Post-matrix product state methods: To tangent space and beyond

Jutho Haegeman,[1] Tobias J. Osborne,[2] and Frank Verstraete[1,3]

[1]*Department of Physics and Astronomy, University of Ghent, Krijgslaan 281 S9, B-9000 Ghent, Belgium*
[2]*Leibniz Universität Hannover, Institute of Theoretical Physics, Appelstrasse 2, D-30167 Hannover, Germany*
[3]*Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Wien, Austria*

We develop in full detail the formalism of tangent states to the manifold of matrix product states, and show how they naturally appear in studying time evolution, excitations, and spectral functions. We focus on the case of systems with translation invariance in the thermodynamic limit, where momentum is a well-defined quantum number. We present some illustrative results and discuss analogous constructions for other variational classes. We also discuss generalizations and extensions beyond the tangent space, and give a general outlook towards post-matrix product methods.

PACS number(s): 05.10.Cc, 75.40.Gb, 03.65.−w

I. INTRODUCTION

The last two decades have witnessed a tremendous cross fertilization between different branches of theoretical physics, including condensed matter physics, quantum information theory, and renormalization group theory, with even some ingredients of quantum gravity added to the picture. One of the major breakthroughs along this line of research was the development of the density matrix renormalization group (DMRG) by White,[1,2] which has quickly evolved to become the standard numerical tool for finding ground states of one-dimensional quantum spin systems. It was later realized that DMRG corresponds to a variational method that optimizes over a class of states which are known as matrix product states (MPS).[3,4] Insight from quantum information theory resulted in the development of many extensions of the original DMRG method. For example, more general *Ansätze* known as tensor network states have been formulated for systems in higher dimensions.[5–7] Another important development was the formulation of the so-called time-evolving block decimation (TEBD) by Vidal[8] for studying real-time evolution of one-dimensional quantum lattice Hamiltonians using the formalism of MPS.

In condensed matter physics and quantum field theory, most interesting quantum systems, aside from quantum dots, are of macroscopic size in at least one spatial dimension, and the main interest is often in the bulk properties of these systems, without any boundary or finite-size influences.[9] Here too, the formalism of MPS has a nice advantage over alternative numerical approximation methods, as they can be formulated directly in the thermodynamic limit and still depend on only a finite number of parameters when the system under study exhibits translation invariance. These states are now known as uniform MPS (uMPS), but had in fact already been formulated before the development of DMRG under the name finitely correlated states,[10] which served as a generalization of the valence bond construction for the spin-1 model by Affleck, Kennedy, Lieb, and Tasaki (AKLT).[11,12] The TEBD with imaginary-time evolution was among the first methods that was actually able to efficiently optimize over these variational parameters,[13] although other algorithms have now been formulated.[14,15]

The low-energy physics of the bulk is not determined by the ground state alone, but requires knowledge of the spectrum of bulk excitations. These excitations are almost always present in the laboratory, either due to finite temperature, because the system is being perturbed by probes in spectroscopy experiments, or because parameters of the system are being quenched. Recently, new algorithms were formulated for studying time evolution[15] and for describing the elementary excitations[16,17] that depend strongly on the concept of tangent vectors to the original variational manifold of uMPS. Analogous results have also been formulated for generic MPS on finite lattices.[18,19] Indeed, it was recently understood that the set of MPS (generic and uniform) with a given bond dimension constitutes a smooth manifold $\mathcal{M}$ embedded in the full Hilbert space $\mathbb{H}$, where the MPS parametrization can be recognized as a principal fiber bundle.[20] More precisely, the set of physical states constitutes a Kähler manifold, which expresses the fact that the complex structure is compatible with the Riemannian geometry that is obtained when inducing the standard Euclidean metric of $\mathbb{H}$ onto $\mathcal{M}$. (Since the metric also defines a closed 2-form, the manifold additionally has a symplectic structure.) This allows us to identify the tangent space of $\mathcal{M}$ with a complex subspace of $\mathbb{H}$. Similar constructions were simultaneously developed in the field of numerical analysis, where real-valued tensor networks are also becoming increasingly popular.[21,22] Whereas Ref. 20 focused on a rigorous study of the differential geometry of $\mathcal{M}_{\mathrm{MPS}}$ and its tangent space, the current paper focuses on the physical relevance of these tangent states.

In Sec. II, we summarize the necessary ingredients from Ref. 20 that will be used throughout the remainder of this paper. Section III explores in full detail the *time-dependent variational principle* (TDVP) which was introduced in Ref. 15. The TDVP is a general prescription for the optimal way to approximate a time-evolving quantum state within a given variational manifold. As such, it does not yet take us outside of the manifold of uMPS. Section IV then explains why the tangent space itself is also useful as a variational subspace for the study of *elementary bulk excitations* of a system, and relates this to linear response theory. We also show how the tangent space can be used to get a quick estimate of spectral functions. The algorithms presented in this section should thus be considered as *post-MPS* methods that allow us extract dynamical information from new variational subspaces that are beyond, but based on, the original MPS manifold. We then illustrate the power of these methods by considering

some explicit examples in Sec. V. We briefly compare our methods to analogous constructions that were developed the context of other variational classes, in particular mean field theory, in Sec. VI. Finally, Sec. VII introduces extensions and generalizations that even takes us beyond the MPS tangent space and gives an outlook on how this will eventually result in the development of a new *Fock space* on top of the MPS vacuum. Finally, the key concepts of this paper are summarized in Sec. VIII.

II. MANIFOLD OF UNIFORM MATRIX PRODUCT STATES IN THE THERMODYNAMIC LIMIT

This section summarizes the definition and key properties of the variational manifold of uMPS, which were derived in full detail in Ref. 20. Readers familiar with that paper can safely skip this section. A proper treatment of quantum states in the thermodynamic limit requires a description in terms of C^* algebras, which is how finitely correlated states were originally constructed.[10,23] We employ the results from these papers, but adopt a physics-style notation which is to be understood as a limiting procedure of finite lattices on which our definitions are meaningful. Even then, this thermodynamic limit requires careful attention due to intricate effects such as infrared divergences and the orthogonality catastrophe.[24] We will emphasize the necessary steps required to avoid these intricacies and obtain consistent results.

A. Definition and properties of the manifold

Consider a one-dimensional lattice $\mathcal{L} \subset \mathbb{Z}$ with $|\mathcal{L}| = 2N + 1$ sites labeled by the integer $n \in \mathcal{L} = \{-N, \ldots, N\}$. Every site n contains a d-dimensional quantum variable, so that local Hilbert space $\mathbb{H}_n \cong \mathbb{C}^d$ is spanned by a basis $\{|s_n\rangle \mid s_n = 1, \ldots, d\}$. The total Hilbert space is given by $\mathbb{H} = \bigotimes_{n=-N}^{N} \mathbb{H}_n$ and is spanned by the product basis

$$|s\rangle = |s_{-N}\rangle_{-N} \otimes \ldots \otimes |s_0\rangle_0 \otimes \ldots \otimes |s_N\rangle_N. \quad (1)$$

In the thermodynamic limit $N \to \infty$, a uniform matrix product state $|\Psi(A)\rangle \in \mathbb{H}$ is defined as

$$|\Psi(A)\rangle \triangleq \sum_{\{s_n\}=1}^{d} \text{tr}[V \ldots A^{s_{-1}} A^{s_0} A^{s_{+1}} \ldots] |s\rangle, \quad (2)$$

and is parametrized by a tensor $A \in \mathbb{A} \cong \mathbb{C}^{D \times d \times D}$ or, in an alternative interpretation, by a set of complex $D \times D$ matrices A^s, for $s = 1, \ldots, d$. The matrix dimension D is known as the bond dimension. The $D \times D$ matrix V encodes the boundary conditions, namely, $V = \mathbb{1}$ corresponds to periodic boundary conditions whereas V should be a rank-1 object (i.e., $V = \boldsymbol{v}_R \boldsymbol{v}_L^\dagger$ with $\boldsymbol{v}_{L,R} \in \mathbb{C}^D$) for a system with open boundary conditions. Based on the results from Refs. 10 and 23, it was shown in Ref. 20 that the expectation value of local observables is well defined and independent of the boundary matrix V when A is an element from the open subset $\mathcal{A} \subset \mathbb{A}$ which is known as the set of *injective* MPS or of pure finitely correlated states. Equivalently, the set $\mathcal{A}$ corresponds to all tensors A for which the transfer matrix $\mathbb{E} = \sum_{s=1}^{d} A^s \otimes \overline{A}^s$ has a single eigenvalue $\omega^{(1)} = \rho(\mathbb{E})$ with

$\rho(\mathbb{E})$ the spectral radius of $\mathbb{E}$, whereas all other eigenvalues $\omega^{(i)}, i = 2, \ldots, D^2$, satisfy $|\omega^{(i)}| < \rho(\mathbb{E})$. Furthermore, the left and right eigenvectors $(l|$ and $|r)$ corresponding to $\omega^{(1)}$, when written as positive-semidefinite Hermitian matrices, should have full rank, i.e., they should be strictly positive definite. To obtain a normalizable state, the tensor $A \in \mathcal{A}$ should then be "renormalized" as $A/\sqrt{\omega^{(1)}}$ so that the spectral radius of the transfer matrix becomes 1.

Let now $\mathcal{M} \subset \mathbb{H}$ denote the set of states $\{|\Psi(A)\rangle \mid A \in \mathcal{A}\}$. The state $|\Psi(A)\rangle$ is invariant under a reparametrization $A \leftarrow A_G$ where $A_G^s = G^{-1} A^s G$ for any invertible matrix $G \in \mathsf{GL}(D, \mathbb{C})$. This invariance is known as gauge invariance, and the map $(A, G) \to A_G$ represents the (right) group action of the gauge group $\mathsf{G} \cong \mathsf{PGL}(D, \mathbb{C})$, where we had to define the gauge group G as the projective linear group $\mathsf{PGL}(D, \mathbb{C})$ obtained by taking the quotient of $\mathsf{GL}(D, \mathbb{C})$ with its center subgroup $\mathsf{GL}(1, \mathbb{C})$ of matrices G which are proportional to the unit matrix, since these choices have the trivial effect $A_G = A$. By restricting to the open subset $\mathcal{A} \subset \mathbb{A}$, one can then show that the group action is free and proper. These properties express that the group action is sufficiently nice to obtain a smooth quotient space $\mathcal{A}/\mathsf{G}$. The injectivity property of the MPS ensures that this quotient space is diffeomorphic (and in fact biholomorphic) to the set of states $\mathcal{M}$, thus also turning the latter one into a smooth (complex) manifold. The MPS representation that maps the tensor $A \in \mathcal{A}$ to the physical state $|\Psi(A)\rangle \in \mathcal{M}$ and exhibits invariance under the action of G can thus be given the structure of a principal fiber bundle, which is a useful identification when also studying the tangent space of $\mathcal{A}$ and $\mathcal{M}$ in the next section.

We now introduce some notations that are used throughout the remainder of this paper. Since we always assume that $A \in \mathcal{A}$, the transfer matrix $\mathbb{E} = \mathbb{E}_A^A$ has a unique eigenvalue $\omega^{(1)}$ that can scaled to be exactly one with corresponding left and right eigenvectors $(l|$ and $|r)$ corresponding to strictly positive-definite matrices l and r that we assume to be normalized as $(l|r) = \text{tr}[lr] = 1$. All other eigenvalues $\omega^{(k)}$, $k > 1$, lie strictly within the unit circle. We also define $\mathbb{S} = |r)(l|$ as a projector onto the eigenspace of eigenvalue 1, and its complement $\mathbb{Q} = \mathbb{1} - \mathbb{S}$. Note that the right action of the transfer matrix on a vector $|x)$ can also be encoded as a completely positive map

$$\mathcal{E} : \mathbb{C}^D \to \mathbb{C}^D : x \mapsto \mathcal{E}(x) = \sum_{s=1}^{d} A^s x A^{s\dagger}. \quad (3)$$

The dual map

$$\widetilde{\mathcal{E}} : \mathbb{C}^D \to \mathbb{C}^D : y \mapsto \widetilde{\mathcal{E}}(y) = \sum_{s=1}^{d} A^{s\dagger} y A^s \quad (4)$$

encodes the left action on a vector $(y|$. The actions of these maps can be computed with computing time that scales $\mathcal{O}(D^3)$, so that l and r (eigenvectors corresponding to the largest eigenvalue) can efficiently be computed using an iterative eigensolver. While we do not assume that the tensor A satisfies any gauge fixing condition in any formula in this paper, it is often convenient to use the left ($l = \mathbb{1}, \sum_s A^{s\dagger} A^s = \mathbb{1}$) or right

$(r = \mathbb{1}, \sum_s A^s A^{s\dagger} = \mathbb{1})$ gauge fixing condition in numerical implementations.

Given a set of local operators $\hat{O}^\alpha$, we can use these definitions to compute the two-point connected correlation function as

$$\Gamma^{(\alpha,\beta)}(n) = (l|\mathbb{E}_{O^\alpha}\mathbb{E}^{n-1}\mathbb{E}_{O^\beta}|r) - (l|\mathbb{E}_{O^\alpha}|r)(l|\mathbb{E}_{O^\beta}|r)$$
$$= (l|\mathbb{E}_{O^\alpha}\mathbb{Q}(\mathbb{Q}\mathbb{E}\mathbb{Q})^{n-1}\mathbb{Q}\mathbb{E}_{O^\beta}|r), \qquad (5)$$

where we have used

$$\mathbb{E}^n - \mathbb{S} = \mathbb{Q}\mathbb{E}^n\mathbb{Q} = \mathbb{Q}(\mathbb{Q}\mathbb{E}\mathbb{Q})^n\mathbb{Q}.$$

The correlation length ξ is then determined by the largest eigenvalue of $\mathbb{Q}\mathbb{E}\mathbb{Q}$ as

$$\xi = -\frac{1}{\ln[\rho(\mathbb{Q}\mathbb{E}\mathbb{Q})]}. \qquad (6)$$

Under the given assumption, $\rho(\mathbb{Q}\mathbb{E}\mathbb{Q}) < 1$ and the correlation length ξ is finite. Hence, all injective uMPS are exponentially clustering.[10] The correlation length is determined by $\rho(\mathbb{Q}\mathbb{E}\mathbb{Q})$, which is equal to the eigenvalue of the transfer matrix $\mathbb{E}$ that is second largest in absolute value.

B. Tangent space and momentum eigenstates

We now discuss the structure of the tangent space of the manifold of uMPS $\mathcal{M}$ and the parametrization thereof. We first introduce the notation

$$|\Phi^{(A)}(B)\rangle$$
$$= |\Phi(B;A)\rangle = B^i \partial_i |\Psi(A)\rangle$$
$$= \sum_{n\in\mathbb{Z}}\sum_{\{s_n\}=1}^d \boldsymbol{v}_L^\dagger \left[\left(\prod_{m<n} A^{s_m}\right) B^{s_n} \left(\prod_{m'>n} A^{s_{m'}}\right)\right]\boldsymbol{v}_R |\boldsymbol{s}\rangle, \qquad (7)$$

where $\partial_i = \partial/\partial A^i$ and the index i is shorthand for a collective index (α,β,s) combining the two virtual indices and the physical index of tensor A, i.e., $A^{(\alpha,\beta,s)} = A^s_{\alpha,\beta}$. The tensor B can take values in the tangent space of parameter space $T_A\mathcal{A}$ at the point A. Since the subset of injective MPS $\mathcal{A}$ is an open subset of the full parameter space $\mathbb{A} = \mathbb{C}^{D \times d \times D}$, we obtain $T_A\mathcal{A} \equiv \mathbb{A}$. The physical states $|\Phi(B;A)\rangle$ for $B \in \mathbb{A} = \mathbb{C}^{D \times d \times D}$ span the tangent space $T_{|\Psi(A)\rangle}\mathcal{M}$. More completely, we can interpret this whole construction as mapping elements $(B;A)$ of the tangent bundle $T\mathcal{A}$ to elements $(|\Phi(B;A)\rangle, |\Psi(A)\rangle)$ of the tangent bundle $T\mathcal{M} \subset \mathbb{H} \times \mathbb{H}$, where Φ should be recognized as the tangent map of Ψ, which is necessarily linear in its first argument.

For the application of the variational principle to the study of translation-invariant phenomena, this tangent space (consisting completely out of translation-invariant states) is sufficient. However, we can also interpret $|\Psi(A)\rangle$ as a special point in the larger class of general MPS with site-dependent matrices, and define generalized tangent vectors

$$|\Phi^{(A)}[\boldsymbol{B}]\rangle = \sum_{n\in\mathbb{Z}}\sum_{\{s_n\}=1}^d \boldsymbol{v}_L^\dagger \left[\left(\prod_{m<n} A^{s_m}\right) B^{s_n}(n)\right.$$
$$\left. \times \left(\prod_{m'>n} A^{s_{m'}}\right)\right]\boldsymbol{v}_R |\boldsymbol{s}\rangle, \qquad (8)$$

where we use square brackets to denote a "functional dependence" on a set of site-dependent tensors $\boldsymbol{B} = \{B(n)\}_{n\in\mathbb{Z}}$. This larger tangent space is denoted as $\mathbb{T}^{|\Psi(A)\rangle}$ and turns out to be important when studying excited states, for which translation invariance is no longer a good assumption. However, when both the Hamiltonian and its ground state are translation invariant, we know that we can label the excited states by a momentum quantum number $p \in [-\pi,\pi)$. The momentum p sector $\mathbb{T}_p^{|\Psi(A)\rangle}$ of the larger tangent space $\mathbb{T}^{|\Psi(A)\rangle}$ is obtained by choosing $B^s(n) = B^s e^{ipn}$, and we define

$$|\Phi_p(B;A)\rangle = |\Phi_p^{(A)}(B)\rangle = \sum_{n\in\mathbb{Z}} e^{ipn} \sum_{\{s_n\}=1}^d \boldsymbol{v}_L^\dagger$$
$$\times \left[\left(\prod_{m<n} A^{s_m}\right) B^{s_n} \left(\prod_{m'>n} A^{s_{m'}}\right)\right]\boldsymbol{v}_R |\boldsymbol{s}\rangle, \quad (9)$$

with thus $|\Phi_0(B;A)\rangle = |\Phi(B;A)\rangle$. Hence, $\Phi_p^{(A)}$ represents a linear map from $\mathbb{A}$ to the momentum p sector of the tangent space $\mathbb{T}^{|\Psi(A)\rangle} \subset \mathbb{H}$ at the translation-invariant point $|\Psi(A)\rangle$. The full tangent space is obtained as $\mathbb{T}^{|\Psi(A)\rangle} = \int_{p\in[-\pi,\pi)}^\oplus \mathbb{T}_p^{|\Psi(A)\rangle}$.

We now repeat in detail some calculations using the tangent states $|\Phi_p(B;A)\rangle$ from Ref. [20] to point out possible divergences that can occur. We introduce the generalized notation $\mathbb{E}_B^A = \sum_{s=1}^d A^s \otimes \overline{B}^s$, which will be used extensively when evaluating expectation values with tangent vectors. We first compute the overlap between a tangent vector and the original uMPS $|\Psi(A)\rangle$ and obtain

$$\langle\Psi(\overline{A})|\Phi_p(B;A)\rangle = \sum_{n\in\mathbb{Z}} e^{ipn}\left(l|\mathbb{E}_A^B|r\right) = 2\pi\delta(p)\left(l|\mathbb{E}_A^B|r\right)$$
$$(10)$$

so that all states $|\Phi_p(B)\rangle$ with $p \neq 0$ are automatically orthogonal to $|\Psi(A)\rangle$ due to the orthogonality of the different momentum sectors. For $p = 0$, $\langle\Psi(\overline{A})|\Phi_0(B;A)\rangle$ is proportional to $(l|\mathbb{E}_A^B|r)$, with a diverging proportionality factor $2\pi\delta(0) = |\mathbb{Z}|$ where the cardinality $|\mathbb{Z}|$ represents the diverging number of lattice sites ($\mathcal{L} = \mathbb{Z}$). It is useful to define an orthogonal complement $\mathbb{T}_p^{(A)\perp}$, which is equal to $\mathbb{T}_p^{|\Psi(A)\rangle}$ if $p \neq 0$ and only contains the tangent vectors for which the tensor B satisfies the linear relation $(l|\mathbb{E}_A^B|r) = 0$ if $p = 0$. We will discuss the physical relevance of restricting to this orthogonal subspace in the following sections.

Next, we compute the overlap between two tangent vectors $\langle\Psi_{p'}(\overline{B'};\overline{A})|\Psi_p(B;A)\rangle$. We have to be very careful with the infinite sums over the positions $n \in \mathbb{Z}$ and $n' \in \mathbb{Z}$ of B and B'. When a diverging result is obtained, it is easily possible to make errors by miscounting. Only when the result is guaranteed to be finite can we freely use index substitutions. We therefore replace every occurrence of $\mathbb{E}^n$ by a "regularized" operator $\mathbb{Q}\mathbb{E}^n\mathbb{Q} = \mathbb{E}^n\mathbb{Q} = \mathbb{Q}\mathbb{E}^n = \mathbb{E}^n - \mathbb{S} = \mathbb{Q}(\mathbb{Q}\mathbb{E}\mathbb{Q})^n\mathbb{Q}$ with $\rho(\mathbb{Q}\mathbb{E}\mathbb{Q}) < 1$ and a "singular" part $\mathbb{S} = |r)(l|$. The reason of this notation becomes clear if we now

evaluate $\langle \Psi_p(\overline{B}; \overline{A}) | \Psi_{p'}(B'; A)\rangle$ as

$$
\begin{aligned}
\langle \Phi_p(\overline{B}; \overline{A}) \mid \Phi_{p'}(B'; A)\rangle &= \overline{B}^{\bar{i}} N_{\bar{i}, j}(p, p'; \overline{A}, A) B'^j = \sum_{n=-\infty}^{+\infty} \sum_{n'=-\infty}^{+\infty} e^{+ip'n'-ipn} \big[\theta(n = n')(l|\mathbb{E}_B^{B'}|r) \\
&\quad + \theta(n' > n)(l|\mathbb{E}_B^A(\mathbb{E})^{n'-n-1}\mathbb{E}_A^{B'}|r) + \theta(n' < n)(l|\mathbb{E}_A^{B'}(\mathbb{E})^{n-n'-1}E_B^A|r) \big] \\
&= \sum_{n_0=-\infty}^{+\infty} e^{i(p'-p)n_0} \sum_{\Delta n=-\infty}^{+\infty} e^{ip\Delta n} \big[\theta(\Delta n = 0)(l|\mathbb{E}_B^{B'}|r) \\
&\quad + \theta(\Delta n > 0)(l|\mathbb{E}_B^A \mathbb{Q}\mathbb{E}^{\Delta n-1}\mathbb{Q}\mathbb{E}_A^{B'}|r) + \theta(\Delta n < 0)(l|\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{-\Delta n-1}\mathbb{Q}\mathbb{E}_B^A|r) \big] \\
&\quad + (l|\mathbb{E}_B^A|r)(l|\mathbb{E}_A^{B'}|r) \sum_{n=-\infty}^{+\infty} \sum_{n'=-\infty}^{n-1} e^{ip'n'-ipn} + (l|\mathbb{E}_A^{B'}|r)(l|\mathbb{E}_B^A|r) \sum_{n=-\infty}^{+\infty} \sum_{n'=n+1}^{+\infty} e^{ip'n'-ipn}.
\end{aligned}
$$

In these expressions, we have introduced a "discrete" Heaviside function θ taking a logical expression as argument and resulting 1 if the argument is true. We have denoted the matrix elements of this overlap with respect to B and B' as $N_{\bar{i}, j}(p, p'; \overline{A}, A)$, where $[N_{\bar{i}, j}]$ is referred to as the *effective norm matrix*. The explicit dependence on both $\overline{A}$ and A is used to indicate that it is not a holomorphic function of A alone. There is no point in trying to compute the overlap between two tangent vectors $|\Phi_p(B; A)\rangle$ and $|\Phi_{p'}(B'; A')\rangle$ at different gauge-inequivalent points A and A', as this overlap is automatically zero, which can be considered as a generalization of Anderson's orthogonality catastrophe.[24] By using the well-known result for the geometric series of an operator with spectral radius smaller than one, we obtain

$$
\sum_{n=0}^{+\infty} \mathbb{Q}\mathbb{E}^n\mathbb{Q} = \sum_{n=0}^{+\infty} \mathbb{Q}(\mathbb{Q}\mathbb{E}\mathbb{Q})^n\mathbb{Q} = \mathbb{Q}(\mathbb{1} - \mathbb{Q}\mathbb{E}\mathbb{Q})^{-1}\mathbb{Q} \tag{11}
$$

and thus

$$
\begin{aligned}
\langle \Phi_p(\overline{B}; \overline{A}) | \Phi_{p'}(B'; A)\rangle &= \overline{B}^{\bar{i}} N_{\bar{i}, j}(p, p'; \overline{A}, A) B'^j = 2\pi\delta(p - p')\overline{B}^{\bar{i}} N_{\bar{i}, j}(p; \overline{A}, A)B'^j \\
&= 2\pi\delta(p' - p)\big[(l|\mathbb{E}_B^{B'}|r) + (l|\mathbb{E}_B^A\mathbb{Q}(\mathbb{1} - e^{ip}\mathbb{Q}\mathbb{E}\mathbb{Q})^{-1}\mathbb{Q}\mathbb{E}_A^{B'}|r) \\
&\quad + (l|\mathbb{E}_A^{B'}\mathbb{Q}(\mathbb{1} - e^{-ip}\mathbb{Q}\mathbb{E}\mathbb{Q})^{-1}\mathbb{Q}\mathbb{E}_B^A|r) + (2\pi\delta(p) - 1)(l|\mathbb{E}_B^{B'}|r)(l|\mathbb{E}_A^A|r)\big].
\end{aligned} \tag{12}
$$

As expected, momentum eigenstates can not be normalized to unity in an infinitely large system, but rather satisfy a δ normalization. For equal momenta, $N_{\bar{i}, j}(p, p; \overline{A}, A)$ contains the diverging prefactor $2\pi\delta(0) = |\mathbb{Z}|$. The remaining part has been denoted as $N_{\bar{i}, j}(p; \overline{A}, A)$ and can be extracted from the terms inside the square brackets. Inside these brackets, the regular part $\mathbb{Q}\mathbb{E}\mathbb{Q}$ produces a finite contribution where B and B' are strongly connected. We therefore also refer to these terms as the *connected contribution*. For $p = 0$, the product $\mathbb{Q}(\mathbb{1} - e^{\pm ip}\mathbb{Q}\mathbb{E}\mathbb{Q})^{-1}\mathbb{Q}$ can be interpreted as the pseudoinverse of the singular superoperator $\mathbb{1} - \mathbb{E}$, which has an eigenvalue zero associated to the left and right eigenvectors $(l|$ and $|r)$. We henceforth define $(\mathbb{1} - e^{\pm ip}\mathbb{E})^{\mathsf{P}} = \mathbb{Q}(\mathbb{1} - e^{\pm ip}\mathbb{Q}\mathbb{E}\mathbb{Q})^{-1}\mathbb{Q}$, so that $(\mathbb{1} - \mathbb{E})^{\mathsf{P}}(\mathbb{1} - \mathbb{E}) = (\mathbb{1} - e^{\pm ip}\mathbb{E})(\mathbb{1} - e^{\pm ip}\mathbb{E})^{\mathsf{P}} = \mathbb{Q} = \mathbb{1} - |r)(l|$. Only for zero momentum does $(\mathbb{1} - e^{\pm ip}\mathbb{E})^{\mathsf{P}}$ denote a true pseudoinverse. For momentum zero, there is an additional divergence inside the square brackets coming from the singular part $\mathbb{S}$. Here, B and B' appear in two separate factors, and this term is henceforth referred to as the *disconnected contribution*. Using Eq. (10), this term can be traced back to the nonzero overlap with the original uMPS $|\Psi(A)\rangle$. It disappears for tangent vectors in $\mathbb{T}_p^{(A)\perp}$. Note also that for momentum zero, $N_{\bar{i}, j}(0, 0; \overline{A}, A)$ can be identified with the metric of the uMPS manifold.

The parametrization of tangent vectors inherits the gauge invariance of the MPS $|\Psi(A)\rangle$, whereby $|\Phi_p(B; A)\rangle = |\Phi_p(B_G; A_G)\rangle$. By fixing the representation A of the base point $|\Psi(A)\rangle$, this "multiplicative" gauge invariance is also fixed. Nevertheless, the linear map $\Phi_p^{(A)} : \mathbb{A} \mapsto \mathbb{T}_p^{|\Psi(A)\rangle}$ still has a nontrivial null space which can be associated with infinitesimal gauge transformations of the MPS $|\Psi(A)\rangle$. Since the full tangent space $\mathbb{T}^{|\Psi(A)\rangle}$ was obtained by interpreting the uMPS $|\Psi(A)\rangle$ as a special point in the space of generic MPS, we have to consider site-dependent gauge transformations that take $A^s(n) = A^s$ to $A_G^s(n) = G(n - 1, \eta)^{-1} A^s G(n, \eta)$ for a one-parameter family of gauge transformations $G(n, \eta) = \exp[\eta x(n)]$. Expressing the invariance of the MPS at first order in η for a choice $x(n) = x \exp(ipn)$ allows us to conclude that $|\Phi_p^{(A)}(B)\rangle = 0$ for $B = \mathcal{N}_p^{(A)}(x)$, where the action of $\mathcal{N}_p^{(A)}$ is given by

$$
\mathcal{N}_p^{(A)} : x \mapsto \mathcal{N}_p^{(A)s}(x) = A^s x - e^{-ip}xA^s, \forall\, s = 1, \dots, d. \tag{13}
$$

The map $\mathcal{N}_p^{(A)}$ establishes an isomorphism between the null space $\mathbb{N}_p^{(A)}$ of the map $\Phi_p^{(A)}$ and the Lie algebra of the gauge group, which is equal to $\mathfrak{gl}(D, \mathbb{C}) = \mathbb{C}^{D \times D}$ for $p \neq 0$ and to $\mathfrak{pgl}(D, \mathbb{C}) = \{x \in \mathbb{C}^{D \times D} | \operatorname{tr}[x] = 0\}$ for $p = 0$. It can easily be checked that this is an isomorphism by trying to determine the null space of the map $\mathcal{N}_p$. By multiplying $\mathcal{N}_{\Phi_p}^s(x) = 0$ ($\forall\, s = 1, \dots, d$) to the left with $(A^s)^\dagger$ and summing over s, we obtain the requirement

$$
\mathbb{E}|xr) = e^{-ip}|xr). \tag{14}
$$

For $p \neq 0$, this equation has no solutions, whereas for $p = 0$, the only solutions are matrices x proportional to the unit matrix $\mathbb{1}_D$, which is not within the algebra $\mathfrak{pgl}(D,\mathbb{C})$ of traceless matrices. For nonzero p, we thus obtain $\dim \mathbb{T}_p = \dim \mathbb{A} - \dim \mathbb{N}_p^{(A)} = (d-1)D^2$, while we obtain $\dim \mathbb{T}_0 = (d-1)D^2 + 1$ for zero momentum. But, of course, $|\Psi(A)\rangle \in \mathbb{T}_0^{|\Psi(A)\rangle}$, and restricting to the part $\mathbb{T}_0^\perp$ that is orthogonal to $|\Psi(A)\rangle$ also reduces the dimension to $\dim \mathbb{T}_0^\perp = (d-1)D^2$. Within the language of principal fiber bundles, the null space $\mathbb{N}_p^{(A)}$ is known as the vertical subspace, and for any $B \in \mathbb{A}$, $B' \in \mathbb{N}_p^{(A)}$ we obtain $|\Phi_p^{(A)}(B + B')\rangle = |\Phi_p^{(A)}(B)\rangle$. Therefore, we sometimes refer to this as an *additive gauge freedom* in the parametrization of MPS tangent vectors. Since the vectors $B = \mathbb{N}_p^{(A)}(x)$ in the null space $\mathbb{N}_p^{(A)}$ are also eigenvectors with zero eigenvalue of the effective normalization matrix $N_{\bar{\imath},j}(p)$, as well as of any other effective operator $O_{\bar{\imath},j}(p)$ that we obtain by restricting a physical operator $\hat{O}$ to the tangent space $\mathbb{T}_p^{|\Psi(A)\rangle}$, we also refer to them as *null modes*. In order to associate a unique parametrization to every tangent vector $|\Phi_p^{(A)}(B)\rangle \in \mathbb{T}_p$, we should define a complementary space $\mathbb{B}_p^{(A)}$ such that $\mathbb{A} = \mathbb{B}_p^{(A)} \oplus \mathbb{N}_p^{(A)}$ and restrict to parametrizations $B \in \mathbb{B}_p^{(A)} \oplus_p$. While there is no unique definition for $\mathbb{B}_p^{(A)}$ (referred to as the horizontal subspace), a gauge-covariant description would be such that $B_G \in \mathbb{B}_p^{(A_G)}$ for any $B \in \mathbb{B}_p^{(A)}$. This is easily accomplished using the machinery of principal bundle connections.[20] We only mention the result. Two different choices for $\mathbb{B}_p^{(A)}$ are obtained as the subspace of solutions of one of the two following linear homogeneous sets of equations, which we can then call gauge fixing conditions for the additive gauge freedom:

  (i) Left gauge fixing condition:

$$\sum_{s=1}^{d} A^{s\dagger} l B^s = 0 \quad \Leftrightarrow \quad (l|\mathbb{E}_A^B = 0. \tag{15}$$

  (ii) Right gauge fixing condition:

$$\sum_{s=1}^{d} B^s r A^{s\dagger} = 0 \quad \Leftrightarrow \quad \mathbb{E}_A^B|r) = 0. \tag{16}$$

Since these conditions are D^2 dimensional, they fix all D^2 linearly independent gauge transformations in $\mathbb{N}_p^{(A)}$ for $p \neq 0$. For $p = 0$, there are only $D^2 - 1$ linearly independent gauge transformations, and these D^2 gauge fixing conditions also include norm preservation, i.e., they imply $\langle\Psi(A)|\Phi_0(B)\rangle = 0$ or thus $|\Phi_0(B)\rangle \in \mathbb{T}_0^{|\Psi(A)\rangle\perp}$. Note that the horizontal subspaces $\mathbb{B}_p^{(A)}$ defined by these equations are momentum independent, even though the vertical subspace $\mathbb{N}_p^{(A)}$ did explicitly depend on the momentum p. Since we will explicitly be using either of the two conditions from Eqs. (15) and (16), the horizontal subspace is henceforth denoted as $\mathbb{B}^{(A)}$ without any reference to the momentum p.

