Unifying time evolution and optimization with matrix product states

Jutho Haegeman,[1] Christian Lubich,[2] Ivan Oseledets,[3,4] Bart Vandereycken,[5,*] and Frank Verstraete[1,6]

[1]*Department of Physics and Astronomy, University of Ghent, Krijgslaan 281 S9, B-9000 Ghent, Belgium*

[2]*Mathematisches Institut, Universität Tübingen, Auf der Morgenstelle 10, D-72076 Tübingen, Germany*

[3]*Skolkovo Institute of Science and Technology, Novaya St. 100, 143025 Skolkovo, Russia*

[4]*Institute of Numerical Mathematics, Russian Academy of Sciences, Gubkina Street, 8, Moscow, Russia*

[5]*Department of Mathematics, Princeton University, Fine Hall, Princeton, New Jersey 08544, USA*

[6]*Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Wien, Austria*

We show that the time-dependent variational principle provides a unifying framework for time-evolution methods and optimization methods in the context of matrix product states. In particular, we introduce a new integration scheme for studying time evolution, which can cope with arbitrary Hamiltonians, including those with long-range interactions. Rather than a Suzuki-Trotter splitting of the Hamiltonian, which is the idea behind the adaptive time-dependent density matrix renormalization group method or time-evolving block decimation, our method is based on splitting the projector onto the matrix product state tangent space as it appears in the Dirac-Frenkel time-dependent variational principle. We discuss how the resulting algorithm resembles the density matrix renormalization group (DMRG) algorithm for finding ground states so closely that it can be implemented by changing just a few lines of code and it inherits the same stability and efficiency. In particular, our method is compatible with any Hamiltonian for which ground-state DMRG can be implemented efficiently. In fact, DMRG is obtained as a special case of our scheme for imaginary time evolution with infinite time step.

I. INTRODUCTION

Tensor network states, and in particular matrix product states (MPS) [1–4], have become increasingly popular and successful for the description of strongly interacting quantum many body systems. While originating from the efficient and robust density matrix renormalization group (DMRG) algorithm [5,6] for finding ground states of one-dimensional (and quasi two-dimensional) spin systems, an undeniable influence for this success has also been the ability to accurately study dynamical properties using time-evolution methods such as the time-evolving block decimation (TEBD) method [7–10]. The reformulation of DMRG in terms of MPS [2,11–15] has been of key importance in this development. Being able to probe and understand quantum dynamics is becoming increasingly valuable for experiments at low temperature or high energy, or for theoretical questions such as thermalization. Nevertheless, quantum dynamics is mostly inaccessible to alternative methods such as Monte Carlo sampling of the partition function.

The key ingredient of TEBD-based methods is a Lie-Trotter-Suzuki [16,17] splitting of the Schrödinger equation according to the individual terms $\hat{h}_i$ of the Hamiltonian $\hat{H} = \sum_i \hat{h}_i$. When these terms are local, they can be applied efficiently and the tensor network can be updated accordingly. The resulting increase of the virtual dimensions of the tensors can be countered by a subsequent truncation step, although the growth of entanglement entropy under time evolution indicates that a net increase of the dimensions is inevitable if an accurate description is required. Unfortunately, the TEBD idea is not applicable to long-range interactions, which appear when using MPS for quasi-2D systems or quantum chemistry applications.

An alternative idea for time evolution, the Dirac-Frenkel time-dependent variational principle (TDVP), was formulated only recently for the variational class of MPS [18,19]. The key ingredient is to project the right-hand side of the Schrödinger equation, $\hat{H} |\Psi\rangle$, onto the tangent space, so that the evolution never leaves the manifold. This approach is independent of the Hamiltonian and can be implemented efficiently for long-range Hamiltonians. The TDVP equations define a simultaneous update for all MPS tensors and, as a consequence thereof, its original implementation depends on inverses of matrices conditioned by the Schmidt coefficients, similar to the original TEBD implementation. This results in the paradoxical situation that the stability of this method deteriorates as the approximation of the exact state (corresponding to the value of the smallest Schmidt coefficient) is improved. More concretely, the TDVP equations suffer from what is known as stiffness in the numerical analysis literature [20] and this results in a severe stepsize restriction when using standard explicit time integrators. Furthermore, the TDVP algorithm has so far not gained widespread acceptance and was in fact criticised to be complex and distinct from DMRG-inspired algorithms in Ref. [21], where another algorithm for time evolution with long-range interactions is proposed, based on an approximation of the evolution operator in terms of a matrix product operator (MPO) [22,23].

This article overcomes such criticism by presenting an alternative integration scheme for the TDVP equations for finite MPS, based on a Lie-Trotter splitting of the tangent space projector. The resulting algorithm resembles DMRG so closely that it can be obtained by merely modifying two lines of code, and the one-site DMRG method is obtained as special case of imaginary time evolution with an infinite time step.

In particular, the time evolution algorithm presented below is compatible with any Hamiltonian for which ground-state DMRG can be executed efficiently, e.g., long-range

---

*Present address: Department of Mathematics, University of Geneva, 2-4 rue du Lièvre, 1211 Geneva, Switzerland.

Hamiltonians, which are written in the form of MPOs. By construction, the TDVP projects the evolution onto the manifold of MPS with fixed bond dimension and avoids the need for a truncation step. As this also has the disadvantage that the bond dimension cannot be dynamically altered in the standard scheme (similar to one-site DMRG), we also present an analogous two-site scheme that does enable a dynamical increase of the bond dimension. We furthermore relate this two-site approach to the more conventional time-dependent DMRG or TEBD methods, which are restricted to local (typically nearest-neighbor) interactions. The resulting ideas will also be relevant for continuous MPS (cMPS) [24,25], which seem incompatible with traditional DMRG- and TEBD-based approaches.

In the next section, we set notation by introducing the definition of a MPS and discussing its various canonical forms. Section III discusses the TDVP and in particular the projector onto the MPS tangent space, whose peculiar form lies at the basis of our new algorithm. The one-site integrator is presented in Sec. IV, together with a discussion of how the one-site DMRG algorithm is obtained as a particular limit. Section V discusses the extension to a two-site algorithm and how this relates to the TEBD algorithm in the special case of nearest neighbor Hamiltonians. We then provide an explicit example that is related to a recent experiment in Sec. VI and finally formulate some conclusions. The appendix provides a detailed derivation of the tangent space projector discusses in Sec. III as well as a pseudocode for implementing the one-site integration scheme of Sec. IV. DMRG practitioners will immediately appreciate the close resemblance to the one-site finite-size DMRG algorithm formulated in terms of MPS.

II. MATRIX PRODUCT STATES AND CANONICAL FORMS

We define a matrix product state (MPS) with open boundary conditions, also known as a tensor train [26] in the numerical mathematics community, as

$$|\psi[A]\rangle = \sum_{\{s_n\}=1}^{d} A^{s_1}(1) A^{s_2}(2) \cdots A^{s_N}(N) |s_1 s_2 \ldots s_N\rangle,$$

where we (ab)used square bracket functional notation [ ] to denote the dependence on *a set of* site-dependent matrices $A^s(n)$ having site-dependent dimensions $D_{n-1} \times D_n$. As the whole

matrix product has to evaluate to a scalar, we have $D_0 = D_N = 1$. While the dimension d of the local Hilbert space can also be site-dependent, this is not explicitly denoted for the sake of brevity. We employ the popular graphical notation of the individual tensors A and the resulting tensor network state in Fig. 1.

