Laurens Vanderstraeten
University of Gent
email: laurens.vanderstraeten@ugent.be

Tangent space methods for matrix product states

In these lecture notes we give a technical overview of tangent-space methods for matrix product states in the thermodynamic limit. We introduce the manifold of uniform matrix product states, show how to compute different types of observables, and discuss the concept of a tangent space. We explain how to optimize ground-state approximations, implement real-time evolution and describe elementary excitations. Also, we provide 'inverse-free' versions of all these algorithms, and conclude with matrix product state approximations for fixed points of transfer matrices.

Contents

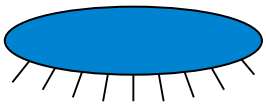
1 Introduction

In the last twenty-five years a set of methods was developed for simulating strongly-correlated quantum systems grouped under the name of 'tensor network methods'. The unifying feature of these methods is their ability to capture the specific entanglement structure of low-energy states, and as such to provide an extremely accurate variational parametrization for describing and simulating quantum many-body physics. Indeed, in the last ten years it has been realized that quantum correlations in ground states of local Hamiltonians are distributed in a very special way [1], and the idea is to design a network of tensors that mimics this special distribution. In this way, tensor network states parametrize the 'tiny corner of Hilbert space', where all the relevant physics is going on.

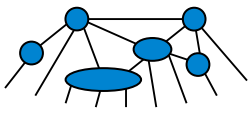
In order to see how a network of tensors can describe a many-body state, take a general state of N spins on an arbitrary lattice

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N} c_{i_1,\ldots,i_N} |i_1,\ldots,i_N\rangle \,.$$

The coefficients $c_{i_1,\ldots,i_N}$ are complex numbers for every input of the indices; alternatively, they can be interpreted as defining a tensor with N indices. If we take the basis states $|i_1,\ldots,i_N\rangle$ as a given, we can say that the tensor describes the state. We can represent the state in a diagrammatic way,

$|\Psi\rangle =$  ,

where a geometric figure always represents a tensor with the legs corresponding to the tensor's indices. Now, a tensor network should be pictured as a decomposition of this one N-leg tensor as a contraction of a number of smaller tensors:

$|\Psi\rangle_{\mathrm{TNS}} =$  .

In this expression, we have introduced the diagrammatic notation for a tensor network contraction: (i) whenever two tensors are connected, the connected indices are identified and summed over, and (ii) unconnected legs remain as indices of the resulting new tensor (if all legs are connected the diagram corresponds to a scalar).

The most important feature of tensor network states is that the dimensions of the virtual indices in these contractions will not be very large, so that the number of parameters that describe these states is small. This implies that such a low-rank tensor-network decomposition of a quantum state is generally not possible, but, of course, it proves to be the case that tensor network states exactly target the low-energy subspace of physical systems. This insight has led to the development of a vast body of theoretical and numerical tools for capturing the physical properties of strongly-correlated quantum matter in the language of tensor network states. [2–4]

In the context of one-dimensional quantum spin systems, the class of matrix product states (MPS) provides an extremely versatile parametrization for ground states of gapped local Hamiltonians [5, 6]. In fact, MPS constitute the variational class underlying the density-matrix renormalization group [7], and the MPS formulation has led to the extension of DMRG towards the simulation of dynamical, finite-temperature and out-of-equilibrium properties of spin chains [8]. Moreover, this realization has led to efficient MPS representations and algorithms that work directly in the thermodynamic limit [9]. Not only are all finite-size effects ruled out, but, as we will see, this has given rise to a very elegant formalism for describing the low-energy dynamics on top of a uniform MPS background.

2 Matrix product states in the thermodynamic limit

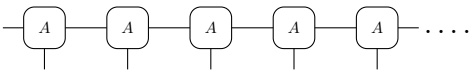
In this first section, we will introduce the class of translation-invariant matrix product states in the thermodynamic limit. As we will see, these *uniform matrix product states* have some very nice properties that allow us to work with them in a computationally efficient way. In the next sections, we will show that, although most state-of-the-art MPS-based methods rather work on finite systems, working directly in the thermodynamic limit has a number of conceptual and numerical advantages.

2.1 Uniform matrix product states, gauge transformations and the canonical form

A translation-invariant matrix product state in the thermodynamic limit is introduced as

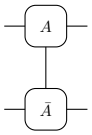
$$|\Psi(A)\rangle = \sum_{\{s\}} \boldsymbol{v_L^\dagger} \left[\prod_{m\in\mathbb{Z}} A^{s_m} \right] \boldsymbol{v_R} \, |\{s\}\rangle \,, \tag{1}$$

or represented diagrammatically as

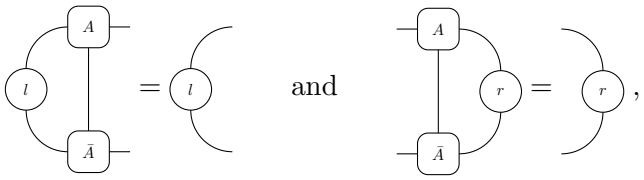
$|\Psi(A)\rangle =$ 

This definition is straightforward: we just repeat the same tensor A on every site in the lattice, giving rise to a state that is translation invariant by construction.[1] In (1) we have also introduced two boundary vectors $\boldsymbol{v_L^\dagger}$ and $\boldsymbol{v_R}$, but, as we work on an infinite system, the boundary conditions will never have any physical meaning. Indeed, we will show that the physical properties (expectation values) of the state $|\Psi(A)\rangle$ only depend on the tensor A.

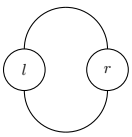
The central object in all our calculations is the transfer operator or *transfer matrix*, defined as

$$E = \sum_{s=1}^{d} A^s \otimes \bar{A}^s =$$ 

which is an operator acting on the space of $D \times D$ matrices. Furthermore, from its definition it follows that the transfer matrix is a completely positive map. In the generic case the transfer matrix has the property that the leading eigenvalue is a positive number η, and can be scaled to 1 by rescaling the uMPS tensor as $A \to A/\sqrt{\eta}$. The corresponding left and right fixed points l and r, i.e. the leading eigenvectors of the eigenvalue equation

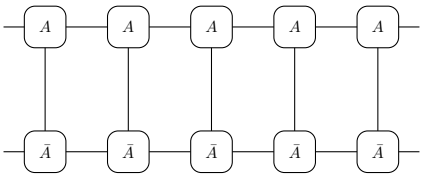


are positive matrices; they are normalized such that $\mathrm{Tr}(lr) = 1$, or, diagrammatically,



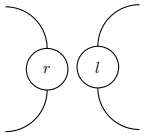
[1] We could introduce states that are translation invariant over multiple sites by working with a repeated unit cell of different matrices $A_1, A_2, \ldots$, and all methods that we will discuss can be extended to the case of larger unit cells.

With these properties in place, the norm of an MPS can be computed as

$\langle \Psi(A)|\Psi(A)\rangle = \ldots$  $\ldots$

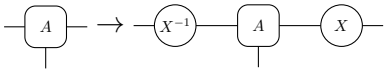
$$= \left(\boldsymbol{v_L} \boldsymbol{v_L}^\dagger\right) \left(\prod_{m\in\mathbb{Z}} E\right) \left(\boldsymbol{v_R} \boldsymbol{v_R}^\dagger\right).$$

The infinite product reduces to a projector on the fixed points,

$\lim_{N\to\infty} E^N =$ 

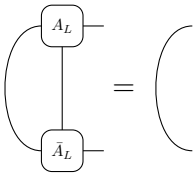
so that the norm reduces to the overlap between the boundary vectors and the fixed points. We will now choose the boundary vectors such that these overlaps equal unity – there is no effect of the boundary vectors on the bulk properties of the MPS anyway – so that the MPS is properly normalized as $\langle \Psi(A)|\Psi(A)\rangle = 1$.

Although the state is uniquely defined by the tensor A, the converse is not true, as different tensors can give rise to the same physical state. This can be easily seen by noting that the *gauge transform*

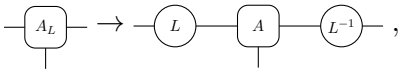


leaves the state (1) invariant. In fact, it can be shown [10, 11] that this is the only set of gauge transformations, so that imposing a unique choice for the invertible matrix X fixes all gauge freedom in the MPS representation.[2]

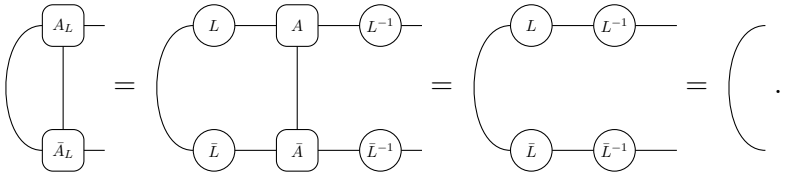
It proves to be the case that fixing the gauge degrees of freedom is crucial for implementing the algorithms that follow in the next sections. First, we can always find a representation of $|\Psi(A)\rangle$ for which the MPS tensor obeys the following condition



Given an arbitrary tensor A, we can construct the tensor A_L by finding the matrix L which transforms A into A_L as

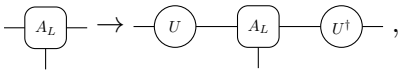


where the matrix L is found by decomposing the fixed point l of A as $l = L^\dagger L$, because indeed:

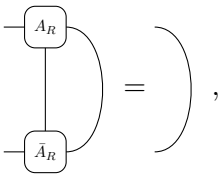


[2]Strictly speaking, this is only true for so-called injective MPS. In generic simulations, this injectivity property will always hold. For more details, see Ref. [11].

The representation of an MPS in terms of a tensor A_L is called the left-canonical form. This gauge condition still leaves room for unitary gauge transformations,



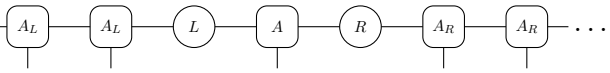
which can be used to bring the right fixed point r in diagonal form. Similarly, a right canonical form A_R can be found such that



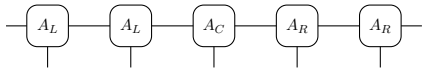
and where the left fixed point l is diagonal.

These left and right canonical forms now allow us to define a mixed canonical form. The idea is that we choose one site, the 'center site', bring all tensors to the left in the left-canonical form, all the tensors to the right in the right-canonical form, and define a new tensor A_C on the center site. Diagrammatically, we obtain the following form

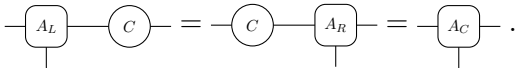
$|\Psi(A)\rangle = \ldots$



$= \ldots$

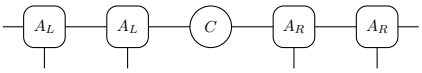


This mixed canonical form has an intuitive interpretation. First of all, we introduce a new tensor $C = LR$ which implements the gauge transform that brings you from the left-canonical form to the right-canonical one, and which defines the center-site tensor A_C:

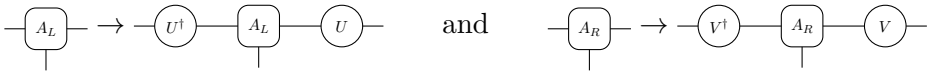


This allows us to rewrite the MPS with only the C tensor on a virtual leg, linking the left- and right canonical tensors,

$|\Psi(A)\rangle = \ldots$



In a next step, the tensor C is brought into diagonal form by performing a singular-value decomposition $C = USV^\dagger$, and taking up U and $V^\dagger$ in a new definition of A_L and A_R – remember that we still had the freedom of unitary gauge transformations on the left- and right-canonical form:



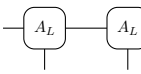
(2)

The above form of the MPS, with a diagonal C, now allows to straightforwardly write down a Schmidt decomposition of the state across an arbitrary bond in the chain:

$$|\Psi(A)\rangle = \sum_{i=1}^{D} C_{ii} \, |\Psi_L^i(A_L)\rangle \otimes |\Psi_R^i(A_R)\rangle \,,$$

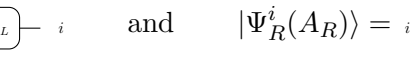
where

$|\Psi_L^i(A_L)\rangle = \ldots$



and

$|\Psi_R^i(A_R)\rangle =$

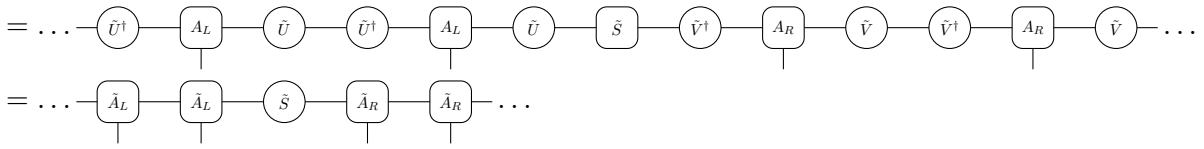


This implies that the singular values of the matrix C in the mixed canonical form are exactly the Schmidt numbers of any bipartition of the MPS.

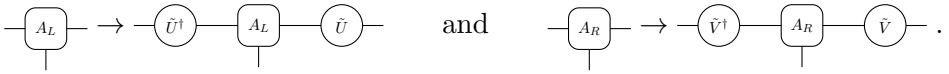
Truncating a uniform MPS

This mixed canonical form also allows us to truncate an MPS efficiently [12]. Indeed, the sum in the above Schmidt decomposition can be truncated, giving rise to a new MPS that has a reduced bond dimension for that bond. This truncation is optimal in the sense that the norm between the original and the truncated MPS is maximized, but the resulting MPS is no longer translation invariant – it has a lower bond dimension on one leg. We can, however, introduce a translation invariant MPS with a lower bond dimension by transforming *every* tensor A_L or A_R as in Eq. 2, but where we have truncated the number of columns in U and V, giving rise to the isometries $\tilde{U}$ and $\tilde{V}$. The truncated MPS in the mixed canonical form is then given by

$|\Psi(A)\rangle_{\text{trunc}}$



with $\tilde{S}$ the truncated singular values of C, and



This procedure is not guaranteed to find the MPS with a lower bond dimension that is globally optimal, in the sense that it minimizes the error on the global (thermodynamic limit) state. A variational optimization of the cost function

$$\left\| |\Psi(A)\rangle - |\Psi(\tilde{A})\rangle \right\|^2$$

would find the optimal truncated MPS tensor A, but the approximate algorithm has, of course, the advantage of being numerically efficient.

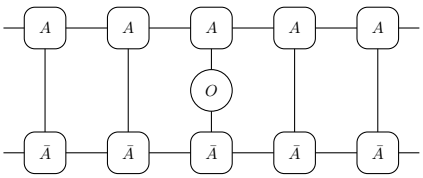
2.2 Computing expectation values

Suppose we want to compute the expectation value of an extensive operator

$$O = \frac{1}{|\mathbb{Z}|} \sum_{n \in \mathbb{Z}} O_n,$$

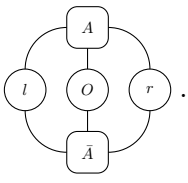
where the extra factor $|\mathbb{Z}|^{-1}$ represents the number of sites, and is introduced to obtain a finite finite value in the thermodynamic limit – in fact, we are evaluating the density corresponding to operator O. Because of translation invariance, we only have to evaluate one term where O acts on an arbitrary site. The expectation value is then – assuming the MPS is already properly normalized

$\langle \Psi(A)| O |\Psi(A)\rangle = \ldots$



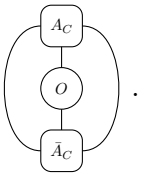
$\ldots.$

We can now use the left and right fixed points of the transfer matrix to contract everything to the left and to the right of the operator, to arrive at the contraction



.

