uniformly evolving stripe pattern in the TEBD2 data imposed over the perturbation light cone. When calculating the Fourier transformations this uniformly evolving phase will cause a $k = \pi$ signal which ultimately may
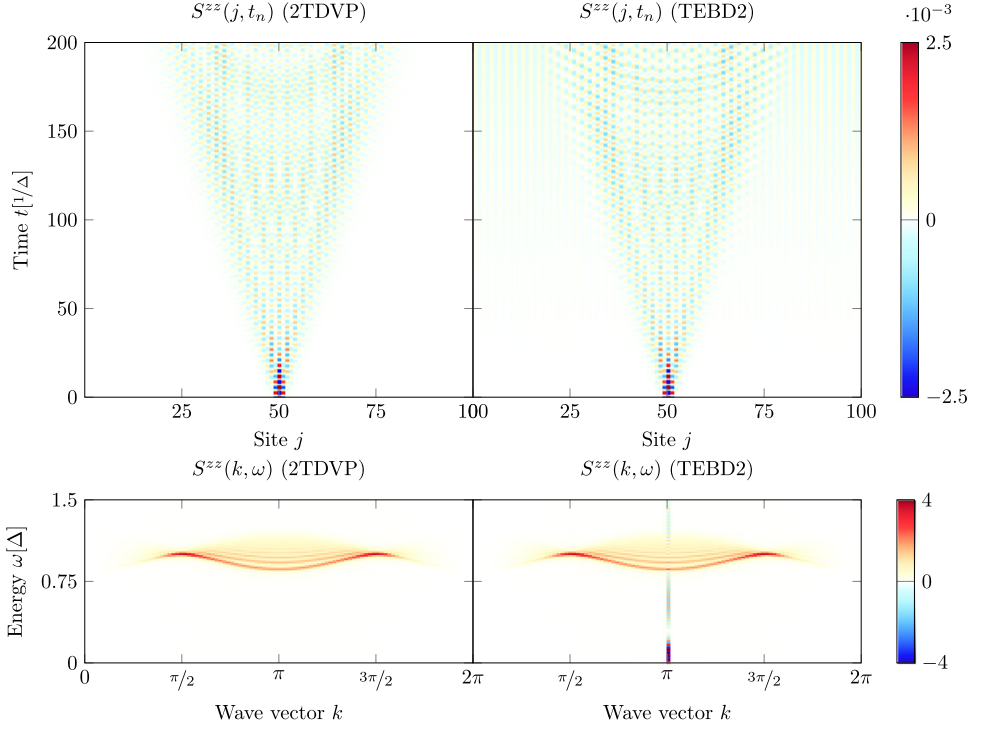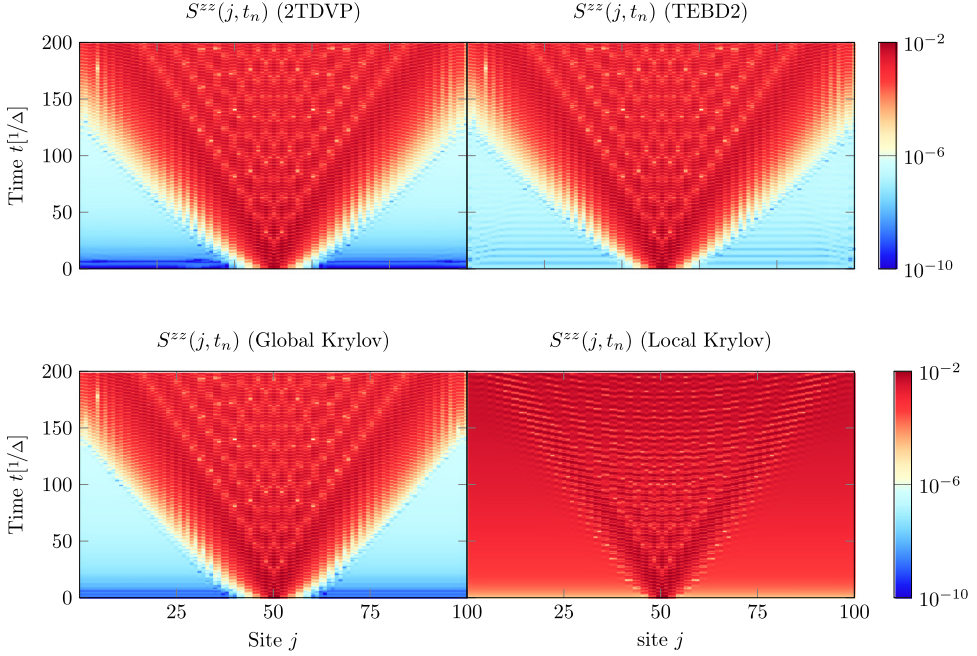
Fig. 22. (DSF) Real-time evolution of the time-dependent correlator $S^{zz}(j, t_n)$ with parameters $J = 0.1$, $J \cdot h = 0.01$ and time step $\delta = 0.1$. Top panels show the evolution of the initial perturbation using SYTEN 2TDVP (left) and SYTEN TEBD2 (right) without damping. Bottom panels show the low-energy part of the dynamical spin structure factor using SYTEN 2TDVP (left) and SYTEN TEBD2 (right) stepper and damping η such that $e^{200\eta} = 0.1$.

overlay the relevant spectral information. In the bottom panels of Fig. 22 we plot the low-energy part of the DSF with damping η applied such that $e^{200\eta} = 0.1$. This plot reveals a dominant $k = \pi$ feature visible in the TEBD2 data which is absent in the 2TDVP data. At least to some extent, this phase shift can be reduced by calculating the explicit time evolution of the ground state $\hat{U}(t_n)|0\rangle \equiv |0(t_n)\rangle$ in the evaluation of $S^{zz}(j, t_n)$.

8.1.2. Spectral functions

More generally, most methods do not evolve the area outside the light cone exactly. That is, while in this region the Hamiltonian should lead to trivial dynamics, small errors are introduced. Given enough time, these errors can accumulate with two major results: (i) The slightly erroneous state is now even less of an eigenstate of the (effective) Hamiltonian and hence acquires dynamics, leading to a runaway effect. (ii) Additional entanglement accumulates and makes the calculation more difficult. The precise way in which this error occurs slightly depends on the method: For the TEBD2 and the MPO W^{II} method, the error is simply a finite time-step error due to the Trotterization or otherwise δ-dependent error terms. The local Krylov method is primarily subject to the error created by the basis transformation after solving each local problem, which is proportional to the step size δ. This error affects every site and, after comparably short times, results in site tensors which have lost the property of being exact eigenstates of the effective Hamiltonian outside the light cone. Altogether, these methods can produce results at time step $\delta = 0.1$, but the results only qualitatively reproduce the spectrum, require a large bond dimension m and also carry a large spectral weight around $\omega = 0$.

Fig. 23. (DSF) Real-time evolution of the time-dependent correlator $S^{zz}(j, t_n)$ with parameters $J = 0.3$, $J \cdot h = 0.1$ and time step $\delta = 0.01$. Panels show the evolution of the initial perturbation using SciPal 2TDVP (top left), SyTen TEBD2 (top right), SyTen global Krylov and SciPal local Krylov without damping.

For the global Krylov method, we had initially selected a Krylov error $< 10^{-6}$, believing that this error should be sufficiently small to provide good results. However, it turns out that an extremely small Krylov error $< 10^{-10}$ is necessary to avoid the runaway effect and the accumulation of errors outside the light cone (observed best in Figs. 25 and 26). For this method, the additional dynamics also make the problem more complicated, hence requiring more Krylov vectors for a precise solution. As these Krylov vectors are typically more strongly entangled, their truncation will be more severe. Conversely, the Krylov error at a fixed number of Krylov vectors will increase, leading to yet larger errors. In effect, the method runs into an entirely artificial exponential wall, built from accumulated errors outside the light cone and stopping the calculation in just a few steps (cf. Fig. 26). Selecting a smaller time step size $\delta = 0.01$ and disabling the extrapolation (cf. Section 5.4.2) – hence forcing a very small Krylov error – leads to a very precise calculation which is also competitive in runtime.

The 2TDVP, in comparison, does evolve the area outside of the light cone exactly − the MPS tensors there are exact eigenstates of the local effective Hamiltonian, as they are the result of a DMRG ground-state search procedure. As such, 2TDVP also produces good results at $\delta = 0.1$ and (like the global Krylov at $\delta = 0.01$) no spectral weight around $\omega = 0$.

Fig. 23 shows the light cones produced by 2TDVP, TEBD2, the global Krylov method and the local Krylov method. 2TDVP and the global Krylov method give small, homogeneous correlations outside the light cone. With sufficiently large frequency resolution, we could therefore find some spectral weight exactly at $\omega = 0$; in practice our resolution is too low and the prefactor too small for this effect to be visible. TEBD2, in contrast, also results in slow dynamics in this region. These dynamics can be seen in some remaining spectral weight close to $\omega = 0$, but the light cone itself is produced very well. The local Krylov method, in contrast, gives immediately large contributions to the correlations outside the light cone. These correlations (i) give large spectral weight around $\omega = 0$ even with $\delta = 0.01$ and (ii) result in a shift of the overall spectral function.
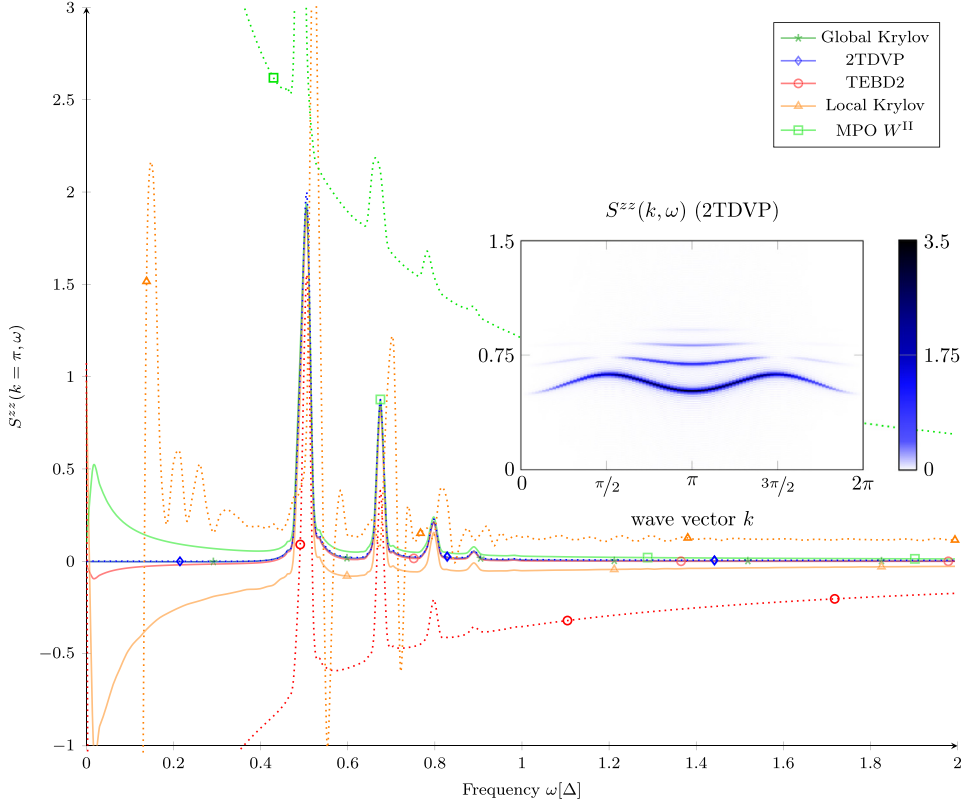
Fig. 24. (DSF) $k = \pi$ cut through the band structure at $J = 0.3$, $J \cdot h = 0.1$ for different methods and step sizes indicated as dotted ($\delta = 0.1$) and solid ($\delta = 0.01$) lines. Only 2TDVP (at $\delta = 0.1$ and $\delta = 0.01$) and the global Krylov method (with $\delta = 0.01$) produce no spectral weight around $\omega = 0$ (their curves are atop of each other for all ω). All other methods produce some spectral weight around $\omega = 0$; the errors of the local Krylov method also produce consistently negative intensities. Inset: Full spectrum as produced by 2TDVP for these parameters. The main graph x-axis corresponds to a vertical line at $k = \pi$ in the inset.