A key ingredient is the observation that the physical state $|\Psi[A]\rangle$ is unchanged under the gauge transformation $A^s(n) \mapsto A^s_G(n) = G(n-1)^{-1} A^s(n) G(n)$. This gauge freedom can be used to impose certain canonical forms on the MPS representation, in terms of left-orthonormal matrices $A^s_L(n)$ or right-orthonormal matrices $A^s_R(n)$ [see Figs. 1(b)–1(e)]. The stability of finite MPS methods originates from the fact that we can transform the original tensors $A^s(n)$ into, e.g., the left orthornormal representation without explicit computation of the inverses $G(n)^{-1}$. Starting from $C(0) = 1$, an orthogonal factorization (e.g., QR-decomposition) allows to write $C(n-1)A^s(n) = A^s_L(n)C(n)$. The final $C(N)$ is a scalar representing the norm of the state, and can be discarded. If the original MPS was already in the right canonical form, the singular values of $C(n)$ correspond to the Schmidt coefficients of $|\Psi\rangle$ for a bipartition of the lattice in $[1:n]$ and $[n+1:N]$. Defining the one-site center block $A^s_C(n) = C(n-1)A^s_R(n) = A^s_L(n)C(n)$ allows to write the state in the mixed canonical form [see also Fig. 1(g)]

$$|\Psi\rangle = \sum_{\alpha, s_n, \beta} \left[A^{s_n}_C(n) \right]_{\alpha,\beta} \left| \Phi^{[1:n-1]}_{L,\alpha} \right\rangle |s_n\rangle \left| \Phi^{[n+1:N]}_{R,\beta} \right\rangle, \quad (1)$$

where new states $\left| \Phi^{[1:n-1]}_{L,\alpha} \right\rangle$ ($\left| \Phi^{[n+1:N]}_{R,\beta} \right\rangle$) constitute an orthonormal basis for the left (right) block of the lattice.

To further introduce the required concepts, we first summarize the finite-size one-site DMRG method algorithm in the MPS framework and refer the excellent review of Ref. [15] for further details. The one-site DMRG method corresponds to an alternating least squares optimization and amounts to sequentially optimizing individual tensors in a number of sweeps until the whole process converges. The individual optimization steps are homogeneous problems and give rise to a generalized eigenvalue equation. By gauging the MPS such that the current tensor being optimized is the center site $A^s_C(n)$, this is transformed into a standard eigenvalue problem with an effective one-site Hamiltonian $H(n)$ defined in Fig. 2(a). The matrix vector product of $H(n)$ can be computed efficiently
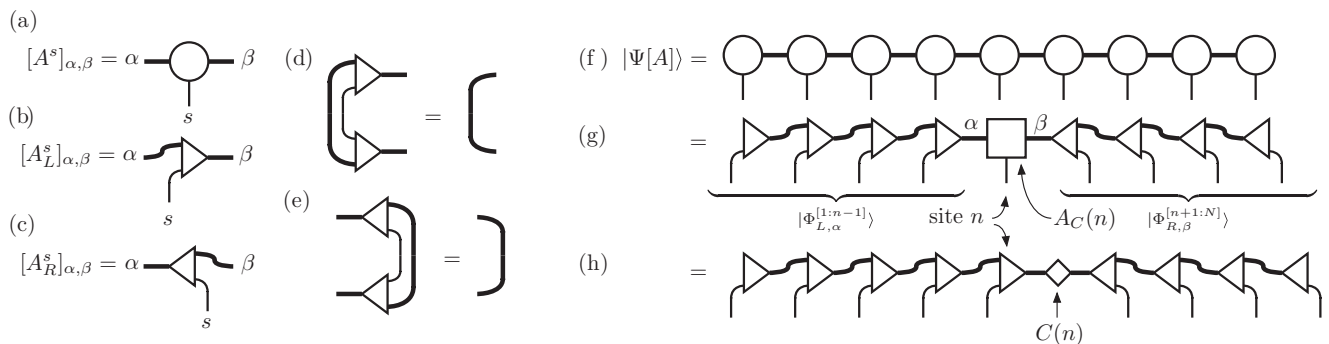

FIG. 1. An arbitrary MPS tensor is represented as a shape with fat legs corresponding to the virtual indices of dimension D and a normal leg corresponding to the physical index of dimension d (a). The left-orthonormal [right-orthonormal] MPS tensors are represented using triangles (b) [(c)] and satisfies the condition (d) [(e)]. An MPS $|\Psi[A]\rangle$ using general tensors (f) can be brought into a mixed-canonical form with a one-site center $A_C(n)$ at site n (g) or even a zero-site center $C(n)$ (h), whose singular values correspond to the Schmidt coefficients of the state.
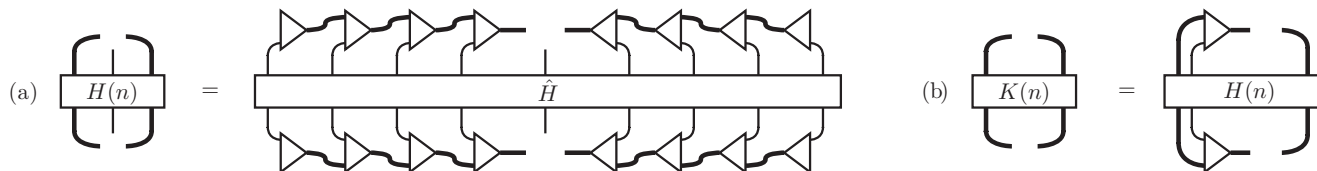
FIG. 2. (a) The one-site effective Hamiltonian $H(n)$ plays a central role in DMRG and the current TDVP integrator, and can be efficiently computer when $\hat{H}$ is, for example, represented as a matrix product operator. (b) Similarly, one can define a zero-site effective Hamiltonian $K(n)$, which can easily be computed from $H(n)$.

for typical physical Hamiltonians having either short-ranged interactions or an MPO form, so that an iterative eigensolver can be used. After updating $A_C^s(n)$ as the lowest eigenvector of $H(n)$, one performs the orthogonal factorization to shift the center site to the left or right, so that the next tensor can be optimized. Reaching one end of the chain, one can then sweep back, so that every orthogonalization step is immediately followed by an optimization step. For further reference, we also define a zero-site effective Hamiltonian $K(n)$ [see Fig. 2(b)] for the degrees of freedom $C(n)$ on the virtual bond between sites n and $n + 1$, as in [Fig. 1(h)]

$$|\Psi\rangle = \sum_{\alpha,\beta} [C(n)]_{\alpha,\beta} \left| \Phi_{L,\alpha}^{[1:n]} \right\rangle \left| \Phi_{R,\beta}^{[n+1:N]} \right\rangle. \quad (2)$$

While $K(n)$ serves no purpose in ground-state DMRG, it will appear in our time evolution algorithm below.

III. TANGENT SPACE PROJECTOR AND TIME-DEPENDENT VARIATIONAL PRINCIPLE

For the purpose of this paper, the TDVP can most easily be understood in a geometric fashion, namely as an orthogonal projection of the evolution vector of the Schrödinger equation $(-iH|\Psi\rangle)$ onto the tangent space of the MPS manifold $\mathcal{M}_{\text{MPS}}$ at the current state

$$\frac{d|\Psi(A)\rangle}{dt} = -i\hat{P}_{T_{|\Psi(A)\rangle}\mathcal{M}_{\text{MPS}}} \hat{H} |\Psi(A)\rangle. \quad (3)$$

By construction, the resulting evolution will never leave the MPS manifold and can therefore be described by time-evolving parameters as $|\Psi(A(t))\rangle$. Indeed, the TDVP can be recast into a complicated set of coupled nonlinear equations for all parameters in the MPS. The algorithm presented in Ref. [18] used the explicit Euler integration scheme to simultaneously update all parameters, which was only possible by using explicit inverses of matrices containing very small singular values.

Here we follow a different approach. Equation (3) explicitly illustrates the central importance of the tangent space projector $\hat{P}_{T_{|\Psi(A)\rangle}}$, which we therefore study in more detail. The tangent
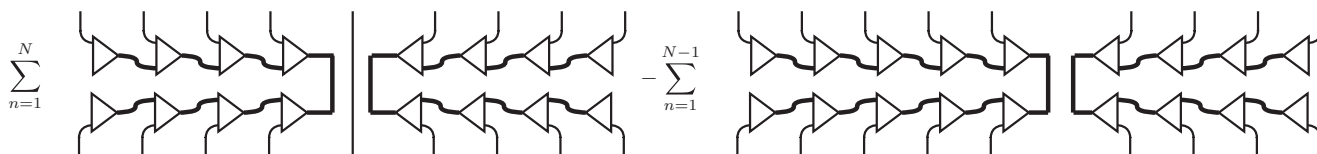
space projector is a generic concept that is independent of the TDVP and can be applied to arbitrary states $|\Theta\rangle$ in the Hilbert space $\mathbb{H}$. The result $\hat{P}_{T_{|\Psi(A)\rangle}}|\Theta\rangle$ is an MPS tangent vector $|\Phi\rangle$ that can be characterized as $\langle\Phi'|\Phi\rangle = \langle\Phi'|\Theta\rangle$ for any other tangent vector $|\Phi'\rangle$. As calculated in Appendix A (see also Ref. [27]), the tangent space projector can be decomposed as

$$\hat{P}_{T_{|\Psi(A)\rangle}\mathcal{M}_{\text{MPS}}} = \sum_{n=1}^{N} \hat{P}_L^{[1:n-1]} \otimes \hat{1}_n \otimes \hat{P}_R^{[n+1:N]}$$

$$- \sum_{n=1}^{N-1} \hat{P}_L^{[1:n]} \otimes \hat{P}_R^{[n+1:N]}, \quad (4)$$

where

$$\hat{P}_L^{[1:n]} = \sum_{\alpha=1}^{D} \left| \Phi_{L,\alpha}^{[1:n]} \right\rangle \left\langle \Phi_{L,\alpha}^{[1:n]} \right|, \quad (5)$$

$$\hat{P}_R^{[n:N]} = \sum_{\beta=1}^{D} \left| \Phi_{R,\beta}^{[n:N]} \right\rangle \left\langle \Phi_{R,\beta}^{[n:N]} \right|. \quad (6)$$