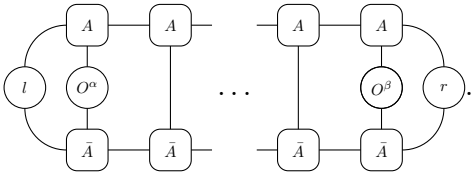
An even easier contraction is obtained by going to the mixed canonical form, and locating the center site where the operator is acting. Indeed, then everything to the left and right is contracted to the identity and we obtain



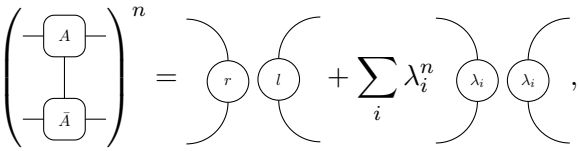
Correlation functions are computed similarly. Let us look at

$$c^{\alpha\beta}(m,n) = \langle\Psi(A)|\,(O_m^\beta)^\dagger O_n^\alpha\,|\Psi(A)\rangle\,,$$

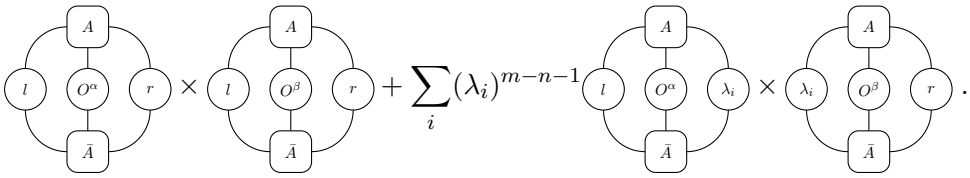
where m and n are abritrary locations in the chain, and, because of translation invariance, the correlation function only depends on the difference $m-n$. Again, we contract everything to the left and right of the operators by inserting the fixed points l and r, so that

$c^{\alpha\beta}(m,n) =$ 

From this expression, we learn that it is the transfer matrix that determines the correlations in the ground state. Indeed, if we apply the eigendecomposition,



we can see that the correlation function reduces to

$c^{\alpha\beta}(m,n) =$ 

The first part is just the product of the expectation values of O^α and O^β, called the disconnected part of the correlation function, and the rest is an exponentially decaying part. This expression implies that connected correlation functions of an MPS *always* decay exponentially, which is one of the reasons why MPS are not well suited for capturing critical states. The largest λ, i.e. the second largest eigenvalue of the transfer matrix, determines the correlation length ξ and the pitch vector of the correlations Q as

$$\xi = -\frac{1}{\log|\lambda_{\max}|} \qquad \text{and} \quad Q = \arg(\lambda_{\max}).$$

2.3 The static structure factor

In experimental situations, one typically has access to the Fourier transform of the correlation functions, called the (static) structure factor. Since we are working in the thermodynamic limit, we can easily compute this quantity with a perfect resolution in the momentum range $[0, 2\pi)$.

The structure factor corresponding to two operatos O^α and O^β is defined as

$$s^{\alpha\beta}(q) = \frac{1}{|\mathbb{Z}|} \sum_{m,n\in\mathbb{Z}} \mathrm{e}^{iq(m-n)} \langle\Psi(A)| (O_n^\beta)^\dagger O_m^\alpha |\Psi(A)\rangle_c \,, \tag{3}$$

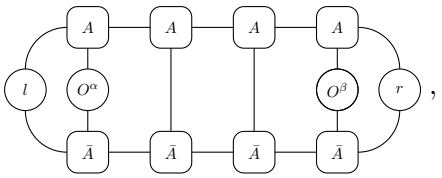
where $\langle\ldots\rangle_c$ denotes that we only take the connected part of the correlation function. This can be implemented by redefining the operators such that their expectation value is zero,

$$O_n^{\alpha,\beta} \to O_n^{\alpha,\beta} - \langle\Psi(A)| O_n^{\alpha,\beta} |\Psi(A)\rangle \,.$$

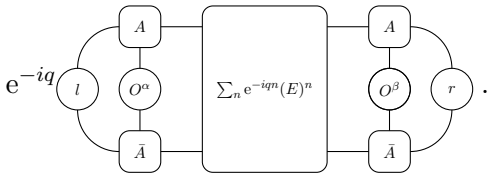
This quantity can be computed directly in momentum space by a number of steps. First, due to translation invariance, every term in (3) only depends on the difference $(m-n)$, so that we can eliminate one of the two terms,

$$s^{\alpha\beta}(q) = \sum_{n\in\mathbb{Z}} \mathrm{e}^{-iqn} \langle\Psi(A)| (O_n^\beta)^\dagger O_0^\alpha |\Psi(A)\rangle_c \,.$$

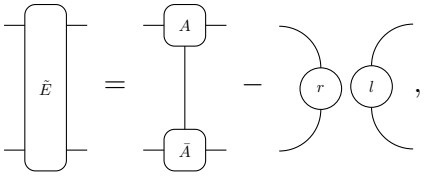
Every term in the sum has the form of a connected correlation function of the form



but we can resum all these diagrams in an efficient way. Indeed, all terms where the operator O^β is to the right of O^α can be rewritten as



If we now define a regularized transfer matrix $\tilde{E}$ by projecting out the component along its fixed points, i.e.

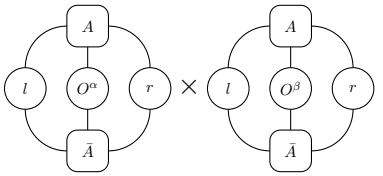


we can take the geometric series of the regularized part E as

$$\begin{aligned} \mathrm{e}^{-iq} \sum_{n=0}^{\infty} \mathrm{e}^{-iqn} \tilde{E}^n &= \mathrm{e}^{-ip} \left(1 - \mathrm{e}^{-ip}\tilde{E}\right)^{-1} \\ &= \mathrm{e}^{-ip} \left(1 - \mathrm{e}^{-ip}E\right)^{P} . \end{aligned}$$

In the last line we have defined the notation $(\ldots)^P$ for 'pseudo-inverse' of an operator, implying that we take the inverse of the operator on its non-zero eigenvalues. Now the part corresponding

to the fixed point projector could be potentially divergent in the thermodynamic limit, but, in fact, vanishes because

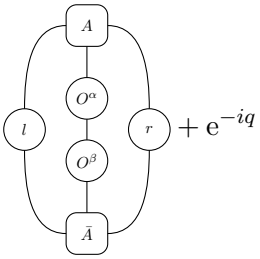


$$= \langle \Psi(A)| O^\alpha |\Psi(A)\rangle \langle \Psi(A)| O^\beta |\Psi(A)\rangle ,$$

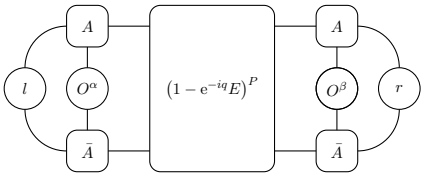
which we have set to zero in the definition of the operators O^α and O^β.

The part where O^β is to the left of O^α is treated similarly, and we also have the term where both are acting on the same site; we obtain the following final expression:

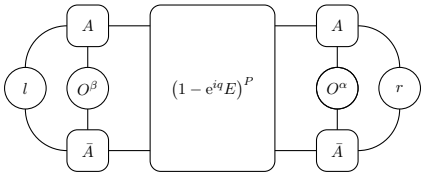
$s(q) =$



$+ \mathrm{e}^{-iq}$

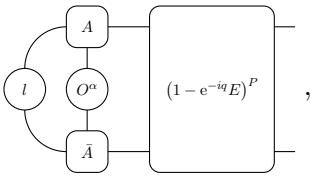


$+ \mathrm{e}^{iq}$



.

The question remains how to compute this quantity efficiently. What we definitely don't want to do, is computing the pseudo-inverse $(1 - \mathrm{e}^{-iq}E)^P$ explicitly – that would entail a computational complexity of $\mathcal{O}(D^6)$. Instead, we will compute e.g. the partial contraction



,

i.e. the action of the pseudo-inverse $(1 - \mathrm{e}^{-iq}E)^P$ on a given left-hand side, which is implemented as the solution of a linear equation of the form $(1 - \mathrm{e}^{-iq}\tilde{E}) \times x = y$. This linear equation can be solved using interative solvers based on e.g. the conjugate-gradient method, where only the action of $(1 - \mathrm{e}^{-iq}\tilde{E})$ on a given vector needs to implemented. This reduces the computational complexity to only $\mathcal{O}(D^3)$.

2.4 The tangent space of the MPS manifold

Let us now introduce the unifying concept of these lecture notes: the MPS tangent space. First, we interpret the set of uniform MPS with a given bond dimension as a manifold [13] within the full Hilbert space of the system we are investigating. The manifold is defined by the map between the set of $D \times D$ tensors A and physical states in Hilbert space $|\Psi(A)\rangle$. This map, and the associated manifold, is non-linear, as any sum of two MPS's with a given bond dimension D clearly does not remain within the manifold. Therefore, it makes sense to associate a tangent

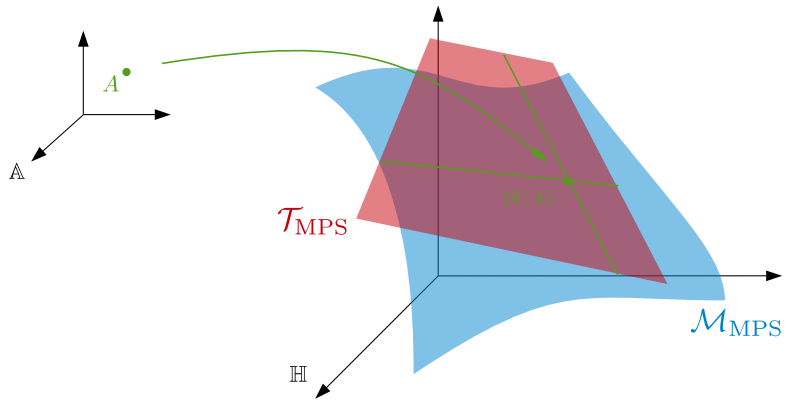
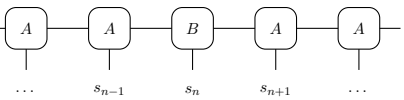


Figure 1: The tangent space of MPS.

space to every point $|\Psi(A)\rangle$ by differentiating with respect to the parameters in A; a general tangent vector is parametrized by a new tensor B as

$$
\begin{aligned}
|\Phi(B;A)\rangle &= \sum_i B_i \frac{\partial}{\partial A_i} |\Psi(A)\rangle \\
&= \sum_{\{s\}} \sum_n (\dots A^{s_{n-2}} A^{s_{n-1}} B^{s_n} A^{s_{n+1}} A^{s_{n+2}} \dots) \, |\dots s_{n-2}, s_{n-1}, s_n, s_{n+1}, s_{n+2}, \dots\rangle \\
&= \sum_n \cdots
\end{aligned}
\tag{4}
$$



where $\sum_i$ sums over all elements of the tensors A and B. The MPS manifold, with a tangent space associated to every point, is represented graphically in Fig. 1.

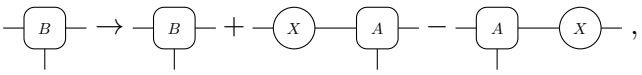
The tangent vectors inherit a set of gauge degrees freedom from the gauge freedom in the manifold. Indeed, the infinitesimal gauge transform $G = \mathrm{e}^{\epsilon X}$ is to first order

$$A^s \to \mathrm{e}^{-\epsilon X} A^s \mathrm{e}^{\epsilon X} = A^s + \epsilon \left(A^s X - X A^s\right) + \mathcal{O}(\epsilon^2).$$

This gauge transform can be brought to the level of states,

$$|\Psi(A)\rangle \to |\Psi(A)\rangle + \epsilon \, |\Phi(B;A)\rangle$$

with $B^s = A^s X - X A^s$. But, since this is a gauge transform in the MPS manifold, the tangent vector $|\Phi(B;A)\rangle$ should be zero. This implies that every transformation on B of the form

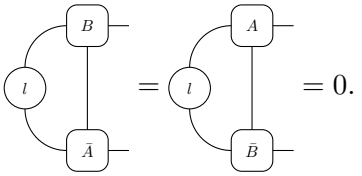


(5)

with X an arbitrary $D \times D$ matrix, leaves the tangent vector $|\Phi(B;A)\rangle$ invariant. Also, this gauge freedom can be easily checked by substituting this form in the state (4), and observing that all terms cancel, leaving the state invariant.

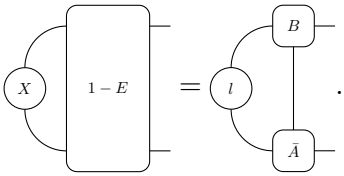
The gauge degrees of freedom can be eliminated by imposing a gauge-fixing condition. The easiest choice is the so-called left gauge-fixing condition (there is of course a right one, too), given

by

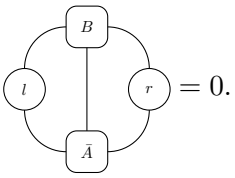


(6)

Any B can always be brought in this left gauge by a gauge transformation of the form (5). Indeed, X should obey the equation

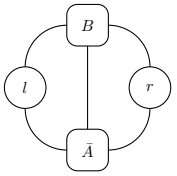


In Sec. 2.3 we have seen how to solve such linear equations by inverting a regularized transfer matrix $(1 - \tilde{E})$. For this procedure to work, we need that the component along the fixed point be zero, i.e. we need that



This is, however, a natural condition, because this is precisely saying that the tangent vector is orthogonal to the original MPS. Indeed, one can easily see that the overlap between an MPS and a tangent vector is given by

$$\langle \Psi(A) | \Phi(B;A) \rangle = 2\pi\delta(0)$$

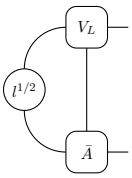


.

The factor corresponds to the system size, which diverges in the thermodynamic limit. We will denote this diverging factor as $2\pi\delta(0)$, inspired by the representation of the δ function as

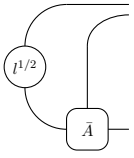
$$\sum_{n\in\mathbb{Z}} \mathrm{e}^{ipn} = 2\pi\delta(p).$$

We can now construct a very simple parametrization for the B tensor that automatically fixes all gauge degrees of freedom, and which has some nice advantages for all later calculations. First, we construct the tensor V_L such that



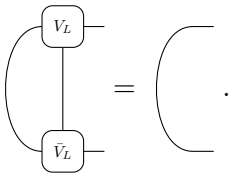
(7)

where the right index of V_L has dimension $n_L = D(d-1)$. Put differently, V_L corresponds to the $D(d-1)$-dimensional null space of the matrix

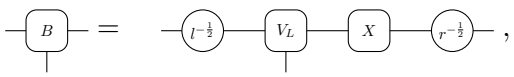


.

We orthonormalize V_L as

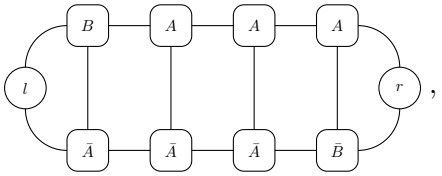


Secondly, the B tensor is expressed in terms of a new matrix X as



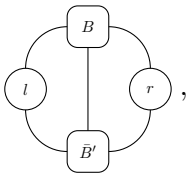
where X is a $(D(d-1)\times D)$-dimensional tensor. We will now use this prescription as an *effective parametrization* for the tangent space. This has the advantage that (i) all gauge degrees of freedom have been fixed, and (ii) we automatically obey the left gauge fixing condition in Eq. (6).

The third advantange concerns the overlap between two tangent vector. The overlap between $|\Phi(B)\rangle$ and $|\Phi(B')\rangle$ is computed similarly to the structure factor in Sec. 2.3: we have two infinite terms, but we can eliminate one sum because of the translation invariance of the ground state; this sum will again result in a factor $2\pi\delta(0)$. There still remains a sum over all relative positions between B and B'. Now the power of the left gauge fixing condition is revealed: all terms vanish, except the term where B and B' are on the same site. Indeed, all terms of the form



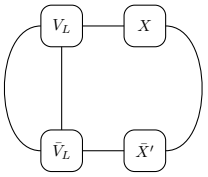
are automatically zero because of Eq. 6. Consequently, the norm reduces to

$$\langle\Phi(B')|\Phi(B)\rangle = 2\pi\delta(0)$$

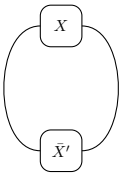


or in terms of the effective parameters in X and X',

$$\langle\Phi(B(X');A)|\Phi(B(X);A)\rangle = 2\pi\delta(0)$$



$$= 2\pi\delta(0)$$



$$= 2\pi\delta(0)\mathrm{Tr}\left((X')X^\dagger\right).$$

The fact the overlap of tangent vectors reduces to the Euclidean inner product on the effective parameters X and X' will prove to be a very nice property in all tangent-space algorithms.