Either choice for $\mathbb{B}^{(A)}$ results in a considerable simplification of the effective norm matrix since all nonlocal terms in Eq. (12) cancel:

$$\langle\Phi_p(\overline{B}; \overline{A})|\Phi_p(B; A)\rangle = 2\pi\delta(p' - p)(l|\mathbb{E}_B^{B'}|r). \tag{17}$$

However, this simplification is only really useful if we can directly parametrize tensors $B \in \mathbb{B}^{(A)}$. A linear parametrization $B = \mathcal{B}^{(A)}(X)$ depending on a $D \times D(d-1)$ matrix X that satisfies Eq. (15) can be constructed. We first define the $D \times Dd$ matrix L as

$$[L]_{\alpha;(\beta,s)} = [A^{s\dagger}l^{1/2}]_{\alpha,\beta} \tag{18}$$

and then construct a $dD \times (d-1)D$ matrix V_L which contains an orthonormal basis for the null space of L, i.e., $L V_L = 0$ and $V_L^\dagger V_L = \mathbb{1}_{(d-1)D}$. Setting $[V_L^s]_{\alpha,\beta} = [V_L]_{(\alpha,s);\beta}$, we then define the representation $\mathcal{B}^{(A)s}(x)$ as

$$\mathcal{B}^{(A)s}(X) = l^{-1/2} V_L^s X r^{-1/2} \tag{19}$$

in order to obtain

$$\langle\Phi_{p'}(\overline{\mathcal{B}}^{(A)}(\overline{X}))|\Phi_p(\mathcal{B}^{(A)}(Y))\rangle = 2\pi\delta(p - p') \,\mathrm{tr}[X^\dagger Y], \tag{20}$$

in combination with the left gauge fixing condition $\sum_{s=1}^{d} A^{s\dagger} l \mathcal{B}^{(A)s}(X) = 0$. An alternative representation for tensors B satisfying the left gauge fixing conditions follows similarly.

III. TIME-DEPENDENT VARIATIONAL PRINCIPLE

After White's formulation of the DMRG, another major breakthrough was the formulation of DMRG-inspired algorithms to study dynamic properties of quantum spin chains. These algorithms can be divided into two classes: some algorithms directly probe spectral functions, whereas other algorithms aim to approximate the full time-evolving wave function. These methods are reviewed by Schollwoeck and White[25] and Jeckelmann.[26] Today's best-known and most powerful method for approximating the time-evolving wave function within the MPS manifold is the time-evolving block decimation (TEBD), which was developed by Vidal.[8] It was later reformulated in order to be compatible with traditional DMRG implementations.[27,28] The TEBD is based on an iterative application of a Lie-Trotter-Suzuki decomposition[29,30] of the exact evolution operator for a small time step dt as

$$\exp(i\hat{H}dt) = \exp(i\hat{H}^{(A)}dt)\exp(i\hat{H}^{(B)}dt) + \mathcal{O}(dt^2). \tag{21}$$

Higher-order decompositions with an error of $\mathcal{O}(dt^p)$ are also possible.[31] $\hat{H}^{(A)}$ and $\hat{H}^{(B)}$ provide a decomposition of the (possibly time-dependent) Hamiltonian $\hat{H} = \hat{H}^{(A)} + \hat{H}^{(B)}$, such that $\hat{H}^{(A)}$ and $\hat{H}^{(B)}$ separately contain local terms that all commute. If necessary, a decomposition into more than two parts is also possible. For a nearest-neighbor Hamiltonian $\hat{H} = \sum_{n\in\mathbb{Z}} \hat{h}_{n,n+1}$, a possible decomposition scheme is into even and odd terms: $\hat{H}^{(A)} = \sum_{n\in\mathbb{Z}} \hat{h}_{2n,2n+1}$ and $\hat{H}^{(B)} = \sum_{n\in\mathbb{Z}} \hat{h}_{2n+1,2n+2}$. The individual operators $\exp(i\hat{H}^{(A)})$ and $\exp(i\hat{H}^{(B)})$ then split into a product of local unitaries that can be dealt with in a parallelized and efficient way. When applied to a generic MPS, the individual evolution operators take the state outside the original manifold since they have the effect of increasing the virtual bond dimension. Once a given maximum bond dimension has been reached, one then approximates the newly obtained state by an MPS with this maximal bond dimension. The best strategy for truncating a single bond dimension is obtained by discarding the smallest Schmidt values. However, since the Hamiltonian

evolution acts on all bonds, several bond dimensions have to be simultaneously truncated. The strategy based on discarding the smallest Schmidt values still serves as a good initial guess but is not optimal. For lattices $\mathcal{L}$ of finite size, an optimal MPS representation with given maximal bond dimension can be obtained by minimizing the norm difference using algorithms inspired by the sweeping process of the finite-size algorithm of the DMRG.[32]

The TEBD can also be applied to translation-invariant systems in the thermodynamic limit. In combination with imaginary-time evolution, this resulted in the first algorithm for finding the best ground-state approximation within the class of uMPS.[13,33] However, no optimal strategy for truncation of the bond dimension is known in the case of infinite lattices. As a variational strategy, the infinite-size time-evolving block decimation then requires a scaling of $dt \rightarrow 0$ as the optimal approximation is approached, in order to correct for the truncation error. Since the exact imaginary-time evolution automatically slows down in the neighborhood of the best ground state (approximation), the need for a decreasing time step induces an additional unfavorable slowing down.

In addition, both for finite and infinite systems, some symmetries of the Hamiltonian $\hat{H}$ might not be inherited by the individual Trotter evolution operators $\exp(i\hat{H}^{(A)}dt)$ and $\exp(i\hat{H}^{(B)}dt)$. In itself, the Lie-Trotter-Suzuki decomposition is symplectic, and under an exact iterative application of the Trotter operators, errors resulting from these broken symmetries would be strongly bound. However, the additional truncation after every evolution step ruins the symplecticity and drifting errors are possible. In particular, for a time-independent Hamiltonian $\hat{H}$, the expectation value of the Hamiltonian is a constant of motion, but will drift away in a simulation based on the TEBD. One last downside of TEBD is that it can not cope with Hamiltonians containing long-range interaction terms.

Recently, a new algorithm for approximating time-evolving quantum states with MPS was proposed[15] based on the TDVP of Dirac.[34–36] The TDVP is a general method that can be formulated for any variational manifold and any Hamiltonian, both with short- and long-range interactions. In combination with the manifold of MPS, the TDVP allows one to overcome the aforementioned shortcomings of the TEBD. Unlike TEBD, which is necessarily formulated using discrete time steps dt, the TDVP is naturally formulated in continuous time. The TDVP transforms the linear Schrödinger equation in the full Hilbert space $\mathbb{H}$ into a nonlinear set of symplectic differential equations in the parameter space of the variational manifold. If we could solve these nonlinear differential equations exactly in continuous time, the only source of errors would be the restriction to the manifold itself. The TDVP describes the best direction in which the quantum state can evolve without leaving the variational manifold in order to approximate the time-dependent Schrödinger equation. Hence, no truncation of any kind is necessary. In addition, there is no need for a Trotter decomposition and thus no corresponding Trotter error. This approach is also perfectly applicable in case of imaginary-time evolution (there is no symplecticity of course). It can also be applied to Hamiltonians with long-range interactions and to generic MPS on finite lattices[19,37] or to a finite subsystem of an infinite lattice.[38] Most importantly, the TDVP can be

implemented for the case of MPS with an efficiency that is comparable to other methods such as DMRG and TEBD [namely $\mathcal{O}(D^3)$]. Of key importance for such an efficient implementation is the use of the gauge fixing conditions for tangent vectors, which was introduced in the previous section.

A numerical integration of the TDVP equation does of course require a discretization of the time variable, resulting in additional errors due to the final time step. However, since many standard numerical integrators can be used, these errors are well controlled and well understood from the general theory of (symplectic) differential equations (on manifolds).[39] We now derive the TDVP equation based on an action formalism for the case of uMPS in the thermodynamic limit. We then compare the resulting equations to the geometric argument that was used in the original publication.[15] While the geometric construction might provide a better visual insight into the approximation made by the TDVP, the action formalism is better suited to derive the properties of the resulting nonlinear differential equation. We discuss the symplectic properties of real-time evolution and discuss convergence and error measures that can be used to assess the approximation error made by confining the evolution to $\mathcal{M}$. Finally, we outline the details of a simple first-order Euler-based algorithm. The numerical implementation of more advanced integration schemes is described elsewhere.[40]

A. Principle of least action

The dynamics of isolated quantum systems are governed by the time-dependent Schrödinger equation (TDSE)

$$i\frac{d}{dt}|\Psi(t)\rangle = \hat{H}(t)|\Psi(t)\rangle, \tag{22}$$

which is a linear first-order differential equation in $\mathbb{H}$. Note that we allow for the Hamiltonian $\hat{H}$ to be time dependent. The TDSE can be derived by applying the variational principle of least action to the action functional

$$S_{\mathbb{H}}[\overline{\Psi},\Psi] = \int_{-\infty}^{+\infty} \left(\frac{i}{2}\langle\Psi(t)|\dot{\Psi}(t)\rangle - \frac{i}{2}\langle\dot{\Psi}(t)|\Psi(t)\rangle \right.$$
$$\left. - \langle\Psi(t)|\hat{H}(t)|\Psi(t)\rangle \right) dt. \tag{23}$$

We now focus on one-dimensional quantum lattice systems with translation invariance. For notational simplicity, we assume that the Hamiltonian contains nearest-neighbor terms only, i.e., $\hat{H} = \sum_{n\in\mathbb{Z}} \hat{T}^n \hat{h} \hat{T}^{-n}$ where $\hat{T}$ is the translation operator that shifts the system by one site and $\hat{h}$ has nontrivial support only on sites zero and one. The generalization to interaction terms on a larger number of neighboring sites or even long-range interaction terms is straightforward since no Trotter-type decomposition of the Hamiltonian is required. An initial state $|\Psi_0\rangle$ that can be encoded as a uMPS $|\Psi(A_0)\rangle$ will in general leave the manifold of uMPS $\mathcal{M}$ under exact time evolution. In order to confine the dynamics to $\mathcal{M}$, we can analogously apply the principle of least action to

$$S_{\mathcal{M}}[\overline{A},A] = \int_{-\infty}^{+\infty} \left(\frac{i}{2}\big[\dot{A}^j(t)\partial_j - \dot{\overline{A}}^{\overline{j}}(t)\partial_{\overline{j}}\big]\langle\Psi(\overline{A}(t))|\Psi(A(t))\rangle \right.$$
$$\left. - \langle\Psi(\overline{A}(t))|\hat{H}(t)|\Psi(A(t))\rangle \right) dt, \tag{24}$$

where the square brackets indicate the functional dependence on the full evolution $A(t)$ and its complex conjugate. Whereas the Schrödinger equation describes a unitary process (norm preserving) that makes the integrand of $S_{\mathbb{H}}$ exactly zero, the extremization of $S_{\mathcal{M}}$ might result in a

flow equation for $|\Psi(A(t))\rangle$ that is not norm preserving and minimizes $S_{\mathcal{M}}$ by converging to a state with zero norm. It is therefore better to define a modified, normalized action which does result in norm-independent dynamics as

$$
\begin{aligned}
\widetilde{S}_{\mathcal{M}}[\overline{A},A] &= \int_{-\infty}^{+\infty} \frac{\frac{i}{2}\big[\dot{A}^j(t)\partial_j - \dot{\overline{A}}^{\overline{j}}(t)\partial_{\overline{j}}\big]\langle\Psi(\overline{A}(t))|\Psi(A(t))\rangle - \langle\Psi(\overline{A}(t))|\hat{H}|\Psi(A(t))\rangle}{\langle\Psi(\overline{A}(t))|\Psi(A(t))\rangle}\, dt \\
&= \int_{-\infty}^{+\infty}\left(\frac{i}{2}\big[\dot{A}^j(t)\partial_j - \dot{\overline{A}}^{\overline{j}}(t)\partial_{\overline{j}}\big]\ln N(\overline{A}(t),A(t)) - H(\overline{A}(t),A(t))\right) dt,
\end{aligned}
\tag{25}
$$

where

$$
N(\overline{A},A) = \langle\Psi(\overline{A})|\Psi(A)\rangle, \quad H(\overline{A},A) = \frac{\langle\Psi(\overline{A})|\hat{H}|\Psi(A)\rangle}{\langle\Psi(\overline{A})|\Psi(A)\rangle}.
\tag{26}
$$

Note that we henceforth omit the explicit time dependence of the Hamiltonian $\hat{H}$ for the sake of simplicity. For a time-dependent Hamiltonian $\hat{H}(t)$, the energy function H would have an explicit time dependence [i.e., $H(\overline{A},A,t)$], but this has no further effect on the resulting expressions. When working with a uMPS $|\Psi(A)\rangle$, we always assume that it has been properly normalized such that $\rho(\mathbb{E}_A^A) = 1$ and we can set $N(\overline{A},A) = 1$. However, this is only true after a normalization of the matrices A^s and in general $N = \langle\Psi(\overline{A})|\Psi(A)\rangle$ depends on A. For the translation-invariant nearest-neighbor Hamiltonian $\hat{H} = \sum_{n\in\mathbb{Z}} \hat{T}^n \hat{h} \hat{T}^{-n}$, we obtain

$$
H(\overline{A},A) = |\mathbb{Z}|h(\overline{A},A),
\tag{27}
$$

where, for a properly normalized uMPS $|\Psi(A)\rangle$,

$$
h(\overline{A},A) = \frac{\langle\Psi(\overline{A})|\hat{h}|\Psi(A)\rangle}{\langle\Psi(\overline{A})|\Psi(A)\rangle} = \big(l|\mathbb{H}_{AA}^{AA}|r\big)
\tag{28}
$$

and we have defined a new superoperator

$$
\mathbb{H}_{A_3 A_4}^{A_1 A_2} = \sum_{s,t,u,v=1}^{d} \langle u,v|\hat{h}|s,t\rangle\, A_1^s A_2^t \otimes \overline{A_3^u A_4^v}.
\tag{29}
$$

If we define the norm-independent action with normalized integrand for the full Hilbert space $\mathbb{H}$ as $\widetilde{S}_{\mathbb{H}}$, then the stability of $\widetilde{S}_{\mathbb{H}}$ with respect to variations $\langle\Psi(t)| \mapsto \langle\Psi(t)| + \langle\delta\Psi(t)|$ requires

$$
\left(\hat{\mathbb{1}} - \frac{|\Psi(t)\rangle\langle\Psi(t)|}{\langle\Psi(t)|\Psi(t)\rangle}\right)\left(i\frac{d}{dt}|\Psi(t)\rangle - \hat{H}|\Psi(t)\rangle\right) = 0.
\tag{30}
$$

Applying the variational principle to the modified action $\widetilde{S}_{\mathbb{H}}$ thus imposes the Schrödinger equation in the plane orthogonal to the vector $|\Psi(t)\rangle$, whereas it leaves the evolution in the direction of the current vector $|\Psi(t)\rangle$ unspecified. Since a nonzero parallel component of the evolution vector ($\langle\Psi(t)|\dot{\Psi}(t)\rangle \neq 0$) results in norm or phase changes, the use of the modified

action unties the restriction to a specific choice of phase and normalization of the state. Norm-independent dynamics within the variational manifold of uMPS are described by the Euler-Lagrange equations of $\widetilde{S}_{\text{uMPS}}$, which are given by

$$
+i\,\widetilde{N}_{\overline{i},j}(\overline{A}(t),A(t))\dot{A}^j(t) = H_{\overline{i}}(\overline{A}(t),A(t))
\tag{31}
$$

and its complex conjugate. Here, we have introduced the gradient

$$
\begin{aligned}
H_{\overline{i}}(\overline{A},A) = \partial_{\overline{i}} H(\overline{A},A) &= \frac{\langle\partial_{\overline{i}}\Psi(\overline{A})|\hat{H}|\Psi(A)\rangle}{N(\overline{A},A)} \\
&\quad - \frac{\langle\partial_{\overline{i}}\Psi(\overline{A})|\Psi(A)\rangle\langle\Psi(\overline{A})|\hat{H}|\Psi(A)\rangle}{N(\overline{A},A)^2}
\end{aligned}
\tag{32}
$$

and the modified metric

$$
\begin{aligned}
\widetilde{N}_{\overline{i},j}(\overline{A},A) = \partial_{\overline{i}}\partial_j \ln N(\overline{A},A) &= \frac{N_{\overline{i},j}(0,0;\overline{A},A)}{N(\overline{A},A)} \\
&\quad - \frac{\langle\partial_{\overline{i}}\Psi(\overline{A})|\Psi(A)\rangle\langle\Psi(\overline{A})|\partial_j\Psi(A)\rangle}{N(\overline{A},A)^2}.
\end{aligned}
\tag{33}
$$

Whenever we need to compute these quantities, we can properly normalize the uMPS $|\Psi(A)\rangle$ and assume that $N(\overline{A},A) = 1$. The metric $N_{\overline{i},j}(0,0;\overline{A},A) = \langle\partial_{\overline{i}}\Psi(\overline{A})|\partial_j\Psi(A)\rangle$ was implicitly defined in Eq. (12). It can easily be seen that the second term in Eq. (33) contains a double divergence (it is proportional to $|\mathbb{Z}|^2$) and cancels exactly with the divergent term within the square brackets in Eq. (12). We thus obtain

$$
\begin{aligned}
\overline{B}^{\overline{i}}\widetilde{N}_{\overline{i},j}(\overline{A},A)(B')^j = |\mathbb{Z}|\big[&(l|\mathbb{E}_B^{B'}|r) + (l|\mathbb{E}_B^A(\mathbb{1} - \mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r) \\
&+ (l|\mathbb{E}_A^{B'}(\mathbb{1} - \mathbb{E})^{-1}\mathbb{E}_B^A|r) \\
&- (l|\mathbb{E}_A^{B'}|r)(l|\mathbb{E}_B^A|r)\big].
\end{aligned}
$$

We can also evaluate the first term in Eq. (32) using the same techniques as in the previous section. We prefer to evaluate the more general expression $\langle\Phi_p(\overline{B})|\hat{H}|\Psi(A)\rangle$, which defines $\langle\partial_{\overline{i}}\Psi(\overline{A})|\hat{H}|\Psi(A)\rangle$ through $\overline{B}^{\overline{i}}\langle\partial_{\overline{i}}\Psi(\overline{A})| = \langle\Phi_0(B)|$. Using the translation-invariant Hamiltonian with nearest-neighbor

interactions $\hat{H} = \sum_{n \in \mathbb{Z}} \hat{T}^n \hat{h} \hat{T}^{-n}$, we obtain

$$
\begin{aligned}
&\langle \Phi_p(\overline{B}) | \hat{H} | \Psi(A) \rangle \\
&= \sum_{n=-\infty}^{+\infty} \sum_{n'=-\infty}^{+\infty} e^{-ipn} \big[\theta(n = n')(l | \mathbb{H}_{BA}^{AA} | r) \\
&\quad + \theta(n = n' + 1)(l | \mathbb{H}_{AB}^{AA} | r) + \theta(n > n' + 1)(l | \mathbb{H}_{AA}^{AA} \\
&\quad \times (\mathbb{E})^{n-n'-2} \mathbb{E}_A^B | r) + \theta(n < n')(l | \mathbb{E}_A^B (\mathbb{E})^{n'-n-1} \mathbb{H}_{AA}^{AA} | r) \big].
\end{aligned}
$$

Repeating the same tricks as for the evaluation of $\langle \Phi_{p'}(B') | \Phi_p(B) \rangle$ leads to

$$
\begin{aligned}
&\langle \Phi_p(\overline{B}; \overline{A}) | \hat{H} | \Psi(A) \rangle \\
&= 2\pi \delta(p) \big[(l | \mathbb{H}_{BA}^{AA} | r) + (l | \mathbb{H}_{AB}^{AA} | r) + (l | \mathbb{H}_{AA}^{AA} (\mathbb{1} - \mathbb{E})^{\mathsf{P}} \\
&\quad \times \mathbb{E}_B^A | r) + (l | \mathbb{E}_B^A (\mathbb{1} - \mathbb{E})^{\mathsf{P}} \mathbb{H}_{AA}^{AA} | r) + (|\mathbb{Z}| - 2) \\
&\quad \times (l | \mathbb{H}_{AA}^{AA} | r)(l | \mathbb{E}_B^A | r) \big].
\end{aligned} \tag{34}
$$

As expected, the translation-invariant state $\hat{H} | \Psi(A) \rangle$ has zero overlap with momentum eigenstates with $p \neq 0$. Indeed, this is a necessary condition in order to be able to approximate time evolution within the translation-invariant manifold of uMPS. Hence, only the tangent vectors with zero momentum feature in the TDVP. For $p = 0$, the overlap is proportional to $2\pi \delta(0) = |\mathbb{Z}|$, which matches with the same factor in $\widetilde{N}_{\bar{\imath},j}(\overline{A}, A)$. As for the metric, there is an additional divergence inside the brackets, which is related to the nonzero overlap of the tangent vector $| \Phi_0 B \rangle$ with the original uMPS $| \Psi(A) \rangle$ [i.e., $(l | \mathbb{E}_B^A | r) \neq 0$]. Here too, this additional divergence cancels with the second term in Eq. (32), resulting in

$$
\begin{aligned}
&\overline{B}^{\bar{\imath}} H_{\bar{\imath}}(\overline{A}, A) \\
&= |\mathbb{Z}| \big[(l | \mathbb{H}_{BA}^{AA} | r) + (l | \mathbb{H}_{AB}^{AA} | r) + (l | \mathbb{H}_{AA}^{AA} (\mathbb{1} - \mathbb{E})^{\mathsf{P}} \mathbb{E}_B^A | r) \\
&\quad + (l | \mathbb{E}_B^A (\mathbb{1} - \mathbb{E})^{\mathsf{P}} \mathbb{H}_{AA}^{AA} | r) - 2(l | \mathbb{H}_{AA}^{AA} | r)(l | \mathbb{E}_B^A | r) \big].
\end{aligned}
$$

A translation-invariant time-evolving quantum state for a one-dimensional lattice system can thus be approximated with a time-evolving uMPS $| \Psi(A(t)) \rangle$ where the evolution of the parametrization $A(t)$ is specified by Eq. (31). However, this equation does not fully specify the time evolution of all degrees of freedom since the modified metric $\widetilde{N}_{\bar{\imath},j}$ has a number of null modes (eigenvectors with zero eigenvalues). First, since $B^i | \partial_i \Psi(A) \rangle = | \Phi_0(B) \rangle = 0$ for any $B = \mathbb{N}_0^{(A)}(x)$ with $x \in \mathfrak{pgl}(D, \mathbb{C})$, the modified metric $\widetilde{N}_{\bar{\imath},j;}(\overline{A}, A)$ inherits the $D^2 - 1$ linearly independent null modes of $N_{\bar{\imath},j}(0, 0; \overline{A}, A)$. These null modes do not render the linear system for $\dot{A}$ in Eq. (31) unsolvable since any null mode $B = \mathbb{N}_0^{(A)}(x)$ also satisfies $\overline{B}^{\bar{\imath}} H_{\bar{\imath}}(\overline{A}, A) = 0$. These modes result in gauge transformations of the parametrization without influencing the physical state. Hence, they are not determined by the dynamics and will have to be fixed by a gauge fixing prescription, which is mathematically equivalent to choosing one of the infinitely many solutions of the linear system for $\dot{A}$ and thus to define a pseudoinverse $\widetilde{N}^{i,\bar{\jmath}}(\overline{A}, A)$ of the modified metric.

However, before being able to do so, we have to take into account all null modes of the modified metric $\widetilde{N}_{\bar{\imath},j}(\overline{A}, A)$. Since $A^i | \partial_i \Psi(A) \rangle = | \Phi_0(A) \rangle \sim | \Psi(A) \rangle$, we have $B = A$ as an additional, linearly independent null mode of $\widetilde{N}_{\bar{\imath},j}(\overline{A}, A)$. Note that also $\overline{A}^{\bar{\imath}} H_{\bar{\imath}}(\overline{A}, A) = 0$. The mode $B = A$ was responsible

for the divergences of order $|\mathbb{Z}|^2$ in $\langle \partial_{\bar{\imath}} \Psi(\overline{A}) | \hat{H} | \Psi(A) \rangle$ and $\langle \partial_{\bar{\imath}} \Psi(\overline{A}) | \partial_j \Psi(A) \rangle$, which were canceled by the second term in Eqs. (32) and (33), respectively. This mode results in physical changes in norm or phase, which are not fixed by the Euler-Lagrange equation of the norm-independent action. This can also be seen from Eq. (30). Hence, we need an additional constraint to be able to invert $\widetilde{N}_{\bar{\imath},j}$ and to fully fix the time evolution of $A(t)$. This boils down to fixing the value of $\langle \Psi(A(t)) | \frac{d}{dt} | \Psi(A(t)) \rangle = \langle \Psi(A(t)) | \Phi(\dot{A}(t)) \rangle$, which was left unspecified by the norm-independent dynamics. Since we would like to keep the norm of the time-evolving uMPS $| \Psi(A(t)) \rangle$ fixed to one, we need to impose at least

$$
\frac{d}{dt} \langle \Psi(A(t)) | \Psi(A(t)) \rangle = 2 \operatorname{Re} \big[\langle \Psi(A(t)) | \Phi(\dot{A}(t)) \rangle \big] = 0. \tag{35}
$$

If we also fix the freedom in phase by

$$
\operatorname{Im}[\langle \Psi(A(t)) | \Phi_0(\dot{A}(t)) \rangle] = 0, \tag{36}
$$

then we effectively restrict to tangent vectors $| \Phi_0(\dot{A}(t)) \rangle \in \mathbb{T}_0^{|\Psi(A)\rangle \perp}$. Hence, we are allowed to restrict to a parametrization $\dot{A}(t) \in \mathbb{B}^{(A)}$ by imposing either the left or right gauge fixing prescriptions defined in Eqs. (15) and (16), which allows us to cancel many terms in the expressions for $\widetilde{N}_{\bar{\imath},j}(\overline{A}, A)$ and $H_{\bar{\imath}}(\overline{A}, A)$. In addition, we can then define a pseudoinverse metric $\tilde{N}^{i,\bar{\jmath}}(\overline{A}, A)$ such that

$$
\widetilde{N}^{i,\bar{\jmath}}(\overline{A}, A) \widetilde{N}_{\bar{\jmath},k}(\overline{A}, A) = (P_{\mathbb{B}^{(A)}})^i_k. \tag{37}
$$

For this gauge fixing prescription, the TDVP equations can then be rewritten as

$$
\dot{A}^i = \widetilde{N}^{i,\bar{\jmath}}(\overline{A}, A) H_{\bar{\jmath}}(\overline{A}, A). \tag{38}
$$

B. Geometric construction

As for the evolution produced by $\widetilde{S}_{\mathbb{H}}$ in the full Hilbert space, one can now define the orthogonal projector onto the space orthogonal to the uMPS $| \Psi(A) \rangle$ as

$$
\hat{P}_0(\overline{A}, A) = \hat{\mathbb{1}} - \frac{| \Psi(A) \rangle \langle \Psi(\overline{A}) |}{N(\overline{A}, A)} \tag{39}
$$

and observe that

$$
\widetilde{N}_{\bar{\imath},j}(\overline{A}, A) = \frac{\langle \partial_{\bar{\imath}} \Psi(\overline{A}) | \hat{P}_0(\overline{A}, A) | \partial_j \Psi(A) \rangle}{N(\overline{A}, A)},
$$

$$
H_{\bar{\imath}}(\overline{A}, A) = \frac{\langle \partial_{\bar{\imath}} \Psi(\overline{A}) | \hat{P}_0(\overline{A}, A) \hat{H} | \Psi(A) \rangle}{N(\overline{A}, A)}.
$$

The TDVP equation [Eq. (31)] can thus be rewritten as

$$
\begin{aligned}
&\langle \partial_{\bar{\imath}} \Psi(\overline{A}(t)) | \hat{P}_0(\overline{A}(t), A(t)) | \partial_j \Psi(A(t)) \rangle \dot{A}^j(t) \\
&= \langle \partial_{\bar{\imath}} \Psi(\overline{A}(t)) | \hat{P}_0(\overline{A}(t), A(t)) \hat{H} | \Psi(A(t)) \rangle,
\end{aligned}
$$

which is the same solution that is obtained if one tries to express that $\dot{A}$ minimizes the norm of the difference between both sides of the Schrödinger equation

$$
\| \hat{P}_0(\overline{A}, A)[i | \partial_j \Psi(A) \rangle \dot{A}^j - \hat{H} | \Psi(A) \rangle] \|,
$$

where one only considers the components orthogonal to the current state $| \Psi(A) \rangle$. Note that, as before, this problem does not have a unique minimum and there are many choices $\dot{A}$ that lead to the same physical state $\frac{d}{dt} | \Psi(A(t)) \rangle$. Restricting to $\dot{A} \in \mathbb{B}^{(A)}$ selects a unique solution. Equivalently, one can just

minimize $\||i\,|\partial_j\Psi(A)\rangle\,\dot{A}^j - \hat{H}\,|\Psi(A)\rangle\|$ for all choices $\dot{A} \in \mathbb{B}^{(A)}$, in which case the orthogonality with respect to $|\Psi(A)\rangle$ is contained in the parametrization and does not have to be included explicitly.