As an example of the effects discussed above, consider $S^{zz}(k = \pi, \omega)$ plotted in Fig. 24. Most importantly, all methods are able to resolve the location of the primary peak at $\omega \approx 0.5$ and also subsequent peaks at $\omega \approx 0.67$, $\omega \approx 0.8$ and $\omega \approx 0.88$. However, only 2TDVP produces the "correct" function at $\delta = 0.1$ and is only joined in this by the global Krylov method at $\delta = 0.01$. All other methods produce additional unphysical features such as an overall shift (local Krylov method at both δ values; TEBD2 and the MPO W^{II} method at $\delta = 0.1$) or spectral weight towards $\omega = 0$ (TEBD2, MPO W^{II} and local Krylov methods at all δ).

8.1.3. Benchmark

Finally, we compare the runtime and growth of the bond dimension in the different methods in Figs. 25 and 26. All calculations where performed single threaded on an Intel Xeon Gold 6150 CPU with 192 GB of RAM and no hard-disk caching. The growth of the bond dimension is slower when the time step used is smaller (especially $\delta = 0.01$), suggesting that at least some part of the entanglement structure is wrong due to a finite time-step error. Still at $\delta = 0.01$, the bond dimension grows slowest during the 2TDVP evolution resulting in a bond dimension at $t = 200$ roughly a factor of two smaller than for the local Krylov and MPO W^{II} methods. The local Krylov method at $\delta = 0.1$ and $\delta = 0.05$ very quickly saturates the maximal bond dimension $m = 200$
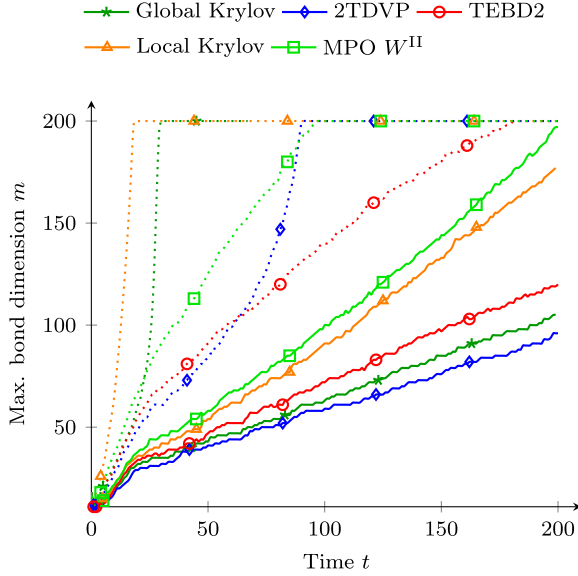
Fig. 25. (DSF) Maximal bond dimensions m during time evolutions of systems with $L = 100$ sites and different time steps $\delta = 0.1$ (dotted) and $\delta = 0.01$ (solid). All methods result in a smaller maximal bond dimension if the calculation is more precise and the step size is smaller.
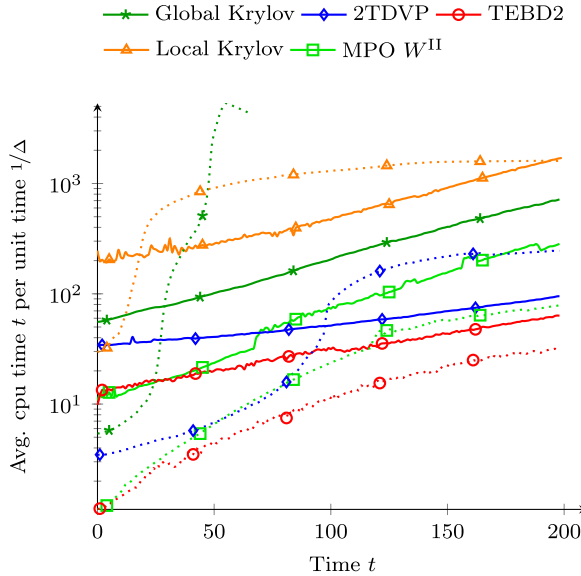
Fig. 26. (DSF) Average CPU time per unit time $\frac{1}{\Delta}$ evolving systems with $L = 100$ sites and different time step sizes $\delta = 0.1$ (dotted) and $\delta = 0.01$ (solid). Note that physically reasonable results at $\delta = 0.1$ are only obtained for 2TDVP.

which is consistent with the picture of the site tensors being pushed away from the ground state. The strong dependence on the step size is related to the error induced by the basis transformation which in turn can be estimated by the overlap between the unevolved and evolved state which is

reduced for smaller time steps. Curiously, the situation is comparable for the global Krylov method even though the underlying reasoning is very different. In this case the strong dependence of the bond dimension on the time step is due to the incurred Krylov error. Once the site tensors are no longer the ground-state solutions of the reduced two site problem, the incurred Krylov error increases even faster as the overall state rapidly evolves away from the global ground state and the small Krylov subspace cannot faithfully represent the action of the operator exponential.

The growth of the bond dimension also translates directly to the computational effort, plotted in Fig. 26. Consistently, we find the local Krylov method to be much slower than the comparable 2TDVP even with the smallest time step. Compared to TEBD2, the third method which produces good data at least in the range $0.4 < \omega < 1$, 2TDVP performance is acceptable though not superb. The global Krylov method also performs comparably well as long as the time step is small enough $\delta = 0.01$ and is completely unsuitable if $\delta = 0.1$.

8.1.4. Conclusion

We calculated the dynamical spin structure factor which can be boiled down to the evaluation of non-equal time correlation functions. In our analysis, we find 2TDVP to provide both the best numerical data and very efficient calculations due to its numerical stability also at larger time steps. In fact, it is the only method which generated a stable time evolution if we set $\delta = 0.1$. The global Krylov method provides equally good numerical results as long as its Krylov errors are sufficiently small, which unfortunately results in a very small time step and no recycling of Krylov subspaces (cf. Section 5.4.2) being possible. Putting the focus on the runtime, we find that TEBD2 can be considered as a method providing a satisfying trade-off between quick MPO–MPS application also at smaller time steps and satisfying numerical results, even though there is a spectral loss at very small energies $\omega \approx 0$ due to the build-up of global phases during the time evolution. Finally, the local Krylov and the MPO W^{II} method are apparently the most susceptible for phase errors with very large errors at $\delta = 0.1$. Even at $\delta = 0.01$, the local Krylov method gives large spectral leakage over the entire frequency range while the MPO W^{II} method gives relatively reasonable data though slower and of worse quality than TEBD2.

An interesting remark is that counterintuitively, higher-precision calculations can be *faster* than lower-precision calculations. That is, as long as the time-evolution method only induces very small errors, the area outside the light-cone stays nearly a (local) eigenstate of the Hamiltonian. If instead large errors are made by the method, these nonphysical errors then need to be evolved also in subsequent steps, which may be much harder than "simply" solving the physical problem. The most obvious example of this behavior is the global Krylov method. In fact, however, *all* methods exhibit larger bond dimensions at larger time step sizes (related to larger Trotterization errors) and a faster growth of bond dimensions, cf. Fig. 25. Without the artificial limit to $m = 200$, we would hence expect an eventual cross-over where the more precise calculation at $\delta = 0.01$ with ten times more steps per unit time is faster than the large-step calculation at $\delta = 0.1$.

8.2. Cooling of a doped Hubbard chain

Let us now consider the cooling of a Hubbard chain from infinite temperature ($\beta = 0$) to a finite temperature $\beta \geq 0$. We will work with a purified state in a canonical ensemble [67] conserving both particle number and spin S^z projection, i.e. we implement the associated $U(1)_N \times U(1)_{S^z}$ symmetries. The local physical dimension is four with each quantum number sector denoted as (N, S^z) only containing one single state (i.e., $(0, 0)$, $(1, -1/2)$, $(1, +1/2)$ and $(2, 0)$). The physical chain has a length of $L = 24$ sites resulting in an MPS with $L' = 48$ tensors. Physical sites will be labeled by indices $1, \ldots, 24$, the auxiliary sites as $a(1), \ldots, a(24)$. The initial state is constructed (cf. Section 2.7) by applying the operator

$$\hat{C}_{\text{tot}}^{\dagger} = \sum_{j=1}^{L} \hat{c}_{j,\uparrow}^{\dagger} \hat{c}_{a(j),\downarrow}^{\dagger} + \hat{c}_{j,\downarrow}^{\dagger} \hat{c}_{a(j),\uparrow}^{\dagger} \tag{206}$$

20 times to a vacuum state, resulting in 5/6 filling. The generated state has a maximal bond dimension $m = 80$ and is used as the initial state for imaginary time evolution under the Hamiltonian

$$\hat{H} = -t \sum_{j=1}^{L-1} \left(\hat{c}_{j,\uparrow}^\dagger \hat{c}_{j+1,\uparrow} + \hat{c}_{j,\downarrow}^\dagger \hat{c}_{j+1,\downarrow} + \hat{c}_{j+1,\uparrow}^\dagger \hat{c}_{j,\uparrow} + \hat{c}_{j+1,\downarrow}^\dagger \hat{c}_{j,\downarrow} \right) + U \sum_{j=1}^{L} \hat{n}_{j,\uparrow} \hat{n}_{j,\downarrow} \tag{207}$$

with $t = 1$ and $U = 4$. To simplify notation in this example, we will time evolve using the evolution operator $e^{-\delta \hat{H}}$ and refer to "times" via the real-valued β variable. That is, we drop the prefactor $-\mathrm{i}$ in δ (which would be necessary to perform an imaginary time evolution using $e^{-\mathrm{i}\delta\hat{H}}$) and we also drop the factor of 1/2 (which relates the target "temperature" β as the result of the evolution to the actual inverse temperature β in the physical system, cf. Section 7.5). In practice, all calculations are done using complex-valued arithmetic and negative imaginary time steps.

In the following, we primarily consider the long-range spin–spin correlator

$$Z_1 = \sum_{j=2}^{L} \langle \hat{s}_1^z \hat{s}_j^z \rangle . \tag{208}$$

The on-site term $\langle \hat{s}_1^z \hat{s}_1^z \rangle$ is about the same order of magnitude as all the long-range terms which is why it is not included here. During the cooling simulation, the value of this correlator evolves monotonously from 0 until it reaches a plateau, ideally at the ground-state value (however, in practice, it may over- or undershoot, see below). In an error-free calculation, the simulated state would then be the ground state of the Hamiltonian and only acquires a global prefactor upon further imaginary time evolution. In practice, due to accumulated errors, particles are able to leave the physical system and move into the auxiliary system of equal size and the state can "tunnel" into the global ground state in the combined physical and auxiliary degrees of freedom, which is not what is looked for. Measuring the particle number in the physical system is, therefore, a measure for the stability of the procedure. This process is monitored by evaluating $\hat{N}_p = \sum_{j=1}^{L} \hat{n}_j$ and comparing it to its true value, which is $\langle \hat{N}_p \rangle = 20$ here.

We run calculations at bond dimensions $m = 100, 200, 300, 400$ and 500 and time step sizes $\delta = 0.01, 0.05$ and 0.1 up to $\beta = 20$. In this problem, the global Krylov method has to be directly disqualified as the highly entangled Krylov vectors generated by it exceed the time and memory resources of the other methods by more than a factor of 20 (RAM) and 100 (CPU time), respectively. The 1TDVP method is also unsuitable here (cf. Section 8.2.3 below). We hence focus on the 2TDVP and MPO W^{II} method (both with SyTen and SciPal) as well as the second-order TEBD and local Krylov methods (with SyTen only). SyTen was configured to apply MPOs (in the MPO W^{II} method and TEBD2) using the zip-up method, whereas the MPO W^{II} method in SciPal is configured to use the zip-up method followed by a few variational optimization sweeps. As always, for the plots we have selected the more suitable of the two implementations.
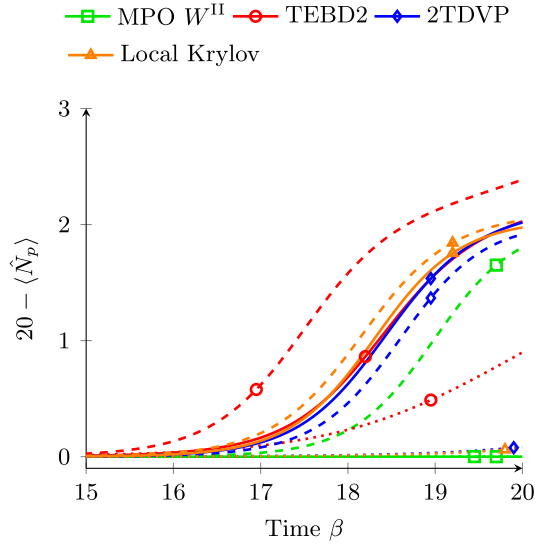
8.2.1. Loss of particles

As we would like to simulate a canonical ensemble, it is quite relevant that loss of particles to the auxiliary system does not occur too early. Studying $\langle \hat{N}_p \rangle$, we find that, first, increasing the step size usually reduces the loss of particles (cf. Fig. 27). Second, there is no clear relation between the selected bond dimension and the onset of loss of particles: E.g. 2TDVP at $\delta = 0.1$ starts losing particles at $m = 400$ around $\beta = 15$ but is stable until $\beta = 17$ for $m = 100, 200, 300$ and 500 (not shown) and while the MPO W^{II} method is stable at $\delta = 0.05$ and $m = 100$ and $m = 500$, it loses particles at $m = 300$ (cf. Fig. 28). As expected, the behavior of the local Krylov method and 2TDVP on this relatively large-scale observable is very comparable.