3 Finding ground states

Now that we have introduced the manifold of matrix product states and the concept of the tangent space, we should explain how to find the point in the manifold that provides the best

approximation for the ground state of a given Hamiltonian H. In these notes, we will only consider nearest-neighbour interactions so that the Hamiltonian is of the form

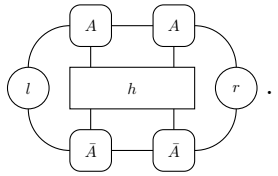
$$H = \sum_n h_{n,n+1},$$

where $h_{n,n+1}$ is a hermitian operator acting non-trivially on the sites n and $n+1$.

As in any variational approach, the variational principle will serve as a guide for finding ground-state approximations, viz. we want to minimize the expectation value of the energy,

$$\min_A \frac{\langle\Psi(A)|\, H\, |\Psi(A)\rangle}{\langle\Psi(A)|\Psi(A)\rangle}.$$

In the thermodynamic limit the energy diverges with system size, but, since we are working with translation-invariant states only, we should rather minimize the energy density. Also, we will restrict to properly normalized states. Diagrammatically, the minimization problem is recast as

$\min_A$ 

.

Typically, this minimization problem is not treated directly, but recourse is taken to imaginary-time evolution using the time-evolving block decimation algorithm [9, 14], or to infinite DMRG methods [15]. In this section, we will rather treat this problem in a more straightforward way, in the sense that we will use numerical optimization strategies for minimizing the energy density directly. This approach has the advantage that it is, by construction, optimal in a *global* way, because we never take recourse to local updates of the tensors – we always use routines that are optimal for the MPS wave function directly in the thermodynamic limit. As a result, we have a convergence criterion on the energy density for the infinite system.

3.1 The gradient

Any optimization problem relies on an efficient evaluation of the gradient, so let us first compute this quantity. The objective function f that we want to minimize is a real function of the complex-valued A, or, equivalently, the independent variables A and $\bar{A}$. The gradient g is then obtained by differentiating $f(\bar{A}, A)$ with respect to $\bar{A}$,[3]

$$\begin{aligned} g &= 2 \times \frac{\partial f(\bar{A}, A)}{\partial \bar{A}} \\ &= 2 \times \frac{\partial_{\bar{A}} \langle\Psi(\bar{A})|\, h\, |\Psi(A)\rangle}{\langle\Psi(\bar{A})|\Psi(A)\rangle} - 2 \times \frac{\langle\Psi(\bar{A})|\, h\, |\Psi(A)\rangle}{\langle\Psi(\bar{A})|\Psi(A)\rangle^2} \partial_{\bar{A}} \langle\Psi(\bar{A})|\Psi(A)\rangle\,, \end{aligned}$$

where we have clearly indicated A and $\bar{A}$ as independent variables. In the implementation we will always make sure the MPS is properly normalized, such that the numerators drop out. By subtracting from every term in the Hamiltonian its expectation value, it can be redefined as

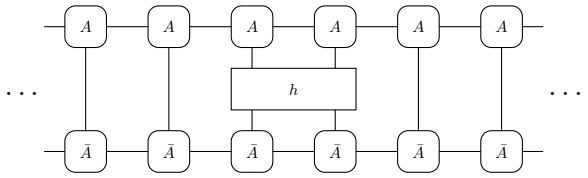
$$h \to h - \langle\Psi(\bar{A})|\, h\, |\Psi(A)\rangle\,,$$

[3]Numerical optimization schemes are typically developed for functions over real parameters. In order to translate these algorithms to complex parameters, we take $x = x_r + ix_i$, and take the gradient $g = g_r + ig_i$ with $g_r = \partial_{x_r} f$ and $g_i = \partial_{x_i} f$, which is equal to $g = 2\partial_{\bar{x}} f(x, \bar{x})$.

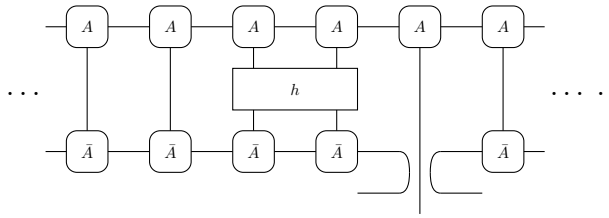
such that the gradient takes on the simple form

$$g = 2 \times \partial_{\bar{A}} \langle \Psi(\bar{A}) | \, h \, | \Psi(A) \rangle \, .$$

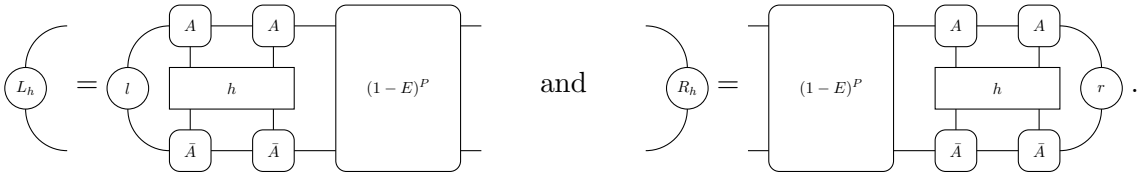
The gradient is obtained by differentating the expression



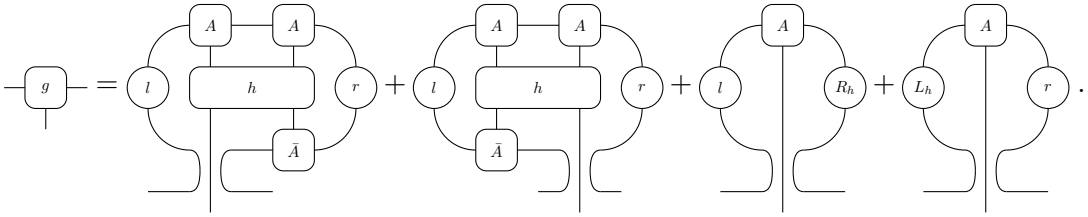
with respect to $\bar{A}$. It is given by a sum over all sites, where in every term we differentiate with one tensor $\bar{A}$ in the bra layer. Differentiating with respect to one $\bar{A}$ tensor amounts to leaving out that tensor, and interpreting the open legs as outgoing ones, i.e. each term looks like



For summing the infinite number of terms, we will use the same techniques as we did for evaluating the structure factor [Sec. 2.3]. Instead of varying the open spot in the diagram, we will vary the location of the Hamiltonian operator h. Then, we first treat all terms where h is either completely to the left or to the right of the open spot, by defining the partial contractions



As we have seen, taking these pseudo-inverses is equivalent to summing the infinite number of terms. These partial contractions are combined with the two contributions where h acts on the open spot, so that we have the final contribution of the gradient



This expression for the gradient is closely connected to a tangent vector, but in order to consistently interpret it as such, we need some additional steps. First, let us note the meaning of the gradient as defined above in terms of the first-order approximation of a change in the tensor $A + \epsilon B$

$$\langle \Psi(A) | \, h \, | \Psi(A) \rangle \to \langle \Psi(A) | \, h \, | \Psi(A) \rangle + \epsilon \boldsymbol{g}^\dagger \boldsymbol{B} + \mathcal{O}(\epsilon^2),$$

where vectorized versions of tensors are denoted in bold.

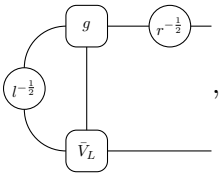
Now we realize that g is a vector in the space of tensors, not a state in the full Hilbert space. So how do we lift the notion of the gradient to the level of a state? Note that an infinitesimal change in the tensor $A + \epsilon B$ corresponds to a tangent vector

$$|\Psi(A + \epsilon B)\rangle \to |\Psi(A)\rangle + \epsilon\, |\Phi(B; A)\rangle + \mathcal{O}(\epsilon^2),$$

so that we would like to write the first-order change in the energy through an overlap between this tangent vector and a 'gradient vector' $|\Phi(G; A)\rangle$

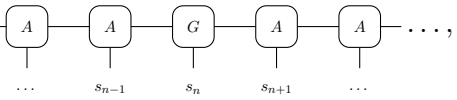
$$\langle\Psi(A)|\, h\, |\Psi(A)\rangle \to \langle\Psi(A)|\, h\, |\Psi(A)\rangle + \epsilon\, \langle\Phi(G; A)|\Phi(B; A)\rangle + \mathcal{O}(\epsilon^2).$$

We know, however, that building $|\Phi(G; A)\rangle$ using the tensor g is not correct, because the overlap $\langle\Phi(G; A)|\Phi(B; A)\rangle \neq \boldsymbol{G}^\dagger \boldsymbol{B}$. Instead, we will have to determine the *tangent-plane gradient* by its reduced parametrization, where, as usual

$X_G =$ 

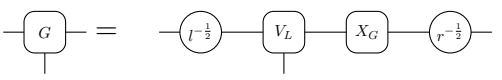
,

so that the tangent-space gradient is given by the usual expression for a tangent vector,

$|\Phi(G; A)\rangle = \sum_n \ldots$ 

$\ldots,$

with the tensor G given by



.

The difference between these two notions of a gradient can be elucidated by looking at the variational manifold from the perspective of differential geometry. The tangent-space gradient would then correspond to a covariant version of the gradient on a manifold with a non-trivial metric [13, 16].

3.2 Optimizing the tensors

Using these expressions for the different types of gradient, we can easily implement a gradient search method for minimizing the energy expectation value. The first obvious option is a steepest descent method, where in every iteration the tensor A is updated in the direction of the gradient:

$$A_{i+1} = A_i - \alpha g.$$

The size of α is determined by doing a line search: we find a value for which the value of the energy has decreased. In principle, we could try to find the optimal value of α, for which we can no longer decrease the energy by taking the direction $-g$ in parameter space. In practice, we will be satisfied with an approximate value of α, for which certain conditions [17] are fulfilled.

Other optimization schemes based on an evaluation of the gradient, such as conjugate gradient or quasi-Newton methods, are more efficient. Even more efficient would be an algorithm that requires an evaluation of the Hessian, which in principle we can also do with the techniques above. Crucially, this way of variationally optimizing an MPS has a clear convergence criterion: we say that we have reached a – possibly local – optimum if the norm of the gradient is sufficiently small.

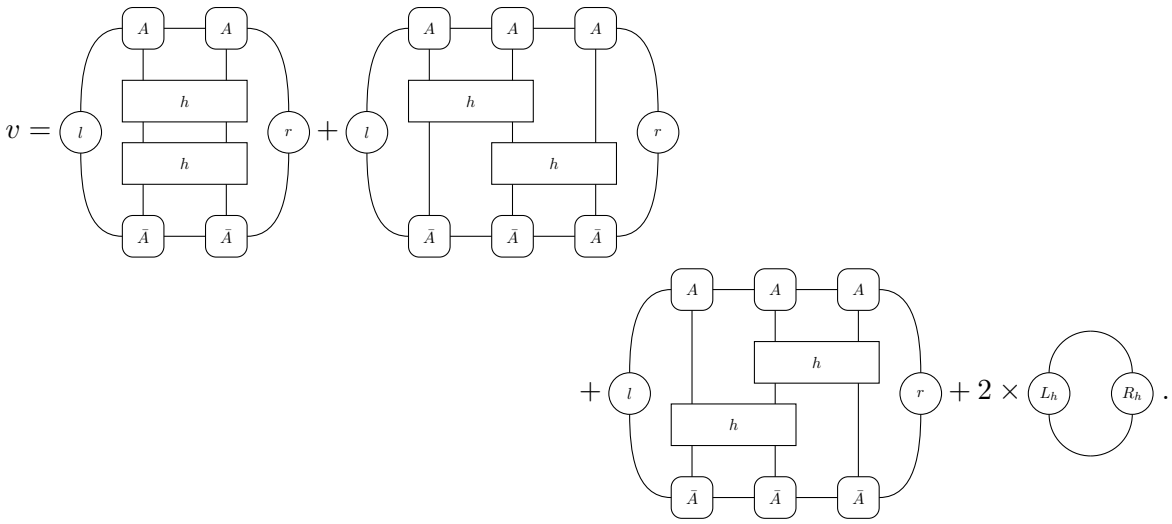
As another road to more efficient optimization schemes we could take the tangent-space gradient a bit more seriously. Indeed, now we compute the tangent vector corresponding to steepest descent in Hilbert space, but then update the A tensor by simply adding them in parameter space. Instead, we would like to do a line search along geodetic paths through the manifold, which would involve integrating the geodesic equation. It remains an open question, however, whether this could lead to more efficient optimization schemes.

3.3 The energy variance

In any variational approach, finding an optimal set of parameters does not guarantee that the state provides a good approximation to the true ground state of the Hamiltonian. We do have access, however, to an unbiased measure of how well the MPS approximates *any* eigenstate of the system, called the variance. It is defined by

$$v = \langle \Psi(A) | \, H^2 \, | \Psi(A) \rangle \, ,$$

where we have subtracted the ground-state energy density from the local nearest-neighbour term in the Hamiltonian, i.e. $h_{n,n+1} \to h_{n,n+1} - \langle \Psi(A) | \, h_{n,n+1} \, | \Psi(A) \rangle$. This quantity can be readily interpreted as a zero-momentum structure factor, so we can apply the formulas from Sec. 2.3. The formulas are a bit more complicated, since we have a two-site operator. In the end, the variance is given by



4 The time-dependent variational principle

Although DMRG was originally developed for finding the ground states, and, possibly, the first low-lying states of a given Hamiltonian, in the last ten years the scope of DMRG simulations has been extended to dynamical properties as well. One of the many new applications has been the simulation of time evolution, where the MPS formalism has been of crucial value for coming up with algorithms such as the time-evolving block decimation [9]. In this section, we discuss another algorithm for simulating time evolution in the thermodynamic limit, based on the time-dependent variational principle (TDVP) [16, 18, 19]. The TDVP has been applied to spin chains [20], gauge theories [21, 22], continuous field theories [23–25], and spin systems with long-range interactions [19, 26–29].

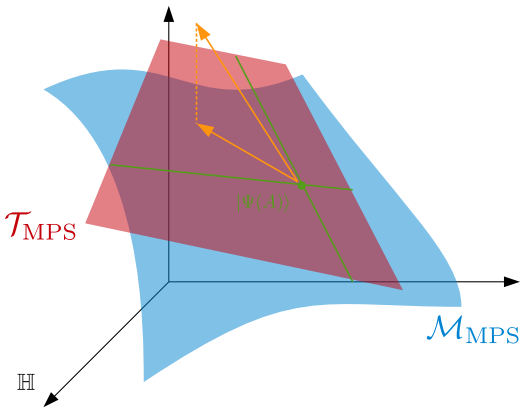


Figure 2: The time-dependent variational principle.

The algorithm relies on the manifold interpretation of uniform matrix product states, and, in particular, the concept of a tangent space. We start from the Schrödinger equation,

$$i\frac{\partial}{\partial t}\left|\Psi(A)\right\rangle = H\left|\Psi(A)\right\rangle,$$

which dictates how a given MPS $|\Psi(A)\rangle$ evolves in time. The problem with this equation is the fact that its integration would bring you out of the MPS manifold. Nonetheless, we would like to find a path inside the manifold $|\Psi(A(t))\rangle$, which approximates the time evolution in an optimal way. The time-derivative of this time-evolved MPS is a tangent vector,

$$i\frac{\partial}{\partial t}\left|\Psi(A(t))\right\rangle = \left|\Phi(\dot{A};A)\right\rangle,$$

but, again, the right-hand side is not. Instead, the vector $H\left|\Psi(A(t))\right\rangle$ points out of the manifold, so that an exact integration of the Schrödinger equation is out of the question. Still, we would like to find the optimal $\dot{A}$ within the tangent space, which would dictate how to integrate the time evolution within the manifold in an optimal way. This optimality condition can be reformulated as minimization problem for $B = \dot{A}$,

$$B = \arg\min \left\| H\left|\Psi(A)\right\rangle - \left|\Phi(B;A)\right\rangle\right\|_2^2.$$

Note that the solution of this minimization problem is equivalent to projecting the time evolution orthogonally onto the tangent space,

$$i\frac{\partial}{\partial t}\left|\Psi(A(t))\right\rangle = P_{|\Psi(A(t))\rangle} H\left|\Psi(A(t))\right\rangle.$$

This projection transforms the linear Schrödinger equation into a highly non-linear differential equation on a manifold, and is illustrated graphically in Fig. 2.