We can also define a projector in Hilbert space that projects onto $\mathbb{T}_0^{|\Psi(A)\rangle\perp}$ as

$$\hat{P}_{\mathbb{T}_0^\perp} = |\partial_i\Psi(A)\rangle\,\widetilde{N}^{i,\bar{j}}(\overline{A},A)\,\langle\partial_{\bar{j}}\Psi(\overline{A})|\,. \tag{40}$$

Note that this projector is in fact independent of the gauge fixing prescription that was used to define $\widetilde{N}^{i,\bar{j}}$. One can check that a sufficient condition such that $\hat{P}_{\mathbb{T}_0^\perp}$ acts as a projector is that $\widetilde{N}^{i,\bar{j}}\widetilde{N}_{\bar{j},k}\widetilde{N}^{k,\bar{l}} = \widetilde{N}^{i,\bar{l}}$, where we assume that only tangent vectors $|\partial_k\Psi(A)\rangle \perp |\Psi(A)\rangle$ are involved and $\langle\Psi(\overline{A})|\Psi(A)\rangle = 1$ such that $\langle\partial_{\bar{j}}\Psi(\overline{A})|\partial_k\Psi(A)\rangle = N_{\bar{j},k} = \widetilde{N}_{\bar{j},k}$. The TDVP equations can now be written in Hilbert space as

$$i\frac{d}{dt}\,|\Psi(A(t))\rangle = \hat{P}_{\mathbb{T}_0^\perp}\hat{H}\,|\Psi(A(t))\rangle\,. \tag{41}$$

C. Symplectic properties of real-time evolution

We now associate to any time-independent operator $\hat{F} \in \mathbb{L}(\mathbb{H})$ the function

$$F : (\overline{A},A) \mapsto \frac{\langle\Psi(\overline{A})|\hat{F}|\Psi(A)\rangle}{\langle\Psi(\overline{A})|\Psi(A)\rangle} \tag{42}$$

that maps the coordinates $(\overline{A},A)$ of a uMPS $|\Psi(A)\rangle$ in the manifold $\mathcal{M}$ to its expectation value. In addition, for any two functions $F(\overline{A},A)$ and $G(\overline{A},A)$, we define a Poisson bracket $\{F,G\}$ as

$$\{F,G\}(\overline{A},A) = -i\partial_i F(\overline{A},A)\widetilde{N}^{i,\bar{j}}(\overline{A},A)\partial_{\bar{j}}G(\overline{A},A)$$
$$+ i\partial_i G(\overline{A},A)\widetilde{N}^{i,\bar{j}}(\overline{A},A)\partial_{\bar{j}}F(\overline{A},A)\,. \tag{43}$$

We can then write the evolution of the expectation value $O(\overline{A}(t),A(t))$ for a solution $A(t)$ of the TDVP equations as

$$\frac{d}{dt}O(\overline{A}(t),A(t)) = \{O,H\}(\overline{A}(t),A(t))\,. \tag{44}$$

The generalization for time-dependent operators $\hat{O}(t)$ is straightforwardly given by

$$\frac{d}{dt}O(\overline{A}(t),A(t),t)$$
$$= \{O,H\}(\overline{A}(t),A(t),t) + \frac{\partial O}{\partial t}(\overline{A}(t),A(t),t)\,. \tag{45}$$

However, we can only really compare this to the symplectic structure of classical Hamiltonian dynamics if the definition in Eq. (43) satisfies all properties of the Poisson bracket, not only the antisymmetry in its arguments but also the Jacobi identity

$$\{F,\{G,H\}\} + \{G,\{H,F\}\} + \{H,\{F,G\}\} = 0\,. \tag{46}$$

Writing out this lengthy equation and doing some index substitutions, the nontrivial part of this identity that needs to be checked is whether

$$F_i G_k H_{\bar{j}}(\widetilde{N}^{i,\bar{j}}\partial_{\bar{j}}\widetilde{N}^{k,\bar{l}} - \widetilde{N}^{k,\bar{j}}\partial_{\bar{j}}\widetilde{N}^{i,\bar{l}}) = 0$$

together with its cyclic permutations. Let us first explain the usual derivation, which is not applicable here as we explain

afterwards. If $\widetilde{N}^{k,\bar{l}}$ would be a proper inverse of $\widetilde{N}_{\bar{l},m}$, we could write

$$\partial_{\bar{j}}\widetilde{N}^{k,\bar{l}} = -\widetilde{N}^{k,\overline{m}}\partial_{\bar{j}}\widetilde{N}_{\overline{m},n}\widetilde{N}^{n,\bar{l}}$$

and use $\partial_{\bar{j}}\widetilde{N}_{\overline{m},n} = \partial_{\overline{m}}\widetilde{N}_{\bar{j},n} = \partial_{\bar{j}}\partial_{\overline{m}}\partial_n\ln(N)$ to show that the expression inside the brackets is identically zero. However, because $\widetilde{N}^{k,\bar{l}}$ is only a pseudoinverse, we can not use the previous expression for $\partial_{\bar{j}}\widetilde{N}^{k,\bar{l}}$. In general, the terms in the round brackets are not zero. It is only when they act on covariant vectors $F_i G_k H_{\bar{j}}$ that originate from functions which are associated with physical operators and which inherit the gauge invariance of the uMPS that we can show that that this expression vanishes. The pseudoinverse $\widetilde{N}^{i,\bar{j}}$ is such that $H_{\bar{j}} = \widetilde{N}_{\bar{l},m}\widetilde{N}^{m,\bar{n}}H_{\bar{n}}$ for gauge-invariant functions H. If we insert this explicitly in the expression above and use $\widetilde{N}^{k,\bar{l}}\widetilde{N}_{\bar{l},m} = (P_{\mathbb{B}})^k{}_m$ [Eq. (37)], we obtain

$$(\partial_{\bar{j}}\widetilde{N}^{k,\bar{l}})\widetilde{N}_{\bar{l},m}\widetilde{N}^{m,\bar{n}} = -\widetilde{N}^{k,\bar{l}}\partial_{\bar{j}}\widetilde{N}_{\bar{l},m}\widetilde{N}^{m,\bar{n}} + (\partial_{\bar{j}}(P_{\mathbb{B}})^k{}_m)\widetilde{N}^{m,\bar{n}}$$

and thus

$$(\widetilde{N}^{i,\bar{j}}\partial_{\bar{j}}\widetilde{N}^{k,\bar{l}} - \widetilde{N}^{k,\bar{j}}\partial_{\bar{j}}\widetilde{N}^{i,\bar{l}})\widetilde{N}_{\bar{l},m}\widetilde{N}^{m,\bar{n}}$$
$$= -(\widetilde{N}^{i,\bar{j}}\widetilde{N}^{k,\bar{l}}\partial_{\bar{j}}\widetilde{N}_{\bar{l},m} - \widetilde{N}^{k,\bar{j}}\widetilde{N}^{i,\bar{l}}\partial_{\bar{j}}\widetilde{N}_{\bar{l},m})\widetilde{N}^{m,\bar{n}}$$
$$+ (\widetilde{N}^{i,\bar{j}}\partial_{\bar{j}}(P_{\mathbb{B}})^k{}_m - \widetilde{N}^{k,\bar{j}}\partial_{\bar{j}}(P_{\mathbb{B}})^i{}_m)\widetilde{N}^{m,\bar{n}}$$
$$= (\widetilde{N}^{i,\bar{j}}\partial_{\bar{j}}(P_{\mathbb{B}})^k{}_m - \widetilde{N}^{k,\bar{j}}\partial_{\bar{j}}(P_{\mathbb{B}})^i{}_m)\widetilde{N}^{m,\bar{n}}\,,$$

where we have now properly used $\partial_{\bar{j}}\widetilde{N}_{\overline{m},n} = \partial_{\overline{m}}\widetilde{N}_{\bar{j},n}$. To evaluate the remaining expression, we need the derivative of the projector $P_{\mathbb{B}}$ onto the horizontal subspace. Note that for any vector B, the action of $P_{\mathbb{B}}$ on B is to replace it by $B + \mathcal{N}_0^{(A)}(X)$ for some X such that $B + \mathcal{N}_0^{(A)}(X) \in \mathbb{B}$. The map $\mathcal{N}_p^{(A)}$ depends only holomorphically on A, and the only dependence on $\overline{A}$ can be in the specific X that was used to make the output satisfy the gauge fixing condition. This implies that for any B, we obtain

$$[\partial_{\bar{j}}(P_{\mathbb{B}})^k{}_m]B^m$$
$$= \partial_{\bar{j}}[(P_{\mathbb{B}})^k{}_m B^m] = \partial_{\bar{j}}[B^k + \mathcal{N}_0^{(A)k}(X)] = \mathcal{N}_0^{(A)k}(\partial_{\bar{j}}X)$$

from which we can infer that the range of $\partial_{\bar{j}}(P_{\mathbb{B}})^k{}_m$ is always in the vertical subspace $\mathbb{N}_0^{(A)}$. The remaining two terms cancel because $F_i\partial_{\bar{j}}(P_{\mathbb{B}})^i{}_m = 0$ and $G_k\partial_{\bar{j}}(P_{\mathbb{B}})^k{}_m = 0$. In conclusion, the Jacobi identity is only satisfied when using gauge-invariant functions. In that case, we are really working on the physical manifold $\mathcal{M} \subset \mathbb{H}$, which unfortunately we only know how to parametrize globally using the overcomplete parametrization in terms of the tensor $A \in \mathcal{A}$.

Indeed, it was shown in Ref. 20 that the manifold of uMPS $\mathcal{M}$ is a Kähler manifold, which implies that its metric also defines a symplectic structure, i.e., a real two-form $\omega = i\tilde{N}_{\bar{i},j}dz^j \wedge d\overline{z}^{\bar{i}}$ that is closed. The fact that it is closed ($d\omega = 0$) relies on $\partial_{\bar{k}}\tilde{N}_{\bar{i},j} = \partial_{\bar{i}}\tilde{N}_{\bar{k},j}$ and thus expresses the essential property for having the Jacobi identity. We can complete the relationship between ω and the Poisson bracket by defining for every (gauge-invariant) function F the *Hamiltonian vector field*

$$X_F(\overline{A},A) = -iF_{\bar{i}}(\overline{A},A)\tilde{N}^{\bar{i}j}(\overline{A},A)\partial_j + iF_j(\overline{A},A)\tilde{N}^{\bar{i}j}(\overline{A},A)\partial_{\bar{i}} \tag{47}$$

and check that

$$\{F,G\} = dF(X_G) = \omega(X_F, X_G). \tag{48}$$

These relations are familiar from classical Hamiltonian mechanics, the only less common ingredient being that the most natural description of phase space is in terms of complex coordinates.

Under exact integration of the TDVP equations for a time-independent Hamiltonian, the antisymmetry of the Poisson bracket results in $\dot{H} = \{H, H\} = 0$, which implies that the energy expectation value $H(\overline{A}(t), A(t))$ is an exact conserved quantity of the TDVP equations. The symplectic properties of the TDVP also conserve other symmetries. Assume that the Hamiltonian is invariant under the action of a unitary symmetry operator $\hat{U}$, such that $[\hat{H}, \hat{U}] = 0$. In order to be able to transfer this symmetry to the uMPS manifold $\mathcal{M}$, we need to assume that for any state $|\Psi(A)\rangle \in \mathcal{M}$, the action of $\hat{U}$ is mapped to a new state $|\Psi(A_U(A))\rangle = \hat{U} |\Psi(A)\rangle \in \mathcal{M}$. One particular class of symmetries that fulfill this condition are those for which the symmetry operators $\hat{U}$ decompose into a product of one-site operators $\hat{U} = \prod_{n \in \mathbb{Z}} \hat{u}_n$ with $\hat{u}$ site independent. We then obtain $A_U^s(A) = \langle s|\hat{u}|t\rangle A^t$. Because of the unitarity of $\hat{U}$, we have $N(\overline{A}_U, A_U) = N(A, A)$, where we omit the explicit dependence of A_U on A for the sake of brevity. By taking the logarithm, followed by differentiating with respect to $\overline{A}^{\overline{l}}$ and A^l, we obtain

$$\partial_{\overline{i}}\overline{A}_U^{\overline{j}}\widetilde{N}_{\overline{j},k}(\overline{A}_U, A_U)\partial_l A_U^k = \widetilde{N}_{\overline{i},l}(\overline{A}, A). \tag{49}$$

The condition $[\hat{H}, \hat{U}] = 0$ also allows us to conclude that $H(\overline{A}_U, A_U) = H(\overline{A}, A)$, from which we find

$$\partial_{\overline{i}}\overline{A}_U^{\overline{j}}H_{\overline{j}}(\overline{A}_U, A_U) = H_{\overline{i}}(\overline{A}, A). \tag{50}$$

The (modified) metric and the gradient thus transform covariantly under the symmetry transformation and can be used to transform the TDVP equation [Eq. (31)] into

$$+i\partial_{\overline{i}}\overline{A}_U^{\overline{j}}(A(t))\widetilde{N}_{\overline{j},k}[\overline{A}_U(\overline{A}(t)), A_U(A(t))]\frac{d}{dt}A_U^k(A(t))$$
$$= \partial_{\overline{i}}\overline{A}_U^{\overline{j}}(A(t))H_{\overline{j}}(\overline{A}_U(\overline{A}(t)), A_U(A(t))].$$

By using the injectivity of the map $A_U(A)$, we can eliminate the Jacobian $\partial_{\overline{i}}\overline{A}_U^{\overline{j}}(A)$ in order to obtain the correct TDVP equation in terms of the new coordinates $A_U(t)$. Hence, it should be straightforward to implement a symmetry-adapted version of the TDVP with increased computational efficiency.

Finally, we study the case where the symmetry operator $\hat{U}$ corresponds to a continuous symmetry generated by the Hermitian generator $\hat{K} \in \mathbb{L}(\mathbb{H})$, with $[\hat{K}, \hat{H}] = 0$. Thus, the expectation value of the generator $\hat{K}$ is conserved under exact evolution. We define a one-parameter family of transformations $\hat{U}(\epsilon) = \exp(i\epsilon\hat{K})$. Since we require that for every uMPS $|\Psi(A)\rangle \in \mathcal{M}$, $\hat{U}(\epsilon)|\Psi(A)\rangle = |\Psi(A_U(A, \epsilon))\rangle \in \mathcal{M}$, we can differentiate this defining relation with respect to ϵ and set $\epsilon = 0$ in order to learn

$$i\hat{K}|\Psi(A)\rangle = \frac{\partial A_U^i}{\partial \epsilon}(A, 0)|\partial_i \Psi(A)\rangle. \tag{51}$$

The action of $\hat{K}$ on a uMPS $|\Psi(A)\rangle$ thus has to be exactly captured in $\mathbb{T}_0^{|\Psi(A)\rangle}$, so that we can write

$$\hat{P}_{\mathbb{T}_{\overline{0}}^{\perp}}\hat{K}'|\Psi(A)\rangle = \hat{K}'|\Psi(A)\rangle,$$

where we have used the definition $\hat{K}' = \hat{K} - K(\overline{A}, A)$. We then obtain

$$\{H, K\}(\overline{A}, A)$$
$$= \frac{\langle\Psi(\overline{A})|\hat{H}'\hat{K}' - \hat{K}'\hat{H}'|\Psi(A)\rangle}{\langle\Psi(\overline{A})|\Psi(A)\rangle}$$
$$= \frac{\langle\Psi(\overline{A})|[\hat{H} - H(\overline{A}, A), \hat{K} - K(\overline{A}, A)]|\Psi(A)\rangle}{\langle\Psi(\overline{A})|\Psi(A)\rangle} = 0. \tag{52}$$

Generators of continuous symmetry transformations are thus constants of motion of the evolution according to the time-dependent variational principle, provided that the symmetry transformation can be captured exactly in the uMPS manifold $\mathcal{M}$.

D. Properties of imaginary-time evolution

The TDVP equation [Eq. (31)] can also be used to simulate imaginary-time evolution by setting $t = -i\tau$. Stationary solutions of the TDVP equation satisfy $H_{\overline{j}}(\overline{A}, A) = 0$, where H is the energy functional that is the central quantity of the time-independent variational principle (TIVP). Hence, stationary solutions of both real- or imaginary-time evolution governed by the TDVP equation correspond to extremal solutions of the TIVP. But, whereas real-time evolutions do typically not converge and are thus not required to end up in such extremal solutions, imaginary-time evolution necessarily has to converge for $\tau \to \infty$ to a solution with $H_{\overline{j}}(\overline{A}, A) = 0$. The reason for this is the rate of change of the energy expectation value given by

$$\frac{d}{d\tau}H = -2H_i\widetilde{N}^{i,\overline{j}}H_{\overline{j}} \leqslant 0, \tag{53}$$

where we have omitted the arguments $(\overline{A}(\tau), A(\tau))$ for the sake of brevity. Hence, the energy expectation value decreases monotonically under imaginary-time evolution. Note that even for our infinite-size system, the previous equation makes sense since H is proportional to the number of sites $|\mathbb{Z}|$, and the pseudoinverse metric contains a factor $|\mathbb{Z}|^{-1}$.

In the full Hilbert space $\mathbb{H}$, imaginary-time evolution will converge any random initial state to the exact ground state, provided that the initial state is not orthogonal to this ground state. Note that imaginary-time evolution in combination with the modified Schrödinger equation [Eq. (30)] does not change the norm of the state. Imaginary-time evolution then describes a continuous version of steepest descent for a convex energy function $H(\overline{\Psi}, \Psi)$ in the convex subspace of constant norm $\langle\Psi|\Psi\rangle$ and thus converges monotonically to the unique minimum. [The full energy function $H(\overline{\Psi}, \Psi)$ in the restriction of $\mathbb{H}$ to a convex subspace of constant norm and phase can have saddle points corresponding to higher eigenstates, but has a single minimum corresponding to the ground state.]

In the restricted manifold of uMPS, the energy functional $H(\overline{A}, A)$ might have many local minima, and there is no guarantee that the flow of the TDVP converges towards

the global minimum (which is assumed to provide the best approximation of the exact ground state). However, if the uMPS manifold $\mathcal{M}$ is able to accurately approximate the exact imaginary-time evolution, one can hope that the flow inherits the global minimizing character of the exact imaginary-time flow and does indeed converge to the global optimum for most random initial states. Note that imaginary-time evolution according to the TDVP equation [Eq. (31)] does not resemble a simple steepest descent in parameter space since the (pseudoinverse) metric explicitly takes the geometry of the manifold into account. It would thus be worthwhile to investigate geometrically covariant formulations of more advanced optimization methods as discussed in Ref. 41 for finding a uMPS ground-state approximation.

E. A simple implementation

We now discuss a simple first-order implementation, which can be considered as an improved version of the algorithm originally introduced in Ref. 15. More advanced numerical integration schemes will be discussed elsewhere.[40]

Let us first discuss how to compute $\tilde{N}^{i,j} H_{\bar{j}}(\overline{A}, A)$. For either choice of the gauge fixing conditions [Eqs. (15) or (16)], one nonlocal term survives in the expression for $\langle \Phi_p(\overline{B}; \overline{A}) | \hat{H} | \Psi(A) \rangle$ [Eq. (34)], which requires the computation of the pseudoinverse of $\mathbb{1} - \mathbb{E}$. An exact computation of $(\mathbb{1} - \mathbb{E})^\mathsf{P}$ would be an operation that scales as $\mathcal{O}(D^6)$, but an iterative strategy is also possible. If we represent B as $\mathcal{B}(X)$ [Eq. (19)], so that the left gauge fixing conditions are fulfilled, we have to compute

$$(K| = (l|\mathbb{H}_{AA}^{AA}(\mathbb{1} - \mathbb{E})^\mathsf{P} = (l|\mathbb{H}_{AA}^{AA}\mathbb{Q}(\mathbb{1} - \mathbb{Q}\mathbb{E}\mathbb{Q})^{-1}\mathbb{Q}, \quad (54)$$

where the last $\mathbb{Q}$ is obsolete and where $(l|\mathbb{H}_{AA}^{AA}\mathbb{Q}$ can be computed efficiently. Since the action of $(\mathbb{1} - \mathbb{Q}\mathbb{E}\mathbb{Q})$ on a vector $(K|$ can also be implemented as an operation with computational efficiency $\mathcal{O}(D^3)$ using the maps $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ [Eqs. (3) and (4)], and since $(\mathbb{1} - \mathbb{Q}\mathbb{E}\mathbb{Q})$ itself is nonsingular, an iterative solver such as the biconjugate gradient stabilized method can be used to compute $(K|$ with a computational cost that scales as $\mathcal{O}(D^3)$. Setting $C^{s,t} = \sum_{u,v=1}^{d} \langle s,t|\hat{h}|u,v \rangle A^u A^v$, we then also define

$$F = \sum_{s,t=1}^{d} V_L^{s\dagger} l^{1/2} C^{s,t} r A^{t\dagger} r^{-1/2}$$

$$+ \sum_{s,t=1}^{d} V_L^{s\dagger} l^{-1/2} A^{t\dagger} l C^{t,s} r^{1/2} + \sum_{s=1}^{d} V_L^{s\dagger} l^{-1/2} K A^s r^{1/2}, \quad (55)$$

in order to obtain

$$\langle \Phi_p(\overline{\mathcal{B}}(\overline{X})) | \hat{H} | \Psi(A) \rangle = 2\pi \delta(p)\, \mathrm{tr}[X^\dagger F]. \quad (56)$$

The solution to the minimization of $\| |\Phi_0(B)\rangle - \hat{H} |\Psi(A)\rangle \|^2$ is then given by $B = \mathcal{B}(F)$.

A simple first-order update scheme is obtained by using the Euler update rule $A(t + dt) = A(t) - i\, dt\, \mathcal{B}(F)$ in case of real-time evolution or $A(\tau + d\tau) = A(\tau) - d\tau\, \mathcal{B}(F)$. F is computed using the current state $A(t)$ or $A(\tau)$ and dt or $d\tau$ is the chosen real- or imaginary-time step. This implementation for uniform matrix product states requires

$\mathcal{O}(N_{\mathrm{iter}} D^3)$ operations, with N_{iter} the number of iterations necessary in the iterative eigensolvers for l, r and in the iterative linear solver for K.

In the case where $A(t)$ also satisfies the left orthonormalization condition $\sum_{s=1}^{d} A^s(t)^\dagger A^s(t) = \mathbb{1}_D$, the left gauge fixing condition for the tangent vector B imposes that this orthonormalization is preserved up to first order: $\sum_{s=1}^{d} A^s(t + dt)^\dagger A^s(t + dt) = \mathbb{1}_D + (dt)^2 \sum_{s=1}^{d} B^{s\dagger} B^s$. The terms of order dt are zero because they correspond exactly to the left-hand side of the left gauge fixing condition [Eq. (15)] and its Hermitian conjugate. We can then use an improved update rule such that $A(t + dt)$ satisfies the left orthonormalization exactly. We therefore define $[A]$ as a matrix representation of A with dimension $Dd \times D$, where the physical index $s = 1, \ldots, d$ and the row index of the matrices A^s is combined: $[A]_{(\alpha s),\beta} = A_{\alpha,\beta}^s$. The left orthonormalization expresses the isometric character of $[A]$ (i.e., $[A]^\dagger[A] = \mathbb{1}_D$). We now state the alternative update rule for real-time evolution

$$[A(t + dt)] = \exp(-i\, dt\{[B][A(t)]^\dagger + [A(t)][B]^\dagger\})[A(t)] \quad (57)$$

or for imaginary-time evolution

$$[A(t + dt)] = \exp(-d\tau\{[B][A(t)]^\dagger - [A(t)][B]^\dagger\})[A(t)]. \quad (58)$$

Expanding the exponential to first order and using $[B]^\dagger[A(t)] = 0$ shows that this update rule is equivalent at first order in dt. Since the argument of the matrix exponential is anti-Hermitian, the isometric character of A is exactly preserved. In case of large local dimension d, it might be disadvantageous to compute a matrix exponential of a $dD \times dD$ matrix. We can expand the exponential into its Taylor series and, using the properties of A and B, resum the series to obtain

$$A^s(t + dt) = A^s(t)\cos(dt|B|) + i B^s |B|^{-1}\sin(dt|B|) \quad (59)$$

in case of real-time evolution, or

$$A^s(\tau + d\tau) = A^s(\tau)\cos(d\tau|B|) - B^s |B|^{-1}\sin(d\tau|B|) \quad (60)$$

in case of imaginary-time evolution. In these equations, the $D \times D$ matrix $|B|$ is given by

$$|B| = ([B]^\dagger[B])^{1/2} = \left(\sum_{s=1}^{d} B^{s\dagger} B^s\right)^{1/2}. \quad (61)$$

Hence, instead of one matrix exponential of a $dD \times dD$ matrix, we have to compute a matrix sine and cosine of a $D \times D$ matrix, which boils down to two matrix exponentials of this $D \times D$ matrix.

F. Error and convergence measures

According to the geometric formulation of the TDVP [Eq. (41)], we can assess the difference between the TDVP evolution and the exact evolution as

$$\epsilon(\overline{A}, A) = \|[\hat{\mathbb{1}} - \hat{P}_{\mathbb{T}_0^\perp}][\hat{H} - H(\overline{A}, A)]|\Psi(A)\rangle\|$$

$$= \sqrt{\Delta H(\overline{A}, A)^2 - \eta(\overline{A}, A)^2}, \quad (62)$$

where we have defined the norm of the TDVP evolution vector as

$$\eta(\overline{A}, A) = \| \hat{P}_{\mathbb{T}_0^\perp} \hat{H} | \Psi(A) \rangle \| \tag{63}$$

and the energy deviation as

$$\Delta H(\overline{A}, A) = \langle \Psi(\overline{A}) | (\hat{H} - H(\overline{A}, A))^2 | \Psi(A) \rangle^{1/2}. \tag{64}$$

For an infinite system, each of these quantities diverges as $|\mathbb{Z}|^{1/2}$. This is an infinitesimal manifestation of the orthogonality catastrophe, which for the case of uMPS expresses the fact that any two injective uMPS that are not gauge equivalent are necessarily orthogonal. Consequently, even the infinitesimal variation from $|\Psi(A)\rangle$ to $|\Psi(A + dA)\rangle$ corresponds to a state $|\Phi_0^{(A)}(dA)\rangle$ with diverging norm. However, we are mostly interested in local properties of systems and should thus use error measures based on regions of finite size. A suitable length scale for such a region which guarantees good global properties of the state is given by the correlation length ξ. For example, for a Hamiltonian $\hat{H} = \sum_{n \in \mathbb{Z}} \hat{T}^n \hat{h} \hat{T}^{-n}$ where $\hat{T}$ is the shift operator and the local interaction term $\hat{h}$ only acts on nearest-neighbor sites, we obtain

$$\Delta H(\overline{A}, A)^2$$
$$= \langle \Psi(\overline{A}) | \left[\sum_{n \in \mathbb{Z}} \hat{T}^n (\hat{h} - h(\overline{A}, A)) \hat{T}^{-n} \right]^2 | \Psi(A) \rangle$$
$$= |\mathbb{Z}| \left(\sum_{n=-1}^{1} \langle \Psi(\overline{A}) | (\hat{h} - h(\overline{A}, A)) \hat{T}^n (\hat{h} - h(\overline{A}, A)) | \Psi(A) \rangle \right.$$
$$\left. + 2(l | \mathbb{E}_{AA}^C (\mathbb{1} - \mathbb{E})^{\mathsf{P}} \mathbb{E}_C^{AA} | r) \right). \tag{65}$$

The first term results in

$$\langle \Psi | (\hat{h} - h)(\hat{h} - h) | \Psi \rangle$$
$$= \sum_{s,t,u,v=1}^{d} \langle u, v | (\hat{h} - h)^2 | s, t \rangle (l | A^s A^t \otimes \overline{A}^u \overline{A}^v | r)$$
$$= \Delta h(\overline{A}, A)^2$$

and

$$\langle \Psi | (\hat{h} - h) \hat{T} (\hat{h} - h) | \Psi \rangle$$
$$= \sum_{r,s,t,u,v,w=1}^{d} \langle u, v, w | (\hat{h} - h) \hat{T} (\hat{h} - h) \hat{T}^{-1} | r, s, t \rangle$$
$$\times (l | A^r A^s A^t \otimes \overline{A}^u \overline{A}^v \overline{A}^w | r),$$

for $n = 0$ and 1, respectively, and in the complex conjugate of the last expression for $n = -1$. By taking out the diverging factor $|Z|^{1/2}$ from $\Delta H(\overline{A}, A)$, we obtain a local and finite measure for the energy deviation which is based on the correlation between any two Hamiltonian terms up to the correlation length. We similarly define a local measure

$$\widetilde{\eta}(\overline{A}, A) = |Z|^{-1/2} \eta(\overline{A}, A)$$
$$= \left[(l | \mathbb{E}_B^{B'} | r) + (l | \mathbb{E}_B^A \mathbb{Q} (\mathbb{1} - e^{ip} \mathbb{Q} \mathbb{E} \mathbb{Q})^{-1} \mathbb{Q} \mathbb{E}_A^{B'} | r) \right.$$
$$\left. + (l | \mathbb{E}_A^{B'} \mathbb{Q} (\mathbb{1} - e^{-ip} \mathbb{Q} \mathbb{E} \mathbb{Q})^{-1} \mathbb{Q} \mathbb{E}_B^A | r) \right],$$

where $(l | \mathbb{E}_A^B | r)$ was assumed. Using the representation $B = \mathcal{B}(x)$ with $x = F$ as constructed in Eq. (55), we obtain

$$\widetilde{\eta}(\overline{A}, A) = \sqrt{\text{tr}[F^\dagger F]} = \| F \|_{\mathrm{F}}, \tag{66}$$

where $\| \ldots \|_{\mathrm{F}}$ denotes the Frobenius norm. A local measure for the error made by the TDVP evolution with respect to the full evolution is then given by

$$\widetilde{\varepsilon}(\overline{A}, A) = \sqrt{\Delta H(\overline{A}, A)^2 / |\mathbb{Z}| - \widetilde{\eta}(\overline{A}, A)^2}. \tag{67}$$

When the exact time evolution is accurately captured in the manifold of (uniform) matrix product states, $\widetilde{\varepsilon}$ contains the difference of two terms which are of comparable size. In addition, the computation of $\Delta H(\overline{A}, A)^2 / |\mathbb{Z}|$ contains four terms that can be both positive and negative and can neutralize each other. This can result in large numerical errors in the computation of these quantities. A better strategy for evaluating $\widetilde{\epsilon}(\overline{A}, A)$ as a sum of strictly positive terms is constructed in Sec. VII B.

For imaginary-time evolution, we expect the evolution to converge to a point $|\Psi(A^*)\rangle$ where $\widetilde{\eta}(\overline{A}^*, A^*) = 0$. We can easily motivate that at any point $|\Psi(A(\tau))\rangle$ in the evolution, $\widetilde{\eta}(\overline{A}(\tau), A(\tau))$ can be used as a local measure for the difference between the current uMPS and the final state $|\Psi(A^*)\rangle$. Indeed, we can show that the change of the expectation value of an operator $\hat{O}$ with support on N sites satisfies

$$\left| \frac{d}{d\tau} O(\overline{A}(\tau), A(\tau)) \right| \leqslant c \| \hat{O} \| (2\xi + N + 2) \widetilde{\eta}(\overline{A}(\tau), A(\tau)),$$

where c is some constant and ξ is the correlation length, set by the second largest eigenvalue of the transfer matrix as in Eq. (6). One special example is when we look at the expectation value of the Hamiltonian density $\hat{h}$, for which we obtain

$$\frac{d}{d\tau} h(\overline{A}(\tau), A(\tau)) = -\widetilde{\eta}(\overline{A}(\tau), A(\tau))^2.$$

Thus, the energy density $h(\overline{A}(\tau), A(\tau))$ converges quadratically as fast as the local measure for the state error $\widetilde{\eta}$, a result that parallels the quadratic convergence of the total energy in the global state error.

Having reached a minimum $|\Psi(A^*)\rangle$ of the uMPS manifold $[\widetilde{\eta}(\overline{A}, A) = 0]$, we can use the local error measure $\widetilde{\epsilon}(\overline{A}^*, A^*) = \Delta H(\overline{A}^*, A^*) / |Z|^{1/2}$ to assess the difference between the variational optimum and the exact ground state.

IV. VARIATIONAL ANSATZ FOR EXCITATIONS

The DMRG was originally developed for finding ground states of strongly correlated quantum lattice systems in one spatial dimension. By applying the variational principle to a state that is enforced to be orthogonal to previously found states, low-lying excited states on finite lattices can be found. Typically, these are not the states that one is interested in. On the finite lattice with open boundary conditions, the momentum quantum number does not exist. Low-lying excited states can easily be related to boundary effects and have no relation to the momentum eigenstates in the bulk of a macroscopic system. In the thermodynamic limit, the suggested approach fails since any two states are likely to be orthogonal due to the orthogonality catastrophe. Even if we were able to construct

a uMPS approximation for the lowest-lying excited state with momentum zero, the finite excitation energy would spread out over an infinite lattice and is undetectable from computing the expectation value of the energy density. States with a different energy density as the ground state contain a finite density (and thus an infinite number) of elementary excitations. On a more mathematical level, we do not expect the class of matrix product states in the thermodynamic limit to have the correct properties for describing elementary excited states since MPS are normalizable, and excited states with definite momentum are not.

Two different strategies for solving this problem emerge. Information about the spectrum of excited states can be obtained from the pole structure of the Fourier transform of dynamic correlation functions. Initially, algorithms for directly evaluating these correlation functions in frequency domain were developed.[26,42–44] But since the development of the TEBD, time evolution can be approximated so well that modern state-of-the-art algorithms first compute the time-dependent correlation function for some finite interval $t \in [0, T]$, and then compute the Fourier transform.[45,46] Starting from a disturbance in the ground state of a large but finite lattice, the time evolution can be computed for any time T below which the information of the disturbance has not yet reached the edges of the lattice. The finite time T results in a broadening of the spectral function, but by combining advanced linear prediction techniques to extend T beyond the computable range with complex statistical machinery for isolating the location of the poles, a fairly accurate determination of the dispersion relation of the elementary excitation in the Heisenberg model was obtained.[46] Because of the (approximately) linear increase of entropy under time evolution,[47,48] very large bond values are required in order to accurately approximate the time evolution all the way up to time T. This is in sharp contrast with the observation that low-lying excited states also satisfy an approximate area law for the scaling of entanglement entropy. For free-field theories, power-law corrections to the area law were found when the field is in a superposition of its ground state and low-lying excited states.[49–51] For large areas, these are negligible and the area law still holds. An area law was also found for low-lying excited states in an integrable one-dimensional lattice model.[52] In Ref. 53, an area law is proven for all low-energy states, not restricted to eigenstates, of short-range interacting lattice models under some technical conditions, including a sufficiently rapid decay of connected correlation functions and an upper bound on the number of low-lying excitations in a subsystem corresponding to a compact spatial region.