These operators are independent of the unitary gauge freedom (basis rotation) that exists in the definition of $|\Phi_{L,\alpha}^{[1:n]}\rangle$ and $|\Phi_{R,\beta}^{[n:N]}\rangle$ and correspond to orthogonal projectors onto the support of the reduced density matrix ρ of the state $|\Psi[A]\rangle$ in the corresponding regions. It is represented graphically in Fig. 3.

Upon applying $\hat{P}_{T_{|\Psi(A)\rangle}\mathcal{M}_{\text{MPS}}}$ onto $\hat{H}|\Psi[A]\rangle$, we can write $|\Psi[A]\rangle$ in the form of Eq. (1) or (2) with a properly chosen location for the center such that we can recognize the one-site or zero-site effective Hamiltonian $H(n)$ or $K(n)$. The result is depicted in Fig. 4

IV. ONE-SITE INTEGRATION SCHEME

The crucial insight leading to our new integration scheme is that, while the TDVP equation (3) with the right hand side of Fig. 4 is an equation in the full Hilbert space, the differential equation obtained for a single term of the right hand side

$$\hat{P}_{T_{|\Psi[A]\rangle}\mathcal{M}_{\text{MPS}}} =$$

FIG. 3. Right-hand side (up to the factor $-i$) of the TDVP equation [Eq. (3)].
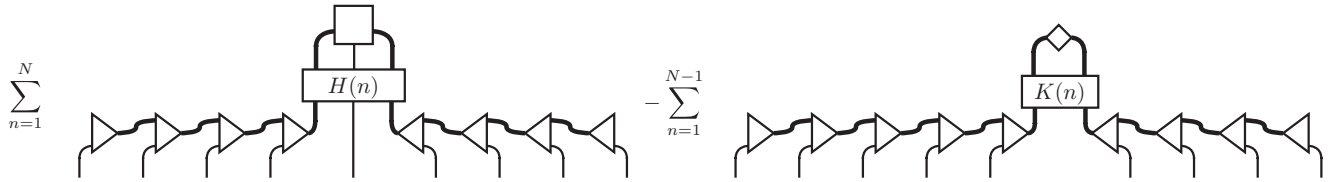
FIG. 4. Right-hand side (up to the factor $-i$) of the TDVP equation [Eq. (3)].

[corresponding to a single term from Eq. (4)] can be integrated exactly [27].

For example, using the single projector $\hat{P}_L^{[1:n-1]} \otimes \hat{1}_n \otimes \hat{P}_R^{[n+1:N]}$ and writing $|\Psi[A]\rangle$ as in Eq. (1) with center site n, the exact solution can be obtained by only making $A_C(n)$ time-dependent and letting it satisfy $\dot{A}_C(n,t) = -\mathrm{i}H(n)A_C(n,t)$ where a bold notation is used for the vector representation of $A_C(n)$. We can thus explicitly integrate this differential equation by only making $A_C(n)$ time dependent as

$$A_C(n,t) = \exp\left[-\mathrm{i}H(n)t\right]A_C(n,0). \tag{7}$$

Similarly, a projector term of the form $-\hat{P}_L^{[1:n]} \otimes \hat{P}_R^{[n+1:N]}$ can be integrated explicitly by writing the $|\Psi\rangle$ as in Eq. (2) and making $C(n)$ time-dependent as

$$C(n,t) = \exp\left[+\mathrm{i}K(n)t\right]C(n,0). \tag{8}$$

Note that opposite sign in the exponentials of Eqs. (7) and (8). The evolution of C in Eq. (8) can be interpreted as an evolution backwards in time. Analogous to the eigenvalue problem in DMRG, both Eqs. (7) and (8) can be integrated using a Lanczos scheme [28] in order to keep the computational cost at $\mathcal{O}(D^3)$. This is a well-controlled approximation for which the error can be made arbitrarily small. We refer to Ref. [29] for estimators for the error resulting from the Lanczos exponentiation.

For this type of differential equation consisting of a sum of integrable parts, it is very natural to use a Lie-Trotter splitting approach [16,29,30], where one evolves according to each integrable part for a small time Δt. There is a natural order similar to the DRMG sweep, which optimizes the computational overhead (and results in robustness in the case of overapproximation [27]). For a right orthogonal MPS, start at $n = 1$ and repeat the following steps: Evolve $A_C(n)$ according Eq. (7) for a time step Δt. Factorize the updated $A_C^s(n) = A_L^s(n)C(n)$. Evolve $C(n)$ backwards in time according to Eq. (8) before absorbing it into the next site to create $A_C^s(n+1) = C(n)A_R^s(n+1)$. Having a Lanczos routine for computing the matrix exponential acting on a vector (e.g., Ref. [31]), this algorithm differs from one-site DMRG by replacing the optimization step for $A_C(n)$ (eigenvalue solver) with the evolution step, and by adding an extra evolution step for $C(n)$ before absorbing it at the next site. Completing a single left-to-right sweep is a first-order integrator that produces an updated $|\Psi\rangle$ at time $t + \Delta t$ with a local integration error of order $\mathcal{O}(\Delta t^2)$. Completing the right-to-left sweep is equivalent to composing this integrator with its adjoint, resulting in a second-order symmetric method [32] so that the state at time $t + 2\Delta t$ has a more favourable error of order $\mathcal{O}(\Delta t^3)$. It is thus natural to set $\Delta t \rightarrow \Delta t/2$ and to define the complete sweep (left and right) as a single integration step. A pseudocode implementation of this algorithm is presented in Appendix B.

It is also possible to obtain higher order integrators by applying composition schemes to this symmetric integrator [30]. Note that the finite time step errors correspond to errors with respect to the exact solution of the TDVP differential equation. The error with respect to the original Schrödinger equation also receives a contribution from the discrepancy between the TDVP evolution and the Schrödinger evolution. As explained in Ref. [29], this source of error can be bound in terms of

$$\|(\hat{1} - \hat{P}_{T_{|\Psi[A]\rangle}\mathcal{M}_{\mathrm{MPS}}})\hat{H}\,|\Psi[A]\rangle\,\|,$$

a quantity that can easily be evaluated [33] and monitored throughout the evolution. To prevent this error from growing too large, the bond dimension can be dynamically increased using the two-site algorithm that is presented in the next section.

Let us conclude this section with a brief discussion of the properties of this algorithm. The proposed integration scheme contains evolution for the one-site and zero-site center blocks $A_C(n)$ and $C(n)$ according to the respective effective Hamiltonian matrices $H(n)$ or $K(n)$, which are Hermitian. Therefore the resulting evolution will have exact norm and energy conservation (in case of a time-independent Hamiltonian), up to the errors in the iterative algorithm used to solve the local problem (which are typically close to machine precision and therefore negligible). This integration scheme can also be combined with imaginary time evolution $t \mapsto -\mathrm{i}\tau$, e.g., for constructing (purifications of) thermal states or for finding ground states. In the latter case, the goal is not to accurately capture the full evolution and the scheme can be varied [34]. In the particular case of $\Delta\tau \rightarrow \infty$, the evolution of $A_C(n)$ will actually force it to be the lowest eigenvector of $H(n)$, exactly as in the optimization step of the one-site DMRG algorithm. One can easily verify that the $C(n)$ that results from then factorizing $A_C(n) = A_L(n)C(n)$ is then also an exact eigenstate of $K(n)$, so that its backwards evolution has no further effect and can thus be omitted. At this point, our algorithm has become identical to variational MPS formulation of the one-site finite-size DMRG algorithm.