The configuration difference in the MPO W^{II} routines to apply the operator to the state between SyTen and SciPal leads to different results here: SyTen was configured to only use the zip-up method whereas SciPal also did a few variational sweeps afterwards to optimize the state. This variational optimization leads to considerably better stability and mostly no loss of particles at all, but is computationally more expensive.
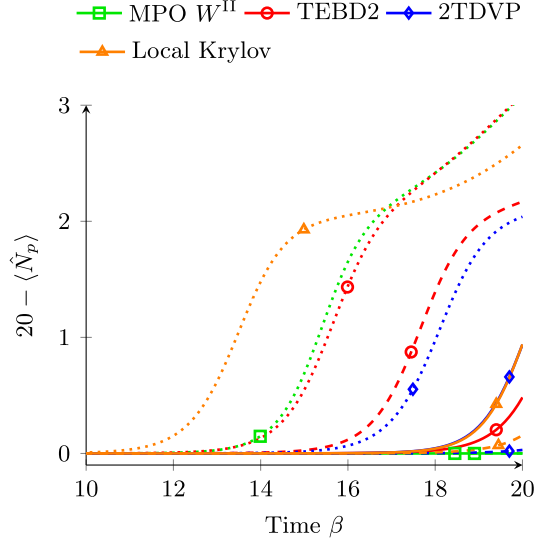
Fig. 27. (Cooling) Loss of particles at $m = 200$ for step sizes $\delta = 0.1$ (solid), $\delta = 0.05$ (dashed) and $\delta = 0.01$ (dotted). At smaller δ, truncation occurs more often and particle loss tends to set on earlier.


Fig. 28. (Cooling) Loss of particles at $\delta = 0.05$ for bond dimensions $m = 100$ (dotted), 300 (dashed) and 500 (solid). No correlation between bond dimension and onset of particle loss can be observed.

8.2.2. Spin–spin correlator

Our proxy for a long-range observable, the sum of correlators Z_1, obtains values comparable to the ground-state value around $\beta = 15$ to $\beta = 18$. This is already deep in the regime where particles tend to be lost to the auxiliary system. Additionally, it is often unnecessary to consider this point when using a finite-temperature method, as we could conceivably use DMRG directly to obtain ground-state properties.
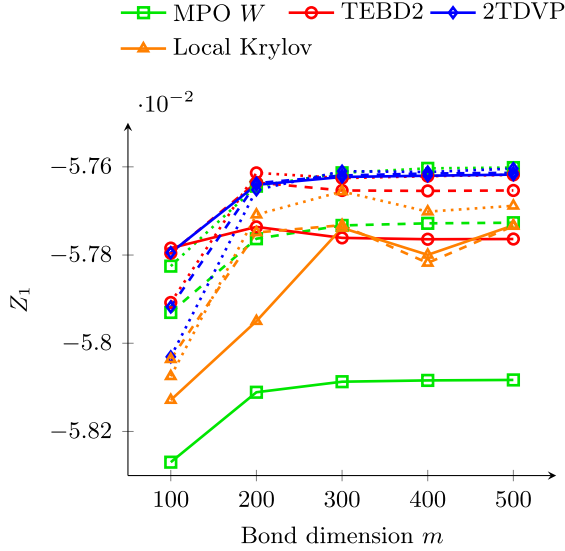
Fig. 29. (Cooling) Convergence of Z_1 with bond dimension at $\beta = 1$ for $\delta = 0.1$ (solid), $\delta = 0.05$ (dashed) and $\delta = 0.01$ (dotted). For $m \geq 300$ and small time steps, three methods converge to the same result, but the time-step error of the W^{II} method and the TEBD2 is very large at $\delta = 0.1$. Even at $\delta = 0.01$, the local Krylov method does not converge to the correct result.

Hence, we will concentrate on a higher temperature at $\beta = 1$ and consider the value of the correlator as a function of m and δ, cf. Fig. 29. The first relevant observation is that 2TDVP, TEBD2 and the MPO W^{II} method converge to the same result once $m \geq 300$ and (for TEBD2 and MPO W^{II}) $\delta = 0.01$. Curiously, the local Krylov method appears to suffer a relatively large time-step error here; only its results with $\delta = 0.01$ are comparable to the other methods. In contrast, 2TDVP provides very good results starting at $m = 200$ which are also nearly independent of the time-step size. Curiously, the remaining dependence of the 2TDVP error on the step size is such that the $\delta = 0.01$ calculation has the highest deviation from the other data points at $m \geq 300$. That is, while the MPO W^{II} and TEBD2 methods suffer from a simple finite time-step error which becomes consistently smaller with smaller time steps, the behavior of 2TDVP is more complicated. It can be understood as the projection error being 2TDVP's primary error source. The projection error acts on each individual time step, hence increasing the number of time steps also increases the projection error. As the initial state is exact at $m = 80$, we cannot increase the MPS manifold to make up for the increased projection error.

Here again we observe that the variational optimization following the zip-up considerably reduces the error in the MPO W^{II} method as used by SciPal compared to the SyTen implementation (not shown). In contrast, 2TDVP provides exactly the same result in both implementations.

8.2.3. Projection error in 1TDVP and 2TDVP

We have excluded 1TDVP from the set of viable time-evolution methods here as the nature of the problem – global cooling and redistribution of particles – does not yield itself to the very restricted manifold available to 1TDVP. Indeed, imaginary time evolution with 1TDVP results in a fixed state around $\beta = 3$ with energy zero. In comparison, a 2TDVP calculation at $\beta = 3$ results in an energy of ≈ -17.2. Additionally, calculating the one-site variance and the two-site variance [86] (initially, they sum to ≈ 45.9) in both cases shows that 1TDVP continuously minimizes the one-site variance but leaves the two-site variance invariant. 2TDVP minimizes the two-site variance *at the first time step* and then slowly minimizes the one-site variance while the two-site variance grows again (from $O(10^{-5})$ to $O(10^{-2})$ at $\beta = 20$). The large value of one- and two-site variance initially suggests that
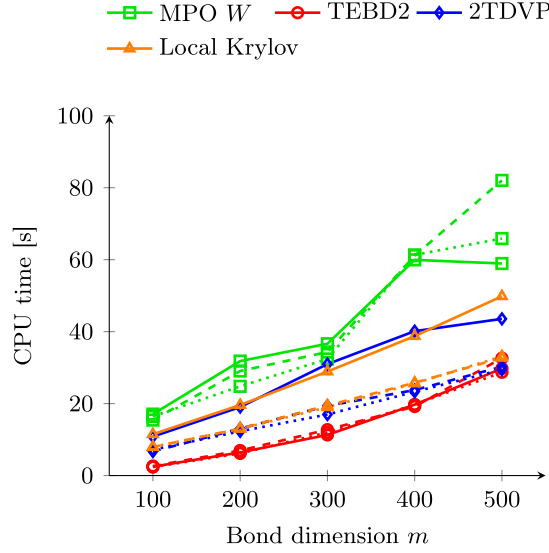
Fig. 30. (Cooling) CPU time required by each method for the single step from $\beta = 1 - \delta$ to $\beta = 1$. Solid lines are for $\delta = 0.1$, dashed for $\delta = 0.05$, dotted for $\delta = 0.01$. Only 2TDVP and the local Krylov method depend on the step size, as larger δ requires more local Krylov vectors.

also the three-site variance is large and 2TDVPs initial abrupt minimization of the two-site variance also suggests that its projection induces some relatively large error which cannot be reduced further by using smaller step sizes.

8.2.4. Runtime comparison

In this experiment, each calculation was run on a single core of a Xeon E5-2630 v4 with 64 GB of RAM and no hard-disk caching. Comparing run times, the major advantage of the 2TDVP due to its very limited time-step error is apparent. Judging from Fig. 29, the MPO W^{II} method and TEBD2 both require a step size of $\delta = 0.01$ to provide data comparable to 2TDVP at $\delta = 0.1$. Supposedly, the local Krylov method would need a yet smaller time step size to provide perfect results. Per individual time step, we compare runtimes in Fig. 30. This runtime per step is lowest for the TEBD2 method and highest for the MPO W^{II} method. The 2TDVP and local Krylov method perform steps similarly quickly. The relatively long runtime of the MPO W^{II} method is largely due to the variational optimization necessary to obtain good results, the MPO W^{II} method in SyTen (using only the zip-up) is approximately a factor of two faster than TEBD2 but also incurs large errors and loss of particles.

Taking into account the larger time step available with 2TDVP compared to TEBD2 and the MPO W^{II} method (as seen in Fig. 29), the 2TDVP method is certainly the fastest approach here. We also confirm that if the number of Krylov vectors is chosen adaptively during the local solution of the 2TDVP problem, the 2TDVP displays slightly longer CPU times per step for larger time steps (as more local Krylov vectors are required). The MPO W^{II} and the TEBD2 CPU times per step generically do not depend on the step size. Only at $m = 500$ the bond dimension is not yet saturated at $\beta = 1$ if too few time steps have been taken, resulting in slightly lower runtimes for larger step sizes.

8.2.5. Conclusion

To summarize the results of this example, we find that for small β, 2TDVP provides adequate results very quickly, with an overall speed-up of 5–10 compared to TEBD2 and the MPO W^{II} method. The local Krylov method incurs large errors which – while having comparable runtime to 2TDVP – make it unsuitable here. At large β, the MPO W^{II} method with variational optimization tends to be the most stable, displaying nearly no particle loss. We did not investigate a possible combination
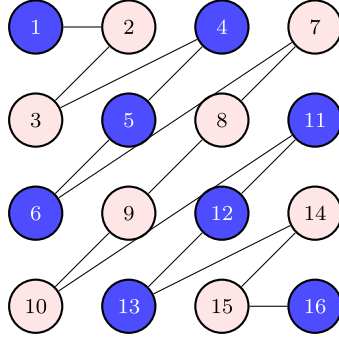
Fig. 31. (2D) Exemplary 4×4 lattice and the mapping of sites to a one-dimensional MPS chain. It was suggested [100] that this tilted z-mapping helps reduce the required bond dimension. Encoded in pink/light and blue/dark are the two sub-lattices *A* and *B*, respectively. In our test, we use an 8×8 lattice with open boundary conditions.

of 2TDVP initially with a later switch to 1TDVP once the bond dimension has grown sufficiently: this would then ensure stability against particle loss at large β combined with the computational efficiency of the TDVP method here.