Given these considerations on the entanglement entropy of excited states, an alternative strategy is thus to construct a variational *Ansatz* that is suited to directly probe the spectrum of excited states. Nevertheless, variational *Ansätze* for excited states based on the matrix product concept have not been very common. Studying energy-momentum dispersion relations seems to automatically redirect us to a lattice with periodic boundary conditions, for which the MPS algorithms are less efficient, unless one can work in the thermodynamic limit. However, given the remarks above, a direct construction in the thermodynamic limit seems far from trivial. The first proposal

for a variational *Ansatz* for excitations was made by Rommer and Östlund.[3,4] In their seminal work on MPS, they also suggested to study excitations with momentum $p = 2\pi n/N$ on a ring of length N using the *Ansatz*

$$|\widetilde{\Phi}_p(x)\rangle = \sum_{n=1}^{N} e^{ipn} \hat{T}^n \sum_{\{s_n\}=1}^{d} \mathrm{tr}[x A^{s_1} A^{s_2} \dots A^{s_N}] |s_1 s_2 \dots s_N\rangle,$$
(68)

which allowed them to get an early estimate of the Haldane gap in the spin-1 Heisenberg chain. The matrices A^s are fixed to the value for which the uniform matrix product state $|\Psi(A)\rangle$ (of finite size N) best approximates the ground state, and one can hope that several branches of the energy-momentum spectrum can be captured by different values of x. The rationale of this *Ansatz* is that low-lying excited states can be described as a momentum superposition of a local disturbance, which is encoded in the virtual system using the virtual operator $x \in \mathbb{L}(\mathbb{C}^D)$. Using an analytic series expansion in the system size N, Rommer and Östlund were even able to extrapolate their results to the thermodynamic limit. A different type of variational class corresponds to the so-called "projected entangled multipartite states," given by the *Ansatz*[54]

$$|\Upsilon_p[A]\rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} e^{ipn} \hat{T}^n \sum_{\{s_n\}=1}^{d} \mathrm{tr}[A^{s_1}(1) A^{s_2}(2) \dots A^{s_N}(N)]$$
$$\times |s_1 s_2 \dots s_N\rangle,$$
(69)

which contains a momentum superposition of the non-translation-invariant MPS. Here, all matrices $A^s(n)$ are variational parameters, and different branches of the spectrum are obtained by creating mutually orthogonal states at a fixed momentum p. This specific superposition is expected to be able to introduce long-range information: writing $|\Upsilon_0[A]\rangle$ as a uMPS $|\tilde{\Psi}(\tilde{A})\rangle$ requires a bond dimension $\tilde{D} = ND$ if D represents the bond dimension of the matrices $A^s(n)$. The computational complexity of this algorithm scales as $\mathcal{O}(N^2 D^5)$, and it is thus restricted to lattices of moderate size and small values of the bond dimension D. This last aspect is partially compensated by the higher entanglement that is allowed in this state.

The idea that low-lying excited states can be regarded as (momentum superpositions of) local disturbances on the ground state is of course inspired by the case of quadratic theories, where creation operators $\hat{a}^\dagger(p)$ can be defined that create elementary excitations when acting on the ground state. For elementary excitations, this pointlike structure is often a good assumption. Bijl, Feynman, and Cohen generalized this concept by acting on the ground state with general operators $\hat{O}(p)$, which represent the Fourier transform of some local operator $\hat{O}$ with compact support, for studying excitations in liquid helium.[55–57] If $\{\hat{O}^\alpha\}$ represents a complete set of local observables, then the Feynman-Bijl operator $\hat{O}$ can be expanded as $\hat{O} = c_\alpha \hat{O}^\alpha$ and $\{c_\alpha\}$ can be treated as the set of variational parameters. This *Ansatz* was first used in the context of spin systems by Arovas, Auerbach, and Haldane[58] and is then referred to as the single-mode approximation. The single-mode approximation was first combined with matrix

product states in Ref. 59, and generalized to local operators acting on up to four sites.[60]

It is now clear that the tangent vectors $|\Phi_p(B)\rangle$ defined in Eq. (9) generalize both the construction of Östlund and Rommer, where the excitation is represented as an operator x in the virtual space (choose $B^s = x A^s$), and the single-mode approximation, where the excitation is represented as an operator $\hat{O}$ in the physical space (choose $B^s = \langle s|\hat{O}|t\rangle A^t$). Feynman-Bijl operators with a larger support of $n > 1$ sites are not strictly included in this variational class, but by transferring information along the virtual space all operators acting on $n \approx 2 \log_q D$ sites are effectively included. We can even hope that the D left and D right Schmidt vectors throw away irrelevant information on the nearest sites in favor of keeping relevant information on sites that are further away. In the final section of this paper, we will also discuss a generalized excitation *Ansatz* where we replace the ground-state matrices A on a contiguous block of several sites. The *Ansatz* $|\Phi_p(B)\rangle$ was used on a ring of N sites in Ref. 16. However, the real power of this *Ansatz* is unleashed by using it in the thermodynamic limit, where we can hopefully reproduce the $\mathcal{O}(D^3)$ scaling that we have grown accustomed to from DMRG and that was also reproduced in the TDVP calculations. This of course requires that we can consistently compute the variational excitation energy directly in the thermodynamic limit. In Sec. IV A we will illustrate how to compute the expectation value of the Hamiltonian $\hat{H}$ in full detail, using the same techniques as in the previous section. Being able to formulate our techniques in the thermodynamic limit also allows us to describe topologically nontrivial excitations, as is sketched in Sec. IV B. In one spatial dimension, topologically nontrivial excitations commonly appear in systems with discrete symmetry breaking as kinks or domain walls that interpolate between two ground states with a different value of the order parameter. This *Ansatz* includes the topologically nontrivial analog of the Feynman-Bijl operators, which are the Mandelstam operators.[61] Section IV C also relates our excitation *Ansatz* to the TDVP from the previous section. Finally, by noting that we have a variational estimate not only for the excitation energies but also for the corresponding wave functions, we look at dynamical correlation functions in Sec. IV D.

A. Topologically trivial states

Let $\hat{H}$ be a given translation-invariant Hamiltonian on an infinite lattice, which we assume to contain only nearest-neighbor interactions for reasons of notational simplicity: $\hat{H} = \sum_{n \in \mathbb{Z}} \hat{T}^n \hat{h} \hat{T}^{-n}$ with $\hat{T}$ the shift operator. We assume that the ground state is well approximated by a uMPS $|\Psi(A)\rangle \in \mathcal{M}$. We henceforth assume that A is the value of at least a local, and hopefully the global, minimum. We again assume that the uMPS $|\Psi(A)\rangle$ is pure ($A \in \mathcal{A}$) and normalized to unity, so that $\mathbb{E}$ has a unique eigenvalue 1 and all the other eigenvalues lie strictly within the unit circle. We now apply the time-independent variational principle to the set of states $|\Phi_p^{(A)}(B)\rangle \in \mathbb{T}_p^{|\Psi(A)\rangle}$. Since we are interested in excited states, we need to impose orthogonality to the ground-state approximation $|\Psi(A)\rangle$. We can thus restrict to $\mathbb{T}^{|\Psi(A)\rangle\perp}$. As in the case of TDVP, this restriction enters naturally. We can

thus recycle the parametrization $B = \mathcal{B}^{(A)}(X)$ from Eq. (19) in terms of the $D(d-1) \times D$ matrix X. With A assumed to be fixed throughout this section, we omit the explicit reference to A and $|\Psi(A)\rangle$ in the notation of the states $|\Phi_p(B)\rangle$, the spaces $\mathbb{T}_p^\perp$, and the representation $\mathcal{B}(X)$. Since our variational manifold for excitations corresponds to a linear subspace $\mathbb{T}_p^\perp$ of Hilbert space $\mathbb{H}$, for which we have a linear representation through the series of linear maps $X \mapsto \mathcal{B}(X) \mapsto |\Phi(\mathcal{B}(X))\rangle$, the variational optimization problem reduces to a Rayleigh-Ritz problem and we will have to solve a generalized eigenvalue equation in X. In fact, because of the way the representation $\mathcal{B}(X)$ was constructed, the effective norm matrix constructed in Eq. (20) is proportional to the unit matrix and we end up with normal eigenvalue problem.

Two remarks are in order. First, the *Ansatz* states $|\Phi_p(B)\rangle$ are momentum eigenstates in an infinite volume and can thus not be normalized to unity. Second, unlike for the ground state, we can not restrict to an evaluation of the energy density expectation value. As explained in the Introduction, the finite excitation energy in a momentum eigenstate is spread out over the complete lattice, and the energy density

$$\frac{\langle \Phi_p(\overline{B})|\hat{h}|\Phi_{p'}(B')\rangle}{\langle \Phi_p(\overline{B})|\Phi_{p'}(B')\rangle}$$

is indistinguishable from its ground-state value $h(\overline{A}, A) = \langle \Psi(\overline{A})|\hat{h}|\Psi(A)\rangle$. We thus have to evaluate the expectation value of the full Hamiltonian $\langle \Phi_p(\overline{B})|\hat{H}|\Phi_{p'}(B')\rangle$, where the excitation energy is present as a finite shift [times the infinite normalization $\langle \Phi_p(\overline{B})|\hat{h}|\Phi_{p'}(B')\rangle$] above a divergent contribution from the extensive ground-state energy $H(\overline{A}, A) = |\mathbb{Z}| h(\overline{A}, A)$ [times the infinite normalization $\langle \Phi_p(\overline{B})|\hat{h}|\Phi_{p'}(B')\rangle$]. Subtracting this ground-state energy can quickly become a source of errors, as we have to subtract precisely $|\mathbb{Z}|$ times the ground-state energy density, and counting errors are easily made. The safest strategy is to subtract $H(\overline{A}, A)$ from $\hat{H}$ from the beginning. Note that, unlike in the evaluation of $\langle \Phi_p(\overline{B})|\hat{H}|\Psi(A)\rangle$ that was required for the TDVP, the ground-state energy contribution is not automatically subtracted by restricting to tangent vectors $|\Phi_p(B)\rangle$ that are orthogonal to $|\Psi(A)\rangle$. We thus redefine $\hat{h} \leftarrow \hat{h} - h(\overline{A}, A)$, where $h(\overline{A}, A) = (l|\mathbb{H}_{AA}^{AA}|r)$ [see Eq. (29)]. With this newly defined $\hat{h}$, we obtain $(l|\mathbb{H}_{AA}^{AA}|r) = 0$.

We are now ready to evaluate the effective Hamiltonian appearing in the Rayleigh-Ritz equation. It corresponds to the restriction of the full Hamiltonian to the subspace $\mathbb{T}_p^\perp$. Expanding $\langle \Phi_p(\overline{B})|\hat{H}|\Phi_{p'}(B')\rangle$ is a lot more involved than either the norm $\langle \Phi_p(\overline{B})|\Phi_{p'}(B')\rangle$ or the TDVP gradient $\langle \Phi_p(\overline{B})|\hat{H}|\Psi(A)\rangle$, as we now have to deal with three infinite sums. The three summation indices indicate the position of B, B', and the first site acted upon by $\hat{h}$. In-between these three positions are transfer matrices $\mathbb{E}$, which can be decomposed into connected contributions coming from $\mathbb{QEQ}$ and disconnected contributions coming from $\mathbb{S} = |r)(l|$. Thanks to the redefinition of the Hamiltonian terms $\hat{h}$, we obtain $(l|\mathbb{H}_{AA}^{AA}\mathbb{S} = 0$ and $\mathbb{S}\mathbb{H}_{AA}^{AA}|r) = 0$ and no disconnected contributions coming from $\mathbb{H}_{AA}^{AA}$ can arise. The connected contributions yield finite results, and we are free to introduce substitutions of the summation indices. Disconnected contributions coming from

$\mathbb{E}_B^A$ and $\mathbb{E}_A^{B'}$ might give rise to additional divergences and should be treated carefully. The total expression is of the general format

$$\langle \Phi_p(\overline{B})|\hat{H}|\Phi_{p'}(B')\rangle = \sum_{n=-\infty}^{+\infty} \sum_{n'=-\infty}^{+\infty} \sum_{n_0=-\infty}^{+\infty} e^{ip'n'-ipn}[B \text{ at site } n, B' \text{ at site } n', \text{ and } \hat{h} \text{ on sites } n_0 \text{ and } n_0+1].$$

We first focus on the terms where everything is connected, thus where all transfer operators have been replaced by their corresponding regularized version $\mathbb{QEQ}$. We can now safely introduce the substitution $n' \leftarrow n_c$, $n \leftarrow n_c + \Delta n$, and $n_0 \leftarrow n_c + \Delta n_0$. The summation over n_c immediately yields the momentum-conserving factor $2\pi\delta(p'-p)$ since the terms within the summation are independent of the global position n_c. If we change Δn to n and Δn_0 to n_0 for notational simplicity and omit the overall factor $2\pi\delta(p'-p)$, we are left with

$$(l|\mathbb{H}_{BA}^{B'A}|r) + (l|\mathbb{H}_{AB}^{AB'}|r) + \sum_{n_0=1}^{+\infty}(l|\mathbb{E}_B^{B'}\mathbb{Q}\mathbb{E}^{n_0-1}\mathbb{Q}\mathbb{H}_{AA}^{AA}|r) + \sum_{n_0=-\infty}^{-2}(l|\mathbb{H}_{AA}^{AA}\mathbb{Q}\mathbb{E}^{-n_0-2}\mathbb{Q}\mathbb{E}_B^{B'}|r)$$

$$+ \sum_{n=-\infty}^{-1} e^{-ipn}\big[\theta(n=-1)(l|\mathbb{H}_{BA}^{AB'}|r) + (l|\mathbb{E}_B^A\mathbb{Q}\mathbb{E}^{n-1}\mathbb{Q}\mathbb{H}_{AA}^{B'A}|r) + (l|\mathbb{H}_{AB}^{AA}\mathbb{Q}\mathbb{E}^{-n-1}\mathbb{Q}\mathbb{E}_A^{B'}|r)$$

$$+ \theta(n<-1)(l|\mathbb{E}_B^A\mathbb{Q}\mathbb{E}^{-n-2}\mathbb{Q}\mathbb{H}_{AA}^{AB'}|r) + \theta(n<-1)(l|\mathbb{H}_{BA}^{AA}\mathbb{Q}\mathbb{E}^{-n-2}\mathbb{Q}\mathbb{E}_A^{B'}|r)$$

$$+ \sum_{n_0=1}^{+\infty}(l|\mathbb{E}_B^A\mathbb{Q}\mathbb{E}^{-n-1}\mathbb{Q}\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{n_0-1}\mathbb{Q}\mathbb{H}_{AA}^{AA}|r) + \sum_{n_0=-\infty}^{n-2}(l|\mathbb{H}_{AA}^{AA}\mathbb{Q}\mathbb{E}^{-n_0+n-2}\mathbb{Q}\mathbb{E}_B^A\mathbb{Q}\mathbb{E}^{-n-1}\mathbb{Q}\mathbb{E}_A^{B'}|r)$$

$$+ \theta(n<-2)\sum_{n_0=n+1}^{-2}(l|\mathbb{E}_B^A\mathbb{Q}\mathbb{E}^{-n+n_0-1}\mathbb{Q}\mathbb{H}_{AA}^{AA}\mathbb{Q}\mathbb{E}^{-n_0-2}\mathbb{Q}\mathbb{E}_A^{B'}|r)\big]$$

$$+ \sum_{n=1}^{+\infty} e^{-ipn}\big[\theta(n=1)(l|\mathbb{H}_{AB}^{B'A}|r) + (l|\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{n-1}\mathbb{Q}\mathbb{H}_{BA}^{AA}|r) + (l|\mathbb{H}_{AA}^{AB'}\mathbb{Q}\mathbb{E}^{n-1}\mathbb{Q}\mathbb{E}_B^A|r)$$

$$+ \theta(n>1)(l|\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{n-2}\mathbb{Q}\mathbb{H}_{AB}^{AA}|r) + \theta(n>1)(l|\mathbb{H}_{AA}^{B'A}\mathbb{Q}\mathbb{E}^{n-2}\mathbb{Q}\mathbb{E}_B^A|r)$$

$$+ \sum_{n_0=n+1}^{+\infty}(l|\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{n-1}\mathbb{Q}\mathbb{E}_B^A\mathbb{Q}\mathbb{E}^{n_0-n-1}\mathbb{Q}\mathbb{H}_{AA}^{AA}|r) + \sum_{n_0=-\infty}^{-2}(l|\mathbb{H}_{AA}^{AA}\mathbb{Q}\mathbb{E}^{-n_0-2}\mathbb{Q}\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{n-1}\mathbb{Q}\mathbb{E}_B^A|r)$$

$$+ \theta(n>2)\sum_{n_0=1}^{n-2}(l|\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{n_0-1}\mathbb{Q}\mathbb{H}_{AA}^{AA}\mathbb{Q}\mathbb{E}^{n-n_0-2}\mathbb{Q}\mathbb{E}_B^A|r)\big].$$

The terms on the first line correspond to $n = 0$, i.e., where B and B' are on the same site. Then, we have all the terms corresponding to $n < 0$ and all the terms corresponding to $n > 0$. For most terms, we can immediately evaluate the geometric series for n_0, followed by an evaluation of the additional geometric series in n for some terms. The only exception are the terms with $\theta(n<-2)$ and $\theta(n>2)$, where it is better to first switch the two sums and express the summation bounds of n in terms of n_0. Then, we first evaluate the geometric series in n, followed by the one in n_0. We obtain

$$(l|\mathbb{H}_{BA}^{B'A}|r) + (l|\mathbb{H}_{AB}^{AB'}|r) + e^{+ip}(l|\mathbb{H}_{BA}^{AB'}|r) + e^{-ip}(l|\mathbb{H}_{AB}^{B'A}|r) + (l|\mathbb{E}_B^{B'}(\mathbb{1}-\mathbb{E})^\mathsf{P}\mathbb{H}_{AA}^{AA}|r) + (l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^\mathsf{P}\mathbb{E}_B^{B'}|r)$$

$$+ e^{+ip}(l|\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^\mathsf{P}\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^\mathsf{P}\mathbb{H}_{AA}^{AA}|r) + e^{-ip}(l|\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^\mathsf{P}\mathbb{E}_B^A(\mathbb{1}-\mathbb{E})^\mathsf{P}\mathbb{H}_{AA}^{AA}|r)$$

$$+ e^{+ip}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^\mathsf{P}\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^\mathsf{P}\mathbb{E}_A^{B'}|r) + e^{-ip}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^\mathsf{P}\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^\mathsf{P}\mathbb{E}_B^A|r)$$

$$+ e^{+ip}(l|\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^\mathsf{P}\mathbb{H}_{AA}^{B'A}|r) + e^{-ip}(l|\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^\mathsf{P}\mathbb{H}_{BA}^{AA}|r)$$

$$+ e^{+ip}(l|\mathbb{H}_{AB}^{AA}(\mathbb{1}-e^{+ip}\mathbb{E})^\mathsf{P}\mathbb{E}_A^{B'}|r) + e^{-ip}(l|\mathbb{H}_{AA}^{AB'}(\mathbb{1}-e^{-ip}\mathbb{E})^\mathsf{P}\mathbb{E}_B^A|r)$$

$$+ e^{+2ip}(l|\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^\mathsf{P}\mathbb{H}_{AA}^{AB'}|r) + e^{-2ip}(l|\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^\mathsf{P}\mathbb{H}_{AB}^{AA}|r)$$

$$+ e^{+2ip}(l|\mathbb{H}_{BA}^{AA}(\mathbb{1}-e^{+ip}\mathbb{E})^\mathsf{P}\mathbb{E}_A^{B'}|r) + e^{-2ip}(l|\mathbb{H}_{AA}^{B'A}(\mathbb{1}-e^{-ip}\mathbb{E})^\mathsf{P}\mathbb{E}_B^A|r)$$

$$+ e^{+3ip}(l|\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^\mathsf{P}\mathbb{H}_{AA}^{AA}(\mathbb{1}-e^{+ip}\mathbb{E})^\mathsf{P}\mathbb{E}_A^{B'}|r) + e^{-3ip}(l|\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^\mathsf{P}\mathbb{H}_{AA}^{AA}(\mathbb{1}-e^{-ip}\mathbb{E})^\mathsf{P}\mathbb{E}_B^A|r).$$

The symbolic notation $(\hat{\mathbb{1}} - e^{\pm ip}\mathbb{E})^\mathsf{P} = \mathbb{Q}(\mathbb{1}-e^{\pm ip}\mathbb{Q}\mathbb{E}\mathbb{Q})^{-1}\mathbb{Q}$ was introduced in the previous section. Only for $p=0$ is this truly a pseudoinverse. For $p\neq 0$, the $\hat{\mathbb{1}} - e^{\pm ip}\mathbb{E}$ is not really singular. Nevertheless, we had to separate the eigenvalue $e^{\pm ip}$ with modulus 1 from the operator $e^{\pm ip}\mathbb{E}$ in order to use the formula for the geometric series.

We now consider the contributions resulting from disconnecting either B or B'. They can not be simultaneously disconnected since this would also imply that $\hat{h}$ is disconnected, which we have excluded above. Whenever B' appears on the complete left (right) side of a term, and is separated from the rest by a transfer operator $\mathbb{E}$, there is such a disconnected contribution. We assume that we can still make the substitution to the global position n_c and the relative positions n and n_0. Only making

substitutions that changes the value of finite bounds in the sum result in a possibility of miscounting contributions and making errors. The summation over the global position again yields the momentum conservation. The total (left and right) contribution from disconnecting B is given by [omitting the momentum-conserving factor $2\pi\delta(p'-p)$]

$$(l|E_B^A|r)\left[(l|\mathbb{H}_{AA}^{B'A}|r)\left(\sum_{n=-\infty}^{-1}e^{-ipn}+\sum_{n=2}^{+\infty}e^{-ipn}\right)+(l|\mathbb{H}_{AA}^{AB'}|r)\left(\sum_{n=-\infty}^{-2}e^{-ipn}+\sum_{n=1}^{+\infty}e^{-ipn}\right)\right.$$

$$+\sum_{n=-\infty}^{-1}e^{-ipn}\sum_{n_0=1}^{+\infty}(l|\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{n_0-1}\mathbb{Q}\mathbb{H}_{AA}^{AA}|r)+\sum_{n=3}^{+\infty}e^{-ipn}\sum_{n_0=1}^{n-2}(l|\mathbb{E}_A^{B'}\mathbb{Q}\mathbb{E}^{n_0-1}\mathbb{Q}\mathbb{H}_{AA}^{AA}|r)$$

$$\left.+\sum_{n=-\infty}^{-3}e^{-ipn}\sum_{n_0=n+1}^{-2}(l|\mathbb{H}_{AA}^{AA}\mathbb{Q}\mathbb{E}^{-n_0-2}\mathbb{Q}\mathbb{E}_A^{B'}|r)+\sum_{n=1}^{+\infty}e^{-ipn}\sum_{n_0=-\infty}^{-2}(l|\mathbb{H}_{AA}^{AA}\mathbb{Q}\mathbb{E}^{-n_0-2}\mathbb{Q}\mathbb{E}_A^{B'}|r)\right].$$

These terms should be treated carefully. We expect them to generate a divergence at $p=0$ through a $2\pi\delta(p)$ since we have not expressed orthogonality with respect to the ground state yet. But, for any other $p\neq0$, they should be finite, as orthogonality to the ground state is automatic. By inserting the result for the finite geometric sums in n_0, we obtain for the terms between the square brackets

$$(l|\mathbb{H}_{AA}^{B'A}|r)\big(2\pi\delta(p)-1-e^{-ip}\big)+(l|\mathbb{H}_{AA}^{AB'}|r)\big(2\pi\delta(p)-1-e^{+ip}\big)+\sum_{n=-\infty}^{-1}e^{-ipn}(l|\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)$$

$$+\sum_{n=3}^{+\infty}e^{-ipn}(l|\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}(\mathbb{Q}-\mathbb{Q}\mathbb{E}^{n-2}\mathbb{Q})\mathbb{H}_{AA}^{AA}|r)+\sum_{n=-\infty}^{-3}e^{-ipn}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}(\mathbb{Q}-\mathbb{Q}\mathbb{E}^{-n-2}\mathbb{Q})\mathbb{E}_A^{B'}|r)$$

$$+\sum_{n=1}^{+\infty}e^{-ipn}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r).$$

Since $(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{Q}=(\mathbb{1}-\mathbb{E})^{\mathsf{P}}$, we can now group the third and fourth terms, as well as the fifth and sixth terms, and complete the sums in n to $\sum_{n=-\infty}^{+\infty}e^{\pm ip}=2\pi\delta(p)$, in order to obtain

$$(l|\mathbb{H}_{AA}^{B'A}|r)(2\pi\delta(p)-1-e^{-ip})+(l|\mathbb{H}_{AA}^{AB'}|r)(2\pi\delta(p)-1-e^{+ip})+(2\pi\delta(p)-1-e^{-ip}-e^{-2ip})(l|\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)$$

$$-\sum_{n=3}^{+\infty}e^{-ipn}(l|\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{Q}\mathbb{E}^{n-2}\mathbb{Q}\mathbb{H}_{AA}^{AA}|r)+(2\pi\delta(p)-1-e^{+ip}-e^{+2ip})(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)$$

$$-\sum_{n=-\infty}^{-3}e^{-ipn}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{Q}\mathbb{E}^{-n-2}\mathbb{Q}\mathbb{E}_A^{B'}|r).$$

Finally, we have to compute two converging geometric series in n. Note that the power of $(\mathbb{Q}\mathbb{E}\mathbb{Q})$ starts at one instead of zero (for $n=3$ on line 2 and for $n=-3$ on line 3). We can absorb the part with factor e^{-2ip} from the last term of the first line, and the part with factor e^{+2ip} from the last term of the second line, respectively, in order to have a geometric series in $(\mathbb{Q}\mathbb{E}\mathbb{Q})$ starting at power zero. We hence obtain for the total disconnected contribution of B:

$$(l|E_B^A|r)\big[(l|\mathbb{H}_{AA}^{B'A}|r)(2\pi\delta(p)-1-e^{-ip})+(l|\mathbb{H}_{AA}^{AB'}|r)(2\pi\delta(p)-1-e^{+ip})+(2\pi\delta(p)-1-e^{-ip})(l|\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)$$

$$-e^{-i2p}(l|\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)+(2\pi\delta(p)-1-e^{+ip})(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)$$

$$-e^{+i2p}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)\big].$$

By adding a similar result from disconnecting B', we obtain the final result

$$\langle\Phi_p(\overline{B})|\hat{H}|\Phi_{p'}(B')\rangle$$

$$=\overline{B}^{\bar{i}}H_{\bar{i},j}(p,p')B^j=2\pi\delta(p'-p)H_{\bar{i},j}(p)2\pi\delta(p-p')\big\{(l|\mathbb{H}_{BA}^{B'A}|r)+(l|\mathbb{H}_{AB}^{AB'}|r)+e^{+ip}(l|\mathbb{H}_{BA}^{AB'}|r)+e^{-ip}(l|\mathbb{H}_{AB}^{B'A}|r)$$

$$+(l|\mathbb{E}_B^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)+(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{E}_B^{B'}|r)+e^{+ip}(l|\mathbb{E}_B^{B'}(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)$$

$$+e^{-ip}(l|\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_B^A(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)+e^{+ip}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)$$

$$+e^{-ip}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_B^A|r)+e^{+ip}(l|\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{B'A}|r)+e^{-ip}(l|\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{BA}^{AA}|r)$$

$$+e^{+2ip}(l|\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AB'}|r)+e^{-2ip}(l|\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AB}^{AA}|r)+e^{+ip}(l|\mathbb{H}_{AB}^{AA}(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)$$

$$+e^{-ip}(l|\mathbb{H}_{AA}^{AB'}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^A|r)+e^{+2ip}(l|\mathbb{H}_{BA}^{AA}(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)+e^{-2ip}(l|\mathbb{H}_{AA}^{B'A}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^A|r)$$

$$+e^{+3ip}(l|\mathbb{E}_B^A(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)+e^{-3ip}(l|\mathbb{E}_A^{B'}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_B^A|r)$$

$$\times (l|\mathbb{E}_B^A|r)\big[\big(2\pi\delta(p)-1-e^{-ip}\big)\big((l|\mathbb{H}_{AA}^{B'A}|r)+(l|\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)\big)+\big(2\pi\delta(p)-1-e^{+ip}\big)\big((l|\mathbb{H}_{AA}^{AB'}|r)$$

$$+(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)\big)-e^{-i2p}(l|\mathbb{E}_A^{B'}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)-e^{+i2p}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}$$

$$\times(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_A^{B'}|r)\big](l|\mathbb{E}_A^{B'}|r)\big[\big(2\pi\delta(p)-1-e^{+ip}\big)\big((l|\mathbb{H}_{BA}^{A}|r)+(l|\mathbb{E}_B^A(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)\big)$$

$$+\big(2\pi\delta(p)-1-e^{-ip}\big)\big((l|\mathbb{H}_{AB}^{AA}|r)+(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}\mathbb{E}_B^A|r)\big)$$

$$-e^{+i2p}(l|\mathbb{E}_B^A(\mathbb{1}-\mathbb{E})^{\mathsf{P}}(\mathbb{1}-e^{+ip}\mathbb{E})^{\mathsf{P}}\mathbb{H}_{AA}^{AA}|r)-e^{-i2p}(l|\mathbb{H}_{AA}^{AA}(\mathbb{1}-\mathbb{E})^{\mathsf{P}}(\mathbb{1}-e^{-ip}\mathbb{E})^{\mathsf{P}}\mathbb{E}_B^A|r)\big]\big\}. \tag{70}$$

This equation defines the effective Hamiltonian $H_{\bar{i},j}(p,p')$. As expected, because of the translation invariance of $\hat{H}$, the δ normalizing factor is obtained. The remaining part of the Hamiltonian has been defined as $H_{\bar{i},j}(p)$, similarly to how the normalization matrices $N_{\bar{i},j}(p,p')=2\pi\delta(p-p')N_{\bar{i},j}(p)$ were defined in Eq. (13).

For momentum zero, the additional divergences $\delta(p)$ in $H_{\bar{i},j}(p)$ signal the need for imposing $(l|\mathbb{E}_B^A|r)=0$ and $(l|\mathbb{E}_A^{B'}|r)=0$, which boils down to restricting to tangent vectors $|\Phi_0(B)\rangle$, $|\Phi_0(B')\rangle \in \mathbb{T}_0^\perp$. For other momenta, there is no need to impose these conditions and these terms (i.e., the terms in the square brackets) are finite. However, thanks to the gauge freedom we can still impose this condition, and even the more general right or left gauge fixing conditions in Eq. (15) or (16), respectively. The terms in the square brackets then all disappear, together with some additional terms (among which the terms with prefactor $e^{\pm i3p}$ on the upper lines). Equation (70) can thus be simplified by a proper choice of the gauge fixing conditions on B.

In particular, by using the representation $B = \mathcal{B}(X)$, the norm matrix for the new variables becomes the identity [Eq. (20)] and we simply have to diagonalize the sum of the remaining terms in the effective Hamiltonian. Since the effective parametrization X has $(d-1)D^2$ components, a complete diagonalization has a computational cost $\mathcal{O}(D^6)$ and is only feasible for low values of D. Since we are mostly interested in the lowest eigenvalues, which correspond to elementary excitations, we can apply an iterative eigensolver. We have already explained how to efficiently compute the gradient $H_{\bar{i}}$ in Sec. III E. An efficient implementation of the matrix vector product $H_{\bar{i},j}B^j$ is a bit more complicated but also possible. Note that we need to iteratively determine the action of the $D^2 \times D^2$ operators $(\mathbb{1}-\mathbb{E})^{\mathsf{P}}$ and $(\mathbb{1}-e^{\pm ip}\mathbb{E})^{\mathsf{P}}$ on a D^2-dimensional vector. An algorithm for efficiently determining the pseudoinverse $(\mathbb{1}-\mathbb{E})^{\mathsf{P}}$ was also sketched in the previous section and is equally applicable to $(\mathbb{1}-e^{\pm ip}\mathbb{E})^{\mathsf{P}}$.