V. TWO-SITE ALGORITHM

As is well known from DRMG simulations, it is often advantageous two work with a two-site algorithm where a block of two successive sites is updated simultaneously. This block will then have to be factorized into two individual tensors using a singular value decomposition, where a new bond dimension can be chosen depending on the singular values (Schmidt coefficients) of the updated state. This allows to dynamically adapt the bond dimension to the a predefined precision or error criterion, and also allows to dynamically

introduce new charge sectors (irreducible representations) when symmetries of the quantum state are explicitly encoded in the MPS representation. As two MPS with different bond dimensions do not belong to the same manifold [35], this idea cannot easily be captured in a smooth evolution described using a differential equation. However, like most numerical integration schemes, the aforementioned algorithm is intrinsically discrete by choosing a time step, and it poses no problem to formulate an analogous two-site algorithm. Thereto, we replace the tangent space projector in Eq. (3) by a projector onto the linear space of two-site variations, which is spanned by the states

$$\sum_{n=1}^{N-1} \sum_{\{s_n\}=1}^{d} A^{s_1}(1) \cdots A^{s_{n-1}}(n-1) B^{s_n s_{n+1}}(n:n+1) A^{s_{n+2}}$$
$$\times (n+2) \cdots A^{s_N}(N) | s_1 \ldots s_n \ldots s_N \rangle.$$

It is now tempting to propose the following two-site integration scheme: evolve a two-site center block $A_C(n:n+1)$ according to its effective Hamiltonian $H(n:n+1)$, factor it into $A_C^{s_n,s_{n+1}}(n:n+1) \rightarrow A_L^{s_n}(n)A_C^{s_{n+1}}(n+1)$ and evolve $A_C^{s_{n+1}}(n+1)$ backwards in time according to $H(n+1)$ before absorbing it in the next two-site block $A_C^{s_{n+1}}(n+1)A_C^{s_{n+2}}(n+2) \rightarrow A_C^{s_{n+1},s_{n+2}}(n+1:n+2)$. Indeed, we validate this approach in Appendix C.

Every two-site block that is factored can increase the local bond dimension. This requires a local truncation based on the singular values in every step or a global approximation with the best MPS with smaller bond dimensions at the end of the sweep [10]. It is straightforward to show that imaginary time evolution with a time step $\Delta\tau \rightarrow \infty$ gives rise to two-site DMRG. The TEBD algorithm also fits in this framework, by observing that for nearest-neighbor interactions, the evolution vector $\hat{H} |\Psi\rangle$ is contained exactly in the space of two-site variations (no projection is necessary) by choosing, e.g., $B^{s_n,s_{n+1}}(n:n+1) = \sum_{t_n,t_{n+1}} \langle s_n, s_{n+1} | \hat{h}_{n,n+1} | t_n, t_{n+1} \rangle A^{t_n}(n)A^{t_{n+1}}(n+1)$ and using a Trotter splitting based on these terms. The representation $B^{s_n,s_{n+1}}(n:n+1)$ is, however, nonunique, and our algorithm based on splitting the projector (even when it acts trivially) gives rise to a slightly different scheme.

VI. EXAMPLE

As an application, we study a one-dimensional model with power-law decay of interactions. There is recently a strong interest in such models as they are naturally realized in experiments with ultra-cold matter. It is thus actively investigated in what regimes they still exhibit a maximal (Lieb-Robinson) velocity [36] and, if not, whether they still exhibit a light cone [37,38]. We consider the particular case of the XY model, given by the Hamiltonian

$$H = \frac{1}{2} \sum_{i<j=1}^{N} \frac{J}{|i-j|^{\alpha}} \left(\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y \right) \qquad (9)$$

on a finite chain of N sites with open boundary conditions. This Hamiltonian was recently realized in an experiment with trapped $^{171}\text{Yb}^+$ ions [39].

In Ref. [38], the time-dependent expectation value $\sigma_r^x(t) = \exp(-iHt)\sigma_r^x \exp(iHt)$ was compared between a fully polarized product state $|\Psi_0\rangle = \otimes_n |\uparrow\rangle_n$ and the perturbed state $U |\Psi_0\rangle$ with $U = \exp(i\pi\sigma_0^y/4)$ as function of r and t for various values $\alpha > 1$. For this choice of $|\Psi_0\rangle$, the quantum many-body problem can be reduced to a single-particle problem. In accordance with the general result of that paper, there is a light cone and the leakage outside the light cone contains a contribution with exponential decay that is determined by the nearest-neighbor interactions and is thus independent of α, as well as a contribution with power-law decay that dominates at larger distances. We repeat this setup but using the fully correlated ground state of H as $|\Psi_0\rangle$, approximated as MPS. We simulate the time evolution with our two-site integrator using a fourth order composition method using a time step $dt = 0.02/J$ for an open chain with $N = 101$ sites (Fig. 5). The long-range interactions of the Hamiltonian were approximated by an MPO [23] with a maximal absolute error smaller than 10^{-8}. The resulting virtual dimensions of the MPO are 28, 28, 24, and 12 for the cases $\alpha = 0.75$, 1.5, 3.0 and 6.0, respectively. We, indeed, also consider a value of $\alpha = 0.75 < 1$, for which the results from Ref. [38] are not applicable. For $\alpha > 1$, we observe a light cone and the leakage outside is in perfect agreement with the theoretical predictions of Ref. [38]. The light cone itself seems to be linear for $\alpha = 3.0$ and $\alpha = 6.0$, but the shape is less clear for $\alpha = 1.5$. For $\alpha = 0.75$, there is no sharp wave front but it still seems to take a finite amount of time to signal over a finite distance, up to power law corrections.

VII. CONCLUSIONS

In conclusion, we have formulated a time integration scheme for the TDVP in the context of MPS, which is based on a Trotter decomposition of the tangent space projector rather than the Hamiltonian terms. This allows to simulate time evolution using an algorithm that is remarkably similar to DMRG and can deal with arbitrary long-range Hamiltonians. This approach generalizes neatly to tree tensor network states and will be equally relevant for continuous MPS [24,25], where algorithms based on a Trotter decomposition of the Hamiltonian terms are not applicable. The main ideas might also help in constructing more robust algorithms for simulations with projected entangled-pair states [40,41].