8.3. Melting of Néel order in the two-dimensional Heisenberg model

In this example we study the melting of Néel order in the ferromagnetic Heisenberg model on the two-dimensional square lattice of side length *L*. The Hamiltonian reads

$$\hat{H} = -\sum_{\langle i,j \rangle} \frac{1}{2} \left(\hat{s}_i^+ \hat{s}_j^- + \hat{s}_j^+ \hat{s}_i^- \right) + \hat{s}_i^z \hat{s}_j^z . \tag{209}$$

where $\langle i, j \rangle$ denotes nearest-neighbor site indices. The initial state at $t = 0$ is chosen to be the Néel state, i.e. spins pointing up and down in a checkerboard pattern

$$|\psi(0)\rangle = \bigotimes_{i \in A} |\downarrow\rangle_i \bigotimes_{j \in B} |\uparrow\rangle_j \tag{210}$$

with sublattices *A* and *B* visualized in Fig. 31. To map the two-dimensional system onto a one-dimensional MPS, we use a tilted z-shaped mapping (cf. Fig. 31) as suggested in Ref. [100]. This mapping transports entanglement between the vertical and horizontal bipartitions of the lattice through $O(L)$ bonds while diagonal bipartitions only cut a single bond. However, nearest-neighbor interactions now have to be carried through $\approx \sqrt{2}L$ other sites.[12] Under real-time evolution, the Néel order of the state is expected to decay, and at the same time the entanglement of this initial product state will grow with time. To monitor this decay, we consider the staggered magnetization per site, which we define as

$$\langle \tilde{m}(t) \rangle = \frac{1}{L^2} \left(\sum_{i \in B} \langle \hat{s}_i^z(t) \rangle - \sum_{j \in A} \langle \hat{s}_j^z(t) \rangle \right) . \tag{211}$$

Initially, $\langle \tilde{m}(0) \rangle = 1/2$. In addition to $\tilde{m}$, we also measure the energy of the state to check the accuracy of energy conservation in the local Krylov and TDVP approaches. Furthermore, we note that the system has an inversion symmetry: Mirroring the initial Néel state on the MPS chain at the central bond results in the same local expectation values. Since the Hamiltonian does not break

---

12 It is not clear a priori, which mapping would be the optimal one for the dynamics of this system; for the sake of simplicity we have chosen this one.

this symmetry, it has to be preserved under real-time evolution. Hence, computing the deviation from this symmetry gives a measure for the error of the time evolution scheme.

In our test, we set $L = 8$, i.e. work with 8×8 lattices of 64 sites with open boundary conditions in both directions. We limit the CPU time for computing the time evolution to one hour on a single core of a Xeon E5-2630 v4 clocked at 2.20 GHz. After this hour, we then evaluate observables on the stored MPS representing the individual time-evolved states. The bond dimension of the states is limited to $m = 200$ and the discarded weight varied between 10^{-8} and 10^{-12}. It appears that the maximal bond dimension limit $m = 200$ is not the leading source of errors at least at short times $t < 1$ – for such short times, calculations running for 24 CPU hours at $m = 2000$ produced essentially identical results. As this is a highly challenging problem for MPS methods, we prefer at this point not to investigate the limitations of our state-of-the-art approaches, but instead compare them for rather restricted bond dimensions focusing on the short time regime. It would be interesting to see, which time scales can be reached when further optimizing our procedure (in particular increase the number of kept states m, which quickly leads to a substantial increase of the needed computational resources, but also further aspects, like the mapping to the chain geometry), and we hope that our considerations here on the short time scales may help future developments of such optimized schemes for treating 2D or quasi-2D systems.

In the following, we choose time steps δ to be either 0.1 or 0.01 to verify approximate convergence and to investigate the error caused by the finite time step.

TDVP and product initial states. The state $|\psi(0)\rangle$ is a product state and can be represented at bond dimension $m = 1$. Attempting to time-evolve this MPS with TDVP or the local Krylov method is pointless, as many of the long-range interactions are lost during the projection step and the resulting time evolution is simply wrong. To avoid this problem, we run an initial DMRG calculation with the Hamiltonian

$$\hat{H}_{\mathrm{DMRG}} = -\alpha \left(\sum_{\langle i,j \rangle} \tfrac{1}{2} \left(\hat{s}_i^+ \hat{s}_j^- + \hat{s}_j^+ \hat{s}_i^- \right) + \hat{s}_i^z \hat{s}_j^z \right) + \sum_{i \in A} \hat{s}^z - \sum_{i \in B} \hat{s}_i^z . \tag{212}$$

We initialize $\alpha = 1$ and then for ten successive sweeps reduce it by a factor of ten each time. Finally, we set $\alpha = 0$ and run five more sweeps. Throughout, we use the subspace expansion to increase the bond dimension of the state up to $m_{\mathrm{init}} = [50, 100, 200]$ and set the discarded weight threshold to zero to keep the state at this artificially large bond dimension. Using the time-evolution Hamiltonian $\hat{H}$ together with the subspace expansion guarantees that the generated additional states have some relevance to the physical problem by being generated from (partial) applications of $\hat{H}$ to $|\psi(0)\rangle$. The DMRG calculation takes up to 120 s, the allowed runtime of the TDVP and local Krylov methods is hence reduced accordingly.

Time-evolving block decimation. The TEBD2 requires splitting the Hamiltonian into internally-commuting parts. To minimize the bond dimensions of the MPOs, each column and each row of the lattice are split into two Hamiltonians. These Hamiltonians then do not contain overlapping gates and have a maximal bond dimension of 4. For optimal inversion symmetry, we apply gates on rows 1 and 8 first, then on rows 2 and 7, rows 3 and 6 and finally rows 4 and 5 (or vice versa) and proceed in the same way with the column terms. These MPOs are applied using the zip-up method which was tested to be as precise as the variational optimization here, but much faster.

The global Krylov method. As can be expected, repeatedly applying $\hat{H}$ to the Néel state generates large amounts of entanglement. This is problematic for the global Krylov method, as its Krylov vectors would ideally have a very large bond dimension. To still proceed with the calculation, we force the truncation of Krylov vectors also to a maximal bond dimension $m = 200$. This truncation can lead to very wrong results when measuring long-range correlators but should be acceptable for $\langle \tilde{m} \rangle$. Given the initial product state, the variational orthogonalization has to work in the two-site variant.
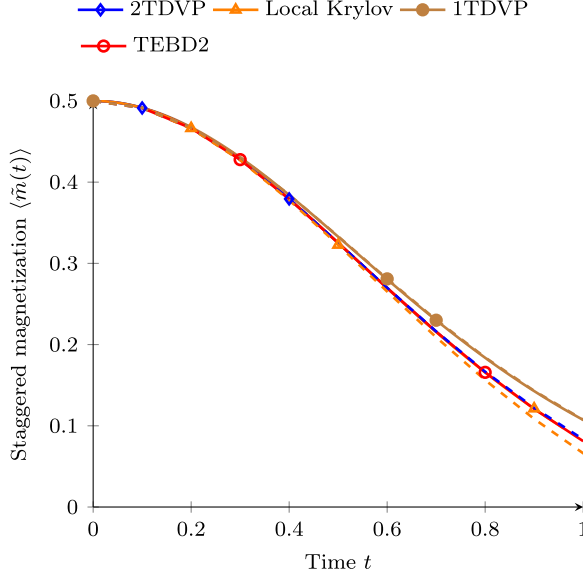
Fig. 32. (2D) Staggered magnetization $\langle \tilde{m}(t) \rangle$ during time evolution with maximal bond dimension $m_{\text{max}} = 200$. Comparison of the dependence on the step size of the 1TDVP, 2TDVP and the local Krylov method; we use $\delta = 0.01$ (solid) and $\delta = 0.1$ (dashed) at initial $m = 100$. 1TDVP and 2TDVP data is step-size independent but the local Krylov results differ slightly between $\delta = 0.1$ and $\delta = 0.01$. Due to the fixed bond dimension, 1TDVP at best provides qualitative results.

8.3.1. The staggered magnetization $\langle \tilde{m}(t) \rangle$

We find that neither 1TDVP nor 2TDVP results change with the step size, but some minimal changes can be seen in the local Krylov results between $\delta = 0.1$ and $\delta = 0.01$ (cf. Fig. 32). Furthermore, we find that the 2TDVP and local Krylov method results for $\langle \tilde{m}(t) \rangle$ do not change with the initial bond dimension $m_{\text{init}} = [50, 100, 200]$ as obtained by DMRG and converge to the same result at least for short times. The 1TDVP results depend on this initial bond dimension (cf. Fig. 33) and even in the best case of $m_{\text{init}} = 200$ only qualitatively reproduce the other results. In the following, we hence use $m_{\text{init}} = 50, \delta = 0.1$ for 2TDVP, $m_{\text{init}} = 100, \delta = 0.1$ for 1TDVP and $m_{\text{init}} = 50, \delta = 0.1$ and $\delta = 0.01$ for the local Krylov method.

The global Krylov results do not change with step size $\delta = 0.01$ or $\delta = 0.1$ nor with the discarded weight $10^{-8,-10,-12}$ and are in both cases in good agreement with the other methods until $t \approx 0.7$. Beyond this point, it provides worse data than 1TDVP due to the very strong truncation of the Krylov vectors. TEBD2 provides data within the range of the other methods at $\delta = 0.1$ with the zip-up algorithm and a discarded weight of 10^{-8}. At smaller δ or when using the variational optimization for operator application, the evolution only obtains times $t \approx 0.5$ (but in agreement with all other methods).

When using the MPO W^{II} method with large time steps $\delta = 0.1$, the results very quickly deviate from the other methods; in the SyTen implementation (using either zip-up or variational application), additional instabilities occur around $t \approx 4$. At $\delta = 0.01$, data is in line with the other methods mostly independent of the discarded weight. The variational operator application is much slower than the zip-up method here, so we use the latter also for the MPO W^{II} method.

Fig. 34 shows the staggered magnetization for the selected configurations of each method. The global Krylov method is handicapped by its need to represent highly-entangled Krylov vectors at $m = 200$, which leads to large errors. The MPO W^{II} method suffers a large time-step error which in turn increases runtime by approximately a factor of ten compared to those methods which can use a larger step size. TEBD2 provides very reasonable data up to $t = 3$ and 2TDVP and the local Krylov
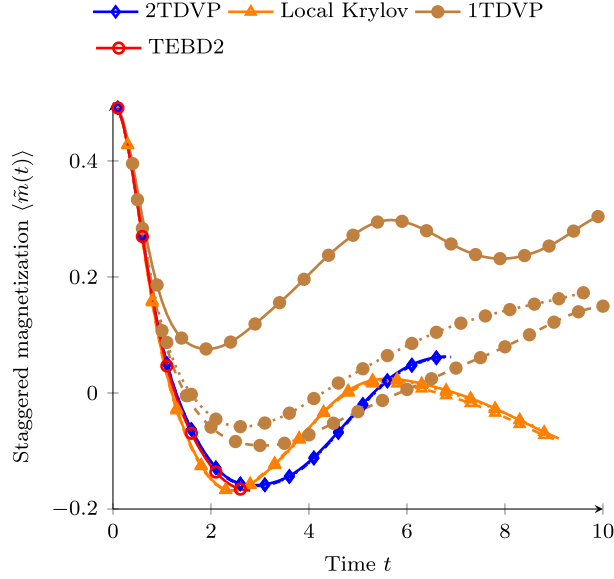
Fig. 33. (2D) Staggered magnetization $\langle \tilde{m}(t) \rangle$ during time evolution with maximal bond dimension $m_{max} = 200$. Comparison of the dependence of the 1TDVP, 2TDVP and the local Krylov method on the initial bond dimension; $m = 50$ (solid), $m = 100$ (dashed) and $m = 200$ (dotted). Step size is $\delta = 0.1$. The best TEBD2 result is plotted as a reference. The 2TDVP and the local Krylov method do not strongly depend on the initial bond dimension.
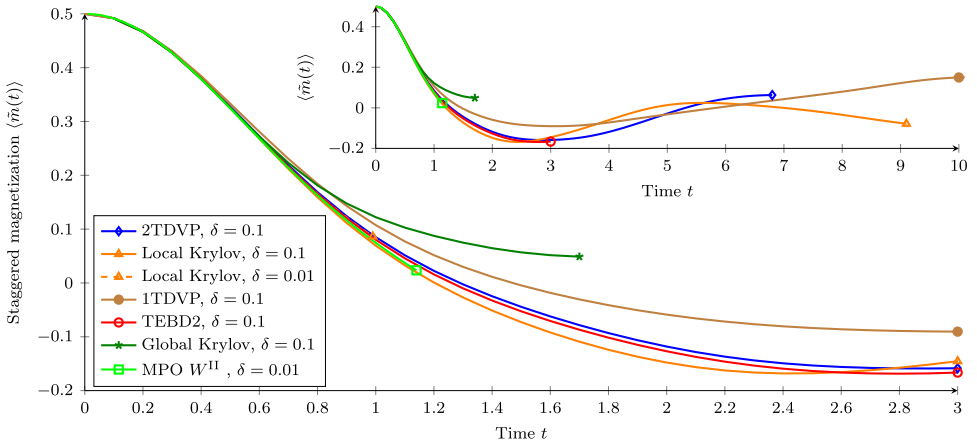