B. Topologically nontrivial states

In a system with discrete symmetry breaking, elementary excitations are often of a topologically nontrivial nature. They appear as domain walls or kinks separating two different ground states at $+\infty$ and $-\infty$. These states are protected from decay into the ground state (or into any other topologically trivial state), as there is an infinitely high energy barrier to tunnel from one ground state into another in a half-infinite region of space.[62] Let $|\Psi(A)\rangle \in \mathcal{M}$ and $|\Psi(\tilde{A})\rangle \in \mathcal{M}$ represent two uMPS with bond dimensions D and $\tilde{D}$ that approximate two different instances from the ground-state subspace of $\hat{H}$. Note that we still require these uMPS to be

pure, which implies that their connected correlation functions are exponentially decaying to zero. This indicates that they are special instances from the ground-state subspace. If $\hat{O}$ is the local order parameter corresponding to the broken symmetry, it is possible to create a symmetric ground state which does not have the exponential clustering property since $\langle O\rangle = 0$ while $\lim_{n\to\infty}\langle\hat{O}\hat{T}^n\hat{O}\hat{T}^{-n}\rangle \neq 0$. Such a state can only be approximated by a linear combination of several pure MPS, which are each individually good ground-state approximations.[63] The pure MPS $|\Psi(A)\rangle$ and $|\Psi(\tilde{A})\rangle$ will always show the symmetry breaking locally [i.e., $\langle\Psi(\overline{A})|\hat{O}|\Psi(A)\rangle \neq 0$].

We now have two transfer operators $\mathbb{E}_A^A$ and $\mathbb{E}_{\tilde{A}}^{\tilde{A}}$, each of which has a unique eigenvalue 1, and we define the corresponding left and right eigenvectors as $(l|, |r)$ and $(\tilde{l}|, |\tilde{r})$, respectively. If the symmetry transformations $\hat{U}_g$ corresponding to every element g from the symmetry group can be written as a product of one-site transformations $\hat{U}_g = \prod_{n\in\mathbb{Z}}\hat{T}^n\hat{u}_g\hat{T}^{-n}$, then we can probably choose $\tilde{A}^s = \sum_{t=1}^d \langle s|\hat{u}_g|t\rangle A^t$ for some $g \in \mathsf{G}$, so that $\tilde{D} = D$ and $\mathbb{E}_{\tilde{A}}^{\tilde{A}} = \mathbb{E}_A^A$, and thus also $\tilde{l} = l$ and $\tilde{r} = r$. We allow for the more general case as well. However, the fact that $|\Psi(A)\rangle$ and $|\Psi(A')\rangle$ are inequivalent (i.e., that there really is symmetry breaking and that they are not just the same state) implies that $\rho(\mathbb{E}_{\tilde{A}}^A) < 1$.

An *Ansatz* for approximating the topologically nontrivial state with momentum p that asymptotically looks like $|\Psi(A)\rangle$ at $-\infty$ and like $|\Psi(\tilde{A})\rangle$ at $+\infty$ (i.e., a kink or domain wall) is given by

$$|\widetilde{\Phi}_p(B; A; \tilde{A})\rangle$$

$$= \sum_{n\in\mathbb{Z}}e^{ipn}\sum_{\{s_n\}=1}^d \boldsymbol{v}_L^\dagger\left[\left(\prod_{m<n}A^{s_m}\right)B^{s_n}\left(\prod_{m'>n}\tilde{A}^{s_{m'}}\right)\right]\boldsymbol{v}_R|\boldsymbol{s}\rangle, \tag{71}$$

with B^s a set of $D \times \tilde{D}$ matrices ($\forall\ s = 1,\ldots,d$). We now impose $(\boldsymbol{v}_L^\dagger r\boldsymbol{v}_L)(\boldsymbol{v}_R^\dagger\tilde{l}\boldsymbol{v}_R) = 1$ so as not to be troubled by the boundary vectors when computing expectation values. In order for this state to have a finite excitation energy, we need to impose $h(\overline{A},A) = h(\overline{\tilde{A}},\tilde{A})$, so that both uMPS approximate their respective ground state equally well. As for the *Ansatz* for topologically nontrivial excitations, the rationale behind the *Ansatz* in Eq. (71) is that the kink itself is a highly localized or pointlike object that is in a momentum superposition. It is not completely restricted to live on a single site since it can spread out along the virtual dimension, and has nontrivial support over at least $\log_d D + \log_d \tilde{D}$ sites. Creating a kink through the action of a physical operator, analogously to the Feynman-Bijl operator, was first attempted by Mandelstam[61] in the context of relativistic quantum field theories. Translated to the lattice case, he proposed to use as operator the Fourier

transform of

$$\hat{O}(n) = \hat{o}_n \prod_{m>n} \hat{u}_n, \tag{72}$$

with $\hat{o}_n$ a completely local operator on site n and $\prod_{m>n} \hat{u}_n$ a string of operators that has the effect to transform the ground state to another ground state for $m > n$, i.e., in our context $\hat{u}_n = \hat{T}^n \hat{u}_g \hat{T}^{-n}$.

The states $|\widetilde{\Phi}_p(B; A, \tilde{A})\rangle$ share many properties with the tangent vectors $|\Phi_p(B, A)\rangle$. Using $\rho(\mathbb{E}_{\tilde{A}}^A) < 1$ even allows for some simplifications. First, $\langle \Psi(\overline{A})|\widetilde{\Phi}_p(B; A, \tilde{A})\rangle = \langle \Psi(\overline{\tilde{A}})|\widetilde{\Phi}_p(B; A, \tilde{A})\rangle = 0$ for all values of the momentum p, including $p = 0$, so that orthogonality with respect to the ground state is automatic and no divergent terms are to be expected. The reason is the appearance of factors $\mathbb{E}_{\tilde{A}}^A$ in a half-infinite space. Second, the linear map

$$\widetilde{\Phi}_p^{(A,\tilde{A})} : \mathbb{C}^{D \times d \times \tilde{D}} \mapsto \mathbb{H} : B \mapsto |\widetilde{\Phi}_p^{(A,\tilde{A})}(B)\rangle = |\widetilde{\Phi}_p(B; A, \tilde{A})\rangle \tag{73}$$

has a nontrivial null space $\widetilde{\mathbb{N}}_p^{(A,\tilde{A})}$. Indeed, it can easily be checked that the map

$$\widetilde{\mathcal{N}}_p^{(A,\tilde{A})} : \mathbb{C}^{D \times \tilde{D}} \mapsto \widetilde{\mathbb{N}}_p^{(A,\tilde{A})} : x \mapsto \mathcal{N}_p^{(A,\tilde{A})}(x) \quad \text{with}$$

$$\times \mathcal{N}_p^{(A,\tilde{A})s}(x) = A^s x - e^{-ip} x \tilde{A}^s, \forall \ s = 1, \ldots, d \tag{74}$$

defines a set of choices $B = \widetilde{\mathcal{N}}_p(x)$ that produce $|\widetilde{\Phi}_p(B)\rangle = 0$. We henceforth omit the explicit notation of A and $\tilde{A}$. It is easy to see that the null space of $\widetilde{\mathbb{N}}_p$ itself is empty $\forall \ p$, including $p = 0$, since $\widetilde{\mathcal{N}}_p^s(x) = 0$ ($\forall \ s = 1, \ldots, d$) requires that $\mathbb{E}_{\tilde{A}}^A |x\tilde{r}\rangle = e^{-ip}|x\tilde{r}\rangle$. But, since $\rho(\mathbb{E}_{\tilde{A}}^A) < 1$, this equation can have no solution. If we define

$$\widetilde{\mathbb{T}}_p = \{|\widetilde{\Phi}_p(B)\rangle \mid B \in \mathbb{C}^{D \times d \times \tilde{D}}\}, \tag{75}$$

then we obtain $\dim \widetilde{\mathbb{T}}_p = \dim \mathbb{C}^{D \times d \times \tilde{D}} - \dim \widetilde{\mathbb{N}}_p = (d - 1)D\tilde{D}, \forall \ p \in [-\pi, +\pi)$. To fix the additive gauge freedom in the parametrization B, we restrict to parametrizations B in the

a horizontal subspace $\widetilde{\mathbb{B}}^{(A,\tilde{A})}$, which is defined as the solution space of either of the following conditions:

(i) left gauge fixing condition:

$$(l|\mathbb{E}_A^B = 0 \quad \Leftrightarrow \quad \sum_{s=1}^d A^{s\dagger} l B^s = 0, \tag{76}$$

(ii) right gauge fixing condition:

$$\mathbb{E}_{\tilde{A}}^B |\tilde{r}) = 0 \quad \Leftrightarrow \quad \sum_{s=1}^d B^s \tilde{r} \tilde{A}^{s\dagger} = 0. \tag{77}$$

Either of these two conditions impose in total $D\tilde{D}$ linearly independent equations that completely fix the gauge freedom in x. A linear parametrization satisfying, e.g., the left gauge fixing condition can easily be obtained as $\widetilde{\mathcal{B}}^s(\tilde{X}) = l^{-1/2} V_L^s \tilde{X} \tilde{r}^{-1/2}$, where $\tilde{X} \in \mathbb{C}^{(d-1)D \times \tilde{D}}$ and V_L was constructed in the context of Eq. (19).

Applying the time-independent variational principle to $\widetilde{\mathbb{T}}_p$ also boils down to solving a Rayleigh-Ritz problem. We first compute the overlap $\langle \widetilde{\Phi}_p(\overline{B})|\widetilde{\Phi}_{p'}(B')\rangle$. Since the superoperators $\mathbb{E}_{\tilde{A}}^A$ and $\mathbb{E}_A^{\tilde{A}}$ that appear between B and B' have a spectral radius smaller than one, they do not need "regularizing" and we can not have disconnected contributions at all. There is then also no need to introduce pseudoinverses. We simply obtain

$$\langle \widetilde{\Phi}_p(\overline{B})|\widetilde{\Phi}_{p'}(B')\rangle$$
$$= 2\pi \delta(p' - p)\big[(l|\mathbb{E}_B^{B'}|\tilde{r}) + (l|\mathbb{E}_B^A \big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{E}_A^{B'}|\tilde{r})$$
$$+ (l|\mathbb{E}_A^{B'} \big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{E}_B^A|\tilde{r})\big]. \tag{78}$$

Similarly, all disconnected contributions that were present in $\langle \Phi_p(\overline{B})|\hat{H}|\Phi_{p'}(B')\rangle$ [terms in square brackets in Eq. (70)] disappear in the evaluation of $\langle \widetilde{\Phi}_p(\overline{B})|\hat{H}|\widetilde{\Phi}_{p'}(B')\rangle$. If we also redefine $\hat{h} \leftarrow \hat{h} - h(\overline{A}, A) = \hat{h} - h(\overline{\tilde{A}}, \tilde{A})$, so that the correct ground-state energy is subtracted and there are no disconnected contributions from $\hat{h}$ either, we immediately obtain

$$\langle \widetilde{\Phi}_p(\overline{B})|\hat{H}|\widetilde{\Phi}_{p'}(B')\rangle = 2\pi \delta(p' - p)$$
$$\times \big[(l|\mathbb{H}_{B\tilde{A}}^{B'\tilde{A}}|\tilde{r}) + (l|\mathbb{H}_{AB}^{AB'}|\tilde{r}) + e^{+ip}(l|\mathbb{H}_{B\tilde{A}}^{AB'}|\tilde{r}) + e^{-ip}(l|\mathbb{H}_{AB}^{B'\tilde{A}}|\tilde{r}) + (l|\mathbb{E}_B^{B'}\big(\mathbb{1} - \mathbb{E}_{\tilde{A}}^{\tilde{A}}\big)^P \mathbb{H}_{\tilde{A}\tilde{A}}^{\tilde{A}\tilde{A}}|\tilde{r}) + (l|\mathbb{H}_{AA}^{AA}\big(\mathbb{1} - \mathbb{E}_A^A\big)^P \mathbb{E}_B^{B'}|\tilde{r})$$
$$+ e^{+ip}(l|\mathbb{E}_B^A\big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{E}_{\tilde{A}}^{B'}\big(\mathbb{1} - \mathbb{E}_{\tilde{A}}^{\tilde{A}}\big)^P \mathbb{H}_{\tilde{A}\tilde{A}}^{\tilde{A}\tilde{A}}|\tilde{r}) + e^{-ip}(l|\mathbb{E}_A^{B'}\big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{E}_B^{\tilde{A}}\big(\mathbb{1} - \mathbb{E}_{\tilde{A}}^{\tilde{A}}\big)^P \mathbb{H}_{\tilde{A}\tilde{A}}^{\tilde{A}\tilde{A}}|\tilde{r})$$
$$+ e^{+ip}(l|\mathbb{H}_{AA}^{AA}\big(\mathbb{1} - \mathbb{E}_A^A\big)^P \mathbb{E}_B^A\big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{E}_{\tilde{A}}^{B'}|\tilde{r}) + e^{-ip}(l|\mathbb{H}_{AA}^{AA}\big(\mathbb{1} - \mathbb{E}_A^A\big)^P \mathbb{E}_A^{B'}\big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{E}_B^{\tilde{A}}|\tilde{r})$$
$$+ e^{+ip}(l|\mathbb{E}_B^A\big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{H}_{\tilde{A}\tilde{A}}^{B'\tilde{A}}|\tilde{r}) + e^{-ip}(l|\mathbb{E}_A^{B'}\big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{H}_{B\tilde{A}}^{\tilde{A}\tilde{A}}|\tilde{r}) + e^{+ip}(l|\mathbb{H}_{AB}^{AA}\big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{E}_{\tilde{A}}^{B'}|\tilde{r})$$
$$+ e^{-ip}(l|\mathbb{H}_{AA}^{AB'}\big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{E}_B^{\tilde{A}}|\tilde{r}) + e^{+2ip}(l|\mathbb{E}_B^A\big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{H}_{\tilde{A}\tilde{A}}^{AB'}|\tilde{r}) + e^{-2ip}(l|\mathbb{E}_A^{B'}\big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{H}_{AB}^{\tilde{A}\tilde{A}}|\tilde{r})$$
$$+ e^{+2ip}(l|\mathbb{H}_{B\tilde{A}}^{AA}\big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{E}_{\tilde{A}}^{B'}|\tilde{r}) + e^{-2ip}(l|\mathbb{H}_{AA}^{B'\tilde{A}}\big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{E}_B^{\tilde{A}}|\tilde{r})$$
$$+ e^{+3ip}(l|\mathbb{E}_B^A\big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{H}_{\tilde{A}\tilde{A}}^{AA}\big(\mathbb{1} - e^{+ip}\mathbb{E}_{\tilde{A}}^A\big)^{-1} \mathbb{E}_{\tilde{A}}^{B'}|\tilde{r}) + e^{-3ip}(l|\mathbb{E}_A^{B'}\big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{H}_{AA}^{\tilde{A}\tilde{A}}\big(\mathbb{1} - e^{-ip}\mathbb{E}_A^{\tilde{A}}\big)^{-1} \mathbb{E}_B^{\tilde{A}}|\tilde{r})\big]. \tag{79}$$

By imposing either the left or the right gauge fixing conditions of Eq. (76) or (77), many terms in Eqs. (78) and (79) cancel. In particular, using the representation $B = \widetilde{\mathcal{B}}(\tilde{X})$, the effective

norm matrix becomes

$$\langle \widetilde{\Phi}_p(\overline{\widetilde{\mathcal{B}}(\overline{\tilde{X}})})|\widetilde{\Phi}_{p'}(\widetilde{\mathcal{B}}(\tilde{X}'))\rangle = 2\pi \delta(p' - p) \, \text{tr}[\tilde{X}^\dagger \tilde{X}'].$$

As for the topologically trivial excitation *Ansatz*, the final problem boils down to finding eigenvalues of a normal eigenvalue problem and can be solved using in iterative eigensolver with a computation complexity that scales as $\mathcal{O}(D^3)$.

One final remark is in order. With two matrices A and $\tilde{A}$ present in the *Ansatz* for topologically nontrivial states $|\widetilde{\Phi}_p(B; A; \tilde{A})\rangle$, both of which can be defined independently from each other, there is some ambiguity present in the definition of the momentum. Suppose we have $A' = e^{i\varphi}A$. We obtain $|\widetilde{\Phi}_p(B; A'; \tilde{A})\rangle = |\widetilde{\Phi}_p(B; e^{i\varphi}A; \tilde{A})\rangle \sim |\widetilde{\Phi}_{p+\varphi}(B; A; \tilde{A})\rangle$, where the similarity sign means the two states are equal up to a global phase. This follows simply from inserting A' in, e.g., Eq. (79). It is ultimately related to the fact that topologically nontrivial excitations can only be defined on systems with open boundary conditions, where momentum is not a good quantum number to start with. This problem did not occur in the *Ansatz* for topologically trivial excitations because $|\Psi_p(B)\rangle$ can be defined starting from a finite lattice with periodic boundary conditions, where p is a good quantum number. In systems with periodic boundary conditions, only pairs of kinks $(|\Psi_{p_1}(B_1; A, \tilde{A})\rangle)$ and antikinks $(|\Psi_{p_2}(B_2; \tilde{A}, A)\rangle)$ with total momentum $p_1 + p_2$ can exist, and the momentum ambiguity is resolved since for $A \leftarrow e^{i\phi}A$ we obtain $p_1 \leftarrow p_1 - \phi$ and $p_2 \leftarrow p_2 + \phi$. Nevertheless, it is often useful to study topologically nontrivial excitations as isolated elementary excitations. Since the momentum is tightly connected to $\mathbb{E}_{\tilde{A}}^A$ in the sense that they appear together as $e^{ip}\mathbb{E}_{\tilde{A}}^A$ in Eq. (79), it makes sense to fix the freedom in the definition of the momentum quantum number by requiring that the relative phase of A and $\tilde{A}$ is chosen so that the eigenvalue with largest modulus of $\mathbb{E}_{\tilde{A}}^A$ is positive.

C. Relation to the time-dependent variational principle

The *Ansatz* for studying excitations starts from an optimal uMPS approximation $|\Psi(A)\rangle$ of the ground state. Thus, A is an optimum of the time-independent variational principle for the ground-state problem in $\mathcal{M}$. Consequently, it is also a station-ary solution of the TDVP equations for the uMPS manifold $\mathcal{M}$, since $H_{\bar{J}}(\overline{A}, A) = 0$, or thus, for any $|\Phi_0(B)\rangle \perp |\Psi(A)\rangle$, $\langle\Phi_0(\overline{B})|\hat{H}|\Psi(A)\rangle = 0$. Translation invariance also dictates that $\langle\Phi_p(\overline{B})|\hat{H}|\Psi(A)\rangle = 0$, so that $|\Psi(A)\rangle$ is also stable against all non-translation-invariant first-order variations.

We can now investigate small perturbations around the stationary and uniform solution A. Note that we have only derived the TDVP equation for the manifold of translation-invariant MPS. The generalization to MPS with site-dependent tensors $A(n)$ is straightforward. We introduce the notation $|\Psi[\mathbf{A}]\rangle$, where the square brackets denote the dependence on an infinite set of tensors $\mathbf{A} = \{A(n)\}_{n \in \mathbb{Z}}$. The corresponding energy is given by $H[\overline{\mathbf{A}}, \mathbf{A}] = \langle\Psi[\overline{\mathbf{A}}]|\hat{H}|\Psi[\mathbf{A}]\rangle$, where the state is assumed to be normalized to $\langle\Psi[\overline{\mathbf{A}}]|\Psi[\mathbf{A}]\rangle = 1$. We also use a big index $I = (n, i) = [n, (\alpha, \beta, s)]$ as a collective index containing also the site index, such that $A^I = A_{\alpha, \beta}^s(n)$, and let ∂_I denote the partial derivative with respect to A^I. The generalized TDVP equation then reads as

$$\langle\partial_{\bar{I}}\Psi[\overline{\mathbf{A}}(t)]|\partial_J\Psi[\mathbf{A}(t)]\rangle \, \dot{\mathbf{A}}^J(t)$$
$$= \langle\partial_{\bar{I}}\Psi[\overline{\mathbf{A}}(t)]|\hat{H} - H[\overline{\mathbf{A}}(t), \mathbf{A}(t)]|\Psi[\mathbf{A}(t)]\rangle . \quad (80)$$

For small perturbations $\mathbf{A}(t) = \mathbf{A}_0 + \epsilon \mathbf{B}(t)$, where $\mathbf{A}_0$ denotes the uniform solution $\{A\}_{n \in \mathbb{Z}}$ and $\mathbf{B}(t) = \{B(n, t)\}_{n \in \mathbb{Z}}$, we can linearize the TDVP equation as

$$\langle\partial_{\bar{I}}\Psi[\overline{\mathbf{A}}_0]|\partial_J\Psi[\mathbf{A}_0]\rangle \, \dot{\mathbf{B}}^J(t)$$
$$= \langle\partial_{\bar{I}}\Psi[\overline{\mathbf{A}}_0]|\hat{H} - H[\overline{\mathbf{A}}_0, \mathbf{A}_0]|\partial_J\Psi[\mathbf{A}_0]\rangle \, \mathbf{B}^J(t)$$
$$+ \langle\partial_{\bar{I}}\partial_{\bar{J}}\Psi[\overline{\mathbf{A}}_0]|\hat{H} - H[\overline{\mathbf{A}}_0, \mathbf{A}_0]|\Psi[\mathbf{A}_0]\rangle \, \overline{\mathbf{B}}^{\bar{J}}(t). \quad (81)$$

More generally, the right-hand side should read as $\partial_{\bar{I}}\partial_J H[\overline{\mathbf{A}}_0, \mathbf{A}_0]\mathbf{B}^J + \partial_{\bar{I}}\partial_{\bar{J}} H[\overline{\mathbf{A}}_0, \mathbf{A}_0]\overline{\mathbf{B}}^{\bar{J}}$. However, using the fact that first-order derivatives of H at $\mathbf{A}_0$ are zero due to stationarity, and that we are only considering variations orthogonal to the original state, only the terms above survive. Note that we had to introduce a second-order derivative $|\partial_I\partial_J\Psi[\mathbf{A}_0]\rangle$ in the right-hand side. For $\mathbf{B}_1 = \{B_1(n)\}_{n \in \mathbb{Z}}$ and $\mathbf{B}_2 = \{B_2(n)\}_{n \in \mathbb{Z}}$, we now introduce the general notation

$$|\Upsilon^{(A)}[\mathbf{B}_1, \mathbf{B}_2]\rangle = \mathbf{B}_1^I \mathbf{B}_2^J |\partial_I\partial_J\Psi[\mathbf{A}_0]\rangle = \sum_{m,n \in \mathbb{Z}} B_1^i(m)B_2^j(n) \left.\frac{\partial^2}{\partial A^i(m)\partial A^j(n)}|\Psi[\mathbf{A}]\rangle\right|_{\mathbf{A}=\mathbf{A}_0}$$

$$= \sum_{m<n \in \mathbb{Z}} \sum_{\{s_n\}=1}^{d} \mathbf{v}_{\mathrm{L}}^\dagger \left[\left(\prod_{k<m} A^{s_k}\right) B_1^{s_m}(m) \left(\prod_{m<k'<n} A^{s_{k'}}\right) B_2^{s_n}(n) \left(\prod_{n<k''} A^{s_{k''}}\right)\right] \mathbf{v}_{\mathrm{R}} |\mathbf{s}\rangle$$

$$+ \sum_{m>n \in \mathbb{Z}} \sum_{\{s_n\}=1}^{d} \mathbf{v}_{\mathrm{L}}^\dagger \left[\left(\prod_{k<w} A^{s_k}\right) B_2^{s_n}(n) \left(\prod_{n<k'<m} A^{s_{k'}}\right) B_1^{s_m}(m) \left(\prod_{n<k''} A^{s_{k''}}\right)\right] \mathbf{v}_{\mathrm{R}} |\mathbf{s}\rangle . \quad (82)$$

These states span the double tangent space $\mathbb{T}^{(2)}$ of the MPS manifold. They were already introduced in Ref. 20 for calculating the Levi-Civita connection.

Returning to the linearized TDVP equation [Eq. (81)], we can now look for specific solutions

$$B(n, t) = B_+ e^{ipn - i\omega t} + B_- e^{-ipn + i\omega t} \quad (83)$$

and contract the free index I of this equation with a test vector $\mathbf{B}' = \{B'(n)\}_{n \in \mathbb{Z}}$ with $B'(n) = Be^{ip'n}$. This essentially boils down to a Fourier transform of the linearized TDVP equation with respect to both space and time, and results in

$$\omega \langle\Phi_{p'}(\overline{B}')|\Phi_p(B_+)\rangle = \langle\Phi_{p'}(\overline{B}')|\hat{H} - H|\Phi_p(B_+)\rangle + \langle\Upsilon_{p',-p}(\overline{B}', \overline{B}_-)|\hat{H} - H|\Psi(A)\rangle - \omega\langle\Phi_{p'}(\overline{B}')|\Phi_{-p}(B_-)\rangle$$
$$= \langle\Phi_{p'}(\overline{B}')|\hat{H} - H|\Phi_{-p}(B_-)\rangle + \langle\Upsilon_{p',p}(\overline{B}', \overline{B}_+)|\hat{H} - H|\Psi(A)\rangle , \quad (84)$$

where $H = H(\overline{A},A) = H[\overline{A}_0,A_0]$ and we have introduced a new basis for the double tangent space as

$$|\Upsilon_{p_1,p_2}(B_1,B_2)\rangle = \sum_{m<n\in\mathbb{Z}} \sum_{\{s_n\}=1}^{d} e^{ip_1m+ip_2n} \boldsymbol{v}_{\mathrm{L}}^\dagger \left[\left(\prod_{k<m} A^{s_k}\right) B_1^{s_m} \left(\prod_{m<k'<n} A^{s_{k'}}\right) B_2^{s_n} \left(\prod_{n<k''} A^{s_{k''}}\right)\right] \boldsymbol{v}_{\mathrm{R}} |\boldsymbol{s}\rangle$$

$$+ \sum_{m>n\in\mathbb{Z}} \sum_{\{s_n\}=1}^{d} e^{ip_1m+ip_2n} \boldsymbol{v}_{\mathrm{L}}^\dagger \left[\left(\prod_{k<w} A^{s_k}\right) B_2^{s_n} \left(\prod_{n<k'<m} A^{s_{k'}}\right) B_1^{s_m} \left(\prod_{n<k''} A^{s_{k''}}\right)\right] \boldsymbol{v}_{\mathrm{R}} |\boldsymbol{s}\rangle . \quad (85)$$

We then also define

$$\langle\Upsilon_{p_1,p_2}(\overline{B}_1,\overline{B}_2)|\hat{H} - H|\Psi(A)\rangle = K_{\bar{i},\bar{j}}(p_1,p_2)\overline{B}_1^{\bar{i}}\overline{B}_2^{\bar{j}} = 2\pi\delta[(p_1+p_2) \bmod 2\pi] K_{\bar{i},\bar{j}}(p_1)\overline{B}_1^{\bar{i}}\overline{B}_2^{\bar{j}}, \quad (86)$$

where the expression for $K_{\bar{i},\bar{j}}(p_1)$ can easily be derived using the techniques that were used for the calculation of $H_{\bar{i},j}(p)$. By taking the complex conjugate of the second line in Eq. (84), we can now write the Fourier-transformed linearized TDVP equation as a generalized eigenvalue equation

$$\omega \begin{bmatrix} N(p) & 0 \\ 0 & -\overline{N}(-p) \end{bmatrix} \begin{bmatrix} B_+ \\ \overline{B}_- \end{bmatrix} = \begin{bmatrix} H(p) & K(p) \\ \overline{K}(-p) & \overline{H}(-p) \end{bmatrix} \begin{bmatrix} B_+ \\ \overline{B}_- \end{bmatrix}. \quad (87)$$

Note that the matrix appearing in the right-hand side of this equation is Hermitian since $H(p)$ and $N(p)$ are Hermitian, and $K(p)$ satisfies $K(p) = K(-p)^{\mathrm{T}}$. If we would have linearized the original TDVP equation for uMPS, we would only have recovered the $p = 0$ case of the above equation. For $p = 0$, the matrix in the right-hand side can easily be recognized as the Hessian of the energy functional $H(\overline{A},A)$ at the point A. We could thus also have denoted $K_{\bar{i},\bar{j}}(0,0)$ as $H_{\bar{i},\bar{j}} = \partial_{\bar{i}}\partial_{\bar{j}}H$. By considering all momenta, we have actually constructed a momentum decomposition of the full Hessian of the energy functional in the manifold of all MPS with site-dependent matrices. Since A is a minimum of the energy functional in the class of uMPS, this Hessian is positive (semi)definite at momentum zero. If the uniform solution is also a minimum in the full class of generic MPS, the corresponding Hessian is positive (semi)definite at any momentum p. In that case, all eigenvalues of the generalized eigenvalue equation in Eq. (87) are real. In addition, for any eigenvector $\omega^{(k)}(p)$ with corresponding eigenvector $(B_+^{(k)}(p),\overline{B}_-^{(k)}(p))$ at momentum p, there exists a dual eigenvalue $\omega^{(-k)}(-p) = -\omega^{(k)}(p)$ with eigenvector $(B_+^{(-k)}(-p) = B_-^{(k)}(p),\overline{B}_-^{(-k)}(-p) = \overline{B}_+^{(k)}(p))$. In particular, at momentum zero, the spectrum is even around zero. Since many systems are invariant under spatial reflections, we expect that $N(p)$, $H(p)$, and $K(p)$ will be gauge equivalent to $N(-p)$, $H(-p)$, and $K(-p)$. In that case, the spectrum of eigenvalues of Eq. (87) is even around zero at all momenta since we then have $\omega^{(-k)}(p) = \omega^{(-k)}(-p) = -\omega^{(k)}(p)$.

From linear response theory, it can be argued that the eigenvalues ω of Eq. (87) correspond to resonances in the system that should be related to the excitation energies of the system.[64] Clearly, the Rayleigh-Ritz problem obtained for our variational *Ansatz* for topologically trivial excitations is equivalent to the linearized TDVP problem of Eq. (87) if the off-diagonal blocks $K(p) = 0$. We now discuss the effect of having $K(p) \neq 0$. The matrix $K(p)$ is obtained as (a momentum decomposition of)

the projection of $\hat{H} - H |\Psi\rangle$ to the double tangent space $\mathbb{T}^{(2)}$, which is spanned by the second-order derivatives $|\partial_I \partial_J \Psi[A_0]\rangle$. We have introduced the notation $|\Upsilon_{p_1,p_2}(B_1,B_2)\rangle$ for momentum states in $\mathbb{T}^{(2)}$ in Eq. (85). We will elaborate on the physical interpretation of this set of states in Sec. VII. For now, the structure of Eq. (87) seems to predict corrections to the variational excitation energies (obtained from the diagonal part) by taking into account a larger part of Hilbert space, namely, the part that is captured by $\mathbb{T}^{(2)}$. The problem with these corrections is that they are not variational in nature, and can therefore be unbounded and unphysical. We illustrate in the next section how Eq. (87) can give rise to spurious excitation energies which are not physical and which do not show up when using the variational *Ansatz* $|\Phi_p(B)\rangle$. Since $K(p)$ is obtained from projecting $\hat{H} - H |\Psi\rangle$ onto the double tangent space, one can show that $\|K(p)\| < \widetilde{\epsilon}(\overline{A},A)$ (some care has to be taken to deal with the diverging prefactors), where $\widetilde{\epsilon}(\overline{A},A) = \Delta H(\overline{A},A)|\mathbb{Z}|^{-1/2}$ is the local measure for the state error. In the limit where the MPS approximation is becoming very accurate, $K(p) \to 0$ and the variational energies will match the eigenvalues of Eq. (87).