Although the original Schrödinger equation is norm preserving, the approximate one is not necessarily, but by imposing that $\langle\Psi(A)|\Phi(B;A)\rangle = 0$, norm preservation is enforced. By now, we know very well that an effective parametrization of the tangent vector in terms of the matrix X can be introduced which automatically enforces orthogonality, so the minimization boils down to

$$\min_X \left\| H\left|\Psi(A)\right\rangle - \left|\Phi(B(X);A)\right\rangle\right\|_2^2.$$

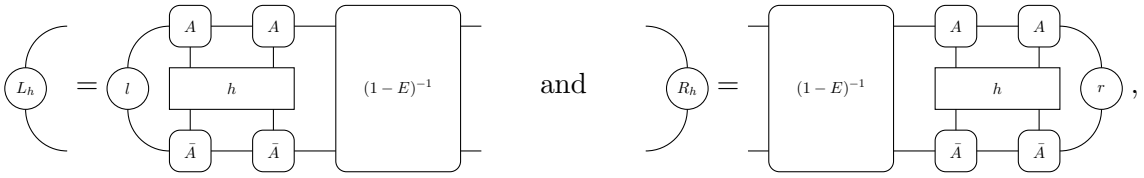
The objective function that we need to minimize is a quadratic function of X and $\bar{X}$, so that we can easily find the solution by differentiating the above expression with respect to $\bar{X}$

$$\partial_{\bar{X}} \langle \Phi(B(X); A) | \Phi(B(X); A) \rangle = \partial_{\bar{X}} \langle \Phi(B(X); A) | H | \Psi(A) \rangle \tag{8}$$

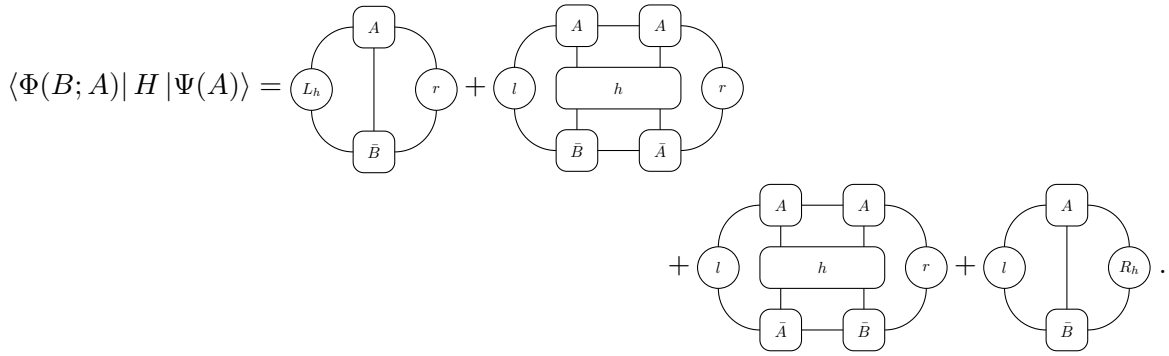
We have earlier shown that the norm of two tangent vectors reduces to the Euclidean norm on the matrices X, i.e.

$$\langle \Phi(B(X); A) | \Phi(B(X); A) \rangle = 2\pi\delta(0)\mathrm{Tr}(X^\dagger X),$$

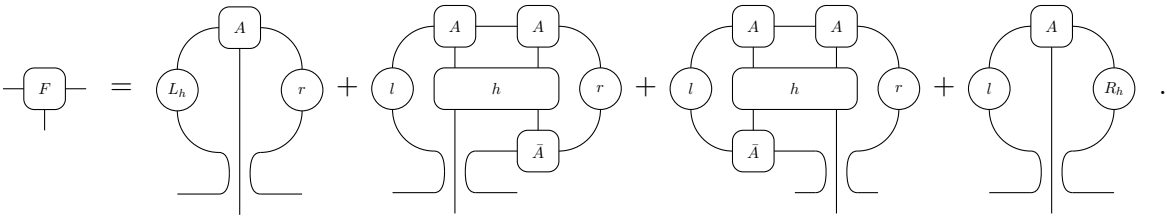
so that the derivative $\partial_{\bar{X}}$ in the left-hand side of Eq. 8 is trivial. In order to compute the right-hand side, we compute the matrix element $\langle \Phi(B; A) | H | \Psi(A) \rangle$ for general B. Again, we have two infinite sums, but one is eliminated because of translation invariance and gives rise to a $2\pi\delta(0)$ factor. Then we need all terms where the Hamiltonian term acts fully to the left and to the right of the B tensor, but this can again be resummed efficiently by introducing pseudo-inverses of the transfer matrix in the following partial contractions:



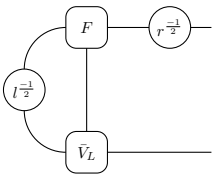
In addition, we also have the terms where the Hamiltonian operator acts directly on the site where the B tensor is located. Putting all contributions together, we obtain



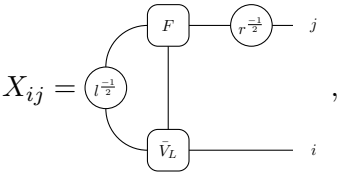
Differentiating this expression with respect to $\bar{B}$ gives us the tensor F,



Through the chain rule, we can compute the derivative with respect to the effective parameters in $\bar{X}$, so that we obtain for the right-hand side of Eq. (8)

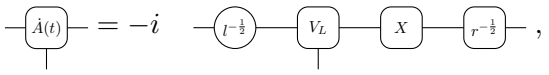


Since the left-hand side of Eq. (8) was trivial – we had $\partial_{\bar{X}} \langle \Phi(B(X); A) | \Phi(B(X); A) \rangle = X$ – we obtain as a solution for the matrix X

$X_{ij} =$ 

,

In conclusion, the path through the MPS manifold that optimally approximates the full time evolution is given by the non-linear differential equation



,

where X is computed according to the above formulas. Here we leave open how to actually integrate this differential equation.

At this point, the attentive reader might already have noticed that these formulas are very similar to the ones that we obtained for the gradient of the energy that appears in a ground-state optimization algorithm – the tangent-space gradient is the same as $|\Phi(\dot{A}; A)\rangle$. The connection is laid bare by noting that another road to a ground-state optimization algorithm is found by implementing an imaginary-time evolution ($t \to -i\tau$) on a random starting point $|\Psi(A_0)\rangle$, confined within the MPS manifold. Indeed, in the infinite-time limit this should result in a projection on the best approximation of the ground state of H,

$$\lim_{\tau\to\infty} \frac{\mathrm{e}^{-H\tau} |\Psi(A_0)\rangle}{\left\| \mathrm{e}^{-H\tau} |\Psi(A_0)\rangle \right\|} = |\Psi(A_{\mathrm{opt}})\rangle \,.$$

If the above equations for the time-dependent variational principle are integrated with a simple Euler scheme,

$$A(\tau + \mathrm{d}\tau) = A(\tau) - \mathrm{d}\tau \dot{A}(\tau),$$

we are effectively doing a steepest-descent optimization with the tangent-space gradient, where in every iteration the line search is replaced by taking a fixed step size $\alpha = \mathrm{d}\tau$.

5 Elementary excitations

We have seen that working directly in the thermodynamic limit has a number of conceptual and numerical advantages over finite-size algorithms, but the real power of the formalism is shown when we want to describe elementary excitations. These show up in dynamical correlation functions [see Sec. 5.4] that can be directly measured in e.g. neutron-scattering experiments. Typically, these experiments allow to probe the spectrum within a certain momentum sector, giving rise to excitation spectra that typically look like the one in Fig. 3. The isolated branches in such a spectrum – these will correspond to δ peaks in the spectral functions, and are seen as very strong resonances in experimental measurements – can be interpreted as quasiparticles, which can be thought of as local perturbations on the ground state, in a plane-wave superposition with well-defined momentum. The rest of the low-energy spectrum can be further built up by summing up the energies and momenta of the isolated quasiparticles – in the thermodynamic limit these quasiparticles will never see each other, so they can be simply superposed. This picture implies that all the low-energy properties should in the end be brought back to the properties of these quasiparticles!

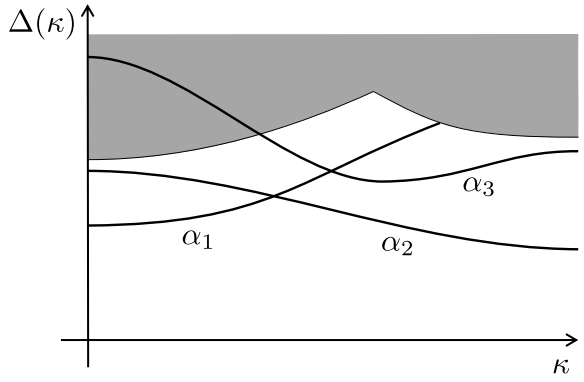


Figure 3: A typical excitation spectrum.

Crucially, this approach differs from standard approaches for describing quasiparticles in strongly-correlated quantum systems. Indeed, typically a quasiparticle is thought of as being defined by starting from a non-interacting limit, and acquires a finite lifetime as interactions are turned on – think of Fermi liquid theory as the best example of this perturbative approach. In contrast, our approach will be variational, as we will approximate exact eigenstates with a variational ansatz. This means that our quasiparticles have an infinite lifetime, and correspond to stationary eigenstates of the fully interacting Hamiltonian.

5.1 The quasiparticle ansatz

It is in fact very natural to construct quasiparticle excitations on top of an MPS ground state in the thermodynamic limit. The variational ansatz that we will introduce is a generalization of the single-mode approximation [30], which appeared earlier in the context of spin chains, and the Feynman-Bijl ansatz [31], which was used to describe the spectrum of liquid helium or quantum Hall systems [32]. In the context of MPS, the ansatz appeared earlier in Refs. [33, 34], but was only recently fully explored in Refs. [16, 35]. In recent years, the ansatz has been succesfully applied to spin chains [36, 37], spin ladders [38], field theories [39], local gauge theories [21] and bose gases [23].

The excitation ansatz is given by

$$|\Phi_p(B)\rangle = \sum_n \mathrm{e}^{ipn} \sum_{\{s\}} \boldsymbol{v}_{\boldsymbol{L}}^\dagger \left[\prod_{m<n} A^{s_m}\right] B^{s_n} \left[\prod_{m>n} A^{s_m}\right] \boldsymbol{v}_{\boldsymbol{R}} \, |\{s\}\rangle \,, \tag{9}$$

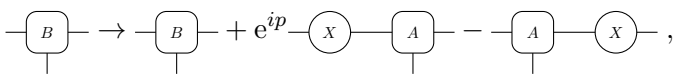
$$= \sum_n \mathrm{e}^{ipn} \ldots - \boxed{A} - \boxed{A} - \boxed{B} - \boxed{A} - \boxed{A} - \,.$$

$$\qquad\qquad\qquad\qquad\quad \ldots \qquad s_{n-1} \qquad s_n \qquad s_{n+1} \qquad \ldots$$

i.e. we change one A tensor of the ground state at site n and make a momentum superposition. The newly introduced tensor B contains all the variational parameters of the ansatz, and perturbs the ground state over a finite region around site n in every term of the superposition – it uses the correlations in the ground state, carried over the virtual degrees of freedom in the MPS to create a lump on the background state. Clearly, these excitations have a well-defined momentum, and, as we will see, a finite (non-extensive) energy above the (extensive) energy density of the ground state.

Before we start optimizing the tensor B, we will investigate the variational space in a bit more detail. First note that the excitation ansatz is, in fact, just a boosted version of a tangent

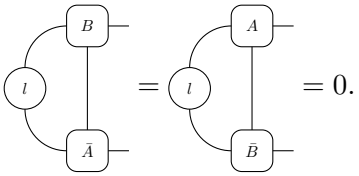
vector, so we will be able to apply all tricks and manipulations of the previous sections. For example, the B tensor has gauge degrees of freedom: the state is invariant under an additive gauge transformation of the form



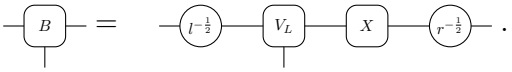
(10)

with X an invertible $D \times D$ matrix. This gauge freedom can be easily checked by substituting this form in the state (9), and observing that all terms cancel, leaving the state invariant.

The gauge degrees of freedom can be eliminated – they correspond to zero modes in the variational subspace, which would make the variational optimization ill-conditioned – by imposing a gauge fixing condition. Again, we can impose the left gauge-fixing condition

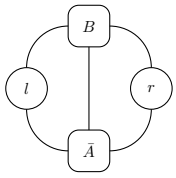


We can reuse the method for parametrizing the B tensor such that it automatically obeys this gauge condition:



As before, this fixing of the gauge freedom entails that the excitation is orthogonal to the ground state, because

$$\langle \Psi(A) | \Phi_p(B) \rangle = 2\pi\delta(p)$$



$$= 0.$$

The overlap between two excitations $|\Phi_p(B)\rangle$ and $|\Phi_{p'}(B')\rangle$ is computed similarly as before: we have two infinite terms, but we can eliminate one sum because of the translation invariance of the ground state. Now this will result in a $2\pi\delta(p - p')$ function,

$$\sum_{n \in \mathbb{Z}} \mathrm{e}^{i(p-p')n} = 2\pi\delta(p - p'),$$

so excitations at different momenta are always orthogonal. Again, the physical norm on the excited states reduces to the Euclidean norm on the effective parameters,

$$\langle \Phi_{p'}(B(X')) | \Phi_p(B(X)) \rangle = 2\pi\delta(p - p') \mathrm{Tr}\left((X')^\dagger X \right).$$

This will prove to be a very nice property for optimizing the variational parameters.

5.2 Computing expectation values

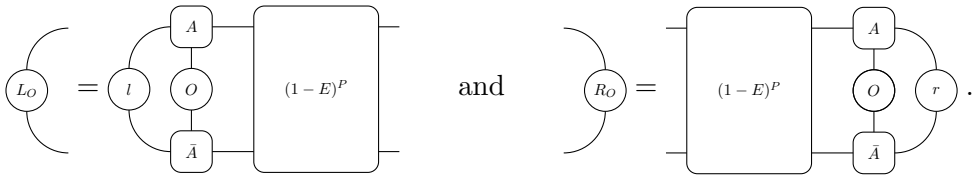
Let us first write down the expressions for evaluating expectation values, or more generally, matrix elements of the form

$$\langle \Phi_{p'}(B') | \, O \, | \Phi_p(B) \rangle \,,$$

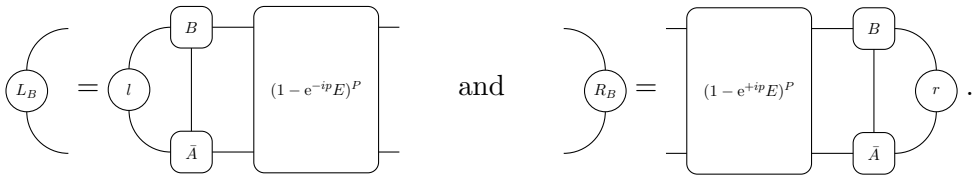
where the operator is of the form $O = \sum_i O_i$, and the ground-state expectation value has already been subtracted, i.e. $\langle \Psi(A) | \, O \, | \Psi(A) \rangle = 0$. This implies that we will look at expectation values

of O relative to the ground state density. As we will see, this will give rise to finite quantities in the thermodynamic limit.