Fig. 34. (2D) Staggered magnetization $\langle \tilde{m}(t) \rangle$ during time evolution with maximal bond dimension $m_{max} = 200$ for all available methods. Only the local Krylov method shows dependence on the time step size. 1TDVP and the global Krylov method show large errors at $t \geq 0.5$. At intermediate times, 2TDVP seems to match the behavior of TEBD2 better than the local Krylov method at $\delta = 0.1$. At long times, both the local Krylov data and the 2TDVP data appear reasonable though not necessarily trustworthy. The 1TDVP, while able to obtain times $t > 100$, produces certainly wrong long-time results.

method are able to go beyond up to $t \approx 7$ and $t \approx 10$ respectively. However, as the truncation error is of the order of 10^{-2} per step size in both methods at larger t, the results will have large quantitative errors at large times, as discussed next.
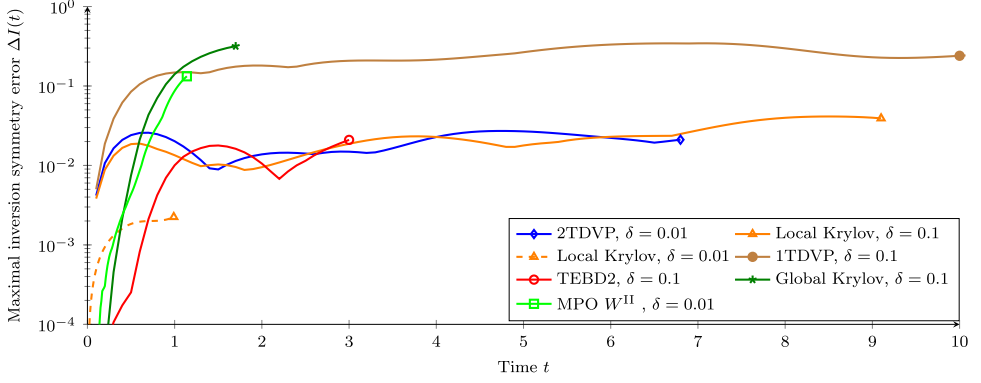
Fig. 35. (2D) Maximal inversion symmetry error $\Delta I(t)$. The simulations using TEBD2, 2TDVP and the local Krylov method result in approximately the same errors at intermediate times. At short times, the local Krylov method with $\delta = 0.01$ and TEBD2 provide results with a relative error of a fraction of 1%.
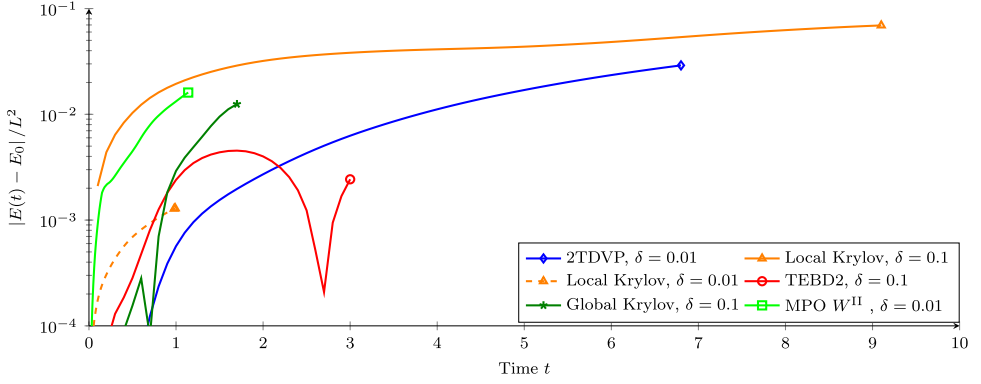
Fig. 36. (2D) Error in energy per site $|E(t) - E_0|/L^2$ during the real-time evolution. Not shown is single-site TDVP with error $< 10^{-7}$ at all times. Comparing 2TDVP and the local Krylov method, the better energy conservation of 2TDVP is obvious. The global Krylov method is good up to $t \approx 0.7$ when the truncation of Krylov vectors becomes problematic.

8.3.2. Energy conservation and inversion symmetry

Under the unitary time evolution, both the energy of the system and the inversion symmetry over the central MPS bond should be conserved. To measure the latter, we consider the maximal deviation

$$\Delta I(t) = \max_{i=1}^{L^2} \left| \langle \hat{s}_i^z(t) \rangle - \langle \hat{s}_{L^2-i+1}^z(t) \rangle \right| . \tag{213}$$

Then considering the same configurations as before, we find in Fig. 35 that the inversion error stays relatively small for 2TDVP, the local Krylov and the TEBD2 methods, while it becomes up to an order of magnitude larger for 1TDVP, the global Krylov and the MPO W^{II} method at intermediate times. At very short times, we find TEBD2 and (with $\delta = 0.01$) the local Krylov method to have the smallest inversion symmetry error. The other methods (run at $\delta = 0.01$) also provide similarly good results (not shown).

Considering $|E(t) - E_0|/L^2$, the error in the energy per site, (cf. Fig. 36) provides nearly the same picture except for a much larger difference between the local Krylov method and 2TDVP – the latter conserves energy to a much higher precision than the former. 1TDVP (not shown) provides accuracy

in energy up to 10^{-7} even at very late times while the global Krylov method is very precise until $t \approx 0.7$, where its error quickly increases.

8.3.3. Conclusion

Obtaining the time evolution in 2D is, even for a non-entangled product initial state, a challenge for all methods. Even though the projection steps in the 2TDVP, the local Krylov, and the 1TDVP methods incur problems with such product initial states, this can be healed by simply increasing the initial bond dimension as discussed above, which is also often done in tDMRG in such situations. Apart from better energy conservation, 2TDVP and the local Krylov method are nearly comparable, with the latter being approximately 20%–30% faster (however, the maximal bond dimension $m = 200$ is here already exhausted at $t < 1$). TEBD2 mostly suffers from slow MPO–MPS products at large bond dimensions but actually has an acceptable error caused by the finite time step, allowing us to choose $\delta = 0.1$. In contrast, the MPO method sports fast MPO–MPS products (due to the smaller $\hat{W}^{II}$) but incurs a large time step error. It hence appears that for longer time scales, the 2TDVP or local Krylov methods are the most promising approaches, while for short times, any of the methods work reasonably well.

8.4. OTOCs for interacting spins

The final example is motivated by the study of information spreading and scrambling in closed quantum many-body systems [94,101,102]. This can be done by studying the dynamics of out-of-time-order correlators (OTOCs) of local observables under an initial perturbation. We consider a periodically-driven Ising chain with a conserved $\hat{S}^z$ quantum number which recently has been shown to exhibit characteristic long-time relaxation behavior: a power law decay in the OTOCs which can be related to a hydrodynamical picture of diffusion [103]. The time evolution is generated by a Floquet driving protocol $\hat{U}(kT) = \left[\hat{U}(T)\right]^k$ with driving period T and

$$\hat{U}(T) = e^{-i\tau\hat{H}_{xy}} e^{-i\tau\hat{H}_{zz}} e^{-i\tau\hat{H}_{xy}} e^{-i\tau\hat{H}_z} \tag{214}$$

$$\hat{H}_{xy} = J \sum_j \hat{s}_j^x \hat{s}_{j+1}^x + \hat{s}_j^y \hat{s}_{j+1}^y \tag{215}$$

$$\hat{H}_z = J_z \sum_j \hat{s}_j^z \hat{s}_{j+1}^z \tag{216}$$

$$\hat{H}_{zz} = J_{zz} \sum_j \hat{s}_j^z \hat{s}_{j+2}^z \ . \tag{217}$$

The driving period is chosen so that $T = 1 = 4\tau$ and the coupling constants are given by

$$J_{xy} = \frac{2\sqrt{3} + 3}{7}, \quad J_z = \frac{\sqrt{3} + 5}{6} \quad \text{and} \quad J_{zz} = \frac{\sqrt{5}}{2} \ . \tag{218}$$

The OTOC we calculate describes correlations in the Heisenberg time evolution between operators $\hat{s}_{L/2}^z(t \equiv 0) \equiv \hat{z}_0$ and $\hat{s}_{L/2+r}^z(t) \equiv \hat{z}_r$ and is given by an ensemble average

$$F_\beta^{\hat{z}_r \hat{z}_0} = \text{Re} \left(\text{Tr} \left\{ \hat{\rho}(\beta) \hat{z}_r^\dagger(t) \hat{z}_0^\dagger \hat{z}_r(t) \hat{z}_0 \right\} \right) \tag{219}$$

which can be interpreted physically as a measure for the causal relation between correlators during time evolution. This is typically done at non-zero temperature so that we need to describe the system in an enlarged Hilbert space $\mathcal{H} = \mathcal{H}_P \otimes \mathcal{H}_A$ to be able to represent mixed states, cf. Section 2.7. Here, we work in the infinite-temperature limit $\beta = 0$ of a canonical ensemble with $S^z = 0$. As carried out in Section 7.5 the time evolution of a purified, (in-)finite temperature state can be manipulated by shifting the purification insertion point so that we can evaluate the OTOC
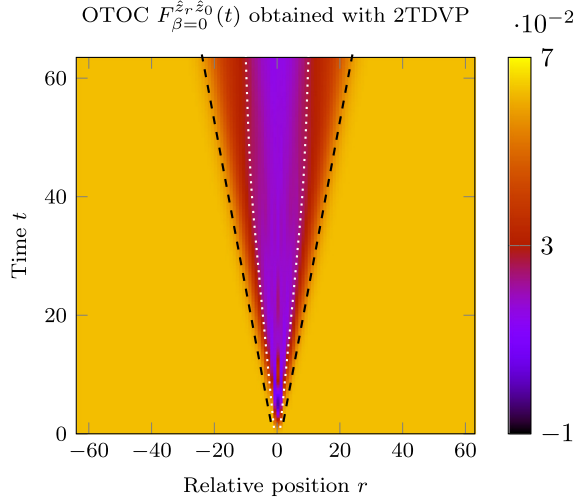
Fig. 37. (OTOC) $F_{\beta=0}^{\hat{z}_r \hat{z}_0}(t)$ for a periodically driven Ising chain with local operators $\hat{z}_r = \hat{S}_{L/2+r}^z$ as obtained with 2TDVP and time step $\delta = 0.01$. The broadening of the lightcone is marked for the envelope of the lightcone front (dashed lines) and for the bulk (dotted lines).

performing two time evolutions and calculate only local expectation values:

$$F_\beta^{\hat{z}_r \hat{z}_0} = \langle Z_0(\beta, t) | \hat{z}_{P;r}^\dagger \otimes \hat{z}_{A;r}^t | Z_0(t) \rangle \tag{220}$$

$$|Z_0(t)\rangle = \left(\hat{U}_P(t) \otimes \hat{U}_A^\star(t) \right) \left(\hat{z}_{P;0}^\dagger \otimes \hat{\mathbf{1}} \right) |0\rangle \tag{221}$$

$$|Z(\beta, t)\rangle = \left(\hat{U}_P(t) \otimes \hat{U}_A^\star(t) \right) \left(\hat{\rho}_P(\beta/2) \otimes \hat{z}_{A;0}^\star \hat{\rho}_A(\beta/2) \right) |0\rangle . \tag{222}$$

We calculated the time evolutions for 32, 64 and 128 physical sites with step sizes $\delta = 0.05, 0.01$ and kept $m = 1000$ as the maximum number of states. We find the global Krylov method to be unsuitable for the current problem. Due to the infinite temperature initial state and time evolutions on both the physical and auxiliary system, the Krylov vectors are already at the beginning highly entangled, forbidding a calculation at reasonable time scales.