Nevertheless, recognizing the right-hand side of Eq. (87) as the Hessian of the energy functional is important for interpreting the variational excitation energies, which correspond to the eigenvalues of the diagonal block $H(p)$. We have already mentioned that $|\Psi(A)\rangle$ is a stationary solution, both in the manifold of uMPS as well as in the manifold of generic MPS. Hence, there are no corrections of first order in ϵ to the energy $H[\overline{A},A]$ when replacing $A \leftarrow A_0 + \epsilon B$. If for some momentum p, the variational energies obtained from $H(p)$ contain negative values, this indicates that condensation of this excitation is possible so as to create a lower-energy MPS approximation of the ground state that is not uniform. Clearly, this should not happen at momentum $p = 0$, as this would indicate that the stationary solution A is not a minimum but rather a saddle point or local maximum. For example, let $H(p)$ have a negative energy $\omega < 0$ with corresponding eigenvector B_+, and let there be reflection invariance such that $H(-p)$ also has eigenvalue ω with corresponding eigenvector B_- (related to B_+ by a gauge transform). For $A \leftarrow A_0 + \epsilon B$, with $B = \{B(n)\}_{n\in\mathbb{Z}}$ and $B(n) = \exp(ipn)B_+ + \exp(-ipn)B_-$, the second-order correction to the energy density will be

$$H[\overline{A},A] = H[\overline{A}_0,A_0] + \frac{\epsilon^2}{2}\mathbb{Z}(\overline{B}_+^{\bar{i}}H_{\bar{i},j}(p)B_+^j + \overline{B}_-^{\bar{i}}H_{\bar{i},j}(-p)$$
$$\times B_-^j + \overline{B}_+^{\bar{i}}\overline{B}_-^{\bar{j}}K_{\bar{i},\bar{j}}(p) + \overline{K}_{i,j}(p)B_+^i B_-^j). \quad (88)$$

The first two terms in the brackets reduce to $\omega < 0$, whereas we can certainly choose the phase of B_+ and B_- in such a way that the last two terms are zero or negative as well. If this happens for $p \neq 0$, it indicates that the uniform solution $|\Psi(A)\rangle$ is only a minimum in the restricted manifold of uniform MPS, but not in the manifold of generic MPS.

Finally, we can also discuss the *Ansatz* $|\widetilde{\Phi}_p(B)\rangle$ for topologically nontrivial excitations. The states $|\widetilde{\Phi}_\kappa(B)\rangle$ do not live in the tangent space $\mathbb{T}$ of $\mathcal{M}$. However, if we construct a larger manifold $\{|\Psi'[A']\rangle\}$ of MPS of bond dimension $D' = D + \tilde{D}$ and define the uniform solution $A'_0 = \{A'\}_{n \in \mathbb{Z}}$ with

$$A'^s = \begin{bmatrix} A^s & 0 \\ 0 & \tilde{A}^s \end{bmatrix}, \quad (89)$$

then the topologically nontrivial excitation $|\widetilde{\Phi}_p(B)\rangle$ is obtained as the tangent vector $|\Phi'[C]\rangle = C^I |\partial_I \Psi'[A'_0]\rangle$ with

$$C^s(n) = \begin{bmatrix} 0 & Be^{ipn} \\ 0 & 0 \end{bmatrix}. \quad (90)$$

Because of the special structure of A', the point $A' = A'_0$ is a stationary point of the energy functional $H[\overline{A'}, A]$. However, A'_0 is not expected to be a true minimum in the variational manifold of MPS with bond dimension $2D$, and the effective Hamiltonian for this variational *Ansatz* can have negative eigenvalues. This would indicate that the uMPS $|\Psi'(A')\rangle$ is not stable at second order against fluctuations that mix the two ground-state approximations $|\psi(A)\rangle$ and $|\psi(\tilde{A})\rangle$ with maximal symmetry breaking. Quantum fluctuations can thus have the tendency to decrease the expectation value of the order parameter with respect to the value obtained with mean field theory, or with an MPS approximation of finite D. As D grows larger and the approximation improves, the negative eigenvalues should eventually disappear.

D. Dynamical correlation functions

We conclude this section by discussing how the tangent space can assist in the computation of dynamical correlation functions. Let the ground state of a system be well approximated by the uMPS $|\Psi(A)\rangle$. As discussed in the Introduction, dynamical correlation functions can either be computed in the time domain, or directly in the frequency domain. Traditional MPS algorithms for computing correlation functions in the time domain had to put the system on a finite lattice, although recently a number of competing algorithms have been developed to deal with localized perturbations in an otherwise translation-invariant background.[38,65–67]

Dynamical correlation functions can be obtained from Fourier transforming the Green's function $G^{(\alpha,\beta)}(p, p', \omega)$ of two operators $\hat{O}^{(\alpha)}$ and $\hat{O}^{(\beta)}$, which is defined as

$$G^{(\alpha,\beta)}(p, p', \omega)$$
$$= \langle \Psi(A) | \hat{O}_p^{(\alpha)\dagger} \frac{1}{\omega - [\hat{H} - H(\overline{A}, A)] + i\eta} \hat{O}_{p'}^{(\beta)} |\Psi(A)\rangle, \quad (91)$$

where the Fourier transform of the operators $\hat{O}$ is defined as

$$\hat{O}_p = \sum_{n \in \mathbb{Z}} e^{ipn} \hat{T}^n \hat{O} \hat{T}^{-n}. \quad (92)$$

We have subtracted the approximate ground-state energy $H(\overline{A}, A)$ so as to get poles at finite excitation energies. Due to the translation invariance, $G^{(\alpha,\beta)}(p, p', \omega)$ is of the form $2\pi \delta(p - p') G^{(\alpha,\beta)}(p, \omega)$ where $G^{(\alpha,\beta)}(p, \omega)$ is finite (aside from its poles due to the excitation energies).

Clearly, for $\hat{O}^{(\alpha)}$ being one-site operators, the state $\hat{O}_m^{(\alpha)} |\Psi(A)\rangle$ exactly corresponds to a tangent vector $|\Phi(B^{(\alpha)})\rangle$ with $B^{(\alpha)s} = \langle s | \hat{O}^{(\alpha)} | t \rangle A^t$. We will generalize this to the case where $\hat{O}^\alpha$ acts on K consecutive sites in Sec. VII. For now, we can, without approximation, write

$$G^{(\alpha,\beta)}(p, p', \omega)$$
$$= \langle \Phi(B^{(\beta)}) | \hat{P}_\mathbb{T} \frac{1}{\omega - [\hat{H} - H(\overline{A}, A)] + i\eta} \hat{P}_\mathbb{T} |\Psi(B^{(\alpha)})\rangle.$$

Note that existing methods for directly computing the Green's function,[26,42–44,68–70] which are almost always formulated for finite systems, try to approximate $\hat{O}^\alpha |\Psi(A)\rangle$, and sometimes even $(\omega - [\hat{H} - H(\overline{A}, A)] + i\eta)^{-1} \hat{O}^\alpha |\Psi(A)\rangle$ as matrix product states. While this has led to successful methods on the finite lattice, the thermodynamic limit makes it clear that these states do not have the normalizable structure of MPS, but rather the structure of tangent vectors to the MPS manifold.

As we illustrate in the next section, the tangent space is typically suited for capturing elementary excitations with a single-particle-like or pointlike structure. If we expect that these excitations are responsible for the dominant contribution to the Green's function $G^{(\alpha,\beta)}(p, \omega)$, then we can replace Eq. (91) with

$$G^{(\alpha,\beta)}(p, p', \omega)$$
$$\approx \langle \Phi(B^{(\beta)}) | \hat{P}_\mathbb{T} \frac{1}{\omega - [\hat{P}_\mathbb{T} \hat{H} \hat{P}_\mathbb{T} - H(\overline{A}, A)] + i\eta} \hat{P}_\mathbb{T} |\Psi(B^{(\alpha)})\rangle,$$

which leads to

$$G^{(\alpha,\beta)}(p, p', \omega)$$
$$\approx 2\pi \delta(p - p') \overline{B}^{(\beta)\overline{i}} \left[\frac{1}{\omega - H(p) + i\eta} \right]_{\overline{i}, j} B^{(\alpha)j}, \quad (93)$$

where the part multiplying the $2\pi \delta(p - p')$ has been defined as $G^{(\alpha,\beta)}(p, \omega)$ above. In practice, we use of course the representation of $H(p)$ and B in terms of the parametrization X, in order to eliminate the null modes. As will be illustrated in the next section, this approach is not able to accurately describe the contribution of the multiparticle continuum. Nevertheless, since $\hat{O}_p^{(\alpha)} |\Psi(A)\rangle$ is exactly in the tangent space, we do not lose any spectral weight.

Finally, we note that we do not need a full inverse of $\omega - H(p)$. If, for example, we are interested in the imaginary part of $G^{(\alpha,\beta)}(p, \omega)$, which is often known as a spectral function or a structure factor, then we can write

$$\text{Im}[G^{(\alpha,\beta)}(p, \omega)] \approx \overline{B}^{(\beta)} [\delta\{\omega - H(p)\}]_{\overline{i}, j} B^{(\alpha)j}. \quad (94)$$

The point is that we do not need to know the full eigenvalue decomposition of $H(p)$ [which would require an $O(D^6)$ calculation] to compute the δ function in the previous equation. Instead, we can decompose the δ function into Chebychev polynomials.[69] Indeed, we can easily and efficiently compute arbitrary polynomials of the effective Hamiltonian $H(p)$ acting on $|\Phi(B^{(\alpha)})\rangle$ without making any further approximation. This

allows us to compute the right-hand side of Eq. (94) at a fixed value of p, but for an arbitrary range of ω values, with a computational efficiency that scales as $\mathcal{O}(ND^3)$, where N is the number of Chebychev moments that is used in the decomposition of the δ function.

V. ILLUSTRATIVE EXAMPLES

We now provide a few selected examples that can be used to illustrate some of the general statements of the previous section regarding the tangent-space framework for excitations and dynamical correlation functions.

A. Variational *Ansatz* for excitations

An interesting model to illustrate the behavior of the variational *Ansatz* for excitations is the bilinear-biquadratic antiferromagnetic Heisenberg model with $S = 1$ spins, which is described by the SO(3)-invariant Hamiltonian

$$\hat{H}_{BB} = J \sum_{n \in \mathbb{Z}} \cos\theta (\hat{\boldsymbol{S}}_n \cdot \hat{\boldsymbol{S}}_{n+1}) + \sin\theta (\hat{\boldsymbol{S}}_n \cdot \hat{\boldsymbol{S}}_{n+1})^2, \quad (95)$$

with an energy scale $J > 0$ and an angle $\theta \in [-3\pi/4, 5\pi/4)$. This model has many interesting phases and phase transitions as a function of θ. There can not be antiferromagnetic order due to the Mermin-Wagner theorem.[71] Ferromagnetic order can exist since the ferromagnetic order parameter commutes with the Hamiltonian, and is indeed present for $\theta \in (\pi/2, 5\pi/4)$. At $\theta = 0$, this Hamiltonian describes the antiferromagnetic Heisenberg model. The ground state is then in a symmetry-protected topologically ordered phase, the Haldane phase, and the lowest-lying excitation is an $S = 1$ triplet with finite mass Δ_{Haldane}, referred to as the Haldane gap.[72–74] The same phase exists throughout $\theta \in (-\pi/4, \pi/4)$. In particular, for $\tan\theta = \frac{1}{3}$, this is the model studied by Affleck, Kennedy, Lieb, and Tasaki,[11,12] which has an exact matrix product state representation with bond dimension $D = 2$. At $\theta = \pm\pi/4$, an exact solution in terms of the Bethe *Ansatz* is possible and the model is critical. The dispersion relation shows nodes for $p = 0$ and π for $\theta = -\pi/4$ (the Takhtajan-Babudjan point), and the model undergoes a second-order phase transition to a dimer phase (only invariant under $\hat{T}^2$), which exists for $\theta \in (-3\pi/4, -\pi/4)$. The existence of a small nematic phase between the dimer phase and the ferromagnetic phase has recently been ruled out.[75] For $\theta = \pi/4$, the dispersion relation of the elementary excitation has nodes at $p = 0$ and $p = \pm 2\pi/3$ for $\theta = \pi/4$. This critical behavior exists throughout the range $\theta \in [\pi/4, \pi/2]$, no trimerization occurs, and the system is a nematic phase. The transition at $\theta = \pi/4$ is supposedly of the Kosterlitz-Thouless type. The whole phase diagram is summarized in Fig. 1 (see Refs. 76 and 77 references therein).

In Ref. 17, the variational *Ansatz* was used at the Heisenberg point $\theta = 0$ in order to compute the value of the Haldane gap $\Delta_{\text{Haldane}}/J$ up to 12 digits of precision. We now turn to the more general bilinear-biquadratic Heisenberg model in the Haldane region $-\pi/4 < \theta < +\pi/4$. As explained in the beginning of this section, the model becomes critical at the end points. Figure 2 depicts the full set of eigenvalues of $H(p)$ for θ decreasing from 0 to $-\pi/4$. Let us first
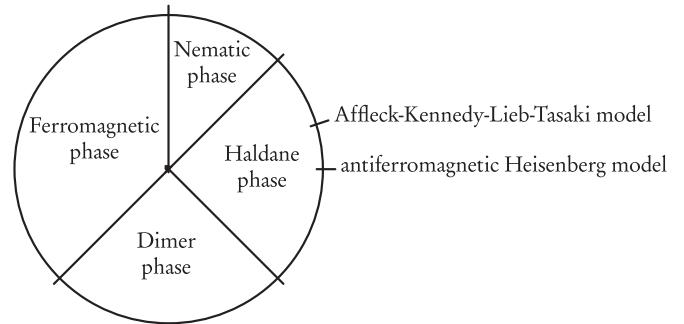
FIG. 1. Phase diagram of the bilinear-biquadratic Heisenberg model (taken from Ref. 77).

discuss some general properties of the spectra obtained with our variational strategy. First, one can observe that we have labeled the different eigenvalues according to their SU(2) spin quantum number S. We did not impose this symmetry explicitly, but were able to read the irreducible representation to which each eigenvalue belonged simply from its degeneracy,
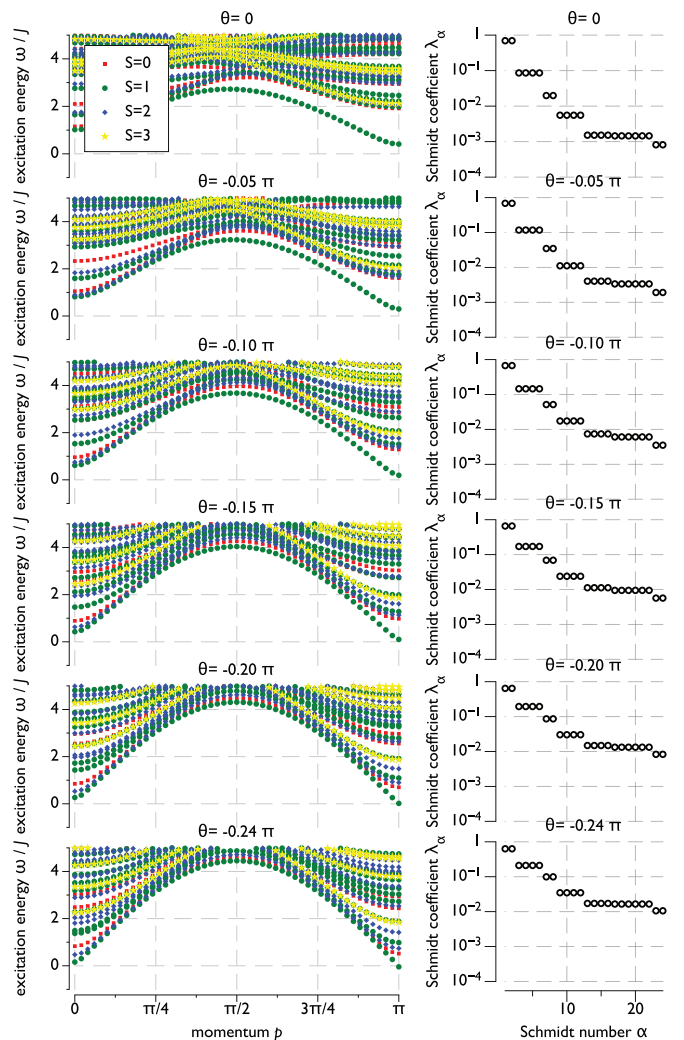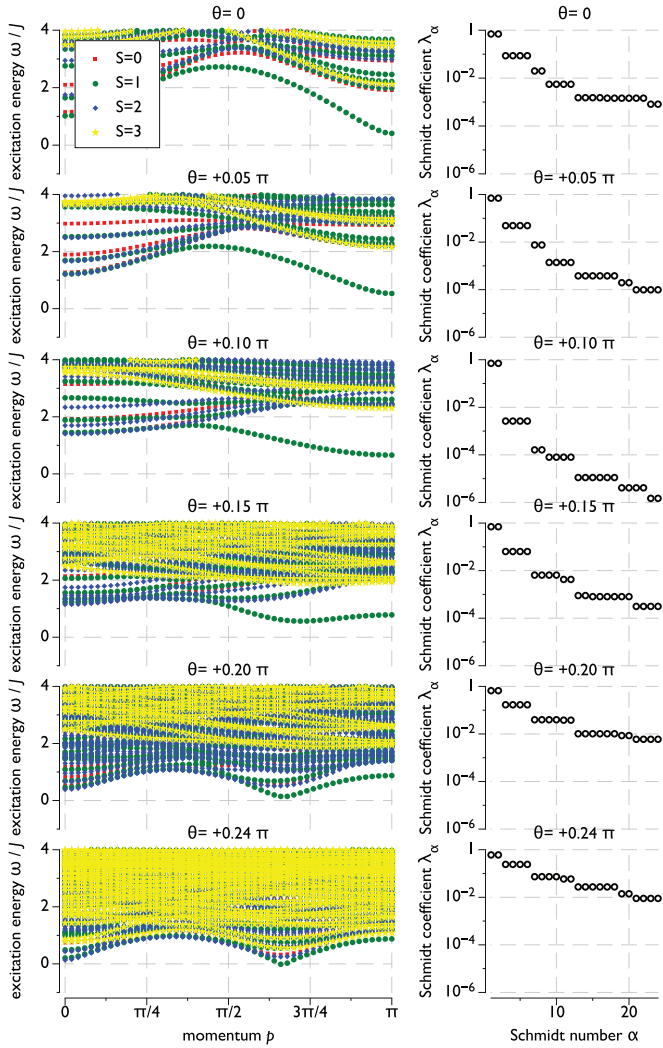

FIG. 2. (Color online) Excitation spectrum (left) and Schmidt spectrum (right) for the bilinear-biquadratic antiferromagnetic Heisenberg model in the region $\theta \in (-\pi/4, 0]$ obtained with a uMPS *Ansatz* with bond dimension $D = 24$.

which was almost perfect up to machine precision. The reason for this is that we started from a ground-state approximation $|\Psi(A)\rangle$ where the bond dimension $D = 24$ was chosen so as to contain a complete number of irreducible representations, and which was then converged, also without imposing the SU(2) symmetry explicitly, up to a state error $\widetilde{\epsilon}(\overline{A}, A) \leqslant 10^{-12}$.

Around momentum $p = \pi$, the lowest excitation energy is separated from the rest of the spectrum and can be identified with the elementary spin-1 magnon excitation, which we can approximate well with our *Ansatz*, as can be inferred from the numerical precision with which the Haldane gap can be estimated.[17] Other excitations all fall within the multimagnon continuum, starting with the 2-magnon ($S = 0,1,2$) or 3-magnon ($S = 0,1,2,3$) bands. Because we are diagonalizing a finite matrix $H(p)$ [dimensions $(d-1)D^2 \times (d-1)D^2$] for each momentum p, we obtain only a finite number of excitation energies. We thus need to interpret the discrete eigenvalues which are contained within the continuous spectrum of the full Hamiltonian. Clearly, the tangent space is not suited as a variational subspace for describing multiparticle excitations, as this would require an *Ansatz* containing several perturbations of the ground state, which move independently with different momenta. Since only the total momentum of a state is a well-defined quantum number, we can, however, create superpositions of these multiparticle states in order to obtain a kind of artificial bound state, where all particles are staying together in a small spatial region. Such states can be described in the uMPS tangent space $\mathbb{T}_p$ and are thus obtained from diagonalizing $H(p)$. They are, however, not good approximations for eigenstates of the full Hamiltonian $\hat{H}$ since they contain a superposition of many exact eigenstates with different energies. The only exception is when the exact dispersion relation of the elementary excitation(s) of the full Hamiltonian H is flat. In that case, all multiparticle excitations should also have a flat dispersion relation. In general, we thus need a different strategy to study multiparticle excitations, on which we elaborate in Sec. VII.

For larger bond dimensions, a full diagonalization of $H(p)$ becomes infeasible. However, we now know that we should only be interested in the lowest excitation energies and can thus resort to an iterative eigensolver, which can be implemented with computational complexity scaling as $\mathcal{O}(D^3)$, as discussed in the previous section. We can also make predictions about the convergence behavior of the variational excitation energies as a function of increasing bond dimension. Note that the excitation energies are no longer truly variational in the most strict sense they can have negative errors. Indeed, there are two competing sources of errors within our variational strategy for studying excitations. First, the variational energy, which is always positive, comes from trying to represent the exact excited state as a uMPS tangent vector. However, since we compute the energy from a redefined Hamiltonian $\hat{H} \leftarrow \hat{H} - H(\overline{A}, A)$, we are also subtracting an approximation to the ground-state energy which is too large. The result of this second effect is a negative error. In fact, both sources produce errors which are infinitely large, unless the uMPS $|\Psi(A)\rangle$ captures the ground state exactly. However, these errors cancel each other because they have opposite sign and are both related to an imperfect representation of the ground state. The result is a finite error, containing a positive contribution

from the variational assumption that the excited state can be obtained from a local perturbation of the ground state, and a negative contribution from subtracting a ground-state energy density which is too large in the region in which this local perturbation has support. For multiparticle excitations, the first error is clearly dominant and the resulting energies decrease as the bond dimension is increased since this has the effect of enlarging the spatial region on which the excitation has support. In addition, new eigenvalues can appear within the multiparticle continuum when D is increased. For elementary excitations or bound states, on the other hand, the locality assumption is often appropriate, and the second error typically dominates. We then observe a variational excitation energy that increases as $D \to \infty$. It can, however, happen that the spatial region on which the excitation has support is too small for small values of D, in which case we also observe a decreasing energy at low bond dimensions. Such effects were indeed observed in Ref. 78, to which we refer for a more elaborate discussion on this subject.

We now return to discuss the elementary magnon excitation in the bilinear-biquadratic antiferromagnetic Heisenberg model. For θ decreasing from 0 to $-\pi/4$, the excitation gap at $p = \pi$ decreases and eventually vanishes (see Fig. 2). Simultaneously, the entanglement increases as can be noticed by the Schmidt coefficients shifting upwards. At $\theta = -0.24\pi$, the elementary magnon has in fact a slightly negative excitation energy at $p = \pi$. While this is of course an artifact of the low value of the bond dimension $D = 24$ since the critical point is not until $\theta = -0.25\pi$, it does indicate a tendency of these magnons to condense, resulting in a ground state that breaks translation invariance down to invariance under shifts over two sites. Indeed, for $\theta < -\pi/4$ the ground state of the system has a dimer configuration. While a zero (negative) excitation energy of the elementary magnon at $p = \pi$ results in zero (negative) excitation energies for some two-magnon excitations with total momentum $p = 0$, the excitation energies found with our *Ansatz* are positive at $p = 0$, as required by the fact that our uMPS approximation is a minimum for the energy functional on our variational manifold. This is a consequence of the aforementioned fact that our *Ansatz* is not suitable for the description of two-particle excitations. The negative energy for the elementary excitation at $p = \pi$ does, however, indicate that we could have found a lower-energy density if we would have used a two-periodic MPS with bond dimension $D = 24$, as was explained at the end of Sec. IV C.

The same analysis can be repeated for θ increasing from 0 to $\pi/4$, which is sketched in Fig. 3. For θ slightly larger than 0, the excitation energy of the elementary magnon around $p = \pi$ increases, resulting in smaller correlation length and a decrease of the entanglement entropy (as indicated by the Schmidt values shifting downwards). Indeed, at $\theta = \arctan(\frac{1}{3}) \approx 0.1024\pi$, this is the AKLT model, for which the ground state has an exact matrix product state representation with $D = 2$. For $\theta = 0.10\pi$, the importance of the Schmidt coefficients λ_α for $\alpha > 2$ has strongly decreased. If θ is increased further, the excitation energy of the elementary magnon starts to decrease around $p = 2\pi/3$ and eventually a null mode develops. Once again, the excitation energy is slightly negative for $\theta = 0.24\pi$, which is an artifact of the small bond dimension $D = 24$. This could again be interpreted

FIG. 3. (Color online) Excitation spectrum (left) and Schmidt spectrum (right) for the bilinear-biquadratic antiferromagnetic Heisenberg model in the region $\theta \in [0,\pi/4)$ obtained with a uMPS *Ansatz* with bond dimension $D = 24$.

as an indication for condensation of elementary magnons with momentum $2\pi/3$ in the phase transition at $\theta = \pi/4$, which would result in the breaking of translation invariance down to invariance under shifts over three sites. However, in the exact solution, no such trimerization occurs and the model remains critical throughout $\theta \in [\pi/4,\pi/2)$. Hence, while the MPS approximation and derived methods for excitations can provide valuable information about the phase of a system and the nature of a phase transition, such information is not always reliable and should be used carefully.

B. Linearized time-dependent variational principle

We now elaborate on the relation between the tangent space as variational *Ansatz* for excitations, and the resonance frequencies that we obtain from linearizing the TDVP equation around the variational optimum $|\Psi(A)\rangle$. For the $S = 1$ Heisenberg antiferromagnet [$\theta = 0$ in Eq. (95)], we perform an exact diagonalization of the generalized eigenvalue problem of Eq. (87) and compare the eigenvalues to the eigenvalues
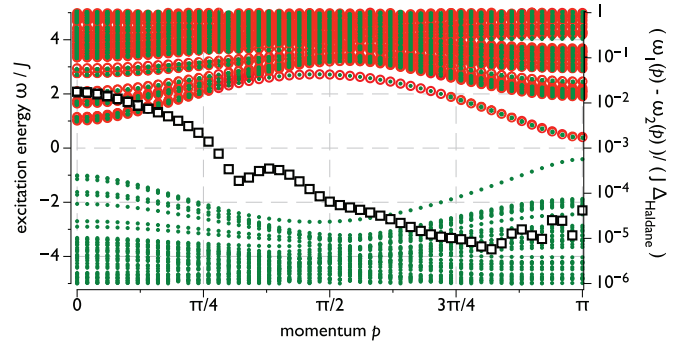
FIG. 4. (Color online) Comparison of the excitation spectrum ω obtained with the linearized TDVP equation [Eq. (87)] (green dots) with the excitation spectrum obtained with the tangent-space variational *Ansatz* for excitations $|\Phi_p(B)\rangle$ (red circles), for the Heisenberg model studied at bond dimension $D = 24$. Also shown is the absolute value of the difference between the dispersion relations (black squares, to be read on the right axis) for the lowest excitation, which corresponds to the elementary spin-1 magnon in the momentum range $\frac{\pi}{4} \lessapprox |p| \leqslant \pi$.

of the first diagonal block $H(p)$. Note that, in the latter case, we have a variational *Ansatz* state $|\Phi_p(B)\rangle$ for the corresponding excitation, whereas in the former, we do not. For bond dimension $D = 24$, a comparative result between both sets of eigenvalues is sketched in Fig. 4. The doubled spectrum of eigenvalues of Eq. (87) is reflection invariant around zero. For the lowest (positive) eigenvalue, we also plot the absolute value of the difference between both approaches.

For $D = 24$, the localized state error $\widetilde{\epsilon}(\overline{A},A)$ is approximately 3×10^{-3}. In the region around $p = \pi$, the lowest eigenvalue obtained with both methods corresponds to the elementary $S = 1$ magnon excitation. It can be observed from Fig. 4 that the difference between both dispersion relations is much smaller than $\widetilde{\epsilon}(\overline{A},A)$ in that region. The second tangent-space correction that is contained in the larger eigenvalue problem of Eq. (87) contributes little to the elementary excitations. Hence, for larger values of the bond dimension D, we expect the difference between the two approaches to be practically inexistent for elementary excitations. However, around $p \approx \pi/4$, the two-magnon continuum becomes lower in energy and the single-magnon excitation ceases to exist.[46] In that region, the difference between both approaches increases, which is to be expected, since the uMPS tangent space $\mathbb{T}$ is no longer a good variational subspace for multiparticle excitations. Corrections obtained from taking the second tangent space into account thus become more substantial.

Note, however, that the larger eigenvalue problem obtained from linearizing the TDVP is not equivalent to a variational *Ansatz* taking the second tangent space into account. This would be a completely different approach, that is far more difficult, as even within a given sector of fixed total momentum p, the second tangent space to the manifold of MPS contains infinitely many degrees of freedom. We return to this question in Sec. VII C. There is no variational *Ansatz* and hence no variational principle associated to the eigenvalues obtained from Eq. (87). This lack of an underlying variational principle results in uncontrollable errors and the possibility for spurious solutions, as we now demonstrate.
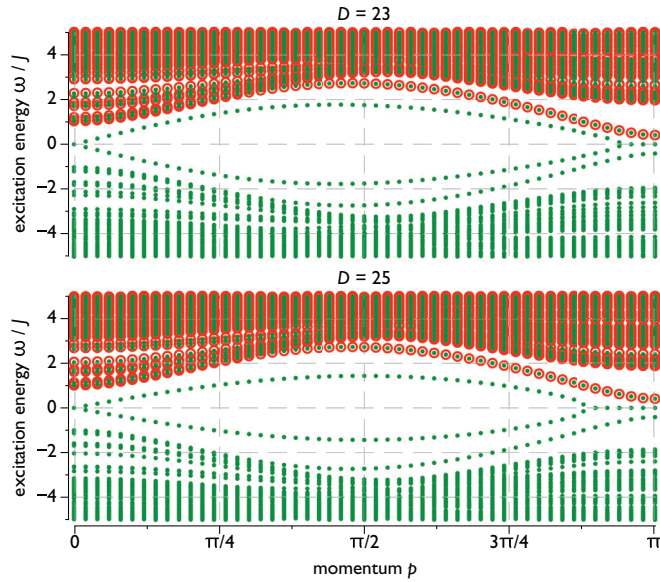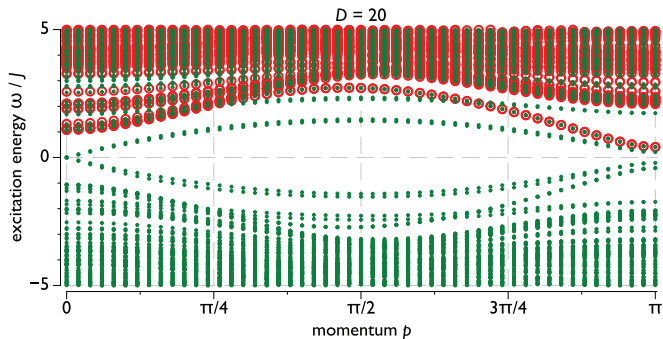
FIG. 5. (Color online) Similar comparison of the excitation spectrum ω obtained with the linearized TDVP equation [Eq. (87)] (green dots) with the excitation spectrum obtained with the tangent-space variational *Ansatz* for excitations $|\Phi_p(B)\rangle$ (red circles), for the Heisenberg model at bond dimension $D = 23$ and 25.

All went well for the results in Fig. 4 since we tuned the bond dimension to be commensurable with the SO(3) degeneracy in the virtual space, i.e., the spectrum of Schmidt coefficients contains an exact number of spin multiplets. As is well known, the Haldane phase is a symmetry-protected topological phase since it only has nontrivial projective representations of SO(3) in its virtual space, i.e., it only has half-integral spin multiplets and every Schmidt value has an even degeneracy.[74] If we choose an odd-valued bond dimension such as $D = 23$ or $D = 25$, the virtual space can not entirely consist out of complete half-integral spin representations, and some vectors of an incomplete multiplet will have to be added. The variational optimum is then no longer unique: there will be a flat valley of equally good choices A that produce the same energy expectation values $\langle\Psi(A)|\hat{H}|\Psi(A)\rangle$. Different values of A correspond to different choices of Schmidt vector associated to the smallest Schmidt coefficient. A first effect of this degenerate valley of energy optima is that the location of the energy optimum with an imaginary-time simulation according to the time-dependent variational principle converges much slower. In addition, the corresponding optimum $|\Psi(A)\rangle$ will have a slight breaking of the SO(3) symmetry of the Hamiltonian, of the same order as the size of of the Schmidt coefficients of the broken multiplet. The effect on the excitation spectrum obtained with the variational *Ansatz* and with the linearized TDVP equation is shown in Fig. 5. The variational *Ansatz* $|\Phi(B; A)\rangle$ still produces a qualitatively correct excitation spectrum, where the degeneracy of the elementary $S = 1$ magnon is slightly lifted. The energy difference between the three states is easily an order of magnitude smaller than the size of the Schmidt coefficients of the broken multiplet, i.e., in the order of 10^{-4} to 10^{-5}. In contrast, when using the linearized TDVP equation around a given point in this optimal value, then it produces a fake null

mode. Indeed, let $A(s)$ be a one-parameter group of tensors A that runs through this valley. Then, $\langle\Psi(A(s))|\hat{H}|\Psi(A(s))\rangle$ is s independent. Since all of these are variational optima, we automatically obtain $\langle\Psi(\overline{A(s)})|\hat{H}|\Phi_p(B; A(s))\rangle = 0$ for any $|\Phi_p(B; A(s))\rangle \in \mathbb{T}_{\text{MPS}}^{\perp}(A(s))$. We also obtain

$$
\frac{d^2}{ds^2}\langle\Psi(A(s))|\hat{H}|\Psi(A(s))\rangle\bigg|_{s=0}
$$
$$
= 2\langle\Phi_0(\overline{B}|\hat{H}|\Phi_0(B)\rangle + \langle\Psi(A)|\hat{H}|\Upsilon_{0,0}(B,B)\rangle
$$
$$
+ \langle\Upsilon_{0,0}(\overline{B},\overline{B})|\hat{H}|\Psi(A)\rangle
$$
$$
= [\,\boldsymbol{B}^{\dagger} \quad \overline{\boldsymbol{B}}^{\dagger}\,]\begin{bmatrix} H(0) & K(0) \\ \overline{K(0)} & \overline{H(0)} \end{bmatrix}\begin{bmatrix} \boldsymbol{B} \\ \overline{\boldsymbol{B}} \end{bmatrix} = 0
$$

with $A = A(0)$ and $B = dA/ds(0)$. Hence, the Hessian of the energy appearing in the left-hand side of Eq. (87) for $p = 0$ has a zero eigenvalue, and so does the generalized eigenvalue equation of Eq. (87) itself. In fact, for the present case, there are two independent null modes. This could be argued to be a good thing since the above proof essentially shows the existence of Goldstone modes whenever the ground state breaks a continuous symmetry. However, the problem with this null mode is that it is always there, irrespective of how good the ground-state approximation $|\Psi(A)\rangle$ is and how small the remaining symmetry breaking is. If the symmetry breaking is only tiny, so is the lift of the degeneracy of the spin multiplets in the eigenvalue spectrum of $H(p)$. In contrast, the spectrum of eigenvalues of Eq. (87) completely loses its multiplet structure as it starts with a twofold degenerate Goldstone mode and the energy difference between the different magnon excitations can be as large as $\mathcal{O}(10^{-2})$. This indicates how fragile the eigenvalue structure of Eq. (87) is.