First we notice that the above matrix element is, in fact, a triple infinite sum. Again, one of the sums can be eliminated due to translation invariance, so that we are only interested in all different relative positions of the operator O, the B tensor in the ket layer, and the B' tensor in the bra layer. Let us first define two partial contractions, corresponding to the orientations where O is to the left and to the right of both B and B',

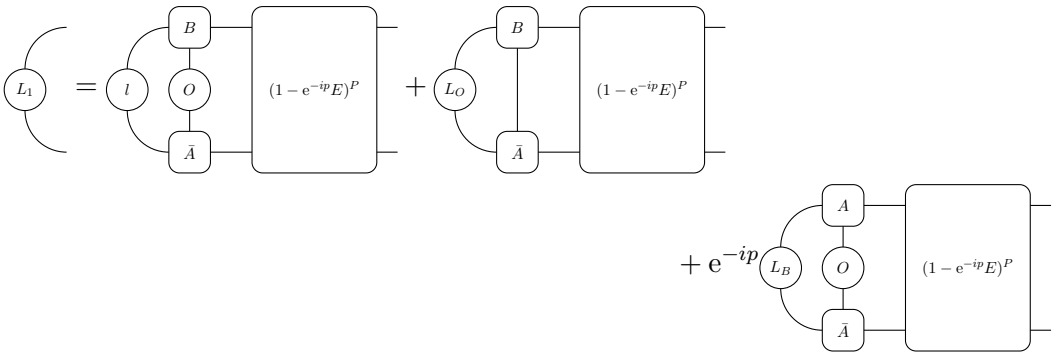


Similarly, we define the partial contractions where B travels to the outer left or right of the chain:

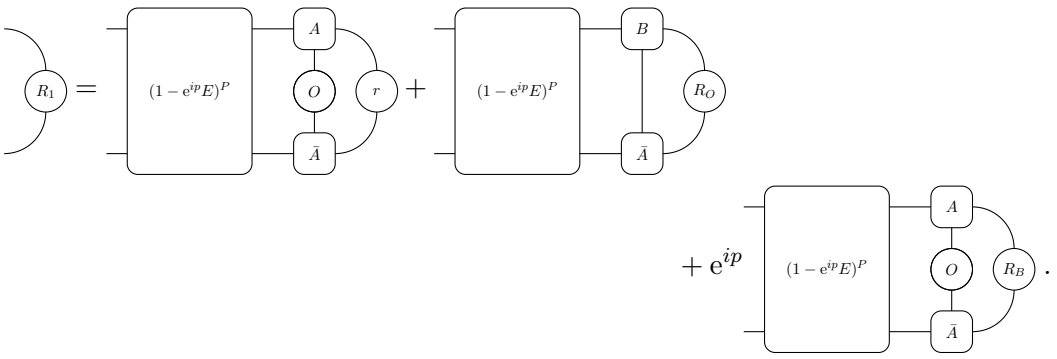


In these expressions, instead of taking pseudo-inverses, we can equally well take the full inverse, under the condition that $p \neq 0$. Indeed, in that case the operator $(1 - \mathrm{e}^{\pm ip}E)$ has spectral radius strictly smaller than one, so that the geometric series always converge and the inverse can be safely taken. In the following, this should always be kept in mind whenever a $(\dots)^P$ appears.

We use the above expressions to define all partial contractions where B and O are both either to the left or to the right of B',



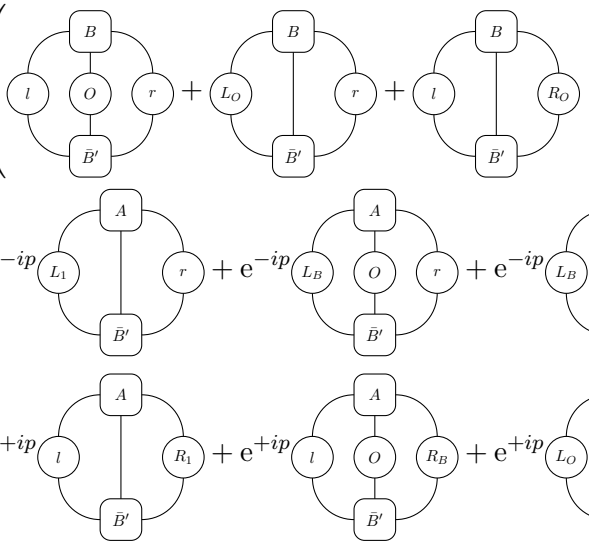
and



The extra $\mathrm{e}^{\pm ip}$ factors are included, because, if these partial contractions are combined in the final expression for the expectation value, the relative positions of the B and B' will require this extra factor.

The final expression is

$$\langle \Phi_{p'}(B')| O |\Phi_p(B)\rangle = 2\pi\delta(p-p') \Bigg(+ \mathrm{e}^{-ip} + \mathrm{e}^{-ip} + \mathrm{e}^{-ip} + \mathrm{e}^{+ip} + \mathrm{e}^{+ip} + \mathrm{e}^{+ip} \Bigg) . \tag{11}$$



5.3 Solving the eigenvalue problem

At this point, we still need to find the algorithm for the variational optimization of the B tensor in the excitation ansatz. We have seen that the effective parametrization in terms of an X matrix (i) fixes all gauge degrees of freedom, (ii) removes all zero modes in the variational subspace, (iii) makes the computation of the norm of an excited state particularly easy, and (iv) makes sure the excitation is orthogonal to the ground state, even at momentum zero. The variational optimization boils down to minimizing the energy function,

$$\min_X \frac{\langle \Phi_p(X)| H |\Phi_p(X)\rangle}{\langle \Phi_p(X)|\Phi_p(X)\rangle}.$$

Because both numerator and denominator are quadratic functions of the variational parameters X, this optimization problem reduces to solving the generalized eigenvalue problem

$$H_{\text{eff}}(p)\boldsymbol{X} = \omega N_{\text{eff}}(p)\boldsymbol{X},$$

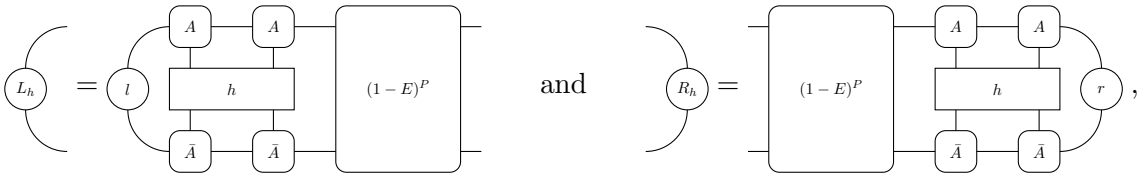
where the effective energy and normalization matrix are defined as

$$\begin{aligned} 2\pi\delta(p-p')(\boldsymbol{X'})^\dagger H_{\text{eff}}(p)\boldsymbol{X} &= \langle \Phi_{p'}(X')| H |\Phi_p(X)\rangle \\ 2\pi\delta(p-p')(\boldsymbol{X'})^\dagger N_{\text{eff}}(p)\boldsymbol{X} &= \langle \Phi_{p'}(X')|\Phi_p(X)\rangle , \end{aligned}$$

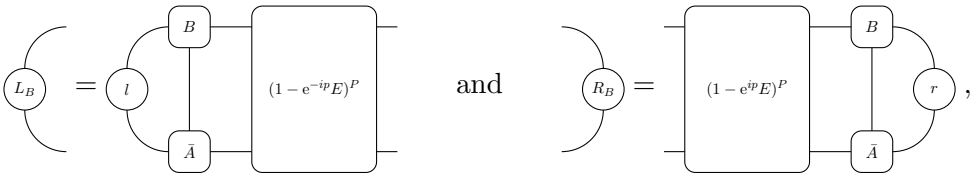
and $\boldsymbol{X}$ is a vectorized version of the matrix X. Now since the overlap between two excited states is of the simple Euclidean form, the effective normalization matrix reduces to the unit matrix, and we are left with an ordinary eigenvalue problem.

Solving the eigenvalue problem requires us to find an expression of H_{eff}, or, rather, the action of H_{eff} on a trial vector. Indeed, since we are typically only interested in finding its lowest eigenvalues, we can plug the action of H_{eff} into an iterative eigensolver. This has great implications on the computational complexity: The full computation and diagonalization of the effective energy matrix would entail a computational complexity of $\mathcal{O}(D^6)$, while the action on an input vector Y can be done in $\mathcal{O}(D^3)$ operations.

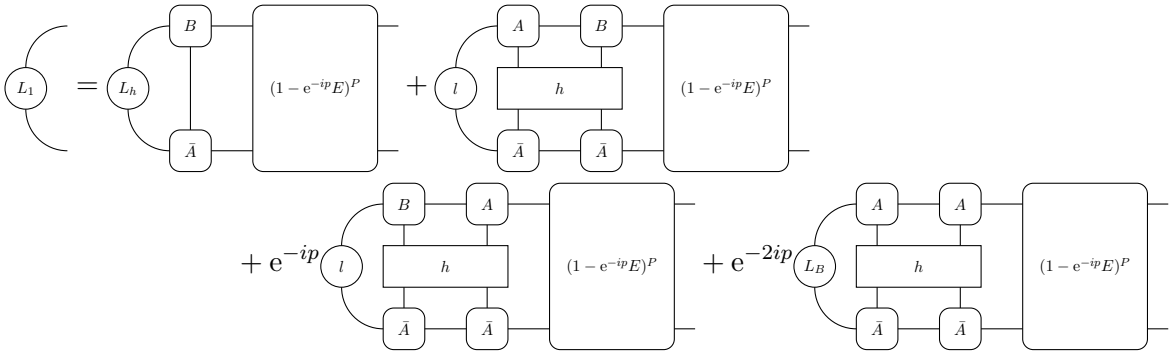
So we need the action of H_{eff} on an arbitray vector $\boldsymbol{Y}$. We first transform the matrix Y to a tensor B in the usual way. Then we need all different contributions that pop up in a matrix element of the form $\langle \Phi_{p'}(B')| \, H \, |\Phi_p(B)\rangle$, i.e. similarly to the expression (11), we need all different orientations of the nearest-neighbour operator of the Hamiltonian, the input B tensor and an output. Because we are confronted with a two-site operator here, the expressions are a bit more cumbersome. Let us again define the following partial contractions



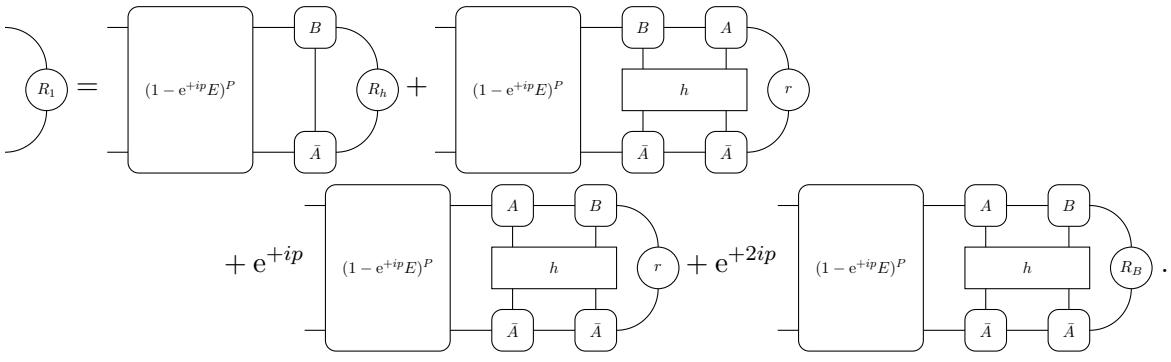
and



which we use for determining

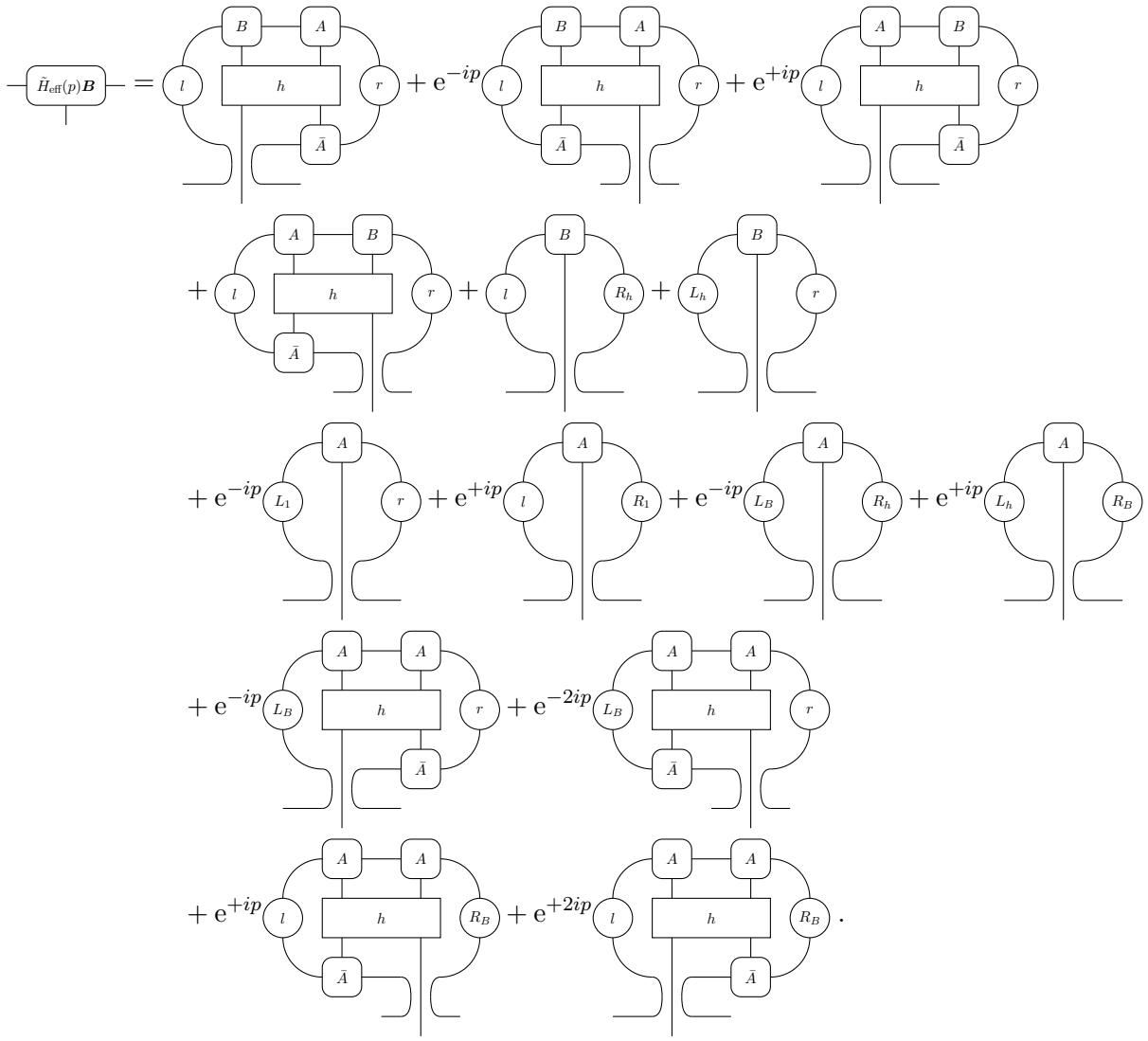


and

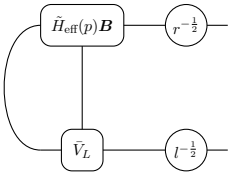


These partial contractions allow us now to implement the action of the effective energy matrix on

a given input vector $\boldsymbol{B}$ as



In the last step, we need the action of $H_{\text{eff}}(p)$ (without the tilde), so we need to perform the last contraction

$H_{\text{eff}}(p)\boldsymbol{X} =$ 

.

All contractions above have a computational complexity of $\mathcal{O}(D^3)$.

By tracing all momenta, we obtain direct access to the full excitation spectrum of the system. Note that the eigenvalue equation has $n_L D$ solutions, but only the few lowest-lying ones have a physical meaning. Indeed, for a given value of the momentum, one typically finds a limited number of excitations living on an isolated branch in the spectrum, whereas all the other solutions fall within the continuous bands. It is not expected that these states are approximated well with the quasiparticle ansatz. The accuracy of the approximation can be assessed by computing the energy variance – just as we did with the ground state in Sec. 3.3 – but, for an excitation this is an involved calculation [38].