8.4.1. Spreading of the light cone

In our setup we considered open boundary conditions so that we need to ensure that excitations from the boundary do not disturb the OTOC. With the chosen set of coupling constants this is fulfilled by restricting the maximal computed time evolution to $kT \leq L/2$ so that in order to study the hydrodynamic tail we will discuss only the results for the largest system with $L = 128$ physical sites. In Fig. 37 the OTOC $F_{\beta=0}^{\hat{z}_r \hat{z}_0}(t)$ obtained with 2TDVP and a step width $\delta = 0.01$ is shown as an example for the time evolution at all relative lattice distances r with respect to the central site of the system. We have marked the envelope of the spreading lightcone front and the bulk for the case of 2TDVP which gives a quantitative picture of the broadening of the lightcone front. Fig. 38 displays the same OTOC obtained with the local Krylov stepper and $\delta = 0.05$. Comparing to the time evolution obtained with 2TDVP we find that both the spreading of the lightcone front and the broadening of the lightcone itself is much less pronounced and disturbed by *rays* in the lightcone which decay only slowly. These rays are generated by structure inside the lightcone which are absent in the 2TDVP datasets. Apart from even–odd effects, we would expect the dynamics inside the lightcone to generate a homogeneous amplitude of the OTOC, strengthening the 2TDVP results. Similar plots can be obtained for the MPO W^{II} method which suffers from the same effects as the local Krylov method.
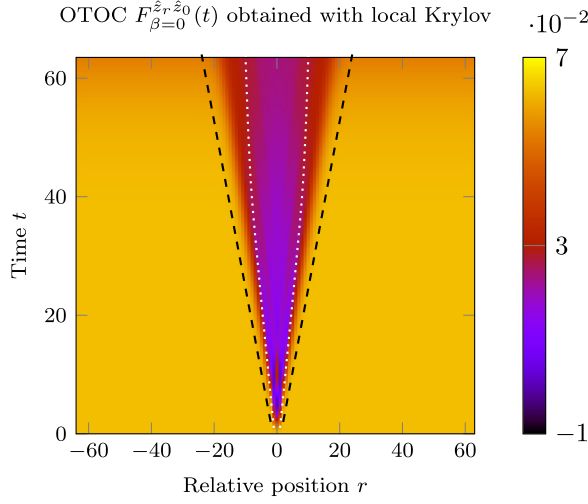
Fig. 38. (OTOC) $F_{\beta=0}^{\hat{z}_r \hat{z}_0}(t)$ for a periodically driven Ising chain with local operators $\hat{z}_r = \hat{S}_{L/2+r}^z$ as obtained with the local Krylov method and time step $\delta = 0.05$. The lightcone broadening seen in the 2TDVP simulations is indicated for the envelope of the lightcone front (dashed lines) and for the bulk (dotted lines).
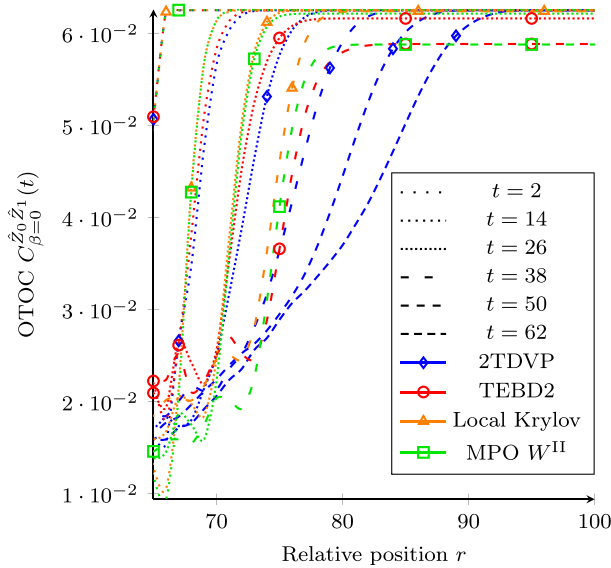
Fig. 39. (OTOC) Spreading of the lightcone front for $F_{\beta=0}^{\hat{z}_r \hat{z}_0}(t)$ of a periodically driven Ising chain with local operators $\hat{z}_r = \hat{S}_{L/2+r}^z$ at constant times and $\delta = 0.05$. 2TDVP appears to resemble the expected spreading, while TEBD2, the local Krylov and the MPO W^{II} methods exhibit only a reduced spreading. Instead, the latter methods create artificial offsets (results from these simulations for time slices $t = 50, 62$ are not shown here for the sake of clarity).

In order to investigate the lightcone spreading more carefully, we show in Fig. 39 the decay of the lightcone at various slices throughout the right half of the system at constant times $t = 2, 14, 26, 38, 50$ and 62 at fixed time step $\delta = 0.05$. Only at the first two time slices all methods (nearly) agree with each other displaying the expected broadening of the lightcone front. The results
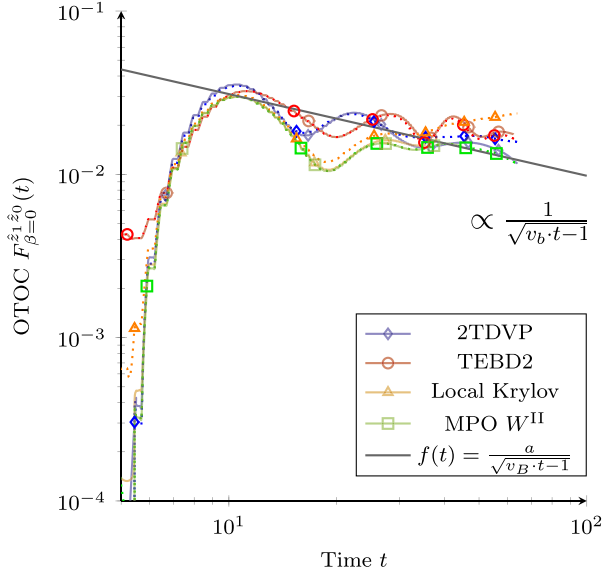
Fig. 40. (OTOC) $F^{\hat{z}_1 \hat{z}_0}_{\beta=0}(t)$ of a periodically driven Ising chain with local operators $\hat{z}_i = \hat{s}^z_{L/2+i}$. Equal colors and symbols refer to the same time evolution method, while the different line shapes indicate different step sizes. Dotted lines represent simulations with $\delta = 0.05$ and solid lines represent $\delta = 0.01$. The black solid line indicates the analytical form of the hydrodynamic tail, which is $f(t) = \frac{1}{\sqrt{v_B \cdot t - 1}}$.

for the 2TDVP method show an overall decay for the slope at the lightcone front at later times, as expected. In turn the local Krylov, TEBD2 and MPO W^{II} methods produce a decay for the slope of the lightcone front under the time evolution which can be attributed to additional ray-like structures in the lightcone, cf. Fig. 38. Furthermore, we find that the saturation value outside the lightcone for these three methods lowers during the time evolution. This can be linked to additional phases built up during the time evolutions between the two states $|Z_0(t)\rangle$ and $|Z(\beta = 0, t)\rangle$, as already observed in Section 8.1.

8.4.2. Hydrodynamic tail

In the case of periodic boundary conditions the long time behavior behind the lightcone, which travels with the butterfly velocity v_B, is expected to exhibit an algebraic power law decay, i.e., for $v_B \cdot t \gg |r|$ the OTOC is dominated by $\frac{1}{\sqrt{v_B \cdot t - |r|}}$. Fig. 40 compares the OTOCs at fixed relative position $r \equiv 1$ obtained with the methods under consideration (2TDVP, TEBD2, MPO W^{II} and local Krylov) and the time steps $\delta = 0.01, 0.05$. 2TDVP qualitatively reproduces the expected hydrodynamic tail as can be seen by comparing to an indicated reference function $f(t) = \frac{a}{\sqrt{v_B \cdot t - 1}}$ where the parameters a and v_B are obtained by a suitable fit of the 2TDVP datasets. Furthermore there are only little deviations when tuning the time steps, strengthening the validity of these results. However, comparing against the local Krylov, MPO W^{II} and TEBD2 datasets there is only agreement until $t \approx 10T$ after which the methods start to deviate and exhibit quite distinct oscillatoric behavior. It seems that at least for the local Krylov (at $\delta = 0.01$) and MPO W^{II} (at $\delta = 0.05, 0.01$) method one can identify the expected hydrodynamic tail but the data is far from being reliable. Since the initial states describe a (perturbed) ensemble at infinite temperature with comparably large initial bond dimensions ($m \sim 150$) we could expect the projection errors in 2TDVP and the local Krylov method, respectively, to be very small because there will be a finite overlap with a large portion of the states in the many-body Hilbert space. Subsequently, the effect of the perturbation $\hat{s}^z_{L/2}$ on the infinite temperature state is then to create quasiparticles from the entire spectrum resulting
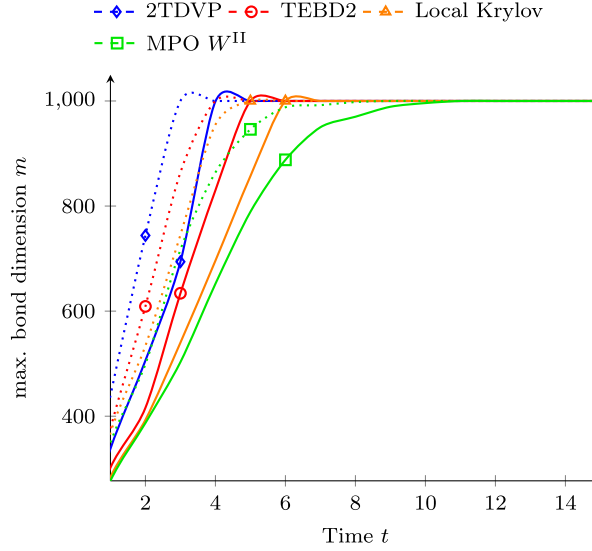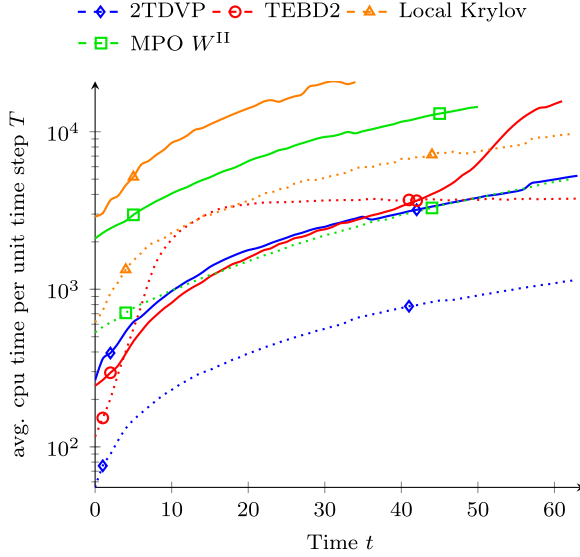
Fig. 41. (OTOC) Maximal bond dimension m during the evolution of $|Z_0(t)\rangle$ plotted over real time t. The datapoints are obtained after a complete driving cycle $t = kT$ and the line styles indicate the used time step: dotted lines correspond to $\delta = 0.05$, solid lines to $\delta = 0.01$. For all methods m grows quickly and reaches the maximal permitted value at times $t < 10T$. 2TDVP exhibits the fastest increase with the maximum m being reached at times $t < 5T$ while MPO W^{II} shows the slowest growth.

in complicated dynamics which is a tough challenge for all time evolution schemes. Nevertheless, since TDVP (1TDVP exact, 2TDVP approximately) fulfills conservations laws for the energy, particles etc. it is not too surprising to find the proper long-time dynamics being captured by TDVP. Even though the datasets created by 2TDVP are physically reasonable, the local Krylov as the other method approximating the action of the exponential suffers from similar phase accumulation as in Section 8.1. Consistently we also find the strongest dependence on the step size δ for the local Krylov method.

8.4.3. Benchmark

Finally we take a look at the benchmark data for the calculated time evolutions at the example of the evolution of the state $|Z_0(t)\rangle$. The calculations were performed single threaded on an Intel Xeon Gold 6150 CPU with 192 GB of RAM and no hard-disk caching. All methods are subject to a rapid increase of the maximal bond dimension m (cf. Fig. 41). 2TDVP exhibits the quickest increase and already reaches the maximal permitted m at times $t < 5T$ while the MPO W^{II} method has the slowest growth and saturates around $t \approx 10T$ which may be attributed to the variational optimization during the MPO application.