APPENDIX A: DERIVATION OF THE TANGENT SPACE PROJECTOR

To derive the main result of this paper, we have to introduce some further notation. We repeat the definition of a general
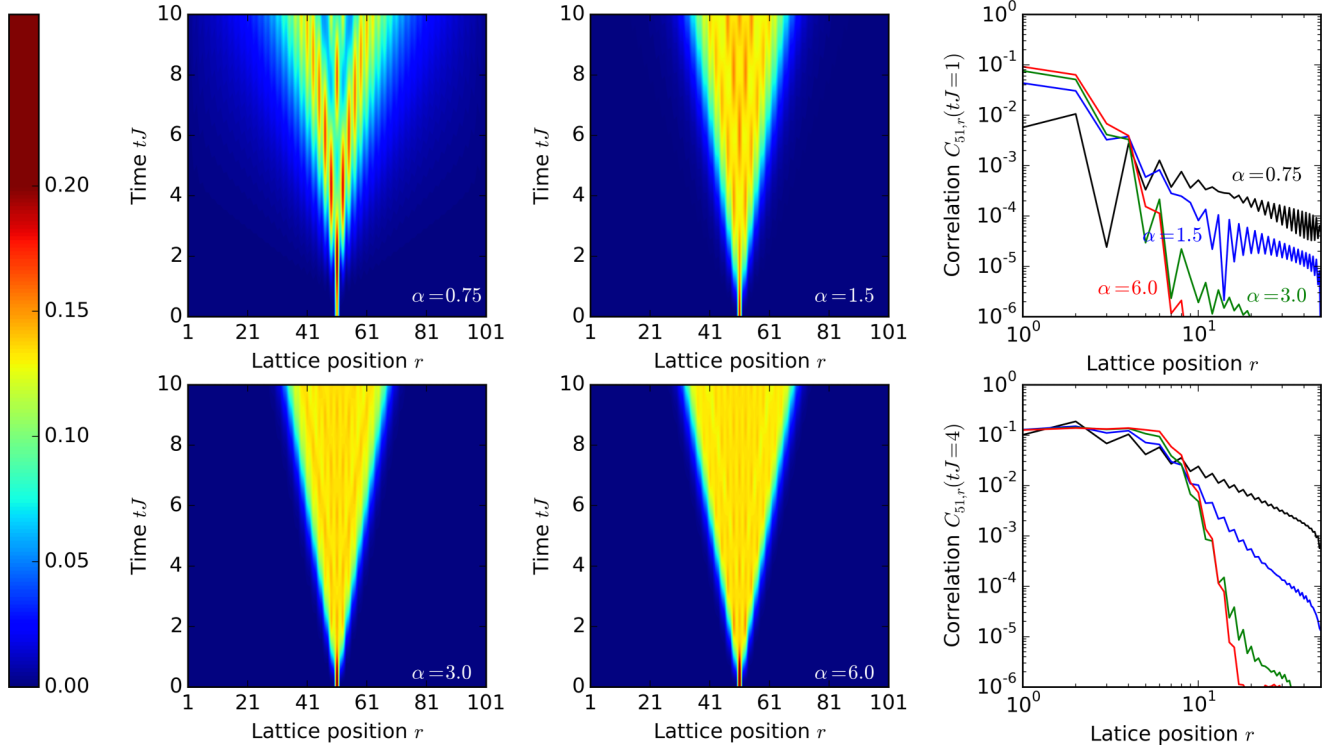
FIG. 5. Absolute value of $\langle \Psi_0 | U^\dagger \sigma_r^x(t) U | \Psi_0 \rangle - \langle \Psi_0 | \sigma_r^x(t) | \Psi_0 \rangle$ with H the XY Hamiltonian in Eq. (9), $|\Psi_0\rangle$ the ground state of H and $U = \exp(i\sigma_{51}^y \pi/4)$ on a chain of $N = 101$ spins, as function of r and t for various values of α.

matrix product state with open boundary conditions, given by

$$|\psi[A]\rangle = \sum_{\{s_n\}=1}^{d} A^{s_1}(1) A^{s_2}(2) \cdots A^{s_N}(N) |s_1 s_2 \ldots s_N\rangle . \quad (A1)$$

If we want to compute expectation values without bringing the MPS in a specific format, we can first define the sets of site-dependent $D_n \times D_n$ density matrices $l(n)$ and $r(n)$ (with $n = 0, \ldots, N$) for the auxiliary system through $l(0) = 1 = r(N)$ and

$$l(n) = \sum_{s=1}^{d} (A^s(n))^\dagger l(n-1) A^s(n), \quad (A2)$$

$$r(n) = \sum_{s=1}^{d} A^s(n+1) r(n+1) (A^s(n+1))^\dagger. \quad (A3)$$

The norm of the state is then given by $r(0) = l(N) = (l(n)|r(n)) = (l(n-1)|\mathbb{E}_{A(n)}^{A(n)}|r(n))$, which we require to be one. Here, we represented the left and right density matrices as D^2-dimensional bra and ket vectors with round brackets via the Choi-Jamiołkowski isomorphism [42–45], and introduced the notation $\mathbb{E}_B^A = \sum_{s=1}^{d} A^s \otimes \overline{B}^s$. For a left-canonical (right-canonical) MPS, where all the tensors are left orthonormalized (right orthonormalized), the left density matrices $l(n)$ [right density matrices $r(n)$] would be identity matrices. The matrices $C(n)$ that allow to go from one canonical form to the other according to $A_L^s(n) C(n) = C(n-1) A_R^s(n)$ can be used to compute the right density matrices $r(n)$ for the left-canonical form as $r(n) = C(n) C(n)^\dagger$, or the left density matrices for the right-canonical form as $l(n) = C(n)^\dagger C(n)$. The eigenvalues

of the density matrices are thus related to the square of the Schmidt coefficients.

To introduce the MPS tangent space, it is required that the set of MPS constitutes a smooth manifold. It was indeed established that the set of full rank MPS—meaning that all density matrices $l(n)$ and $r(n)$ as defined in the main text have full rank—constitute a smooth manifold in Refs. [46,47] (for real-valued MPS with open boundary conditions) and in Ref. [48] (for complex-valued finite MPS with open boundary conditions and uniform MPS with periodic boundary conditions or in the thermodynamic limit). Finite MPS for which some $l(n)$ and/or $r(n)$ do not have full rank, or uniform MPS which are not injective, correspond to singular points where the tangent space is not well defined. We briefly elaborate on this in the sections on the two-site integrators. In the proofs of Refs. [47,48], the mapping between the MPS parameters [the site dependent tensors $A(n) \in \mathbb{C}^{D_{n-1} \times d \times D_n}$] and the state $|\Psi[A]\rangle$ is identified as a principal fiber bundle, resulting from the gauge redundancy of MPS as discussed in the main text.

The MPS tangent $T_{|\Psi[A]\rangle} \mathcal{M}_{\text{MPS}}$ space is spanned by the partial derivatives of $|\Psi[A]\rangle$ with respect to all entries $A_{\alpha,\beta}^s(n)$ for every site $n = 1, \ldots, N$. We now denote a general variation as B^i with a collective index $i = (\alpha, s, \beta, n)$, such that the most general MPS tangen vector can be written as

$$B^i |\partial_i \psi\rangle = \sum_{n=1}^{N} \sum_{\{s_n\}=1}^{d} A^{s_1}(1) \cdots B^{s_n}(n) \cdots A^{s_N}(N)$$
$$\times |s_1 \ldots s_n \ldots s_N\rangle . \quad (A4)$$

The gauge redundancy of the MPS parametrization reflects itself in tangent space by the fact that not all linearly independent choices B^i produce independent tangent vectors $B^i |\partial_i \Psi[A]\rangle$, i.e., the basis of partial derivatives is overcomplete. Put differently, the linear map from the parameters B^i to the states $B^i |\partial_i \Psi[A]\rangle$ has a nontrivial kernel. In particular, the space of infinitesimal gauge transformations (also known as the vertical subspace) $B^s(n) = \mathcal{N}^s[X](n) = X(n-1)A^s(n) - A^s(n)X(n)$ with $X(n) \in \mathbb{C}^{D_n \times D_n}$ result in $B^i |\partial_i \psi(A)\rangle = 0$, as can easily be checked by explicit insertion. Finding a unique parametrization for every tangent vector is equivalent to specifying a complement of the vertical subspace (the so-called horizontal subspace), for which there is no unique prescription. Within the framework of principal fibre bundles, a natural way to define the horizontal subspace is via a principal bundle connection. We simply refer to the result as a gauge fixing prescription for the tangent vectors and present two possible choices. One can describe any tangent vector by a parametrization B that satisfies the "left gauge fixing condition,"

$$(l(n-1)|\mathbb{E}^{B(n)}_{A(n)} = 0, \qquad \forall n = 1, \ldots, N-1, \qquad (A5)$$

or alternatively the "right gauge fixing condition,"

$$\mathbb{E}^{B(n)}_{A(n)}|r(n)) = 0, \qquad \forall n = 2, \ldots, N. \qquad (A6)$$

That these two gauge fixing conditions are related to a principal bundle connection implies that they transform covariantly under gauge transformations on the original MPS. If we transform $A^s(n) \to A^s_G(n) = G(n-1)^{-1}A^s(n)G(n)$ then $B^s(n)$ has to follow the identical transformation law to $B^s_G(n) = G(n-1)^{-1}B^s(n)G(n)$ in order to still satisfy the same gauge fixing condition. Extending the left gauge fixing condition to $n = N$ or the right gauge fixing condition to $n = 1$ is equivalent to restricting to those variations of $|\Psi[A]\rangle$ that preserve norm and phase, i.e., those tangent vectors which are orthogonal to $|\Psi[A]\rangle$. In combination with the left, respectively right orthonormal form for the MPS tensors, these gauge fixing conditions express that the orthonormality constraint is preserved to first order.