5.4 Dynamical correlations

As we have mentioned before, the excitation spectrum determines the dynamical correlation functions or spectral functions. We will use the following definition of the spectral function:

$$S^{\alpha\alpha}(q,\omega) = \int_{-\infty}^{+\infty} \mathrm{d}t\, \mathrm{e}^{i\omega t} \sum_{n\in\mathbb{Z}} \mathrm{e}^{iqn} \left\langle \Psi(A)\right| O_n^\alpha(t) O_0^\alpha(0) \left|\Psi(A)\right\rangle .$$

where the time-evolved operator $O_n^\alpha(t) = \mathrm{e}^{iHt} O_n^\alpha(0) \mathrm{e}^{-iHt}$ is introduced. By projecting the time evolution on all excited states of H, we obtain the following representation

$$S^{\alpha\alpha}(q,\omega) = \sum_\gamma \int_{-\infty}^{+\infty} \mathrm{d}t\, \mathrm{e}^{i\omega t} \mathrm{e}^{-i(E_\gamma - E_0)t} \sum_{n\in\mathbb{Z}} \mathrm{e}^{iqn} \left\langle \Psi(A)\right| O_n^\alpha(0) \left|\gamma\right\rangle \left\langle \gamma\right| O_0^\alpha(0) \left|\Psi(A)\right\rangle ,$$

where γ labels all excited states of the system with excitation energies $E_\gamma - E_0$. Let us now take only the one-particle excitations into account (the excitations corresponding to isolated branches in the excitation spectrum), for which we know that they can be described by the excitation ansatz. For these states, which have a well-defined momentum, the sum is rewritten as

$$\sum_{\gamma,1\mathrm{p}} \left|\gamma\right\rangle \left\langle\gamma\right| = \sum_{\gamma\in\Gamma_1} \int_{\mathcal{R}_\gamma} \frac{\mathrm{d}p}{2\pi} \left|\Phi_p^\gamma(B)\right\rangle \left\langle\Phi_p^\gamma(B)\right| ,$$

where we have introduced Γ_1 as the set of all isolated branches in the spectrum, $\mathcal{R}_\gamma$ as the momentum region where every branche γ exists. Because of translation invariance, we have

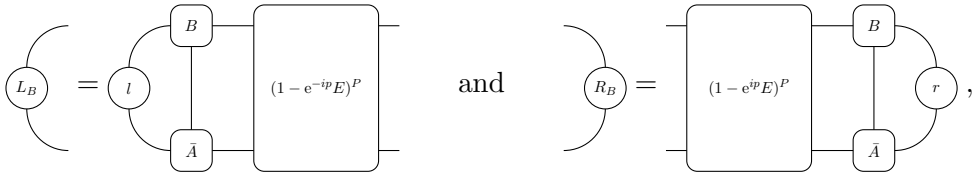
$$\sum_n \mathrm{e}^{iqn} \left\langle \Psi(A)\right| O_n^\alpha(0) \left|\Phi_p^\gamma(B)\right\rangle = 2\pi\delta(p-q) \left\langle \Psi(A)\right| O_0^\alpha(0) \left|\Phi_p^\gamma(B)\right\rangle ,$$

so that we obtain for the one-particle part of the spectral function

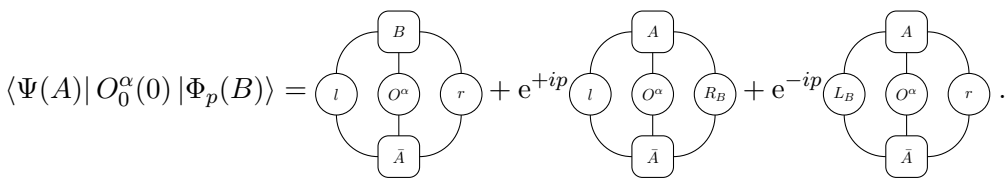
$$S_{1\mathrm{p}}^{\alpha\alpha}(q,\omega) = \sum_{\gamma\in\Gamma_1(q)} 2\pi\delta\left(\omega - \omega_\gamma(q)\right) \left|\left\langle \Psi(A)\right| O_0^\alpha(0) \left|\Phi_q^\gamma(B)\right\rangle\right|^2 ,$$

where $\Gamma_1(q)$ denotes the set of one-particle states in the momentum sector q and $\omega_\gamma(p)$ is the excitation energy of that mode.

The spectral weights are easily computed. First, we again define the following partial contractions



so that we have the following contractions



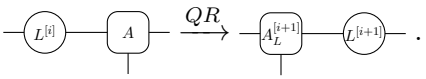
6 Inverse-free formulations

All the above tangent-space algorithms have one common issue involving the appearance of the matrices $l^{-1/2}$ and $r^{-1/2}$. The problem is that MPS approximations for physical states rely on the fact that the Schmidt spectrum decays fast; indeed, a finite-D approximation of a given state is only warranted if the Schmidt values that were neglected are small enough. This implies that the inversion of the matrices l and r should be ill-conditioned, because they contain the Schmidt numbers as their singular values. Therefore, it seems a bit of a paradox that one should have to invert these matrices to implement all the above algorithms. In fact, this problem does not occur in finite DMRG/MPS algorithms, because one can always implement the right gauge transformations by absorbing the transformation matrices into the tensor on the next site [8]. In the infinite case, this is not possible, because our state always has to remain translation-invariant. Recently, however, partly inspired by the finite-system approach, an inverse-free formulation of the TDVP was developed [19] for the case of MPS on a finite system, and the generalization to the thermodynamic limit was introduced for fixed point algorithms of transfer matrices [40]. In this section, we discuss the inverse-free versions of the TDVP and the quasiparticle ansatz.

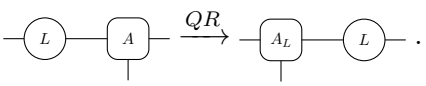
6.1 A single-layer algorithm for finding canonical forms

The trouble with ill-conditioned inversions starts with finding the canonical forms for the MPS tensors, which, indeed, requires taking the inverse square root of the fixed points. In addition to the potentially ill-conditioned inversion, there is an additional difficulty: the precision of the fixed points is necessarily bad, because they are found as a square root of another matrix. Indeed, if we find the smallest eigenvalues of l with a precision ϵ, the precision on the ones of L is only $\sqrt{\epsilon}$. Both problems are resolved by taking recourse to single-layer algorithms, i.e. algorithms that only work on the level of the MPS tensors in the ket layer, and never consider operations for which contractions with the bra layer are needed.

Suppose we are given an MPS tensor A, and we want to find the left-canonical tensor A_L and the matrix L, such that $A_L = L^{-1}AL$.[4] The idea is to solve the equation $LA_L = AL$ iteratively, where in every iteration (i) we start from a matrix $L^{[i]}$, (ii) we construct the tensor $L^{[i]}A$, (iii) we take a QR decomposition to obtain $A_L^{[i+1]}L^{[i+1]} = L^{[i]}A$, and (iv) we take $L^{[i+1]}$ to the next iteration. The QR decomposition is represented diagramatically as



Because the QR decomposition is unique – in fact, it is made unique by the additional condition that the diagonal elements of the triangular matrix be positive – this iterative procedure is bound to converge to a fixed point for which $L^{[i+1]} = L^{[i]} = L$:



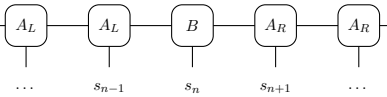
6.2 The tangent-space projector

In order to arrive an inverse-free formulation of the time-dependent variational principle, we need to find a representation of the projector onto the tangent space that avoids the factors $l^{-1/2}$ or

[4] We apply a slight abuse of notation here: The expressions AX and XA, with A an MPS tensor and X a matrix are meant as A^sX, $\forall s$.

$r^{-1/2}$. The way forward is by writing a different representation for a tangent vector:

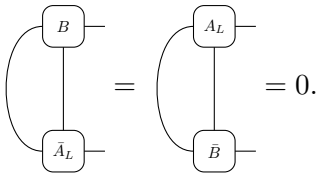
$$|\Phi(B; A_L, A_R)\rangle = \sum_n \ldots$$



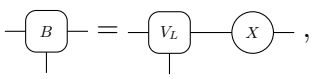
,

i.e. we bring in every term all tensors to the left of the B tensors in the left gauge, and vice versa for the tensors to the right. The transformation matrices can be easily absorbed into the definition of B. The crucial difference with the standard form of the tangent vector is that the elements of B are now not directly related to derivatives with respect to the parameters in the MPS tensors A_L and A_R, and that we need to derive the projector onto the tangent space in a slightly more involved way.

The first step, however, is analoguous to what we did previously: we impose the left-gauge fixing condition, which has the simpler form

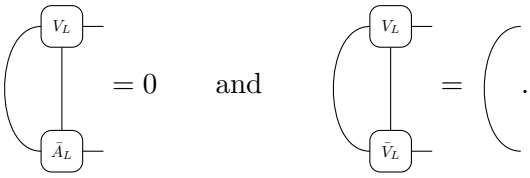


We define the effective parametrization of the tangent vector in terms of the matrix X as



, (12)

where the tensor V_L obeys the usual conditions



.

Suppose now we have an abritrary state $|\Psi\rangle$, which we want to project on a given tangent vector $|\Phi(B; A_L, A_R)\rangle$. We realize that an orthogonal projection on a linear subspace of Hilbert space is equivalent to minimizing

$$\min_X \big\| |\Psi\rangle - |\Phi(B(X); A_L, A_R)\rangle \big\|^2 = \min_X \Big(\langle \Phi(B(X); A_L, A_R) | \Phi(B(X); A_L, A_R)\rangle - \langle \Psi | \Phi(B(X); A_L, A_R)\rangle - \langle \Phi(B(X); A_L, A_R) | \Psi\rangle \Big).$$

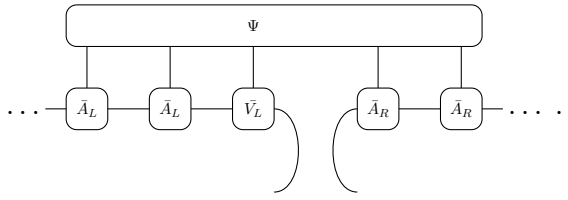
As this minimization problem is quadratic in X and $\bar{X}$, the solution is given by $X = \partial_{\bar{X}}(\ldots)$. The parametrization of Eq. (12) was chosen in a smart way, because we have

$$\langle \Phi(B(X); A_L, A_R) | \Phi(B(X); A_L, A_R)\rangle = \mathrm{Tr}(X^\dagger X),$$

so that the solution of the minimization problem is found as

$$X = \frac{\partial}{\partial \bar{X}} \langle \Phi(B(X); A_L, A_R) | \Psi\rangle \,,$$

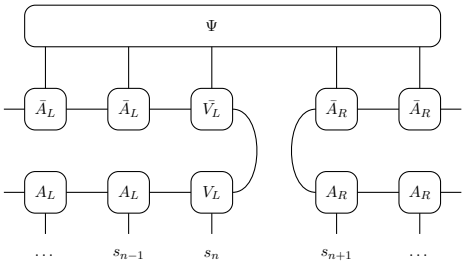
which is given by



.

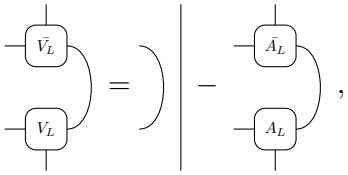
The corresponding tangent vector is

$$|\Phi(B; A_L, A_R)\rangle = \sum_n \cdots$$



$\cdots$.

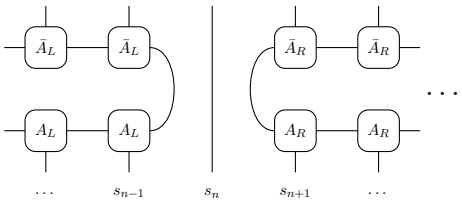
We now already see the form of the projector, but we take one more step. We rewrite the projector on V_L as



,

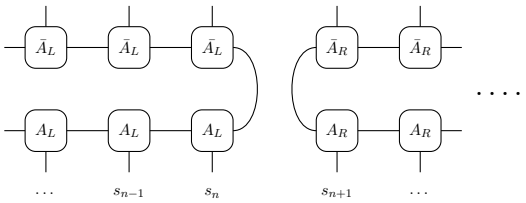
so that the final form of the ground state projector is given by

$$P_{|\Psi(A)\rangle} = \sum_n \cdots$$



$\cdots$

$$- \sum_n \cdots$$



$\cdots$.

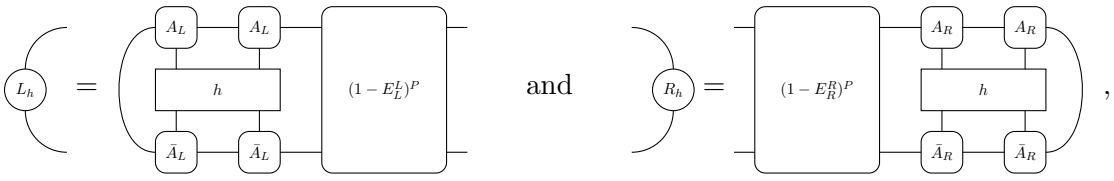
In contrast to the simpler form of the previous section, in the mixed canonical representation the tangent-space projector has two different terms, but, precisely because we work with a mixed canonical form, the projector does not require the inversion of potentially ill-conditioned matrices $l^{-1/2}$ and $r^{-1/2}$.

6.3 Inverse-free TDVP

We can now use this tangent-space projector to write down the TDVP equation for an MPS in the mixed canonical form. We have explained that the optimal way for implementing real-time evolution within the MPS manifold is by projecting the exact time derivative onto the tangent space at every point, i.e.

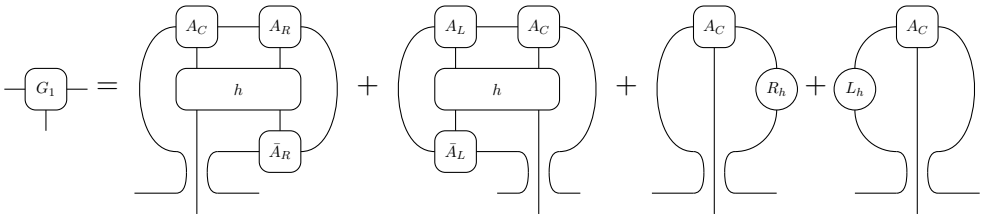
$$-i\frac{\partial}{\partial t}|\Psi(A_L, A_C, A_R)\rangle = P_{|\Psi(A)\rangle} H |\Psi(A_L, A_C, A_R)\rangle$$

For both parts of the tangent-space projector, the differential equation can be integrated straightforwardly. Let us take the first part, for which we first define the partial contractions



,

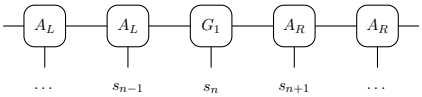
which capture the contributions where the Hamiltonian is completely to the left and to right of the open spot in the projector. Again, we have used pseudo-inverses for resumming the infinite number of terms. These are combined into



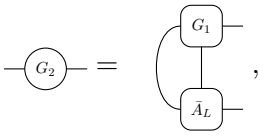
(13)

such that

$$P^1_{|\Psi(A)\rangle} H\,|\Psi(A_L, A_C, A_R)\rangle = \sum_n \ldots$$



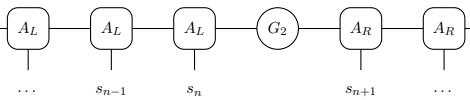
In order to obtain the second part, we need to contract the above G_1 tensor another time with $\bar{A}_L$,



(14)

in order to arrive at

$$P^2_{|\Psi(A)\rangle} H\,|\Psi(A_L, A_C, A_R)\rangle = \sum_n \ldots$$



The two parts of the differential equation can be solved separately, but in different representations of the MPS. Indeed, if we write the MPS in the mixed canonical form *with* center site, the first equation is simply

$$\dot{A}_C = -iG_1(A_C),$$

where $G^1(A)$ is interpreted as a linear operator working on A_C according to Eq. (13); the solution is simply $A_C(t) = \mathrm{e}^{-iG^1 t} A_C(0)$. Alternatively, if the MPS is written in the mixed canonical form *without* center site, the second equation is

$$\dot{C} = +iG_2(C),$$

where now $G_2(A)$ is seen as a linear operator acting on C according to Eqs. (13) and (14); again, the solution is $C(t) = \mathrm{e}^{+iG_2 t} C(0)$. These exponentials can be evaluated efficiently by using iterative procedures.

Integrating the TDVP equations

The meaning of the TDVP equations is slightly different in this mixed canonical form, and a correct interpretation starts from considering the case of a finite lattice. There the meaning is clear: every site in the lattice has a different MPS tensor attached to it, and performing one time step amounts to doing one sweep through the chain. For every step in the sweep at site n, we

- start from a mixed canonical form with center site tensor $\hat{A}_C(n)$, all tensors $\tilde{A}_L(n-2)$, $\tilde{A}_L(n-1)$, etc, have already been updated, while tensors $A_R(n+1)$ and $A_R(n+2)$, etc., are still waiting for their update,

- we update the center-site tensor as $\tilde{A}_C(n) = \mathrm{e}^{-iG_1(n)\delta t}\hat{A}_C(n)$,
- we do a QR decomposition, $\tilde{A}_C(n) = \tilde{A}_L(n)\tilde{C}(n)$,
- we update the center matrix as $\hat{C}(n) = \mathrm{e}^{+iG_2(n)\delta t}\tilde{C}(n)$
- we absorb this center matrix into the tensor on the right to define a new center-site tensor $\hat{A}_C(n+1) = \hat{C}(n)A_R(n+1)$.