In Fig. 42 we plot the CPU time required per unit time T averaged over one driving cycle. We find the best performance for 2TDVP followed by TEBD2 and the MPO W^{II}. Interestingly, the performance of the MPO W^{II} method at $\delta = 0.05$ is very competitive with TEBD2 at the same step size but when reducing the step size towards $\delta = 0.01$ the variational update in the MPO application, which is absolutely necessary to ensure numerical stability in this testcase, renders the simulations extremely slow. The large difference between the runtimes for 2TDVP and the local Krylov method can be explained by inspecting the number of local Krylov vectors which are required to achieve the target precision of 10^{-10}. TDVP typically requires only 5 iterations of the local Krylov solver to converge while the local Krylov method reaches the maximum number of permitted iterations (12) already after a few time steps. Consequently the local Krylov method is unsuitable for small time steps $\delta = 0.01$.

Fig. 42. (OTOC) Elapsed CPU time per unit time T during real-time evolution of $|Z_0(t)\rangle$ plotted over real time t. The datapoints are obtained after a complete driving cycle $t = kT$ and the line styles indicate the used time step: dotted lines correspond to $\delta = 0.05$, solid lines to $\delta = 0.01$. By far, 2TDVP shows the best performance while the TEBD2, MPO W^{II} and local Krylov methods at least permit for reasonable calculation of the time-evolution at time steps $\delta = 0.05$.

8.4.4. Conclusion

We have tested four time evolution methods: 2TDVP, TEBD2, local Krylov and MPO W^{II} for their ability to describe the time evolution of out-of-time-order correlators for a periodically-driven Ising chain. The best results are provided by 2TDVP with respect to both physical accuracy and performance. TEBD2 again yields a competitive runtime at least for larger step sizes but the data fails to reproduce the broadening of the lightcone and even the hydrodynamic tail of the OTOC is only qualitatively reproduced. Both the local Krylov and MPO W^{II} methods fail to adequately reproduce the broadening of the lightcone front employing the numerically feasible step size $\delta = 0.05$ and, more severely, pick up global offsets in the OTOC at long runtimes. The hydrodynamic tail may be identified in these datasets but the reliability should be strengthened by a very careful analysis with more time steps and different maximal bond dimensions m. Since fulfilling conservation laws seems to be advantageous for capturing the complicated dynamics inside the lightcone the possibly best strategy would be to perform 2TDVP time evolution until the bulk bond dimensions have saturated to the permitted maximum value and then switch to 1TDVP. In addition to the results presented here using the purification approach, it was recently shown that TDVP when used with METTS can also provide reasonable data for OTOCs [104].

9. Future developments

While one should of course not lose sight of new alternative methods to evaluate excitation spectra [105–109] or Green's functions [19], the direct evaluation of real-time observables will always find interesting applications. For the future, at least three distinct approaches for new developments of real-time evolution methods should be considered:

Improved MPS-local methods. Following the spirit of td-DMRG, the local Krylov and TDVP may provide the best avenue towards hydrodynamics with the 1TDVP method already enforcing complete energy conservation though still suffering from a limited bond dimension [87]. Combining e.g. the 2TDVP method with a truncation scheme which *also* ensures energy conservation [74] may prove particularly fruitful.

Easier and more accurate construction of $\hat{U}(\delta)$. At the moment one has to use either the TEBD2 or the MPO $W^{\text{I,II}}$ method to construct an MPO representation of $\hat{U}(\delta)$. Both incur relatively large time-step errors and require more than simply a set of tensors describing the MPO as input. One potential alternative here is a construction as $(1 - \mathrm{i}\delta'\hat{H})^N$ with $\delta'N = \delta$ and $N \gg 1$. The ingredients can all be represented within the MPO framework and MPO compression may provide a very helpful tool to conserving as much accuracy as possible during the construction [110]. Alternatively, combining different approaches, the newest example being a large-scale Trotter decomposition combined with a small-scale Krylov method, may also lead to flexible time steppers which only suffer from minimal time step-size errors [79].

"Better" Krylov vectors. The major drawback of the global Krylov method at the moment is the fast growth of entanglement in the Krylov vectors. While it is used where absolutely necessary – due to its flexibility in the choice of time step sizes – alternative methods provide much faster long-time evolution. In addition to this flexibility, it also only requires an MPO representation as input, making it potentially useful in applications where decompositions are not straightforward. Hence, an approach to use it also in complex settings would be highly appreciated. This could be done by using less entangled Krylov-like subspaces whose basis vectors can be represented efficiently as matrix-product states. Alternatively, one can envision a second basis transformation finding minimally-entangled basis vectors for the standard Krylov subspace.

Acknowledgments

Funding information

References

[1] S.R. White, Phys. Rev. L 69 (1992) 2863–2866, http://dx.doi.org/10.1103/PhysRevLett.69.2863.
[2] S.R. White, Phys. Rev. B 48 (1993) 10345–10356, http://dx.doi.org/10.1103/PhysRevB.48.10345.
[3] G. Vidal, Phys. Rev. Lett. 91 (2003) 147902, http://dx.doi.org/10.1103/PhysRevLett.91.147902.
[4] G. Vidal, Phys. Rev. Lett. 93 (2004) 040502, http://dx.doi.org/10.1103/PhysRevLett.93.040502.
[5] S.R. White, A.E. Feiguin, Phys. Rev. Lett. 93 (2004) 076401, http://dx.doi.org/10.1103/PhysRevLett.93.076401.
[6] A.J. Daley, C. Kollath, U. Schollwöck, G. Vidal, J. Stat. Mech. 2004 (04) (2004) P04005, http://dx.doi.org/10.1088/1742-5468/2004/04/P04005.
[7] M. Zwolak, G. Vidal, Phys. Rev. Lett. 93 (2004) 207205, http://dx.doi.org/10.1103/PhysRevLett.93.207205.
[8] F. Verstraete, J.J. García-Ripoll, J.I. Cirac, Phys. Rev. L 93 (2004) 207204, http://dx.doi.org/10.1103/PhysRevLett.93.207204.
[9] S. Rommer, S. Östlund, Phys. Rev. B 55 (1997) 2164–2181, http://dx.doi.org/10.1103/PhysRevB.55.2164.
[10] J. Dukelsky, M.A. Martín-Delgado, T. Nishino, G. Sierra, Europhys. Lett. 43 (4) (1998) 457, http://dx.doi.org/10.1209/epl/i1998-00381-x.
[11] I.P. McCulloch, J. Stat. Mech. 2007 (10) (2007) P10014, http://dx.doi.org/10.1088/1742-5468/2007/10/P10014.
[12] N. Schuch, M.M. Wolf, F. Verstraete, J.I. Cirac, Phys. Rev. L 100 (2008) 030504, http://dx.doi.org/10.1103/PhysRevLett.100.030504.
[13] I.V. Oseledets, SIAM J. Sci. Comput. 33 (5) (2011) 2295–2317, http://dx.doi.org/10.1137/090752286.

[14] U. Schollwöck, Ann. Phys. 326 (1) (2011) 96–192, http://dx.doi.org/10.1016/j.aop.2010.09.012.
[15] A. Weiße, G. Wellein, A. Alvermann, H. Fehske, Rev. Modern Phys. 78 (2006) 275–306, http://dx.doi.org/10.1103/RevModPhys.78.275.
[16] A. Holzner, A. Weichselbaum, I.P. McCulloch, U. Schollwöck, J. von Delft, Phys. Rev. B 83 (2011) 195115, http://dx.doi.org/10.1103/PhysRevB.83.195115.
[17] F.A. Wolf, I.P. McCulloch, O. Parcollet, U. Schollwöck, Phys. Rev. B 90 (2014) 115124, http://dx.doi.org/10.1103/PhysRevB.90.115124.
[18] J.C. Halimeh, F. Kolley, I.P. McCulloch, Phys. Rev. B 92 (2015) 115130, http://dx.doi.org/10.1103/PhysRevB.92.115130.
[19] H.D. Xie, R.Z. Huang, X.J. Han, X. Yan, H.H. Zhao, Z.Y. Xie, H.J. Liao, T. Xiang, Phys. Rev. B 97 (2018) 075111, http://dx.doi.org/10.1103/PhysRevB.97.075111.
[20] T.D. Kühner, S.R. White, Phys. Rev. B 60 (1999) 335–343, http://dx.doi.org/10.1103/PhysRevB.60.335.
[21] E. Jeckelmann, Phys. Rev. B 66 (2002) 045114, http://dx.doi.org/10.1103/PhysRevB.66.045114.
[22] E. Jeckelmann, Progr. Theoret. Phys. Suppl. 176 (2008) 143–164, http://dx.doi.org/10.1143/PTPS.176.143.
[23] C. Moler, C. Van Loan, SIAM Rev. 45 (1) (2003) 3–49, http://dx.doi.org/10.1137/S00361445024180.
[24] Y. Saad, Iterative Methods for Sparse Linear Systems, second ed., SIAM, 2003, http://www-users.cs.umn.edu/saad/IterMethBook_2ndEd.pdf.
[25] T.J. Park, J.C. Light, J. Chem. Phys. 85 (10) (1986) 5870–5876, http://dx.doi.org/10.1063/1.451548.
[26] R. Kosloff, J. Phys. Chem. 92 (8) (1988) 2087–2100, http://dx.doi.org/10.1021/j100319a003.
[27] J.J. García-Ripoll, New J. Phys. 8 (12) (2006) 305, http://dx.doi.org/10.1088/1367-2630/8/12/305.
[28] P.E. Dargel, A. Wöllert, A. Honecker, I.P. McCulloch, U. Schollwöck, T. Pruschke, Phys. Rev. B 85 (2012) 205119, http://dx.doi.org/10.1103/PhysRevB.85.205119.
[29] M.L. Wall, L.D. Carr, New J. Phys. 14 (12) (2012) 125015, http://dx.doi.org/10.1088/1367-2630/14/12/125015.
[30] C. Lanczos, An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators, United States Governm. Press Office, Los Angeles, CA, 1950.
[31] P. Schmitteckert, Phys. Rev. B 70 (2004) 121302, http://dx.doi.org/10.1103/PhysRevB.70.121302.
[32] A.E. Feiguin, S.R. White, Phys. Rev. B 72 (2005) 020404, http://dx.doi.org/10.1103/PhysRevB.72.020404.
[33] S.R. Manmana, A. Muramatsu, R.M. Noack, AIP Conf. Proc. 789 (1) (2005) 269–278, http://dx.doi.org/10.1063/1.2080353.
[34] K. Rodriguez, S.R. Manmana, M. Rigol, R.M. Noack, A. Muramatsu, New J. Phys. 8 (8) (2006) 169, http://dx.doi.org/10.1088/1367-2630/8/8/169.
[35] E. Ronca, Z. Li, C.A. Jimenez-Hoyos, G.K.-L. Chan, J. Chem. Theory Comput. 13 (2017) 5560–5571, http://dx.doi.org/10.1021/acs.jctc.7b00682.
[36] J. Haegeman, J.I. Cirac, T.J. Osborne, I. Pižorn, H. Verschelde, F. Verstraete, Phys. Rev. Lett. 107 (2011) 070601, http://dx.doi.org/10.1103/PhysRevLett.107.070601.
[37] J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, F. Verstraete, Phys. Rev. B 94 (2016) 165116, http://dx.doi.org/10.1103/PhysRevB.94.165116.
[38] J.B. Keller, D.W. McLaughlin, Amer. Math. Monthly 82 (5) (1975) 451–465, http://dx.doi.org/10.1080/00029890.1975.11993855.
[39] M. Suzuki, Comm. Math. Phys. 51 (2) (1976) 183–190, http://dx.doi.org/10.1007/BF01609348.
[40] M.P. Zaletel, R.S.K. Mong, C. Karrasch, J.E. Moore, F. Pollmann, Phys. Rev. B 91 (2015) 165112, http://dx.doi.org/10.1103/PhysRevB.91.165112.
[41] S.C. Kramer, J. Hagemann, ACM Trans. Parallel Comput. 1 (2) (2015) 15:1–15:31, http://dx.doi.org/10.1145/2686886.
[42] S. Paeckel, T. Köhler, The SYMMPS toolkit. https://symmps.eu.
[43] C. Hubig, F. Lachenmaier, N.-O. Linden, T. Reinhard, L. Stenzel, A. Swoboda, The SYTEN toolkit. https://syten.eu.
[44] C. Hubig, Symmetry-Protected Tensor Networks (Ph.D. thesis), LMU München, 2017, https://edoc.ub.uni-muenchen.de/21348/.
[45] A. Wietek, A.M. Läuchli, Phys. Rev. E 98 (2018) 033309, http://dx.doi.org/10.1103/PhysRevE.98.033309.
[46] A.W. Sandvik, J. Kurkijärvi, Phys. Rev. B 43 (1991) 5950–5961, http://dx.doi.org/10.1103/PhysRevB.43.5950.
[47] A. Georges, G. Kotliar, W. Krauth, M.J. Rozenberg, Rev. Modern Phys. 68 (1996) 13–125, http://dx.doi.org/10.1103/RevModPhys.68.13.
[48] U. Schollwöck, Rev. Modern Phys. 77 (2005) 259–315, http://dx.doi.org/10.1103/RevModPhys.77.259.
[49] J. Eisert, M. Cramer, M.B. Plenio, Rev. Modern Phys. 82 (2010) 277–306, http://dx.doi.org/10.1103/RevModPhys.82.277.
[50] H.N. Phien, G. Vidal, I.P. McCulloch, Phys. Rev. B 86 (2012) 245107, http://dx.doi.org/10.1103/PhysRevB.86.245107.
[51] M. Binder, T. Barthel, Phys. Rev. B 92 (2015) 125119, http://dx.doi.org/10.1103/PhysRevB.92.125119.
[52] I.P. McCulloch, M. Gulácsi, Europhys. Lett. 57 (6) (2002) 852, http://dx.doi.org/10.1209/epl/i2002-00393-0.
[53] I.P. McCulloch, Collective Phenomena in Strongly Correlated Electron Systems (Ph.D. thesis), Australian National University, 2002, http://physics.uq.edu.au/people/ianmcc/aachen/public_html/thesis.pdf.
[54] S. Singh, R.N.C. Pfeifer, G. Vidal, Phys. Rev. A 82 (2010) 050301, http://dx.doi.org/10.1103/PhysRevA.82.050301.
[55] S. Singh, R.N.C. Pfeifer, G. Vidal, Phys. Rev. B 83 (2011) 115125, http://dx.doi.org/10.1103/PhysRevB.83.115125.
[56] S. Singh, G. Vidal, Phys. Rev. B 86 (2012) 195114, http://dx.doi.org/10.1103/PhysRevB.86.195114.
[57] A. Weichselbaum, Ann. Phys. 327 (12) (2012) 2972–3047, http://dx.doi.org/10.1016/j.aop.2012.07.009.
[58] C. Hubig, Sci. Post Phys. 5 (2018) 47, http://dx.doi.org/10.21468/SciPostPhys.5.5.047.
[59] G.K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, S.R. White, J. Chem. Phys. 145 (1) (2016) 014102, http://dx.doi.org/10.1063/1.4955108.