If we have unwillingly broken an unknown symmetry of the Hamiltonian, no matter how small the symmetry breaking is, the linearized TDVP equation [Eq. (87)] will always force a zero-energy mode upon us. It is unable to detect whether the symmetry breaking is artificial, i.e., the optimal valley of A's map to states $|\Psi(A)\rangle \in \mathcal{M}$ which are very close in Hilbert space and are centered around one unique direction $|\Psi_0\rangle$ (the true ground state) that is not exactly contained in the manifold (or whether the symmetry breaking is physical), i.e., the optimal valley of A's map to states $|\Psi(A)\rangle$ that are lying in completely different direction of Hilbert space and are close to the different directions in the exact degenerate ground-state subspace. In the latter case, if we can well approximate the different ground state by uMPS $|\Psi(A)\rangle$, we expect that $H(p)$ itself contains very small eigenvalues corresponding to Goldstone modes since the terms $K(0)$ vanish for the exact ground state anyway. As a final example, we plot the same comparison for bond dimension $D = 20$ in Fig. 6. Even though this value of the bond dimension is even, it still cuts through a spin multiplet in the virtual space since the Schmidt coefficients with number 17 to 22 all correspond to an $S = \frac{5}{2}$ multiplet. For $D = 23$ and 25, we had two Goldstone modes that became massless both at $p = 0$ and π, but aside from those, Eq. (87) did reproduce a nearly degenerate gapped magnon excitation around the same energy as the variational *Ansatz*. For $D = 20$, the situation is completely different. The two Goldstone modes are only massless at momentum $p = 0$, but mix with other excitations near momentum π. Indeed, aside

FIG. 6. (Color online) Similar comparison of the excitation spectrum ω obtained with the linearized TDVP equation [Eq. (87)] (green dots) with the excitation spectrum obtained with the tangent-space variational *Ansatz* for excitations $|\Phi_p(B)\rangle$ (red circles), for the Heisenberg model at bond dimension $D = 20$.

from the magnon excitation and the two Goldstone modes, we find other excitations below the multiparticle continuum around momentum π.

C. Spectral functions

We close this section by showing a simple result for the spectral function obtained by computing Eq. (94). For the $S = 1$ Heisenberg antiferromagnet, where we use $\hat{O}^{(\alpha)} = \hat{O}^{(\beta)} = \hat{S}^x$, we obtain the result in Fig. 7. If we look at absolute scale, we note that the spin operators have a large overlap with the single magnon excitation around momentum $p = \pi$, but that this overlap goes to zero very quickly as the momentum gets closer to the point where the single magnon excitation vanishes in the two-particle continuum. By looking at the logarithm of the spectral function, we show that part of the spectral weight is also distributed over the discrete eigenvalues in the multimagnon continuum. Since the tangent space $\mathbb{T}$ captures $\hat{S}_p^{(x)}$ acting on the ground state exactly, no spectral weight is lost. Since we accurately reproduce the single magnon dispersion relation,[17] we assume that the corresponding variational estimate for the eigenstate $|\Phi_p(B)\rangle$ is sufficiently accurate to correctly capture the spectral overlap with the spin operators. Then, we necessarily also have an accurate estimate for the fraction of the spectral weight that
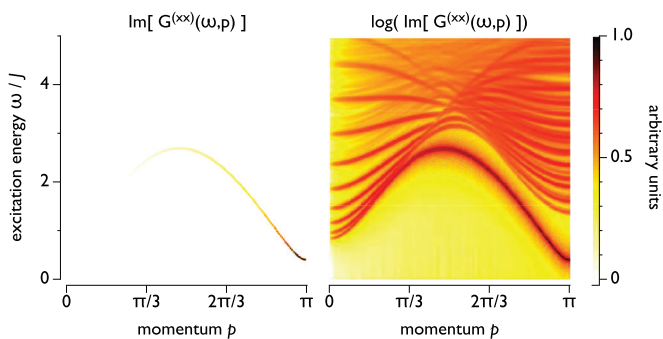
FIG. 7. (Color online) Spectral function $\mathrm{Im}[G^{(xx)}(\omega, p)]$ for the $S = 1$ Heisenberg antiferromagnet with operators $\hat{O}^{(\alpha)} = \hat{O}^{(\beta)} = \hat{S}^x$, as obtained from Eq. (94). We have applied the Chebychev decomposition of the δ function using the first $N = 1000$ polynomials. The bond dimension was chosen as $D = 192$.
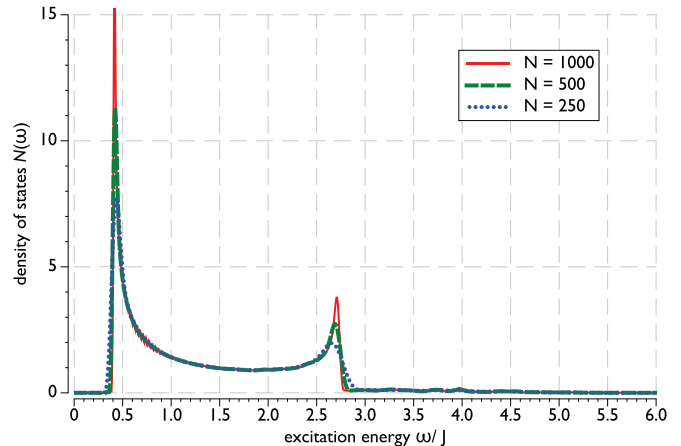
FIG. 8. (Color online) Density of states $N(\omega)$ obtained from integrating $\mathrm{Im}[G^{(xx)}(\omega, p)]$ over $p \in [0, 2\pi)$ for the $S = 1$ Heisenberg antiferromagnet. We show results where we have applied the Chebychev decomposition of the δ function using a varying number $N = 250, 500,$ or 1000 polynomials. The bond dimension was chosen as $D = 192$.

is lost to the 2- and 3-magnon continuum, but instead of this spectral weight being spread out in a continuous distribution over ω, it is localized at the artificial, discrete set of eigenvalues that is supported within the tangent space. Hence, while a tangent-space-based computation of spectral functions is acceptable if the only interest is in the contribution of the elementary excitations, it can not accurately reproduce the contribution of multiparticle excitations. A more advanced strategy is required, the foundations of which are discussed in Sec. VII C.

Finally, we can also integrate $\mathrm{Im}[G^{(xx)}(\omega, p)]$ over the momentum to obtain the density of states $N(\omega)$. This is illustrated in Fig. 8 and compares well to the results in Figs. 5 and 6 of Ref. 49. Clearly, the tangent space offers a very efficient method to directly generate spectral functions or the density of states once we have obtained the uMPS approximation $|\Psi(A)\rangle$ of the ground state. With the tangent-space approach, we did not need any kind of statistical procedures to extract or improve the result, nor did we need to take any kind of boundary effects or finite-size effects into account. Note that, as the number of terms N in the Chebychev decomposition of the δ function increases, the peaks become sharper and higher, and will eventually develop into singularities for $N \to \infty$. The fringes that appear for $N = 1000$ indicate that the discretization of the momentum p (namely, $dp = \pi/200$) is too rough, which is why the contribution of individual momentum states in the integrated quantity $N(\omega)$ become visible.

VI. ANALOGIES WITH OTHER VARIATIONAL METHODS

The TDVP is a general method that can be applied to any variational manifold for any quantum problem, be it a single-particle or a many-particle quantum system. However, it is not guaranteed that for any variational manifold the tangent space provides a suitable *Ansatz* for elementary excitations of the system. In fact, it is a remarkable property of the manifold of MPS that its tangent space contains states which seem to have the correct structure for well approximating

elementary excitations. The best motivation for recognizing that this is indeed the correct structure is not via the linearized TDVP equation, but by observing that the MPS tangent vectors generalize the Feynman-Bijl variational *Ansatz* for excitations. In fact, this observation allows us to simply generalize the excitation *Ansatz* to states which are no longer in the MPS tangent space, as described in the next section.

Nevertheless, we now try to compare the methods constructed in the previous sections to similar developments with other variational methods that are used in the context of quantum many-body systems. The most prominent example is without doubt Hartree-Fock theory, i.e., mean field theory for a system of N-fermionic particles.[79–83] The variational manifold is the set of all Slater determinants, and can be identified with a Grassmann manifold. Applying the TDVP for $D > 1$ results in time-dependent Hartree-Fock theory, and by linearizing these equations around the stationary solution, one obtains the so-called random phase approximation.[84,85] In this context, the random phase approximation can also be derived using Green's function techniques[86] or by using the equations-of-motion approach.[87]

In line with what we have illustrated in the previous section, is it well known in quantum chemistry and nuclear physics that the random phase approximation can be plagued by instabilities. In that case, it is also better to ignore the off-diagonal blocks in the generalized eigenvalue equation, which boils down to a diagonalization of the Hamiltonian within the tangent space[86] and which is known as the Tamm-Dancoff approximation.[88,89] However, unlike as for MPS, which has a refinement parameter D (the bond dimension), there is no systematic way to improve the Hartree-Fock approximation for the ground state, and the random phase approximation very often provides the only way of taking correlations in the ground state into account.

For Hartree-Fock theory, the tangent space consists of all single particle-hole excitations, and there is one-to-one mapping between derivatives with respect to the variational parameters and physical excitation operators creating the same tangent state. Similarly, the second tangent space of the Hartree-Fock manifold corresponds to all states with two particle-hole excitations on top of the Slater determinant. We discuss the possibility for excitation operators that are in one-to-one correspondence to the tensors B for the case of MPS tangent vectors in the next section, as these are required for further developments that take us beyond the tangent space. For a more detailed comparison between Hartree-Fock theory and matrix product states on finite lattices, we refer the reader to Ref. 90. This paper also describes and implements an MPS analog of the post-Hartree-Fock method CISD, which is an abbreviation for configuration interaction with singles and doubles, and boils down to an exact diagonalization of the Hamiltonian within the subspace obtained by combining the optimal MPS, its tangent space, and its double tangent space. However, such a method breaks down in the thermodynamic limit, as it lacks extensivity. Indeed, the MPS tangent space and double tangent space have a different normalization structure and can not lower the energy density. Constructing an MPS analog of an extensive post-Hartree-Fock method such as coupled cluster theory, on the other hand, could also be relevant for calculations in the thermodynamic limit.

Mean field theory for bosons can start from the N-particle wave function $\Psi(\vec{r}_1, \ldots, \vec{r}_N) = \prod_{n=1}^{N} \varphi(\vec{r}_n)$ or, for systems in second quantization, from the coherent state $|\Psi\rangle = \exp\left[\int \varphi(\vec{r})\hat{\psi}^\dagger(\vec{r}) \, d\vec{r}\right] |\Omega\rangle$. In both cases, the variational parameters correspond to the choice of single-particle function $\phi(\vec{r})$. Applying the TDVP results in the Gross-Pitaevskii equation for $\varphi(\vec{r})$.[91,92] Linearization around the stationary point results in the Bogoliubov–de Gennes equation.[93] As in Hartree-Fock theory, there is a clear link between the tangent vectors and the action of physical operators since

$$\frac{\delta}{\delta\varphi(\vec{r})} |\Psi\rangle = \frac{\delta}{\delta\varphi(\vec{r})} \exp^{\int \varphi(\vec{r})\hat{\psi}^\dagger(\vec{r}) \, d\vec{r}} |\Omega\rangle$$
$$= \hat{\psi}^\dagger(\vec{r}) |\Psi\rangle$$

and the Bogoliubov–de Gennes equation can also be obtained from a bosonic expansion of the Hamiltonian up to second order, which clearly shows its nonvariational character (since higher-order terms of the expansion are being ignored). The relation between the Gross-Pitaevskii theory and our framework is best studied with continuous MPS.[94–96] Indeed, a continuous MPS with bond dimension $D = 1$ corresponds exactly to the mean field coherent state $|\Psi\rangle$, whereas correlations are systematically included for larger D. Applying the TDVP for $D > 1$ results in a noncommutative version of the Gross-Pitaevskii equation.[97]

VII. BEYOND THE TANGENT SPACE

Most of this paper so far has been focused on the MPS tangent space $\mathbb{T}$, although we also briefly encountered states in the double tangent space $\mathbb{T}^{(2)}$. We now discuss several extensions that go beyond the strict MPS tangent space $\mathbb{T}$. As there is no natural reason to stick to the tangent space when constructing a variational *Ansatz* for excitations, we discuss the effect of using tensors B that act on several sites in the first subsection. While the tangent space did appear naturally for time evolution, we show in the second subsection that there can also be a role for momentum zero states with a tensor B acting on several sites. Finally, in the last subsection, we give a general outlook on how successively replacing more and more A matrices corresponds to building multiparticle states, eventually leading to the construction of a complete Fock space in which we can define an effective low-energy theory. Therefore, we need to link the action of replacing tensors A to physical operators.

A. Spreading blocks over several sites

As a variational *Ansatz* for excitations, the tangent vectors $|\Phi_p(B)\rangle$ constructed at the base point $|\Psi(A)\rangle$ were motivated because they capture exactly the Feynman-Bijl *Ansatz* for one-site operators. We expect that these states can accurately capture the effect of operators with bigger support as well since, via the virtual dimension, the tensor B can have have an effect over a distance approximately $\log_d D$ away from the site on which B is living. However, there is no intrinsic reason why we should stick to the tangent space, and we can define a generalized excitation *Ansatz* where we replace the tensors A

TABLE I. Variational excitation energy for the single magnon excitation at momentum $p = \pi$ in the AKLT model, obtained by using the variational *Ansatz* $|\Phi_{p=\pi,K}(B)\rangle$, where the tensor B replaces the matrices A on K consecutive sites.

K	E_{magnon}
1	0.370370370370370
2	0.350634581086138
3	0.350165202217295
4	0.350129173076820
5	0.350124768941854
6	0.350124225439426
7	0.350124164567491
8	0.350124158096952
9	0.350124157417523
10	0.350124157346044
11	0.350124157338490
12	0.350124157337687

on a block of K consecutive sites by a single tensor B as

$$|\Phi_{p,K}(B;A)\rangle = \sum_{n\in\mathbb{Z}} e^{ipn} \sum_{\{s_n\}=1}^{d} \boldsymbol{v}_{\text{L}}^{\dagger} \left[\left(\prod_{m<n} A^{s_m} \right) B^{s_n s_{n+1}\ldots s_{n+K-1}} \right.$$
$$\left. \times \left(\prod_{m'\geqslant n+K} A^{s_{m'}} \right) \right] \boldsymbol{v}_{\text{R}} |\boldsymbol{s}\rangle. \quad (96)$$

One particular example where this construction is extremely useful is when the ground state is an exact MPS, as for the AKLT model. In that case, it is impossible to systematically improve the tangent-space *Ansatz* $|\Phi_p(B)\rangle = |\Phi_{p,1}(B)\rangle$ since there is no point in increasing the bond dimension. Note that in this case the error on the approximate excitation energies is purely variational, i.e., the variational energies are an upper bound for the exact value. By increasing the spatial support of B, the variational energies can systematically be lowered and will thus converge to the exact result. We can show that this convergence is exponentially fast.[98] Table I shows the value for the magnon gap at momentum $p = \pi$ in the AKLT model, obtained with this *Ansatz* for K ranging from 1 to 12, and it can be observed that the energy converges at a rate of approximately 1 digit per extra site. The $K = 1$ case corresponds to $E_{\text{magnon}}^{(K=1)} = \frac{10}{27}$, as was calculated analytically in Ref. 58.

For systems where the ground state is not an exact MPS, it can of course also be useful to combine the scaling in D with a scaling in K, in particular for approximating excitations of which we expect that they might be spread out over a large number of sites, such as bound states with a very weak binding energy.

For dynamic correlation functions, it can also be useful to work in the larger subspace $\mathbb{T}_{p,K}$, in particular when trying to evaluate the Green's function of operators $\hat{O}^{(\alpha)}$ which have support on more than one site. As long as the size of the support of $\hat{O}^{(\alpha)}$ is smaller than or equal to K, the subspace $\mathbb{T}_{p,K}$ can capture $\hat{O}_p^{(\alpha)}$ [defined in Eq. (92)] exactly and no spectral weight is lost if $\hat{H}$ is being replaced by the effective Hamiltonian on $\mathbb{T}_{p,K}$ in Eq. (91). If, however, K is smaller than the number of sites on which $\hat{O}^{(\alpha)}$ acts, then a truncation occurs and part of the spectral weight will be lost in the approximation.

In order to be able to work with the variational subspace $\mathbb{T}_{p,K}$ efficiently, we need to be able to transfer some of the techniques of parametrizing tangent vectors to this larger space. For example, we can now discuss the null space $\mathbb{N}_{p,K}$ of the linear map $\Phi_{p,K} : \mathbb{C}^{D \times d^K \times D} \to \mathbb{H}$. While we can not infer the form of the null modes from the principal fibre bundle construction of $\mathcal{M}$, we can easily generalize the map $\mathcal{N}_p^{(A)}$ defined in Eq. (13) and just check explicitly that $|\Phi_{p,K}(B)\rangle = 0$ for any B which is of the form

$$B^{s_1 s_2 \ldots s_{K-1} s_K} = A^{s_1} x^{s_2 \ldots s_{K-1} s_K} - e^{-ipK} x^{s_1 s_2 \ldots s_{K-1}} A^{s_K},$$
$$\forall \ X \in \mathbb{C}^{D \times d^{K-1} \times D}.$$

The dimension of the null space $\mathbb{N}_{p,K}$ of $\Phi_{p,K}$ is thus given by the number of linearly independent tensors B that can be obtained from this construction. Expressing that $B = 0$, multiplying this equation with $(A^{s_1} \ldots A^{s_K})^{\dagger} l$ and summing over all physical indices results in the condition $(l|\mathbb{E}_{A\ldots A}^x (\mathbb{1} - e^{-iKp}\mathbb{E}) = 0$. We have introduced the shorthand notation

$$\mathbb{E}_{A\ldots A}^x = \sum_{s_1=1}^{d} \cdots \sum_{s_{K-1}=1}^{d} x^{s_1 \ldots s_{K-1}} \otimes \overline{A}^{s_1} \ldots \overline{A}^{s_{K-1}}.$$

Clearly, this condition can not be satisfied for $p \neq 0$, while for $p = 0$ the one-dimensional solution space is given by $x^{s_1 \ldots s_{K-1}} \sim A^{s_1} \ldots A^{s_{K-1}}$. However, expressing that $|\Phi_{p,K}(B)\rangle \perp |\Psi(A)\rangle$ imposes an additional condition at momentum zero, so that the subspaces $\mathbb{T}_{p,K}^{\perp}$ which are orthogonal to the ground state have a physical dimension $(d-1)d^{K-1}D^2$.

One can then also show that the a complement to the null space $\mathbb{N}_{p,K}$ can be constructed as the subspace $\mathbb{B}_K$ of solutions to either the left or right gauge fixing condition

$$\sum_{s=1}^{d} A^{s\dagger} l B^{s s_2 \ldots s_K} = 0, \quad \forall \ s_2, \ldots, s_K = 1, \ldots, d$$

$$\text{or} \sum_{s=1}^{d} B^{s_1 s_2 \ldots s_{K-1} s} r A^{s\dagger} = 0, \quad \forall \ s_1, \ldots, s_{K-1} = 1, \ldots, d.$$
$$(97)$$

We can then build a linear representation $B^{s_1 s_2 \ldots s_K} = \mathcal{B}_K(X) = l^{-1/2} V_L^{s_1} X^{s_2 \ldots s_K} r^{-1/2}$, where the free parameters are represented by a tensor $X \in \mathbb{C}^{(d-1)D \times d^{K-1} \times D}$, which automatically satisfies the left gauge fixing condition and for which

$$\langle \Phi_{p,K}(\overline{\mathcal{B}}_K(\overline{X})) | \Phi_{p',K}(\mathcal{B}_K(X')) \rangle$$
$$= 2\pi \delta(p-p') \sum_{s_2=1}^{d} \cdots \sum_{s_K=1}^{d} \text{tr}[X^{s_2 \ldots s_K \dagger} X'^{s_2 \ldots s_K}]. \quad (98)$$

Hence, in terms of the degrees of freedom in X, the effective normalization matrix is the unit matrix. We can then define the effective Hamiltonian by generalizing the computation from Sec. IV and diagonalize it iteratively to obtain a method that scales as $\mathcal{O}(d^K D^3)$. Clearly, the exponential scaling in K limits the block sizes that we can achieve with this approach. However, if for describing some excitation we would like to have access to larger block sizes, the tensor B itself can also be written as matrix product decomposition $B^{s_1 s_2 \ldots s_k} = B^{s_1}(1) B^{s_2}(2) \ldots B^{s_K}(K)$.

Finally, we elaborate on the obvious statement that $\mathbb{T}_p^\perp = \mathbb{T}_{p,1}^\perp \subset \mathbb{T}_{p,2}^\perp \subset \ldots \subset \mathbb{T}_{p,K}^\perp$. We can represent any $|\Phi_{p,K-1}(\tilde{B})\rangle$ as a vector $|\Phi_{p,K}(B)\rangle \in \mathbb{T}_{p,K}^\perp$. Note that we are restricting our discussion to the tangent spaces orthogonal to the uMPS $|\Psi(A)\rangle$ since this is a restriction that we will always want to make in applications. One way to represent this embedding in the parametrization is by setting $B^{s_1 \ldots s_{K-1} s_K} = \tilde{B}^{s_1 \ldots s_{K-1}} A^{s_K}$, which has the favorable effect that if $\tilde{B}$ satisfies the left gauge fixing condition [first equation in Eq. (97)], then so will B. Now, for any $K \geqslant 2$, we can also try to define the space $\mathbb{V}_{p,K}$ as the orthogonal complement of $\mathbb{T}_{p,K-1}^\perp$ in $\mathbb{T}_{p,K}^\perp$, i.e., $\mathbb{V}_{p,K}$ contains all states $|\Phi_{p,K}(C)\rangle$ which are orthogonal to all vectors $|\Phi_{p,K-1}(B)\rangle$. Note that we can easily parametrize $\mathbb{V}_{p,K}$, by setting

$$C^{s_1,\ldots,s_K} = \mathcal{V}^{s_1,\ldots,s_K}(Y) = l^{-1/2} V_L^{s_1} Y^{s_2,\ldots,s_{K-1}} V_R^{s_K} r^{-1/2},$$

where $Y \in \mathbb{C}^{(d-1)D \times d^{K-2} \times (d-1)D}$. Here, we have introduced the set of $(d-1)D \times D$ matrices V_R^s ($s = 1, \ldots, d$) which is such that $\sum_{s=1}^d V_R^{s\dagger} r^{1/2} A^s = 0$ and $\sum_{s=1}^d V_R^{s\dagger} V_R^s = \mathbb{1}_{(d-1)D}$. This matrix can be created analogously to the construction for V_L as discussed in the context of Eq. (19). One would need V_R to build a parametrization of tangent vectors that automatically satisfies the right gauge fixing condition. Note that the number of parameters in Y corresponds to the dimension of $\mathbb{V}_{p,K}$. Indeed, since $\mathbb{T}_{p,K}^\perp$ has a dimension $d^{K-1}(d-1)D^2$, the dimension of $\mathbb{V}_{p,K}$ is given by $d^{K-1}(d-1)D^2 - d^{K-2}(d-1)D^2 = d^{K-2}(d-1)^2 D^2$. It can also by checked that $\langle \Phi_{p,K-1}(B)|\Phi_{p,K}(C)\rangle = 0$ for any $C = \mathcal{V}(Y)$, and for any $|\Phi_{p,K-1}(B)\rangle \in \mathbb{T}_{p,K-1}^\perp$. In addition, we also obtain

$$\langle \Phi_{p,K}(\overline{\mathcal{V}}(\overline{Y}))|\Phi_{p',K}(\mathcal{V}(Y'))\rangle = 2\pi \delta(p - p') \, \mathrm{tr}[Y^\dagger Y'] \quad (99)$$

so that this parametrization corresponds to an effective normalization matrix which is the unit matrix. There is at least one simple application of this construction. The error measure $\epsilon(\overline{A}, A)$ was defined in Eq. (3) as the norm of the part of $[\hat{H} - H(\overline{A}, A)]|\Psi(A)\rangle$ that was lost in the projection onto the MPS tangent space $\mathbb{T}_p$. If $\hat{H}$ is a nearest-neighbor Hamiltonian $\hat{H} = \sum_{n\in\mathbb{Z}} \hat{T}^n \hat{h} \hat{T}^{-n}$, with $\hat{h}$ acting on two sites, we can exactly represent $[\hat{H} - H(\overline{A}, A)]|\Psi(A)\rangle$ in $\mathbb{T}_{p=0,K=2}^\perp$. Hence, the error measure corresponds to the norm of the projection of $[\hat{H} - H(\overline{A}, A)]|\Psi(A)\rangle$ onto $\mathbb{V}_{p=0,K=2}$. Using the parametrization $\mathcal{V}$, we can easily obtain that $\hat{P}_{\mathbb{V}_{p=0,K=2}}[\hat{H} - H(\overline{A}, A)]|\Psi(A)\rangle = |\Phi_{p=0,K=2}(\mathcal{V}(Y))\rangle$, where Y is given as

$$Y = \sum_{s,t,u,v=1}^D \langle s,t|\hat{h}|u,v\rangle \, V_L^{s\dagger} l^{1/2} A^u A^v r^{1/2} V_R^{t\dagger}. \quad (100)$$

We then obtain for the localized error measure $\tilde{\epsilon}(\overline{A}, A) = \|\mathbb{Z}\|^{-1/2} \epsilon(\overline{A}, A) = \|\mathbb{Z}\|^{-1/2} \||\Phi_{p=0,K=2}(\mathcal{V}(Y))\rangle\| = \sqrt{\mathrm{tr}[Y^\dagger Y]} = \|Y\|_F$. Hence, we do not have to compute a difference of terms which are almost equal in magnitude. This is a a much quicker and more stable algorithm for computing $\tilde{\epsilon}(\overline{A}, A)$. If $\hat{H}$ contains only local interactions up to K sites, we can easily compute the norm of the projection of $[\hat{H} - H(\overline{A}, A)]|\Psi(A)\rangle$ onto all spaces $\mathbb{V}_{p=0,k}$ for $k = 2, \ldots, K$. The local error measure $\tilde{\epsilon}(\overline{A}, A)$ is then obtained as the square root of the sum of the squares of all

these norms, and is thus computed as a sum of positive values, without any risk for contributions that neutralize each other.

B. Dynamic expansion of the manifold

For studying time evolution, the MPS tangent space appeared naturally. If we want to approximate a time-evolving quantum state as a time-evolving MPS with bond dimension D, the time derivative of the state has to be an element of the MPS tangent space at that particular point, and the TDVP is prescribing which tangent vector is optimal. Nevertheless, it also interesting to ask oneself would it imply to try to evolve a uMPS $|\Psi(A)\rangle$ in a direction given by a state $|\Phi_{0,K}(B)\rangle \in \mathbb{T}_{0,K}^{|\Psi(A)\rangle}$ for $K > 1$. In particular, for nearest-neighbor Hamiltonians $\hat{H} = \sum_{n\in\mathbb{Z}} \hat{T}^n \hat{h} \hat{T}^{-n}$ with $\hat{h}$ acting on two sites, we can exactly represent $\hat{H}|\Psi(A)\rangle$ in $\mathbb{T}_{p=0,K=2}$. The way to go beyond to tangent space is by noting that for a variational class such as MPS, which has a refinement parameter D, it is not necessary to have a fixed value for D throughout the whole evolution. Suppose that at some point during the evolution, the error measure $\tilde{\epsilon}(\overline{A}, A)$, which measures the difference between the TDVP evolution and the exact evolution given by the Schrödinger equation, exceeds some predefined tolerance value. We then might try to reduce the error by expanding the variational manifold by increasing the bond dimension from its original value D to some new value $\tilde{D}$. We can easily embed a uMPS $|\Psi(A)\rangle \in \mathcal{M}$ with bond dimension D into the uMPS manifold $\tilde{\mathcal{M}}$ with bond dimension $\tilde{D}$ by writing it as $|\tilde{\Psi}(\tilde{A})\rangle$ with

$$\tilde{A}^s = \begin{bmatrix} A^s & 0 \\ 0 & 0 \end{bmatrix}. \quad (101)$$

Note, however, that $\tilde{A}$ is not in the set of injective MPS $\tilde{\mathcal{A}}$, which has some undeniable consequences. For example, if we follow the TDVP spirit in the most strict sense, we set $\tilde{A}(t = 0) = \tilde{A}$ and

$$\frac{d}{dt}\tilde{A}(t) = \tilde{B} = \begin{bmatrix} B_{1,1}^s & B_{1,2}^s \\ B_{2,1}^s & B_{2,2}^s \end{bmatrix}, \quad (102)$$

with $B_{1,1}^s$ a $D \times D$ matrix, $B_{1,2}^s$ a $D \times (\tilde{D} - D)$ matrix, $B_{2,1}^s$ a $(\tilde{D} - D) \times D$ matrix, and $B_{2,2}^s$ a $(\tilde{D} - D) \times (\tilde{D} - D)$ matrix. We then find the $\tilde{B}$ that minimizes $\| |\tilde{\Phi}(\tilde{B}, \tilde{A})\rangle + i\hat{H}|\tilde{\Psi}(\tilde{A})\rangle \|$. By construction, $|\tilde{\Psi}(\tilde{A})\rangle = |\Psi(A)\rangle$. However, we also find $|\tilde{\Phi}(\tilde{B}, \tilde{A})\rangle = |\Phi(B_{1,1}, A)\rangle$ and the newly added directions $B_{2,1}$, $B_{1,2}$, and $B_{2,2}$ do not feature to first order. The TDVP is very keen to restricting its evolution to the original manifold $\mathcal{M}$. Put differently, because $\tilde{A}$ is not in the set of injective MPS $\tilde{\mathcal{A}}$, it corresponds to a singular region where the tangent space $\tilde{\mathbb{T}}_0^{|\tilde{\Psi}(\tilde{A})\rangle}$ does not have dimensions $(d-1)\tilde{D}^2$ but rather $(d-1)D^2$.