The version for the infinite system can be derived by starting this procedure at an arbitrary site n in the chain – say $n \to -\infty$, so that we will never notice the effect of this abrupt operation in the bulk of the system – and start applying exactly the same procedure until it converges. In this context, convergence at site n would mean that the center-site that we obtain for the next site, $\hat{A}_C(n+1)$, would give us the same as the one we started from, $\tilde{A}_C(n) = A_C(n+1)$. Our real interest, however, goes out to the converged value of $\tilde{A}_L$, because this allows us to obtain $\tilde{A}_R$, $\tilde{A}_C$ and $\tilde{C}$. Only after we have obtained convergence in this sense, we have concluded the integration of one time step δt.

Imaginary-time evolution

In the case of imaginary-time evolution, where we are interested in the infinite-time limit, integrating these equations with full convergence in every time step would be very costly. Let us therefore take another approach, and see what convergence in every time step would imply. As before, we start from $-\infty$, and sweep through the lattice; in every step we

- start from a center-site tensor $\hat{A}_C$, an updated $\tilde{A}_L$, and the A_R from the previous time step,
- evolve the center-site tensor as $\tilde{A}_C = \mathrm{e}^{-G_1\delta\tau}\hat{A}_C$,
- define $\tilde{A}_L$ and $\tilde{C}$ from a QR decomposition of $\tilde{A}_C$,
- evolve the center matrix as $\hat{C} = \mathrm{e}^{+G_2\delta\tau}\tilde{C}$,
- and, redefine $\hat{A}_C = \hat{C}A_R$.

If this would be a fixed point, we would have $\hat{A}_C = A_C$ and $\hat{C} = C$. But this would imply that we can reverse the time integration according to G_2 and write $\tilde{C} = \mathrm{e}^{-G_1\delta\tau}C$. Further turning things around would lead us to a procedure where we start from a given $\{A_L, A_R, A_C, C\}$, evolve *both* A_C and C *forward* in imaginary time, and find a new value of $\tilde{A}_L$ and $\tilde{A}_R$ from the updated $\tilde{A}_C$ and $\tilde{C}$.

One troublesome step is finding the updated $\tilde{A}_L$ and $\tilde{A}_R$ from $\tilde{A}_C$ and $\tilde{C}$. One obvious solution would be $\tilde{A}_L = \tilde{C}^{-1}\tilde{A}_C$, but this would again involve the inversion of the center matrix, which we were out to avoid. The problem is recast as a minimization problem

$$\min_{\tilde{A}_L} \left\| \tilde{A}_C - \tilde{A}_L\tilde{C} \right\|^2,$$

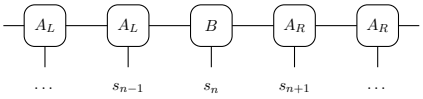
for which the exact solution is obtained by taking the singular-value decomposition $USV^\dagger = \tilde{A}_C\tilde{C}^\dagger$, and setting $\tilde{A}_L = UV^\dagger$. However, if $\tilde{A}_C$ would indeed be equal to $\tilde{A}_L\tilde{C}$, we would be computing $\tilde{A}_L\tilde{C}^\dagger\tilde{C}$, for which the singular values correspond to the squares of the ones of $\tilde{C}$. Again, we would lose precision on the small singular values, which we want to avoid. An alternative way of solving the minimization problem consists of taking the QR decompositions $\tilde{A}_C = Q_1R_1$ and $\tilde{C} = Q_2R_2$, and taking $\tilde{A}_L = Q_1Q_2^\dagger$. Indeed, in order to get an exact equality $Q_1R_1 = \tilde{A}_LQ_2R_2$, we would need $R_1 = R_2$, so that taking $\tilde{A}_L = Q_1Q_2^\dagger$ should be a good approximation.

6.4 Inverse-free excitations

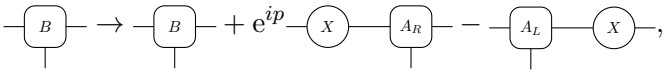
Lastly, we can reformulate the excitation ansatz in an inverse-free form. Remember that we had to invert the left and right fixed forms l and r in the effective parametrization of the B tensor, which, in the case of a ground state with very small Schmidt values, is an ill-defined operation. Yet, we can easily write down an exciation ansatz that doesn't have this issue by bringing all MPS tensor to the left of the B tensor in the left gauge, and vice versa to the right,

$$|\Phi_p(B)\rangle = \sum_n \mathrm{e}^{ipn} \sum_{\{s\}} \boldsymbol{v}_{\boldsymbol{L}}^\dagger \left[\prod_{m<n} A_L^{s_m}\right] B^{s_n} \left[\prod_{m>n} A_R^{s_m}\right] \boldsymbol{v}_{\boldsymbol{R}} \,|\{s\}\rangle\,.$$

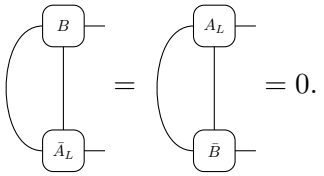
$$= \sum_n \mathrm{e}^{ipn} \ldots$$



Again, the variational ansatz has a number of additive gauge degrees of freedom,



which can be eliminated by imposing the left gauge fixing condition



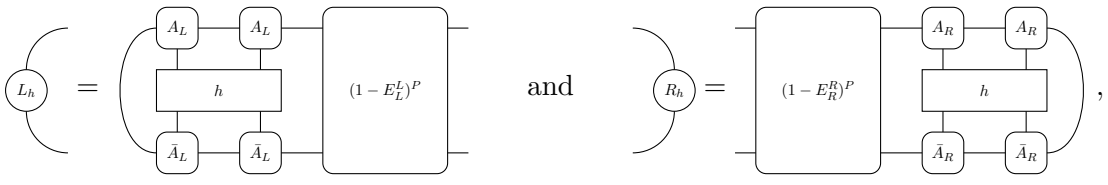
This is effectively imposed by reusing the parametrization of the tangent plane in Eq. (12) in terms of a matrix X, such that the overlap between excited states reduces to the Euclidean inner product,

$$\langle \Phi_{p'}(B(X'))|\Phi_p(B(X))\rangle = 2\pi\delta(p-p')\mathrm{Tr}\left((X')^\dagger X\right).$$

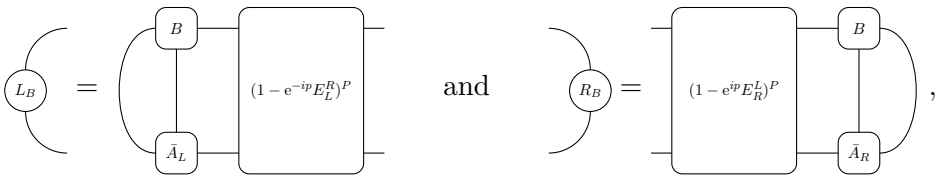
The rest of the implementation is similar, and optimizing the variational parameters boils down to finding the lowest eigenvalues and vectors of the eigenvalue problem

$$H_{\mathrm{eff}}(p)\boldsymbol{X} = \omega \boldsymbol{X}.$$

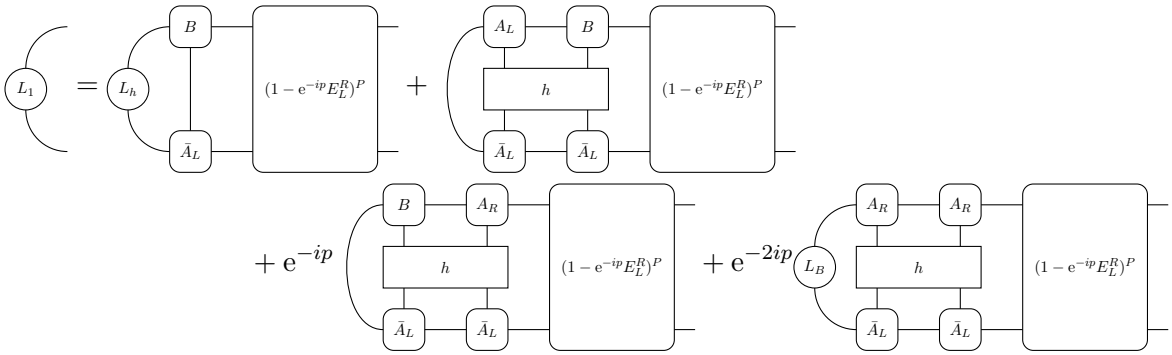
This requires us to implement the action of $H_{\mathrm{eff}}(p)$ on an input vector Y. First we transform Y to a tensor B, then we define the following partial contractions



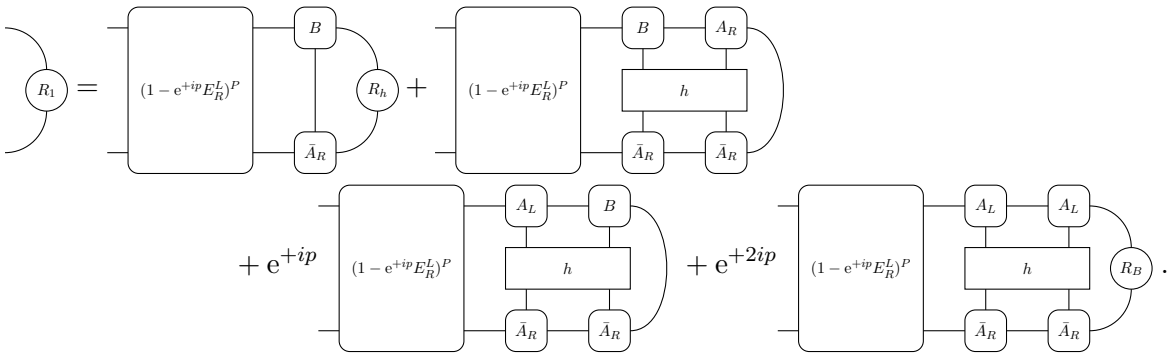
and



which we use for determining

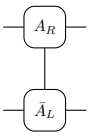


and

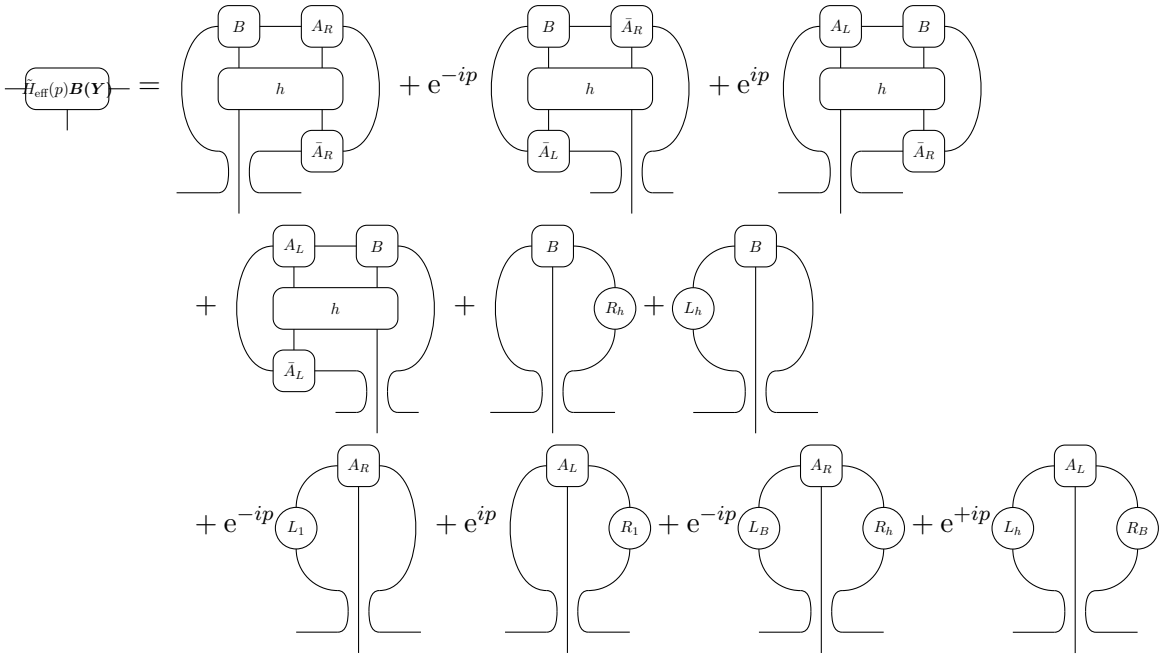


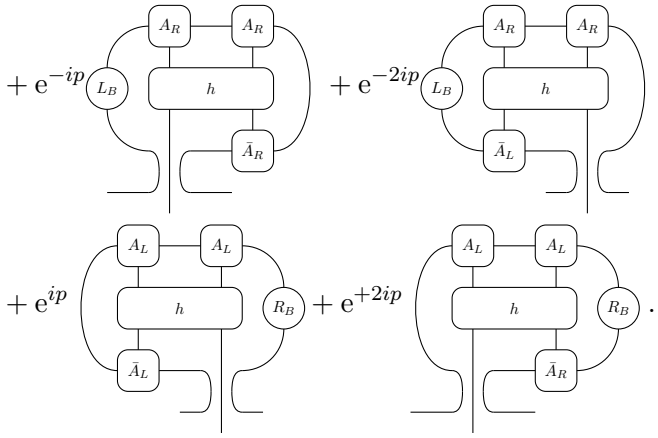
Note that we have introduced mixed transfer matrices of the form

$$E_L^R = \quad .$$

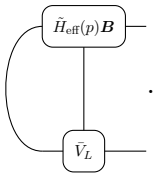


These partial contractions are then used to implement the action of the effective energy matrix on a given input vector as





In the last step, we need the action of $H_{\text{eff}}(p)$ (without the tilde), so we need to perform the last contraction

$H_{\text{eff}}(p)\boldsymbol{X} =$ 

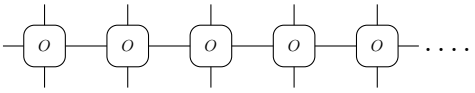
As promised, this implementation does not require us to take inverses of fixed points anywhere, so that it is potentially better conditioned.

7 Transfer matrices and fixed points

Matrix product states have been used extensively as variational ansatz for ground states of local Hamiltonians, but in the last years it has been observed that they can also provide accurate approximations for fixed points of transfer matrices. Whereas locality of the Hamiltonian is important for approximating a ground state, it appears that a transfer matrix should have the form of matrix product operator (MPO) such that the fixed point can be approximated by an MPS. A one-dimensional MPO in the thermodynamic limit is written as

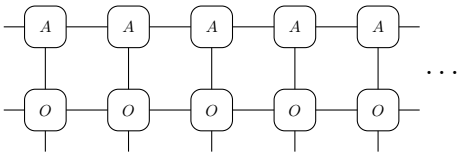
$$T(O) = \sum_{\{i\}\{j\}} \left(\dots O^{i_{n-1},j_{n_1}} O^{i_{n-1},j_{n_1}} O^{i_{n-1},j_{n_1}} \dots \right) \dots |i_{n-1}\rangle \langle j_{n-1}| \otimes |i_n\rangle \langle j_n| \otimes |i_{n+1}\rangle \langle j_{n+1}| \dots ,$$

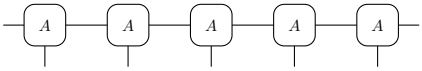
or represented diagrammatically as

$T(O) = \dots$ 

$\dots$

We now make the ansatz that the fixed point (leading eigenvector) of this operator is an MPS, such that it obeys the eigenvalue equation

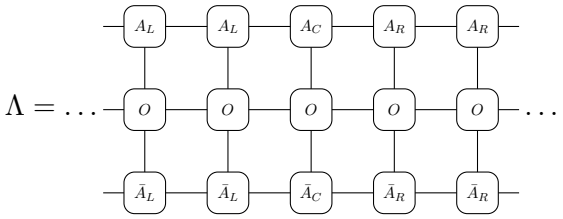
$\dots$  $\dots$

$\propto \dots$  $\dots$.