To derive the inverse-free action of the tangent space projector, we now switch to a different representation of tangent vectors, given by

$$|\Theta[B]\rangle = \sum_{n=1}^{N} \sum_{\{s_n\}=1}^{d} A^{s_1}_L(1) \cdots A^{s_{n-1}}_L(n-1)B^{s_n}(n)A^{s_{n+1}}_R(n+1)$$
$$\cdots A^{s_N}_R(N)|s_1 \ldots s_n \ldots s_N\rangle$$
$$= \sum_{n=1}^{N} \sum_{\{\alpha,\beta,s_n\}} B^{s_n}_{\alpha,\beta}(n) \big|\Phi^{[1:n-1]}_{L,\alpha}\big\rangle |s_n\rangle \big|\Phi^{[n+1:N]}_{R,\beta}\big\rangle. \quad (A7)$$

It is clear how this representation can be obtained from Eq. (A4) by applying left orthonormalization and right orthonormalization in every term separately, and absorbing the resulting factors in the new definition of $B^{s_n}(n)$. In this new representation, however, there is no easy way to relate B to the basis of partial derivatives. The vertical subspace is now given by choices $B^s(n) = \mathcal{N}^s[X](n) = X(n-1)A^s_R(n) - A^s_L(n)X(n)$ and allows to impose, e.g., the "left gauge fixing condition" $\sum_s A^s_L(n)B^s(n) = 0, \forall n =$

$1, \ldots, N-1$. This ensures that the overlap of two tangent vectors only contains diagonal contributions, as in

$$\langle\Theta[B_1]|\Theta[B_2]\rangle = \sum_{n=1}^{N} \sum_{s_n} \mathrm{tr}\left[B^{s_n}_1(n)^\dagger B^{s_n}_2(n)\right]. \qquad (A8)$$

This corresponds to the standard Euclidean inner product and thus to the choice of an orthonormal basis. Note that imposing the "left gauge fixing condition" is essential, since the basis composed of the states

$$\left\{\big|\Phi^{[1:n-1]}_{L,\alpha}\big\rangle |s_n\rangle \big|\Phi^{[n+1:N]}_{R,\beta}\big\rangle, \forall\alpha,\beta,s_n, \forall n = 1, \ldots, N\right\}$$

might seem orthonormal at first, but is not (consider the overlap between two states for different n) and is in fact overcomplete.

Let us now discuss action of the tangent space projector $P_{T_{|\Psi[A]\rangle}\mathcal{M}_{\mathrm{MPS}}}$ onto an arbitrary vector $|\Xi\rangle$ in the Hilbert space $\mathbb{H}$. The resulting vector $P_{T_{|\Psi[A]\rangle}\mathcal{M}_{\mathrm{MPS}}} |\Xi\rangle$ should be an element of the tangent space $T_{|\Psi[A]\rangle}\mathcal{M}_{\mathrm{MPS}}$ of the manifold $\mathcal{M}_{\mathrm{MPS}}$ of MPS at the base point $|\Psi[A]\rangle$. It can thus be represented in the form $|\Theta[B]\rangle$. The standard Euclidean inner product of $\mathbb{H}$ guarantees that the equation $|\Theta[B]\rangle = P_{T_{|\Psi[A]\rangle}\mathcal{M}_{\mathrm{MPS}}} |\Xi\rangle$ is also a solution of

$$\min_B \| \, |\Theta[B]\rangle - |\Xi\rangle \, \|^2. \qquad (A9)$$

We can reformulate the minimization problem of Eq. (A9) as

$$\min_B \sum_{n=1}^{N} \sum_s \left[B^s(n)B^s(n)^\dagger - B^s(n)F^s(n)^\dagger - F^s(n)B^s(n)^\dagger\right]$$

together with the constraints

$$\sum_s A^s_L(n)^\dagger B^s(n) = 0, \quad \forall n = 1, \ldots, N-1.$$

The tensor $F(n)$ can easily be read from omitting tensor $B(n)$ in the overlap between the nth term of Eq. (A7) and the general state $|\Xi\rangle$, and corresponds to the overlap with the basis vectors

$$F^{s_n}_{\alpha,\beta}(n) = \langle\Phi^{[1:n-1]}_{L,\alpha}, s_n, \Phi^{[n+1:N]}_{R,\beta}|\Xi\rangle. \qquad (A10)$$

The solution to this constrained optimization problem is given by

$$B^s(n) = \sum_t (\delta_{s,t} - A^s_L(n)A^t_L(n)^\dagger)F^t(n)$$

for all $n = 1, \ldots, N-1$, where first factor is a projector that imposes the left-gauge fixing condition. Tensor $B(N)$ is not subject to any constraint and is given by $B^s(N) = F^s(N)$. For $n = 1, \ldots, N-1$, we also introduce the notation $G(n) = \sum_t A^t_L(n)^\dagger F^t(n)$, or thus

$$G_{\alpha,\beta}(n) = \langle\Phi^{[1:n]}_{L,\alpha}, \Phi^{[n+1:N]}_{R,\beta}|\Xi\rangle, \qquad (A11)$$

so that we can rewrite $B^s(n) = F^s(n) - A^s_L(n)G(n)$. Inserting this solution into $|\Theta[B]\rangle$ gives immediate rise to

$$|\Theta[B]\rangle = P_{T_{|\Psi[A]\rangle}\mathcal{M}_{\mathrm{MPS}}} |\Xi\rangle = \left(\sum_{n=1}^{N} \hat{P}^{[1:n-1]}_L \otimes \hat{1}_n \otimes \hat{P}^{[n+1:N]}_R\right.$$
$$\left. - \sum_{n=1}^{N-1} \hat{P}^{[1:n]}_L \otimes \hat{P}^{[n+1:N]}_R\right) |\Xi\rangle, \qquad (A12)$$

where we can recognize the contents of the brackets as the decomposition of the tangent space projector into a sum of orthogonal projectors, where $\hat{P}_L^{[1:n]}$ and $\hat{P}_R^{[n+1:N]}$ where defined in the main text.

This derivation differs from previous papers in using the new representation $|\Theta[B]\rangle$. Previous implementations of the TDVP required parameters that were directly defined with respect to the basis of partial derivatives $|\partial_i \Psi[A]\rangle$, which then specified how to update the parameters A in the time evolution. For the new representation $|\Theta[B]\rangle$, the parameters of B cannot easily be related to basis of partial derivatives with respect to all parameters $A^s(n)$, $\forall n = 1, \ldots, N$. The splitting scheme presented in the main text nevertheless allows us to update the MPS $|\Psi\rangle$ using this representation.

APPENDIX B: FULL ALGORITHM FOR THE SYMMETRIC ONE-SITE INTEGRATOR

For completeness, and to indicate the strong similarity to a typical DMRG implementation, we here present the implementation of a full step of the symmetric one-site integration scheme, which was presented in the main text and is based on a Lie-Trotter splitting scheme according to the decomposition of the tangent space projector as discussed in the previous section.

Algorithm 1: one-site symmetric integration step. Starting from a right-orthonormal MPS parametrization $\{A_C(1,t), A_R(2,t), \ldots, A_R(N,t)\}$ for the state at time t, compute a right-orthonormal MPS parametrization $\{A_C(1,t + \Delta t), A_R(2,t + \Delta t), \ldots, A_R(N,t + \Delta t)\}$ for the state at time $t + \Delta t$ by applying the following steps.

(1) Repeat for $n = 1 : N - 1$ # left-to-right sweep. (a) Evolve $A_C(n,t)$ forward in time according to

$$A_C(n,t + \Delta t/2) = \exp[-iH(n)\Delta t/2]A_C(n,t),$$

where $H(n)$ is computed using $A_L(1,t + \Delta t/2)$, ..., $A_L(n - 1,t + \Delta t/2)$, $A_R(n + 1,t)$, ..., $A_R(N,t)$. (b) Perform an orthogonal decomposition of $A_C(n,t + \Delta t/2)$ into $A_L(n,t + \Delta t/2)$ and $\tilde{C}(n,t + \Delta t/2)$. (c) Evolve $\tilde{C}(n,t + \Delta t/2)$ backwards in time according to

$$\tilde{C}(n,t) = \exp[+iK(n)\Delta t/2]\tilde{C}(n,t + \Delta t/2),$$

where $K(n)$ is computed using $A_L(1,t + \Delta t/2),\ldots,A_L(n,t + \Delta t/2)$, $A_R(n + 1,t)$, ..., $A_R(N,t)$. (d) Absorb $\tilde{C}(n,t)$ into $A_R(n + 1,t)$ to obtain $A_C(n + 1,t)$.