[60] C. Hubig, I.P. McCulloch, U. Schollwöck, Phys. Rev. B 95 (2017) 035129, http://dx.doi.org/10.1103/PhysRevB.95.035129.
[61] S. Paeckel, T. Köhler, S.R. Manmana, Sci. Post Phys. 3 (2017) 035, http://dx.doi.org/10.21468/SciPostPhys.3.5.035.
[62] I. Peschel, X. Wang, M. Kaulke, K. Hallberg (Eds.), Density Matrix Renormalization - a New Numerical Method in Physics, Springer Verlag, Berlin, 1999.
[63] G.M. Crosswhite, D. Bacon, Phys. Rev. A 78 (2008) 012356, http://dx.doi.org/10.1103/PhysRevA.78.012356.
[64] E.M. Stoudenmire, S.R. White, New J. Phys. 12 (5) (2010) 055026, http://dx.doi.org/10.1088/1367-2630/12/5/055026.
[65] U. Fano, Rev. Modern Phys. 29 (1957) 74–93, http://dx.doi.org/10.1103/RevModPhys.29.74.
[66] A.E. Feiguin, S.R. White, Phys. Rev. B 72 (2005) 220401, http://dx.doi.org/10.1103/PhysRevB.72.220401.
[67] T. Barthel, Phys. Rev. B 94 (2016) 115157, http://dx.doi.org/10.1103/PhysRevB.94.115157.
[68] S.M. Barnett, B.J. Dalton, J. Phys. A: Math. Gen. 20 (2) (1987) 411, http://dx.doi.org/10.1088/0305-4470/20/2/026.
[69] C. Karrasch, J.H. Bardarson, J.E. Moore, Phys. Rev. Lett. 108 (2012) 227206, http://dx.doi.org/10.1103/PhysRevLett.108.227206.
[70] C. Karrasch, J.H. Bardarson, J.E. Moore, New J. Phys. 15 (8) (2013) 083031.
[71] T. Barthel, New J. Phys. 15 (7) (2013) 073010, http://dx.doi.org/10.1088/1367-2630/15/7/073010.
[72] J. Hauschild, E. Leviatan, J.H. Bardarson, E. Altman, M.P. Zaletel, F. Pollmann, Phys. Rev. B 98 (2018) 235163, http://dx.doi.org/10.1103/PhysRevB.98.235163.
[73] E.M. Stoudenmire, MPO-MPS Multiplication: Density Matrix Algorithm http://tensornetwork.org/mps/algorithms/denmat_mpo_mps/.
[74] C.D. White, M. Zaletel, R.S.K. Mong, G. Refael, Phys. Rev. B 97 (2018) 035127, http://dx.doi.org/10.1103/PhysRevB.97.035127.
[75] C.C. Paige, B.N. Parlett, H.A. van der Vorst, Numer. Linear Algebra Appl. 2 (2) (1995) 115–133, http://dx.doi.org/10.1002/nla.1680020205.
[76] R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, H. van der Vorst, Soc. Ind. Appl. Math. (1994) http://dx.doi.org/10.1137/1.9781611971538.
[77] M. Hochbruck, C. Lubich, SIAM J. Numer. Anal. 34 (5) (1997) 1911–1925, http://dx.doi.org/10.1137/s0036142995280572.
[78] T. Barthel, Y. Zhang, Optimized Lie-Trotter-Suzuki decompositions for two and three non-commuting terms http://arxiv.org/abs/1901.04974.
[79] T. Hashizume, I.P. McCulloch, J.C. Halimeh, Dynamical phase transitions in the two-dimensional transverse-field ising model http://arxiv.org/abs/1811.09275.
[80] M. Urbanek, P. Soldán, Parallel implementation of the time-evolving block decimation algorithm for the bose–hubbard model 199 (2016) 170–177. http://dx.doi.org/10.1016/j.cpc.2015.10.016.
[81] C. Hubig, I.P. McCulloch, U. Schollwöck, F.A. Wolf, Phys. Rev. B 91 (2015) 155115, http://dx.doi.org/10.1103/PhysRevB.91.155115.
[82] T. Koffel, M. Lewenstein, L. Tagliacozzo, Phys. Rev. Lett. 109 (2012) 267203, http://dx.doi.org/10.1103/PhysRevLett.109.267203.
[83] P. Hauke, L. Tagliacozzo, Phys. Rev. Lett. 111 (2013) 207202, http://dx.doi.org/10.1103/PhysRevLett.111.207202.
[84] S. Goto, I. Danshita, Performance of the time-dependent variational principle for matrix product states in long time evolution of a pure state http://arxiv.org/abs/1809.01400.
[85] E. Leviatan, F. Pollmann, J.H. Bardarson, D.A. Huse, E. Altman, Quantum thermalization dynamics with Matrix-Product States http://arxiv.org/abs/1702.08894.
[86] C. Hubig, J. Haegeman, U. Schollwöck, Phys. Rev. B 97 (2018) 045125, http://dx.doi.org/10.1103/PhysRevB.97.045125.
[87] B. Kloss, Y.B. Lev, D. Reichman, Phys. Rev. B 97 (2018) 024307, http://dx.doi.org/10.1103/PhysRevB.97.024307.
[88] M.J. Hartmann, J. Prior, S.R. Clark, M.B. Plenio, Phys. Rev. Lett. 102 (2009) 057202, http://dx.doi.org/10.1103/PhysRevLett.102.057202.
[89] T. Prosen, M. Žnidarič, J. Stat. Mech. Theory Exp. 2009 (02) (2009) P02035.
[90] D. Kennes, C. Karrasch, Comput. Phys. Comm. 200 (2016) 37–43, http://dx.doi.org/10.1016/j.cpc.2015.10.019.
[91] P. Vaidyanathan, The Theory of Linear Prediction, in: Synthesis Lectures on Engineering Series, Morgan & Claypool, 2008.
[92] S.R. White, I. Affleck, Phys. Rev. B 77 (2008) 134437, http://dx.doi.org/10.1103/PhysRevB.77.134437.
[93] T. Barthel, U. Schollwöck, S.R. White, Phys. Rev. B 79 (2009) 245101, http://dx.doi.org/10.1103/PhysRevB.79.245101.
[94] A. Bohrdt, C.B. Mendl, M. Endres, M. Knap, New J. Phys. 19 (6) (2017) 063001, http://dx.doi.org/10.1088/1367-2630/aa719b.
[95] C. Zhang, E. Jeckelmann, S.R. White, Phys. Rev. Lett. 80 (1998) 2661–2664, http://dx.doi.org/10.1103/PhysRevLett.80.2661.
[96] C. Guo, A. Weichselbaum, J. von Delft, M. Vojta, Phys. Rev. Lett. 108 (2012) 160401, http://dx.doi.org/10.1103/PhysRevLett.108.160401.
[97] C. Brockt, F. Dorfner, L. Vidmar, F. Heidrich-Meisner, E. Jeckelmann, Phys. Rev. B 92 (2015) 241106, http://dx.doi.org/10.1103/PhysRevB.92.241106.
[98] F.A.Y.N. Schröder, A.W. Chin, Phys. Rev. B 93 (2016) 075105, http://dx.doi.org/10.1103/PhysRevB.93.075105.
[99] F. Dorfner, Numerical Methods for Strongly Correlated Many-Body Systems with Bosonic Degrees of Freedom (Ph.D. thesis), LMU Munich, 2017, http://nbn-resolving.de/urn:nbn:de:bvb:19-205528.
[100] B. Bruognolo, Z. Zhu, S.R. White, E.M. Stoudenmire, Matrix product state techniques for two-dimensional systems at finite temperature http://arxiv.org/abs/1705.05578.

[101] X. Chen, T. Zhou, D.A. Huse, E. Fradkin, Ann. Phys. 529 (7) (2017) 160033, http://dx.doi.org/10.1002/andp.201600332, arXiv:1610.00220.
[102] C.W. von Keyserlingk, T. Rakovszky, F. Pollmann, S.L. Sondhi, Phys. Rev. X 8 (2018) 021013, http://dx.doi.org/10.1103/PhysRevX.8.021013.
[103] T. Rakovszky, F. Pollmann, C.W. von Keyserlingk, Phys. Rev. X 8 (2018) 031058, http://dx.doi.org/10.1103/PhysRevX.8.031058.
[104] K. Hémery, F. Pollmann, D.J. Luitz, Tensor network approaches to operator spreading in ergodic quantum systems, in prep.
[105] J. Haegeman, B. Pirvu, D.J. Weir, J.I. Cirac, T.J. Osborne, H. Verschelde, F. Verstraete, Phys. Rev. B 85 (2012) 100408, http://dx.doi.org/10.1103/PhysRevB.85.100408.
[106] J. Haegeman, T.J. Osborne, F. Verstraete, Phys. Rev. B 88 (2013) 075133, http://dx.doi.org/10.1103/PhysRevB.88.075133.
[107] L. Vanderstraeten, Tensor Network States and Effective Particles for Low-Dimensional Quantum Spin Systems (Ph.D. thesis), University of Ghent, 2017, http://dx.doi.org/10.1007/978-3-319-64191-1_4.
[108] L. Vanderstraeten, J. Haegeman, F. Verstraete, Simulating excitation spectra with projected entangled-pair states http://arxiv.org/abs/1809.06747.
[109] L. Vanderstraeten, J. Haegeman, F. Verstraete, Sci. Post Phys. Lect. Notes (2019) 7, http://dx.doi.org/10.21468/SciPostPhysLectNotes.7.
[110] S.M. Greene, V.S. Batista, J. Chem. Theory Comput. 13 (9) (2017) 4034–4042, http://dx.doi.org/10.1021/acs.jctc.7b00608, pMID: 28763215.