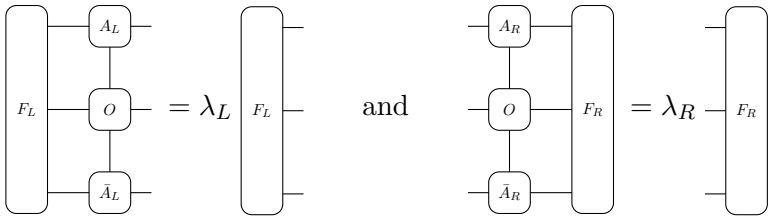
Let us first try to find a way to properly define this eigenvalue equation. Suppose we have indeed found an MPS representation $|\Psi(A)\rangle$ of the fixed point of $T(O)$, then the eigenvalue is given by

$$\Lambda = \langle\Psi(A)|\, T\, |\Psi(A)\rangle\,.$$

In order to determine Λ, we bring $|\Psi(A)\rangle$ in the mixed canonical form, such that



Contracting this infinite network requires that we find F_L and F_R, the fixed points of the left and right channel operators and T_L and T_R, which are represented diagrammatically as



The eigenvalues λ_L and λ_R are necessarily the same value λ, so that Λ is given by

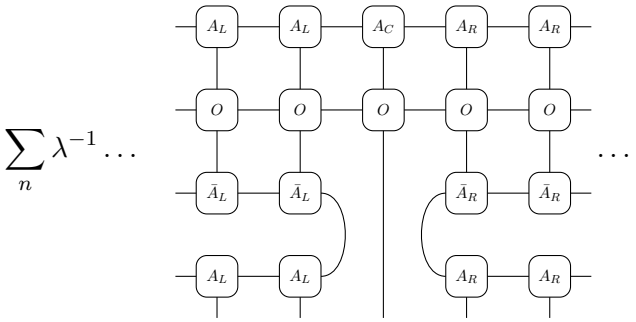
$$\Lambda = \lim_{N\to\infty} \lambda^N,$$

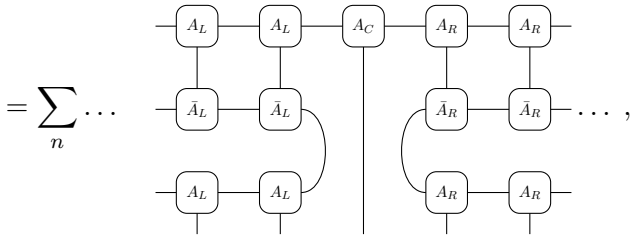
where N is the diverging number of sites. From a physical point of view, it is the 'free energy density' $f = \frac{1}{N}\log\Lambda = \log\lambda$ that is the most important quantity. In the case that we want to normalize the MPO, such that the leading eigenvalue is equal to one (or $f = 0$), we can just divide by λ: $O \to O/\lambda$.

The next step towards an algorithm is stating an optimality condition for $|\Psi(A)\rangle$ such that it can serve as an approximate eigenvector of $T(O)$. Inspired by all the above tangent-space algorithms, we will require that the projection of the residual onto the tangent space is zero:

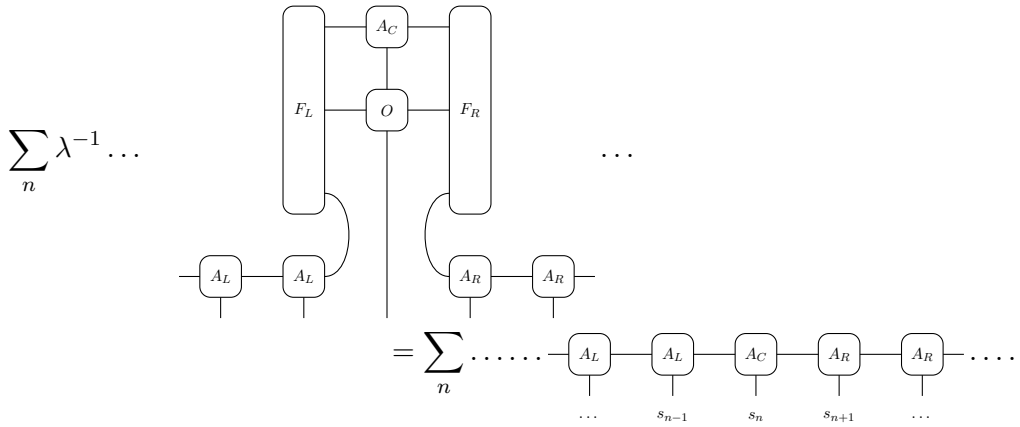
$$P_{|\Psi(A)\rangle}\left(T(O)\,|\Psi(A)\rangle - \Lambda\,|\Psi(A)\rangle\right) = 0.$$

In the mixed canonical form, the tangent-space projector consists of two parts. The first part gives us

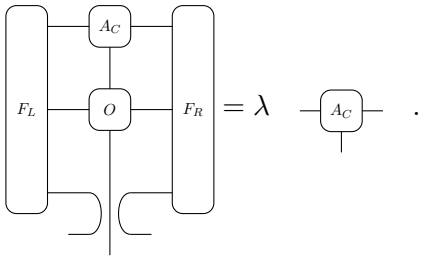




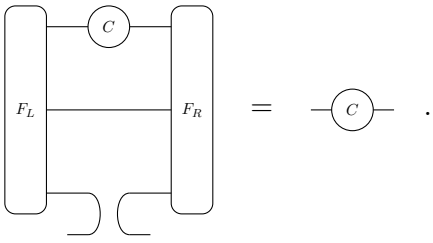
or,



The extra factor λ^{-1} arises because we have one extra MPO tensor on the left-hand side. By projecting on the $\bar{A}_L$'s and $\bar{A}_R$'s on the left and right, we end up with an eigenvalue equation for A_C,



The same procedure leads to an eigenvalue equation for C,



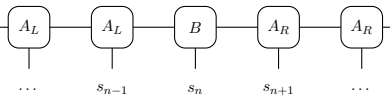
Now the condition for having an optimal MPS representation for the fixed point of $T(O)$ can be recast as the condition that the two above eigenvalue equations for A_C and C are satisfied, because no further decrease of the residual is possible without leaving the MPS manifold.

The final step towards an efficient algorithm is straightforward, given the inverse-free TDVP algorithm that we presented above. In every iteration of our algorithm, we (i) start from a given MPS $\{A_L, A_R, A_C, C\}$, (ii) determine F_L and F_R, (iii) solve the two eigenvalue equations obtaining $\tilde{A}_C$ and $\tilde{C}$, and (iv) determine the $\tilde{A}_L$ and $\tilde{A}_R$ that minimize $\|\tilde{A}_C - \tilde{A}_L\tilde{C}\|^2$ and $\|\tilde{A}_C - \tilde{C}\tilde{A}_R\|^2$.

Excited states of an MPO

For completeness, we note that we can also apply the excitation ansatz to compute 'excitations' of a transfer matrix. The algorithms for computing dispersion relations are quite similar to the case of Hamiltonians, which we have studied extensively. In a first step, we renormalize the MPO such that the eigenvalue λ of the fixed point equation equals one. Then we use the excitation ansatz,

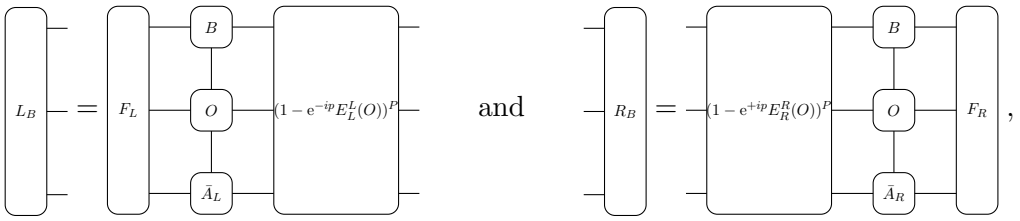
$$|\Phi_p(B)\rangle = \sum_n \mathrm{e}^{ipn} \ldots$$



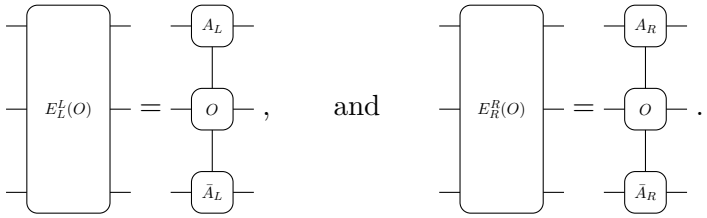
to find the subleading eigenvectors. Optimizing the variational parameters boils down to solving the eigenvalue equation,

$$T_{\text{eff}}(p)\boldsymbol{X} = \omega \boldsymbol{X},$$

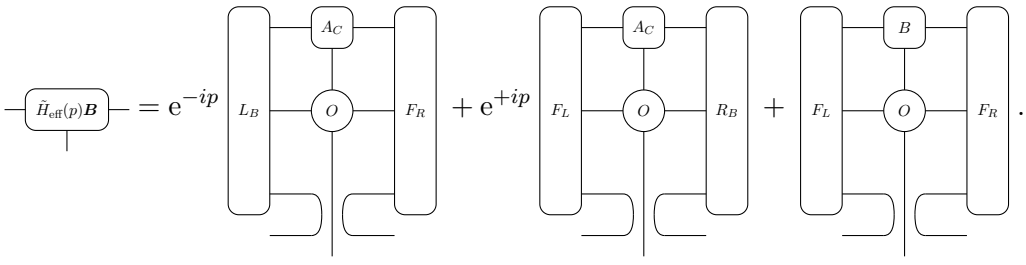
where T_{eff} is implemented iteratively on a general input tensor B. We define the partial contractions,



where the channel operators are defined as



Again – if everything is properly normalized – these operators have a leading eigenvalue equal to one (with F_L and F_R as fixed points), so they should be regularized in order to define the inverses at momentum $p = 0$. The action of $T_{\text{eff}}(p)$ is then given by



References

[1] J. Eisert, M. Cramer, and M. B. Plenio, "Colloquium: Area laws for the entanglement entropy," *Reviews of Modern Physics* **82**, 277 (2010).

[2] R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Annals of Physics* **349**, 117 (2014).

[3] N. Schuch, "Condensed Matter Applications of Entanglement Theory," *arXiv* , 1306.5551 (2013).

[4] J. Eisert, "Entanglement and tensor network states," *arXiv* , 1308.3318 (2013).

[5] M. B. Hastings, "Solving gapped Hamiltonians locally," *Physical Review B* **73**, 085115 (2006).

[6] F. Verstraete and J. I. Cirac, "Matrix product states represent ground states faithfully," *Physical Review B* **73**, 094423 (2006).

[7] F. Verstraete, D. Porras, and J. I. Cirac, "Density Matrix Renormalization Group and Periodic Boundary Conditions: A Quantum Information Perspective," *Physical Review Letters* **93**, 227205 (2004).

[8] U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states," *Annals of Physics* **326**, 96 (2011).

[9] G. Vidal, "Classical Simulation of Infinite-Size Quantum Lattice Systems in One Spatial Dimension," *Physical Review Letters* **98**, 070201 (2007).

[10] D. Pérez-García, F. Verstraete, M. M. Wolf, and J. I. Cirac, "Matrix Product State Representations," *arXiv* , 0608197 (2006).

[11] D. Pérez-García, M. Wolf, M. Sanz, F. Verstraete, and J. I. Cirac, "String Order and Symmetries in Quantum Spin Lattices," *Physical Review Letters* **100**, 167202 (2008).

[12] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, "Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems," *Physical Review Letters* **93**, 207204 (2004).

[13] J. Haegeman, M. Mariën, T. J. Osborne, and F. Verstraete, "Geometry of matrix product states: Metric, parallel transport, and curvature," *Journal of Mathematical Physics* **55**, 021902 (2014).

[14] R. Orús and G. Vidal, "Infinite time-evolving block decimation algorithm beyond unitary evolution," *Physical Review B* **78**, 155117 (2008).

[15] I. P. McCulloch, "Infinite size density matrix renormalization group, revisited," *arXiv* , 0804.2509 (2008).

[16] J. Haegeman, T. J. Osborne, and F. Verstraete, "Post-Matrix Product State Methods: To tangent space and beyond," *Physical Review B* **88**, 075133 (2013).

[17] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical Optimization* (Springer-Verlag Berlin Heidelberg, 2006).

[18] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pizorn, H. Verschelde, and F. Verstraete, "Time-Dependent Variational Principle for Quantum Lattices," *Physical Review Letters* **107**, 070601 (2011).

[19] J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, and F. Verstraete, "Unifying time evolution and optimization with matrix product states," *Physical Review B* **94**, 165116 (2016).

[20] A. Milsted, J. Haegeman, T. J. Osborne, and F. Verstraete, "Variational matrix product ansatz for nonuniform dynamics in the thermodynamic limit," *Physical Review B* **88**, 155116 (2013).

[21] B. Buyens, J. Haegeman, K. Van Acoleyen, H. Verschelde, and F. Verstraete, "Matrix Product States for Gauge Field Theories," *Physical Review Letters* **113**, 091601 (2014).

[22] B. Buyens, J. Haegeman, H. Verschelde, F. Verstraete, and K. Van Acoleyen, "Confinement and string breaking for QED$_2$ in the Hamiltonian picture," (2015), arXiv:1509.00246 .

[23] D. Draxler, J. Haegeman, T. J. Osborne, V. Stojevic, L. Vanderstraeten, and F. Verstraete, "Particles, Holes, and Solitons: A Matrix Product State Approach," *Physical Review Letters* **111**, 020402 (2013).

[24] J. Haegeman, D. Draxler, V. Stojevic, J. I. Cirac, T. J. Osborne, and F. Verstraete, "Quantum Gross-Pitaevskii Equation," , 4 (2015), arXiv:1501.06575 .

[25] D. Draxler, J. Haegeman, F. Verstraete, and M. Rizzi, "Atomtronics - a continuous matrix product state approach," (2016), arXiv:1609.09704 .

[26] T. Koffel, M. Lewenstein, and L. Tagliacozzo, "Entanglement Entropy for the Long-Range Ising Chain in a Transverse Field," *Physical Review Letters* **109**, 267203 (2012).

[27] P. Hauke and L. Tagliacozzo, "Spread of correlations in long-range interacting quantum systems." *Physical review letters* **111**, 207202 (2013).

[28] J. C. Halimeh, V. Zauner-Stauber, I. P. McCulloch, I. de Vega, U. Schollwöck, and M. Kastner, "Prethermalization and Persistent Order in the Absence of a Thermal Phase Transition," (2016), arXiv:1610.01468 .

[29] J. C. Halimeh and V. Zauner-Stauber, "Enriching the dynamical phase diagram of spin chains with long-range interactions," (2016), arXiv:1610.02019 .

[30] D. P. Arovas, A. Auerbach, and F. D. M. Haldane, "Extended Heisenberg models of antiferromagnetism: Analogies to the fractional quantum Hall effect," *Physical Review Letters* **60**, 531 (1988).

[31] R. P. Feynman, "Atomic Theory of the Two-Fluid Model of Liquid Helium," *Physical Review* **94**, 262 (1954).

[32] S. M. Girvin, A. MacDonald, and P. Platzman, "Collective-Excitation Gap in the Fractional Quantum Hall Effect," *Physical Review Letters* **54**, 581 (1985).

[33] S. Östlund and S. Rommer, "Thermodynamic limit of the density matrix renormalization for the spin-1 Heisenberg chain," *Physical Review Letters* **75**, 13 (1995).

[34] S. Rommer and S. Östlund, "Class of ansatz wave functions for one-dimensional spin systems and their relation to the density matrix renormalization group," *Physical Review B* **55**, 2164 (1997).

[35] J. Haegeman, B. Pirvu, D. J. Weir, J. I. Cirac, T. J. Osborne, H. Verschelde, and F. Verstraete, "Variational matrix product ansatz for dispersion relations," *Physical Review B* **85**, 100408 (2012).

[36] V. Zauner, D. Draxler, L. Vanderstraeten, M. Degroote, J. Haegeman, M. M. Rams, V. Stojevic, N. Schuch, and F. Verstraete, "Transfer matrices and excitations with matrix product states," *New Journal of Physics* **17**, 053002 (2015).

[37] L. Vanderstraeten, J. Haegeman, F. Verstraete, and D. Poilblanc, "Quasiparticle interactions in frustrated Heisenberg chains," *Physical Review B* **93**, 235108 (2016).

[38] L. Vanderstraeten, F. Verstraete, and J. Haegeman, "Scattering particles in quantum spin chains," *Physical Review B* **92**, 125136 (2015).

[39] A. Milsted, J. Haegeman, and T. J. Osborne, "Matrix product states and variational methods applied to critical quantum field theory," *Physical Review D* **88**, 085030 (2013).

[40] J. Haegeman and F. Verstraete, *In preparation* (2016).