(2) Evolve $A_C(N,t)$ forward in time according to

$$A_C(N,t + \Delta t) = \exp[-iH(N)\Delta t]A_C(N,t),$$

where $H(N)$ is computed using $A_L(1,t + \Delta t/2)$, ..., $A_L(N - 1,t + \Delta t/2)$.

(3) Repeat for $n = N - 1 : -1 : 1$ # right-to-left sweep. (a) Perform an orthogonal decomposition of $A_C(n + 1,t + \Delta t)$ into $\tilde{C}(n,t + \Delta t)$ and $A_R(n + 1,t + \Delta t)$. (b) Evolve $\tilde{C}(n,t + \Delta t)$ backwards in time according to

$$\tilde{C}(n,t + \Delta t/2) = \exp[+iK(n)\Delta t/2]\tilde{C}(n,t + \Delta t),$$

where $K(n)$ is computed using $A_L(1,t + \Delta t/2),\ldots,A_L(n,t + \Delta t/2)$, $A_R(n + 1,t + \Delta t)$, ..., $A_R(N,t + \Delta t)$. (c) Absorb $\tilde{C}(n,t + \Delta t/2)$ into $A_L(n,t + \Delta t/2)$ to obtain $A_C(n + 1,t +$

$\Delta t/2)$. (d) Evolve $A_C(n,t + \Delta t/2)$ forward in time according to

$$A_C(n,t + \Delta t) = \exp[-iH(n)\Delta t/2]A_C(n,t + \Delta t/2),$$

where $H(n)$ is computed using $A_L(1,t + \Delta t/2)$, ..., $A_L(n - 1,t + \Delta t/2)$, $A_R(n + 1,t + \Delta t)$, ..., $A_R(N,t + \Delta t)$.

Let us briefly elaborate on the philosophy behind this integration scheme. If we separate the evolution in step 2 into two consecutive evolutions with time step $\Delta t/2$, referred to as step 2(a) and 2(b), respectively, then one can recognize the right-to-left sweep [step 2(b) and everything in step 3] as the adjoint method to the left-to-right sweep [everything in step 1 combined with step 2(a)]. For a general finite-step integrator $\mathcal{I}_{\Delta t}$ with time step Δt, the adjoint method $\mathcal{I}_{\Delta t}^*$ is defined as the inverse map of $\mathcal{I}_{-\Delta t}$. The combined scheme $\mathcal{I}_{\Delta t/2}^* \circ \mathcal{I}_{\Delta t/2}$ is symmetric by construction. If $\mathcal{I}_{\Delta t}$ is a first order integrator (meaning that the local error scales as $\mathcal{O}(\Delta t^2)$, the combined scheme becomes second order. However, it is not often the case that the adjoint method $\mathcal{I}^*$ can exactly be evaluated. For example, the adjoint method of the explicit Euler method is the implicit Euler method. That the adjoint method for the finite MPS integrator can be evaluated explicitly can be traced back to the fact that the individual terms obtained from splitting the tangent space projector are exactly integrable. The underlying reason for this is that the effective Hamiltonian $H(n)$ [or $C(n)$] for the one-site [or zero-site] center degrees of freedom $A_C(n)$ [or $C(n)$] does not depend on these degrees of freedom, but only on the tensors on the neighboring sites, so that the differential equations corresponding to the individual terms are first-order linear differential equations.

APPENDIX C: DERIVATION OF THE TWO-SITE INTEGRATION SCHEME

For general variational manifold, the TDVP principle of projecting the evolution vector onto the tangent space is the only strategy that ensures that the evolution can be described within the set of states that are assumed tractable, i.e., the manifold itself. Matrix product states as variational ansatz are special in the sense that they have (a set of) refinement parameters corresponding to the virtual dimensions. There is thus not a single variational manifold, but rather a hierarchy of manifolds, where the ones with lower bond dimension actually correspond to the singular regions of the ones with higher bond dimension. This raises the question whether we can generalize the "TDVP philosophy" so as to obtain an evolution within this hierarchy of states. From DMRG and TEBD, we know that such an evolution typically involves acting on a center block of two sites. The two-site evolution will take the state out of the manifold of MPS with fixed bond dimensions. While one typically truncates back to the original manifold, the advantage is that such an approach allows to truncate to a different manifold with higher (or lower) bond dimension when this is still computationally feasible. For the remainder, we assume that the truncation scheme is based on discarding singular values smaller than some tolerance level ε. Unfortunately, such a strategy cannot easily be formulated as a differential equation, as it necessarily requires a finite time step Δt. We can nevertheless get inspired by the normal TDVP derivation.

We start by defining the linear space of two-site variations to the current MPS $|\Psi[A]\rangle$. This is the linear space contains states of the form

$$
\sum_{n=1}^{N-1} \sum_{\{s_n\}=1}^{d} A_L^{s_1}(1) \cdots A_L^{s_{n-1}}(n-1) B^{s_n s_{n+1}}(n:n+1) A_R^{s_{n+2}}
$$
$$
\times (n+2) \cdots A_R^{s_N}(N) |s_1 \ldots s_n \ldots s_N\rangle . \quad \text{(C1)}
$$

Let us call this space $T^{[2]}_{|\Psi[A]\rangle}\mathcal{M}_{\text{MPS}}$. Note that this notation should not be confused with the double tangent space, which contains the span of all single-site variations on two arbitrary sites. As for the tangent space, the linear map from the representation $\{B^{s,t}(n:n+1); n=1,\ldots,N-1\}$ to the physical states in $T^{[2]}_{|\Psi[A]\rangle}\mathcal{M}_{\text{MPS}}$ has a nontrivial kernel corresponding to choices

$$
B^{s,t}(n:n+1) = A_L^s(n)X^t(n+1) - X^s(n)A_R^t(n+1)
$$

for any $\{X(n) \in \mathbb{C}^{D_n \times d \times D_{n+1}}, n=2,\ldots,N-1\}$. This freedom allows to impose a gauge-fixing condition of the form

$$
\sum_s A_L^s(n)^\dagger B^{s,t}(n:n+1) = 0, \quad \forall n=1,\ldots,N-2. \quad \text{(C2)}
$$

The orthogonal projector $\hat{P}_{T^{[2]}_{|\Psi[A]\rangle}\mathcal{M}_{\text{MPS}}}$ that maps any state $|\Phi\rangle \in \mathbb{H}$ onto the subspace $T^{[2]}_{|\Psi[A]\rangle}\mathcal{M}_{\text{MPS}}$ can then be constructed and decomposed—using manipulations similar to those of the previous subsection—as a sum of orthogonal projections:

$$
\hat{P}_{T^{[2]}_{|\Psi[A_L]\rangle}\mathcal{M}_{\text{MPS}}} = \sum_{n=1}^{N-1} \hat{P}_L^{[1:n-1]} \otimes \hat{1}_n \otimes \hat{1}_{n+1} \otimes \hat{P}_R^{[n+2:N]}
$$
$$
- \sum_{n=2}^{N-1} \hat{P}_L^{[1:n-1]} \otimes \hat{1}_n \otimes \hat{P}_R^{[n+1:N]}. \quad \text{(C3)}
$$

In particular, when applying $\hat{P}_{T^{[2]}_{|\Psi[A_L]\rangle}\mathcal{M}_{\text{MPS}}}$ onto $\hat{H}|\Psi[A]\rangle$, one obtains

$$
\hat{P}_{T^{[2]}_{|\Psi[A_L]\rangle}\mathcal{M}_{\text{MPS}}} \hat{H}|\Psi[A]\rangle = \sum_{n=1}^{N-1} \sum_{\alpha,\beta,s_n,s_{n+1}} B_{\alpha,\beta}^{s_n s_{n+1}}(n:n+1) \left| \Phi_{L,\alpha}^{[1:n-1]}, s_n, s_{n+1}, \Phi_{R,\beta}^{[n+2:N]} \right\rangle
$$
$$
- \sum_{n=2}^{N-1} \sum_{\alpha,\beta,s_n,s_{n+1}} B_{\alpha,\beta}^{s_n}(n) \left| \Phi_{L,\alpha}^{[1:n-1]}, s_n, \Phi_{R,\beta}^{[n+2:N]} \right\rangle, \quad \text{(C4)}
$$

where $\boldsymbol{B}(n) = H(n)\boldsymbol{A}_C(n)$ and now also $\boldsymbol{B}(n:n+1) = H(n:n+1)\boldsymbol{A}_C(n:n+1)$, with $H(n)$ and $H(n:n+1)$ the effective one-site and two-site Hamiltonians defined in the context of DMRG in the main text. This result justifies the two-site integration scheme that was proposed in the main text.