A solution for this problem is to use a kind of singular perturbation theory. We set $\tilde{A}(dt) = \tilde{A} + \tilde{dA}$ with

$$\tilde{dA}^s = \begin{bmatrix} dt \, B_{1,1}^s & (dt)^{1/2} B_{1,2}^s \\ (dt)^{1/2} B_{2,1}^s & B_{2,2}^s \end{bmatrix}. \quad (103)$$

We then obtain to first order in dt that

$$|\tilde{\Psi}(\tilde{A} + \tilde{dA})\rangle - |\tilde{\Psi}(\tilde{A})\rangle = |\tilde{d\Psi_0}\rangle + |\tilde{d\Psi_1}\rangle + \mathcal{O}(dt^2) \quad (104)$$

with

$$|\widetilde{d\Psi}_0\rangle = dt\,|\Phi_{p=0,K=1}(B_{1,1};A)\rangle \qquad (105)$$

and

$$
\begin{aligned}
|\widetilde{d\Psi}_1\rangle &= dt \sum_{n\in\mathbb{Z}} \sum_{\{s_n\}=1}^{d} \sum_{m=0}^{+\infty} \boldsymbol{v}_{\mathrm{L}}^{\dagger} \left(\prod_{k<n} A^{s_k}\right) \\
&\quad \times B_{1,2}^{s_n} B_{2,2}^{s_{n+1}} \dots B_{2,2}^{s_{n+m}} B_{2,1}^{s_{n+m+1}} \left(\prod_{k>n+m+1} A^{s_k}\right) \boldsymbol{v}_{\mathrm{R}} |\boldsymbol{s}\rangle.
\end{aligned}
$$
$$(106)$$

In particular, if we set $B_{2,2} = 0$, we obtain $|\widetilde{d\Psi}_1\rangle = dt\,|\Phi_{p=0,K=2}(B;A)\rangle$ with $B^{s_1 s_2} = B_{1,2}^{s_1} B_{2,1}^{s_2}$. Since we already have $|\widetilde{d\Psi}_1\rangle \in \mathbb{T}_{p=0,K=1}$, we can restrict to $|\widetilde{d\Psi}_2\rangle \in \mathbb{V}_{p=0,K=2}$, so that we can parametrize

$$B_{1,2}^{s} = l^{-1/2} V_L^s Z_{1,2} \quad \text{and} \quad B_{2,1}^{s} = Z_{2,1} V_R^s r^{-1/2}, \quad (107)$$

where $Z_{1,2} \in \mathbb{C}^{(d-1)D\times(\widetilde{D}-D)}$ and $Z_{2,1} \in \mathbb{C}^{(\widetilde{D}-D)\times(d-1)D}$. We then automatically have $\langle \widetilde{d\Psi}_1|\widetilde{d\Psi}_2\rangle = 0$. If we now try to imitate the geometric strategy of the time-dependent variational principle, we have to optimize the parameters $B_{1,1}$, $B_{1,2}$, and $B_{2,1}$ so as to minimize

$$\||\widetilde{d\Psi}_0\rangle + |\widetilde{d\Psi}_1\rangle + i\,dt\,\hat{H}\,|\widetilde{\Psi}(\widetilde{A})\rangle\|^2. \qquad (108)$$

Because of the orthogonality of $|\widetilde{d\Psi}_1\rangle$ and $|\widetilde{d\Psi}_2\rangle$, the optimization of the parameters $B_{1,1}$ decouples from the optimization of the parameters $B_{1,2}$ and $B_{2,1}$. For $B_{1,1}$, we can use the standard TDVP prescription, i.e., if we use the left gauge fixing condition, then $B_{1,1} = \mathcal{B}(F)$ as described in Sec. III E. For the remaining parameters, we need to extremize

$$
\begin{aligned}
&\langle \widetilde{d\Psi}_1|\widetilde{d\Psi}_1\rangle + i\,dt\,\langle \widetilde{d\Psi}_1|\hat{H}|\Psi(A)\rangle - i\,dt\,\langle \Psi(\overline{A})|\hat{H}|\widetilde{d\Psi}_1\rangle \\
&= (dt)^2 \left\{ \mathrm{tr}[(Z_{1,2}Z_{2,1})^{\dagger}(Z_{1,2}Z_{2,1})] + \mathrm{tr}[(Z_{1,2}Z_{2,1})^{\dagger}Y] \right. \\
&\quad \left. + \mathrm{tr}[Y^{\dagger}(Z_{1,2}Z_{2,1})] \right\},
\end{aligned}
$$

where Y was defined in Eq. (100) for the case of a Hamiltonian with nearest-neighbor interactions. We are thus looking for the optimal matrix $Z_{1,2}Z_{2,1}$ of rank $\widetilde{D}-D$ such that $\|Z_{1,2}Z_{2,1} - Y\|_{\mathrm{F}}$ is minimized, with $\|\dots\|_{\mathrm{F}}$ the Frobenius norm or Hilbert-Schmidt norm. The best approximation can thus found by performing a singular value decomposition of the $(d-1)D \times (d-1)D$ matrix Y and retaining the largest $\widetilde{D}-D$ singular values. If $\widetilde{D} = dD$, we can exactly capture one time step of the Schrödinger evolution for a nearest-neighbor Hamiltonian exactly. This compares well to TEBD, where an exact application of the lowest-order Trotter decomposition of a nearest-neighbor Hamiltonian also increases the bond dimension from D to $\widetilde{D} = dD$.

Clearly, we can also generalize this construction to be able to evolve according to states in $\mathbb{T}_{p=0,K}$ by using

$$
\widetilde{dA} = (dt)^{1/K}
\begin{bmatrix}
0 & B_{1,2} & 0 & \dots & 0 \\
0 & 0 & B_{2,3} & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \dots & B_{K-1,K} \\
B_{K,1} & 0 & 0 & \dots & 0
\end{bmatrix}
\qquad (109)
$$

in which case $|\widetilde{\Psi}(\widetilde{A}+\widetilde{dA})\rangle - |\widetilde{\Psi}(\widetilde{A})\rangle = dt\,|\Phi_{p=0,K}(B;A)\rangle + \mathcal{O}(dt^2)$ with $B^{s_1 s_2 \dots s_K} = B_{1,2}^{s_1} B_{2,3}^{s_2} \dots B_{K-1,K}^{s_K}$.

C. Towards a Fock-space construction

As a variational *Ansatz*, the generalized subspaces $\mathbb{T}_{p,K}$ are still only useful for studying elementary excitations with a single-particle-like nature, which might include bound states of several fundamental particles of the system. If we want to be able to describe true unbound multiparticle excitations, we need a different variational *Ansatz* that supports two independent perturbations of the ground state that can be arbitrarily far apart. There is not much interesting information in the energy of multiparticle states, as we expect that in a system with local interactions, we should be able to add the momentum (modulus 2π) and energy of any two eigenstates and find a new eigenstate at that point in the energy-momentum diagram. For finite systems, there can be finite-size corrections to the energy, but this relation should hold true exactly in the thermodynamic limit, as long as the number of excitations remains finite (i.e., the density of excitations is zero). However, being able to describe multiparticle excitations is relevant for a more accurate description of dynamic correlation functions, as well as to understand the scattering properties of elementary particles and thus the interactions that exist between them. Eventually, the goal is to create a new Fock space on top of the interacting MPS vacuum, in which we can build an effective Hamiltonian that describes the low-energy behavior of the system. This effective Hamiltonian can then be used to study the thermodynamics of the system, i.e., the behavior of the system under all kinds of perturbations, such as adding a nonzero temperature, applying external fields, quenching some parameters.

We now assume that we have a complete description of all elementary particles in the system, labeled by some index α, in terms of states $|\Phi_{p,K}(B^{(\alpha)})\rangle$, which are normalized as $\langle \Phi_{p,K}(\overline{B}^{(\alpha)})|\Phi_{p',K}(B^{(\beta)})\rangle = 2\pi\delta(p-p')\delta_{\alpha,\beta}$. In general, the tensors $B^{(\alpha)}$ can depend on the momentum p. However, we expect the momentum dependence to be weak if we expand around minima in the dispersion relation $\omega^{(\alpha)}(p)$, and we ignore this subtlety in the present discussion. The span of these states thus defines our single-particle space. The logical next step is the definition of two-particle space. The double tangent space $\mathbb{T}^{(2)}$ that we encountered in the context of Sec. IV C seemed to have the correct structure to capture two independent disturbances of the ground state. We can now generalize this for tensors B acting on K consecutive sites by first defining for every $n_2 > n_1 + K$ the states

$$
\begin{aligned}
|\Upsilon_{n_1,n_2,K}(B_1,B_2)\rangle &= \boldsymbol{v}_{\mathrm{L}}^{\dagger} \left(\prod_{k<n_1} A^{s_k}\right) B_1^{s_{n_1}\dots s_{n_1+K-1}} \\
&\quad \times \left(\prod_{n_1+K\leqslant k<n_2} A^{s_k}\right) B_2^{s_{n_2}\dots s_{n_2+K-1}} \\
&\quad \times \left(\prod_{n_2+K\leqslant k} A^{s_k}\right) \boldsymbol{v}_{\mathrm{R}} |\boldsymbol{s}\rangle. \qquad (110)
\end{aligned}
$$

To construct a full basis of states that is complete but not overcomplete, we have started from the set of states where the

perturbations encoded by B_1 and B_2 are localized on sites n_1 and n_2, respectively. It is clear that we can only define these states for $n_2 > n_1 + K$ or, alternatively, if we would like to put B_2 on a position $n_2 < n_1 - K$, we should just write it as $|\Upsilon_{n_2,n_1,K}(B_2,B_1)\rangle$. Hence, we do not expect to be able to label the double tangent space by two completely independent momentum numbers without overcounting. However, by setting $n_1 = n$ and $n_2 = n + m$, we can take the Fourier transform with respect to n and obtain a new basis of states with a well-defined total momentum (for $m \geqslant K$)

$$|\Upsilon_{p,m,K}(B_1,B_2)\rangle = \sum_{n\in\mathbb{Z}} e^{ipn} \boldsymbol{v}_{\mathrm{L}}^\dagger \left(\prod_{k<n} A^{s_k}\right) B_1^{s_n\cdots s_{n+K-1}}$$

$$\times \left(\prod_{n+K<k<m} A^{s_k}\right) B_2^{s_{n_2}\cdots s_{n_2+K}}$$

$$\times \left(\prod_{n_2+K<k} A^{s_k}\right) \boldsymbol{v}_{\mathrm{R}} |\boldsymbol{s}\rangle. \qquad (111)$$

For fixed p and for all $m \geqslant K$, these states span a space $\mathbb{T}_{p,K}^{(2)}$. They should be considered as the product state basis because, by taking linear combinations of states $|\Upsilon_{p,m,K}(B_1,B_2)\rangle$ with different B_1's and B_2's, we will create "entanglement" between the two excitations.

So far, we have only fixed the total momentum, whereas ideally, we would like to put both excitation in an individual momentum superposition by building a state such as

$$|p - q, q, \alpha, \beta\rangle = \sum_{m>K} e^{iqm} |\Upsilon_{p,m,K}(B_1^{(\alpha)},B_2^{(\beta)})\rangle. \quad (112)$$

The asymptotic part of this state ($m \gg K$) looks like an excitation $B^{(\alpha)}$ with momentum $p - p_2$ and an excitation $B^{(\beta)}$ with momentum q. One can indeed show that the energy of this state is dominated by the asymptotic part and is given by $\omega^{(\alpha)}(p - q) + \omega^{(\beta)K}(q)$. However, this is in itself not a good eigenstate since it is not even allowed to transform the (internal) position coordinate m, which is only labeling sites on a half infinite lattice $m > K$, to momentum space. A proper approximation for a stationary scattering state of the full Hamiltonian has an asymptotic contribution of every other pair of particles with total momentum p and the same energy, in particular from the state $|q, p - q, \beta, \alpha\rangle$. The coefficients with respect to the different states appearing in the asymptotic region determine the content of the scattering matrix. In one dimension, the scattering matrix is determined by the simultaneous effect of interactions and statistics and there is no unambiguous way to distinguish between both. Free particles with one type of statistics have an equivalent description as interacting particles with a different type of statistics. In any way, to get an accurate approximation of the scattering coefficients, we also need an accurate description of the scattering state in the regime where the two excitations

are close to each other and are interacting with each other. Given that the $\mathbb{T}_{p,K}^{(2)}$ corresponds to a half infinite lattice, where every site contains the internal degrees of freedom of the two excitations B_1 and B_2, the subspace $\mathbb{T}_{p,K}^{(2)}$ is infinite dimensional and there is no straightforward way to diagonalize the Hamiltonian. We discuss how to tackle this scattering problem elsewhere.[99]

On an intuitive level, it is clear that putting more and more B tensors into the uMPS $|\Psi(A)\rangle$ corresponds to having more particles. However, several complications have to be taken into account. In order to build a consistent framework, we have to check that the vacuum and the elementary excitations are good approximations of true eigenstates of the Hamiltonian. This requires that, e.g., the Hamiltonian does not (or only very weakly) couples the two-particle space with the single-particle space [or even with the vacuum state, which is expressed by the matrices $K(p)$ in Eq. (87)]. If this is the case, then we can start building a Fock space in which we can construct an effective Hamiltonian describing the interactions between the elementary excitations of the system. The two-particle interaction is already computed in Ref. 99 in order to solve the scattering problem. There can also be three-particle couplings, or couplings that do not respect the particle number and transform three excitations into two excitations, and so forth. To know how these excitations couple to physical perturbations of the system, it is also required to identify the excitations with the action of physical operators, such that $|\Phi_{p,K}(B^{(\alpha)})\rangle = \hat{A}_p^{(\alpha)\dagger}|\Psi(A)\rangle$. It can be shown that, if K is equal to or larger than the injectivity length ℓ,[100] then we can always find an operator $\hat{A}_p^{(\alpha)\dagger}$ that creates the excitation by acting on exactly K sites. If $K < \ell$, the corresponding operator might have to act on ℓ sites. If we want to identify $\hat{A}_p^{(\alpha)\dagger}$ with a creation operator, we should also require that $\hat{A}_p^{(\alpha)}$ annihilate the ground state. This requires that $\mathbb{E}_{AA\ldots A}^B$ (with K matrices A) is zero, which is completely different and not compatible with the kind of gauge fixing conditions that we have been imposing so far. It remains an open question as to whether we can build such a theory consistently and efficiently.

VIII. SUMMARY

In this section, we summarize some of the key concepts and accompanying equations that were discussed throughout this paper. While the implementational details will be discussed in greater detail elsewhere,[40] this summary sketches a workflow by which one should be able to implement the different algorithms required to reproduce the results presented in this paper. To facilitate the interpretation of the central equations, we now present them in a graphical format, which is well known within the tensor network community. Our starting point is the notion of a uniform MPS, discussed in Sec. II. It is given by
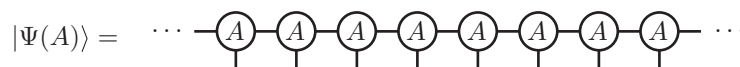
We have not drawn the boundary vectors, as they are completely irrelevant anyway. The main properties that we need of the uniform MPS are its transfer matrix $\mathbb{E}$, whose largest eigenvalue should be 1 for reasons of normalization, and the corresponding

left and right eigenvectors l and r corresponding to this eigenvalue. They are represented as

Tensors A with the physical index pointing upwards are automatically implied to be complex conjugated; this will not be denoted explicitly in the graphical representation. If the uniform MPS $|\Psi(A)\rangle$ is injective, all other eigenvalues of the transfer matrix lie within the unit circle, i.e., their magnitude is smaller than one. In all expectation values of local observables $\hat{o}$, the half infinite string of transfer matrices to the left and right produce the fixed points l and r, such that these expectation values can be written as

$$\langle\Psi(A)|\hat{O}|\Psi(A)\rangle = $$

The first question is how to find a uniform MPS description of the ground state of a given translation-invariant Hamiltonian $\hat{H} = \sum_{n\in\mathbb{Z}} \hat{T}^n \hat{h} \hat{T}^{-n}$, where $\hat{h}$ is assumed to be a nearest-neighbor Hamiltonian acting on sites 0 and 1 for reasons of simplicity. Here, imaginary-time evolution using the TDVP provides an efficient algorithm. The TDVP requires that we introduce the tangent space of uniform MPS, which consists of the states $|\Phi(B;A)\rangle = |\Phi_0(B;A)\rangle$ given as

$$|\Phi(B;A)\rangle = \sum_{n\in\mathbb{Z}} \cdots$$

Infinitesimal gauge transformations establish an equivalence $|\Phi(\tilde{B};A)\rangle = |\Phi(B;A)\rangle$ for any $\tilde{B}$ given by

This freedom in the parametrization allows us to restrict to a set of parameters B that satisfy a certain gauge fixing condition, such as the left or right gauge fixing condition B, represented as

These gauge fixing conditions also express that we have restricted to tangent vectors $|\Phi(B;A)\rangle \perp |\Psi(A)\rangle$ since

$$\langle\Psi(\overline{A})|\Phi(B)\rangle \sim$$

hence they can not be imposed for tangent vectors for which $\langle\Psi(\overline{A})|\Phi(B;A)\rangle \neq 0$. However, for the remainder, we are only interested in orthogonal tangent vectors $|\Phi(B;A)\rangle \in \mathbb{T}^\perp$. A linear parametrization $B = \mathcal{B}(X)$ that automatically satisfies the left gauge fixing can be obtained by finding an orthonormal basis V_L of solutions for the equation below

Here, the thick bond represents an index of dimension $D(d-1)$, which is the size of the solution space of the first system of equations. The representation $\mathcal{B}(X)$ was constructed such that, with respect to the new variables X, the effective normalization becomes the unit matrix, i.e.,

$$\langle\Phi(\overline{\mathcal{B}(X)})|\Phi(\mathcal{B}(Y))\rangle = $$

The time-dependent variational principle, described in Sec. III, gives a prescription for the direction $\dot{A}$ in which A has to be varied in order to approximate time evolution according to the Schrödinger equation, either real- or imaginary-time evolution

depending on the presence of a factor i. The optimal direction B is obtained by projection $(\hat{H} - \langle \hat{H} \rangle) |\Psi(A)\rangle$ onto the tangent space of the manifold of uniform MPS, where we always subtract the expectation value of the Hamiltonian with respect to the current state, in order to avoid infinite phase or normalization factors. The tangent-space projection $|\Phi(\mathcal{B}(F))\rangle = \hat{P}_{\mathbb{T}^\perp} \hat{H} |\Psi(A)\rangle$ [see Eq. (41)] is then captured by the equation $\langle \Phi(\overline{\mathcal{B}(X)}) | \Phi(\mathcal{B}(F)) \rangle = \langle \Phi(\overline{\mathcal{B}(X)}) | \hat{H} | \Psi(A) \rangle$, $\forall~X$, which is represented as
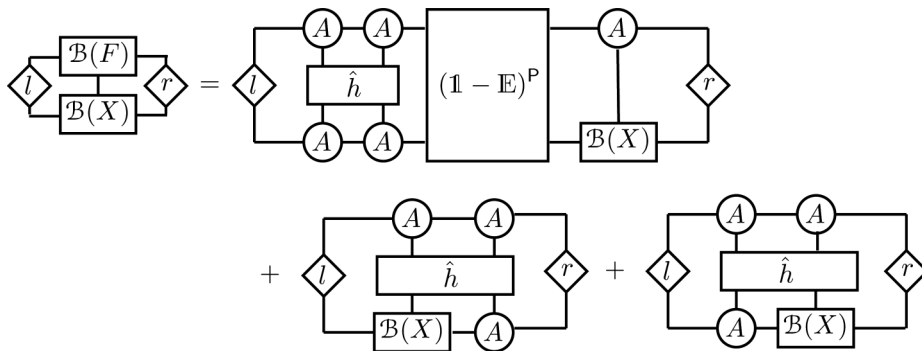
By inserting the definition for $\mathcal{B}(X)$ and $\mathcal{B}(F)$ and the derivative with respect to $\overline{X}$, which effectively opens the indices connected to X and eliminates X, an explicit expression for F is obtained, from which the corresponding update direction $\dot{A} = -(i)\mathcal{B}(F)$ for the tensor A is obtained. This expression contains the pseudoinverse of $(\mathbb{1} - \mathbb{E})$, which was introduced in Sec. II B. For finding the ground state, one can then integrate this set of coupled nonlinear differential equations using a simple first-order Euler scheme, or using one of the slightly more advanced schemes discussed in Sec. III E. Note that the TDVP intrinsically stays within the manifold of uniform MPS with fixed bond dimension. If some application requires a dynamical increase of the bond dimension, the technique presented in Sec. VII B can be used.

Having found a uniform MPS approximation $|\Psi(A)\rangle$ of the ground state of $\hat{H}$, which we now assume to be such that $\langle \Psi(\overline{A}) | \hat{H} | \Psi(A) \rangle = 0$, we now go on to discuss how to study excitations. Hereto, we introduce the larger tangent space with site-dependent variations $B(n)$, which can be decomposed into momentum sectors by selected $B(n) = B \exp(ipn)$ for $p \in [-\pi, \pi)$. In fact, we can even use the slightly more general *Ansatz* introduced in Sec. VII A, where the perturbation B acts on several sites. These result in the states

Observing that there is again a parametrization redundancy $|\Phi_{p,K}(B)\rangle = |\Phi_{p,K}(\tilde{B})\rangle$ given by

we can now restrict to parametrizations B satisfying

for which we also construct a representation $\mathcal{B}(X)$ using the same basis V_L defined above. The new set of parameters X now has a virtual bond of dimension $D(d-1)$, $K-1$ physical indices of dimension d and another virtual bond of dimension D. This representation is chosen such that the effective normalization matrix is still the unit matrix. A variational approximation for the excitation energies is then obtained by diagonalizing the corresponding effective Hamiltonian, whose matrix elements can be read from Eq. (70) for the case of $K = 1$. For $K > 1$, this expression becomes even slightly more complex, but it can still be calculated using the techniques described in this paper. Topologically nontrivial excitations such as kinks and domain walls can be described by using a different tensor $\tilde{A}$ on one of both sides of the perturbation B, where $|\Psi(\tilde{A})\rangle$ corresponds to a linearly independent but equally good ground state (approximation). The computation of the spectral weight of the variational excitations $|\Phi_{p,K}(B)\rangle$ with respect to some operator $\hat{o}$ closely resembles the computation of the TDVP projection and can be computed analogously.

IX. CONCLUSIONS AND OUTLOOK

This paper discusses in great detail some recent algorithms to study the dynamics of quantum lattice systems that rely on the concept of the tangent space to the manifold of

uMPS. The need for understanding the uMPS tangent space arises naturally when applying the TDVP to the manifold of uMPS, in order to capture time-evolving quantum states within this manifold. However, it then turns out that the tangent space is also very convenient in the formulation of "post-MPS" methods for studying elementary excitations of the quantum lattice Hamiltonian and to compute spectral functions. This situation is reminiscent to mean field theory for fermions (Hartree-Fock) and bosons (Gross-Pitaevskii), where similar constructions have been well developed and used extensively.

However, unlike mean field theory, (u)MPS have a refinement parameter D that can be used to improve the ground-state approximation if necessary, and we have in this paper refrained from studying whether "post-MPS" also imply corrections to the ground state. Rather, we assume the uMPS description of the ground state to be "quasi"-exact, and we want to obtain a better understanding of the elementary excitations in the system and the interactions that exist between them. We can easily generalize the tangent-space *Ansatz* in order to obtain a more accurate description of elementary excitations, and have given an outlook on how we can start building a new Fock space on top of the MPS vacuum, in which an effective Hamiltonian can be created that describes the quantum dynamics of the elementary excitations in the system. The formalism in this paper should also be applicable to the case of continuous MPS, as well as to other tensor networks for higher-dimensional quantum systems. This program should enable us in the long term to construct an effective description of the low-energy behavior of strongly interacting quantum systems, which can be used to describe the complete thermodynamics in the quantum regime.

Note added. Recently, we learned about Ref. 90 where related concepts were introduced for MPS on finite lattices and were compared in detail to similar constructions for Hartree-Fock theory. The contents of this paper and its relation to our paper were briefly discussed in Sec. VI.

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

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

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

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

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

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

[7] J. I. Cirac and F. Verstraete, J. Phys. A: Math. Theor. **42**, 504004 (2009).

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

[9] The topic of edge excitations is crucial for systems with topological order.

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

[11] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987).

[12] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Commun. Math. Phys. **115**, 477 (1988).

[13] G. Vidal, Phys. Rev. Lett. **98**, 070201 (2007).

[14] I. P. McCulloch, arXiv:0804.2509.

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

[16] B. Pirvu, J. Haegeman, and F. Verstraete, Phys. Rev. B **85**, 035130 (2012).

[17] J. Haegeman, B. Pirvu, D. J. Weir, J. I. Cirac, T. J. Osborne, H. Verschelde, and F. Verstraete, Phys. Rev. B **85**, 100408 (2012).

[18] J. M. Kinder, C. C. Ralph, and G. Kin-Lic Chan, arXiv:1103.2155.

[19] T. Koffel, M. Lewenstein, and L. Tagliacozzo, Phys. Rev. Lett. **109**, 267203 (2012).

[20] J. Haegeman, M. Mariën, T. J. Osborne, and F. Verstraete, arXiv:1210.7710.

[21] S. Holtz, T. Rohwedder, and R. Schneider, Numer. Math. **120**, 701 (2011).

[22] A. Uschmajew and B. Vandereycken, MATHICSE Technical Report, 2012 (unpublished).

[23] M. Fannes, B. Nachtergaele, and R. F. Werner, J. Phys. A: Math. Gen. **24**, L185 (1991).

[24] P. W. Anderson, Phys. Rev. Lett. **18**, 1049 (1967).

[25] U. Schollwoeck and S. R. White, in *Effective Models for Low-Dimensional Strongly Correlated Systems*, edited by G. G. Batrouni and D. Poilblanc (AIP, Melville, New York, 2006), p. 155.

[26] E. Jeckelmann, Prog. Theor. Phys. Suppl. **176**, 143 (2008).

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

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

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

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

[31] M. Suzuki, Phys. Lett. A **146**, 319 (1990).

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

[33] R. Orús and G. Vidal, Phys. Rev. B **78**, 155117 (2008).

[34] P. A. M. Dirac, Math. Proc. Cambridge Philos. Soc. **26**, 376 (1930).

[35] J. Frenkel, *Wave Mechanics, Advanced General Theory* (Clarendon, Oxford, 1934).

[36] P. Kramer and M. Saraceno, *Geometry of the Time-Dependent Variational Principle in Quantum Mechanics*, Lecture Notes in Physics, Vol. 140 (Springer, Berlin, 1981).

[37] P. Hauke and L. Tagliacozzo, arXiv:1304.7725.

[38] A. Milsted, J. Haegeman, T. J. Osborne, and F. Verstraete, arXiv:1207.0691.

[39] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure Preserving Algorithms for Ordinary Differential Equations*, Springer Series in Computational Mathematics, Vol. 31 (Springer, Berlin, 2004).

[40]J. Haegeman *et al.* (unpublished).

[41]P. A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds* (Princeton University Press, Princeton, NJ, 2009).

[42]K. A. Hallberg, Phys. Rev. B **52**, R9827 (1995).

[43]T. D. Kühner and S. R. White, Phys. Rev. B **60**, 335 (1999).

[44]E. Jeckelmann, Phys. Rev. B **66**, 045114 (2002).

[45]R. G. Pereira, S. R. White, and I. Affleck, Phys. Rev. Lett. **100**, 027206 (2008).

[46]S. R. White and I. Affleck, Phys. Rev. B **77**, 134437 (2008).

[47]S. Bravyi, M. B. Hastings, and F. Verstraete, Phys. Rev. Lett. **97**, 050401 (2006).

[48]J. Eisert and T. J. Osborne, Phys. Rev. Lett. **97**, 150404 (2006).

[49]S. Das and S. Shankaranarayanan, Phys. Rev. D **73**, 121701 (2006).

[50]S. Das and S. Shankaranarayanan, Classical Quantum Gravity **24**, 5299 (2007).

[51]S. Das, S. Shankaranarayanan, and S. Sur, Phys. Rev. D **77**, 064013 (2008).

[52]V. Alba, M. Fagotti, and P. Calabrese, J. Stat. Mech. (2009) P10020.

[53]L. Masanes, Phys. Rev. A **80**, 052104 (2009).

[54]D. Porras, F. Verstraete, and J. I. Cirac, Phys. Rev. B **73**, 014410 (2006).

[55]A. Bijl, J. de Boer, and A. Michels, Physica (Amsterdam) **8**, 655 (1941).

[56]R. P. Feynman, Phys. Rev. **94**, 262 (1954).

[57]R. P. Feynman and M. Cohen, Phys. Rev. **102**, 1189 (1956).

[58]D. P. Arovas, A. Auerbach, and F. D. M. Haldane, Phys. Rev. Lett. **60**, 531 (1988).

[59]E. Bartel, A. Schadschneider, and J. Zittartz, Eur. Phys. J. B **31**, 209 (2003).

[60]S. G. Chung and L. Wang, Phys. Lett. A **373**, 2277 (2009).

[61]S. Mandelstam, Phys. Rev. D **11**, 3026 (1975).

[62]R. Rajaraman, *Solitons and Instantons: An Introduction to Solitons and Instantons in Quantum Field Theory* (North-Holland, Amsterdam, 1982).

[63]B. Nachtergaele, Commun. Math. Phys. **175**, 565 (1996).

[64]J. J. Dorando, J. Hachmann, and G. K.-L. Chan, J. Chem. Phys. **130**, 4111 (2009).

[65]H. N. Phien, G. Vidal, and I. P. McCulloch, Phys. Rev. B **86**, 245107 (2012).

[66]V. Zauner, M. Ganahl, H. G. Evertz, and T. Nishino, arXiv:1207.0862.

[67]H. N. Phien, G. Vidal, and I. P. McCulloch, arXiv:1207.0678.

[68]P. E. Dargel, A. Honecker, R. Peters, R. M. Noack, and T. Pruschke, Phys. Rev. B **83**, 161104 (2011).

[69]A. Holzner, A. Weichselbaum, I. P. McCulloch, U. Schollwöck, and J. von Delft, Phys. Rev. B **83**, 195115 (2011).

[70]P. E. Dargel, A. Wöllert, A. Honecker, I. P. McCulloch, U. Schollwöck, and T. Pruschke, Phys. Rev. B **85**, 205119 (2012).

[71]N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966).

[72]F. D. M. Haldane, Phys. Lett. A **93**, 464 (1983).

[73]F. D. M. Haldane, Phys. Rev. Lett. **50**, 1153 (1983).

[74]F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, Phys. Rev. B **81**, 064439 (2010).

[75]T. Grover and T. Senthil, Phys. Rev. Lett. **98**, 247202 (2007).

[76]A. Läuchli, G. Schmid, and S. Trebst, Phys. Rev. B **74**, 144426 (2006).

[77]S. Rachel and M. Greiter, Phys. Rev. B **78**, 134415 (2008).

[78]D. Draxler, J. Haegeman, T. J. Osborne, V. Stojevic, L. Vanderstraeten, and F. Verstraete, arXiv:1212.1114.

[79]D. R. Hartree, Math. Proc. Cambridge Philos. Soc. **24**, 89 (1928).

[80]D. R. Hartree, Math. Proc. Cambridge Philos. Soc. **24**, 111 (1928).

[81]J. C. Slater, Phys. Rev. **35**, 210 (1930).

[82]V. Fock, Z. Phys. A **61**, 126 (1930).

[83]V. Fock, Z. Phys. A **75**, 622 (1932).

[84]D. J. Rowe, A. Ryman, and G. Rosensteel, Phys. Rev. A **22**, 2362 (1980).

[85]P. Ring and P. Schuck, *The Nuclear Many-Body Problem* (Springer, Berlin, 2004).

[86]W. H. Dickhoff and D. Van Neck, *Many-Body Theory Exposed! Propagator Description of Quantum Mechanics in Many-Body Systems* (World Scientific, Singapore, 2005).

[87]D. J. Rowe, Rev. Mod. Phys. **40**, 153 (1968).

[88]I. Tamm, J. Phys. (Moscow) **9**, 449 (1945).

[89]S. M. Dancoff, Phys. Rev. **78**, 382 (1950).

[90]S. Wouters, N. Nakatani, D. Van Neck, and G. K.-L. Chan, Phys. Rev. B **88**, 075122 (2013).

[91]E. P. Gross, Il Nuovo Cimento **20**, 454 (1961).

[92]L. P. Pitaevskii, Zh. Eksp. Teor. Fiz. **40**, 646 (1961) [Sov. J. Exper. Theor. Phys. **13**, 451 (1961)].

[93]P. G. de Gennes, *Superconductivity of Metals and Alloys* (Benjamin, New York, 1966).

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

[95]T. J. Osborne, J. Eisert, and F. Verstraete, Phys. Rev. Lett. **105**, 260401 (2010).

[96]J. Haegeman, J. I. Cirac, T. J. Osborne, and F. Verstraete, arXiv:1211.3935.

[97]F. Verstraete *et al.* (unpublished).

[98]J. Haegeman, S. Michalakis, B. Nachtergaele, T. J. Osborne, N. Schuch, and F. Verstraete, arXiv:1305.2176 [Phys. Rev. Lett. (to be published)].

[99]L. Vanderstraeten *et al.* (unpublished).

[100]D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Quantum Inf. Comput. **7**, 401 (2007).