The error analysis of this two-site integrator is more complicated, as there does not exist an underlying differential equation that describes the evolution projected down to $T^{[2]}_{|\Psi[A_L]\rangle}\mathcal{M}_{\text{MPS}}$. With respect to the full Schrödinger equation, we expect a competition between the finite time step error, a truncation error ε, and a projection error that can be estimated in terms of $\|(\hat{1} - \hat{P}_{T^{[2]}_{|\Psi[A_L]\rangle}\mathcal{M}_{\text{MPS}}})\hat{H}|\Psi[A]\rangle\|$. Without truncation ($\varepsilon = 0$), one can easily verify that, after applying the complete sweep (left-to-right and back) with a time step Δt to an MPS $|\Psi[A]\rangle$, the original physical state $|\Psi[A]\rangle$ is obtained by reapplying the scheme with a time step $-\Delta t$, although $|\Psi[A]\rangle$ will now be encoded as an MPS with a larger bond dimension (and matrices $l(n)$ or $r(n)$, which do not have full rank). Hence, without the truncation, the integrator is still symmetric and one expects a finite time step error that scales as $\mathcal{O}(\Delta t^3)$. This should still hold for $\epsilon > 0$, since any error of lower order in Δt should scale at least proportional to ϵ

and is thus negligible in comparison to the contribution of $\mathcal{O}(\epsilon)$ that is introduced by the truncation anyway. Instead of truncating after every step in the algorithm, we could also find the best MPS approximation with lower bond dimension after a complete sweep without truncation [10].

In the case of a nearest-neighbor Hamiltonian, the projector error is absent as the action of $\hat{H}$ on $|\Psi[A]\rangle$ is exactly contained in $T^{[2]}_{|\Psi[A_L]\rangle}\mathcal{M}_{\text{MPS}}$. It is then also possible to apply a Lie-Trotter splitting scheme to the individual terms in the Hamiltonian. This is exactly the scheme that gives rise to the TEBD algorithm. In contrast, our scheme is based on the splitting of the projector in Eq. (C3) and can be applied to arbitrary Hamiltonians (for which $H(n)$ and $H(n:n+1)$ can be computed efficiently) with the same computational cost. The difference between both schemes can be related to redundancy in the representation of states in $T^{[2]}_{|\Psi[A_L]\rangle}\mathcal{M}_{\text{MPS}}$. In particular, by directly applying a nearest-neighbor Hamiltonian $\hat{H} = \sum_{n=1}^{N-1} \hat{h}(n:n+1)$ to $|\Psi[A]\rangle$, one obtains a state of the form of Eq. (C1), which is represented by $B^{s_1,s_2}(n:n+1) = \sum_{t_1,t_2} \langle s_1,s_2|\hat{h}(n:n+1)|t_1,t_2\rangle A^{s_1}(n)A^{s_2}(n)$. This is different from the representation in Eq. (C4), even though both describe the same physical state $\hat{H}|\Psi[A]\rangle$.

[1] M. Fannes, B. Nachtergaele, and R. F. Werner, Commun. Math. Phys. **144**, 443 (1992).

[2] S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995).

[3] F. Verstraete, V. Murg, and J. I. Cirac, Adv. Phys. **57**, 143 (2008).

[4] J. I. Cirac and F. Verstraete, J. Phys. A **42**, 504004 (2009).

[5] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).

[6] S. R. White, Phys. Rev. B **48**, 10345 (1993).

[7] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004).

[8] S. R. White and A. E. Feiguin, Phys. Rev. Lett. **93**, 076401 (2004).

[9] A. J. Daley, C. Kollath, U. Schollwöck, and G. Vidal, J. Stat. Mech. (2004) P04005.

[10] F. Verstraete, J. J. Garcia-Ripoll, and J. I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004).

[11] S. Rommer and S. Östlund, Phys. Rev. B **55**, 2164 (1997).

[12] F. Verstraete, D. Porras, and J. I. Cirac, Phys. Rev. Lett. **93**, 227205 (2004).

[13] H. Saberi, A. Weichselbaum, and J. von Delft, Phys. Rev. B **78**, 035124 (2008).

[14] A. Weichselbaum, F. Verstraete, U. Schollwöck, J. I. Cirac, and J. von Delft, Phys. Rev. B **80**, 165117 (2009).

[15] U. Schollwöck, Ann. Phys. **326**, 96 (2011).

[16] H. F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959).

[17] M. Suzuki, Commun. Math. Phys. **51**, 183 (1976).

[18] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pizorn, H. Verschelde, and F. Verstraete, Phys. Rev. Lett. **107**, 070601 (2011).

[19] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken, SIAM J. Matrix Anal. Appl. **34**, 470 (2013).

[20] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, Springer Series in Computational Mathematics (Springer-Verlag, Berlin, Heidelberg, 1996).

[21] M. P. Zaletel, R. S. K. Mong, C. Karrasch, J. E. Moore, and F. Pollmann, Phys. Rev. B **91**, 165112 (2015).

[22] I. P. McCulloch, J. Stat. Mech. (2007) P10014.

[23] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, New J. Phys. **12**, 025012 (2010).

[24] F. Verstraete and J. I. Cirac, Phys. Rev. Lett. **104**, 190405 (2010).

[25] J. Haegeman, J. I. Cirac, T. J. Osborne, and F. Verstraete, Phys. Rev. B **88**, 085118 (2013),

[26] I. V. Oseledets, SIAM J. Sci. Comput. **33**, 2295 (2011).

[27] C. Lubich, I. Oseledets, and B. Vandereycken, SIAM J. Numer. Anal. **53**, 917 (2015).

[28] M. Hochbruck and C. Lubich, SIAM J. Numer. Anal. **34**, 1911 (1997).

[29] C. Lubich, *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis* (European Mathematical Society, Zürich, 2008).

[30] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure Preserving Algorithms for Ordinary Differential Equations*, 2nd ed., Vol. 31 (Springer-Verlag, Berlin, Heidelberg, 2006).

[31] R. B. Sidje, ACM Trans. Math. Software (TOMS) **24**, 130 (1998).

[32] A symmetric integration scheme has the property that, applying it with time step Δt and then with time step $-\Delta t$, the initial state is recovered .

[33] J. Haegeman, T. J. Osborne, and F. Verstraete, Phys. Rev. B **88**, 075133 (2013).

[34] For example, the backward evolution of $C(n)$, which actually increases the energy but potentially allows to avoid local minima, can be omitted.

[35] The set of MPS with all possible bond dimensions constitutes a stratification .

[36] E. H. Lieb and D. W. Robinson, Commun. Math. Phys. **28**, 251 (1972).

[37] P. Hauke and L. Tagliacozzo, Phys. Rev. Lett. **111**, 207202 (2013).

[38] Z.-X. Gong, M. Foss-Feig, S. Michalakis, and A. V. Gorshkov, Phys. Rev. Lett. **113**, 030602 (2014)

[39] P. Richerme, Z.-X. Gong, A. Lee, C. Senko, J. Smith, M. Foss-Feig, S. Michalakis, A. V. Gorshkov, and C. Monroe, Nature (London) **511**, 198 (2014).

[40] F. Verstraete and J. I. Cirac, arXiv:cond-mat/0407066.

[41] V. Murg, F. Verstraete, and J. I. Cirac, Phys. Rev. A **75**, 033605 (2007).

[42] E. C. Sudarshan, P. M. Mathews, and J. Rau, Phys. Rev. **121**, 920 (1961).

[43] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).

[44] M.-D. Choi, Linear Algebra Appl. **10**, 285 (1975).

[45] P. Arrighi and C. Patricot, Ann. Phys. **311**, 26 (2004).

[46] S. Holtz, T. Rohwedder, and R. Schneider, Numer. Math. **120**, 701 (2012).

[47] A. Uschmajew and B. Vandereycken, Linear Algebra Appl. **439**, 133 (2013).

[48] J. Haegeman, M. Mariën, T. J. Osborne, and F. Verstraete, J. Math. Phys. **55**, 021902 (2014)