Optimization at the boundary of the tensor network variety

Christandl, Matthias; Gesmundo, Fulvio; França, Daniel Stilck; Werner, Albert H.

Published in:
Physical Review B

DOI:
10.1103/PhysRevB.103.195139

Publication date:
2021

Document version
Publisher's PDF, also known as Version of record

Document license:
Other

Citation for published version (APA):
Optimization at the boundary of the tensor network variety

Matthias Christandl, 1,* Fulvio Gesmundo, 1,2,† Daniel Stilck França, 1,4 and Albert H. Werner 1,3,§

1 QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark
2 Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany
3 NBIA, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark

(Received 6 July 2020; revised 16 March 2021; accepted 13 April 2021; published 18 May 2021)

Tensor network states form a variational ansatz class widely used, both analytically and numerically, in the study of quantum many-body systems. It is known that if the underlying graph contains a cycle, e.g., as in projected entangled pair states, then the set of tensor network states of given bond dimension is not closed. Its closure is the tensor network variety. Recent work has shown that states on the boundary of this variety can yield more efficient representations for states of physical interest, but it remained unclear how to systematically find and optimize over such representations. We address this issue by defining an ansatz class of states that includes states at the boundary of the tensor network variety of given bond dimension. We show how to optimize over this class in order to find ground states of local Hamiltonians by only slightly modifying standard algorithms and code for tensor networks. We apply this method to different models and observe favorable energies and runtimes when compared with standard tensor network methods.

DOI: 10.1103/PhysRevB.103.195139

I. INTRODUCTION

Tensor network states are quantum states obtained by contracting tensors, placed on vertices of a graph, according to the edges of the graph that identify indices of the tensors. They are featured in successful approaches to the study of quantum and classical many-body systems, and in particular, they provide an efficient ansatz class for quantum many-body states satisfying an area law [1–13]. For a fixed graph, the expressive power of the ansatz class is determined by an integer parameter $D$, called bond dimension. It is known that when the underlying graph contains cycles, the set of tensor network states of bond dimension at most $D$ is not closed in the standard Euclidean topology [14]; its closure is an algebraic variety called the tensor network variety. In [15] the authors show that states of physical interest may belong to the boundary of the variety, i.e., their bond dimension is strictly higher than $D$, but they can be approximated arbitrarily well by states of bond dimension $D$. Moreover, they showed how to exploit these approximate representations to obtain more efficient representations of the target state. However, it remained unclear how to make a variational method out of such states.

In this article, we define an ansatz class which includes states on the boundary. This class can describe certain states that arise as ground states of local Hamiltonians more efficiently than the standard tensor network ansatz class. Although by definition states on the boundary can be approximated using standard methods, we argue that this gives rise to ill-conditioned tensors. This leads to a slower convergence in variational methods and requires a precision in the computations that scales with system size. Thus, although tensor networks are often only an effective approximation to many-body states and variational methods converge exponentially fast to a good approximation to the ground-state interior, the convergence will be slow for boundary states. Such slow convergence has already been observed for frustrated systems on the Kagome lattice [16], suggesting that the underlying states ground states lie on the boundary. In contrast, our methods remain stable, and our examples show that they still converge fast to the target state while the traditional tensor network ansatz does not.

Standard numerical methods for tensor networks can be applied to the setting, with only slight modification to algorithms [10,13] and code: in particular, we show how to use the ansatz class to find better approximate representations of ground states compared to a standard tensor network ansatz. We demonstrate our methods numerically in different directions. Our ansatz class achieves smaller energies with less runtime for Hamiltonians for which there is a provable separation for the bond dimension required to represent the ground state. In addition, the methods achieve smaller energies than an MPS ansatz with the same number of parameters for the Heisenberg chain on a ring.

Thus, our findings indicate that this ansatz class can be used to obtain better numerical results for models of physical interest, besides further advancing our understanding of the geometry of tensor network states.

The paper is organized as follows. We will first discuss the structure of the states on the boundary. In the next step we define an ansatz class which includes those states. We then discuss how to perform computations with states from this
ansatz class before discussing the efficiency of their representation. The final part of the paper discusses the variation over this class and numerical examples comparing its performance to standard tensor network methods. A concluding discussion is provided at the end.

II. THE STRUCTURE OF STATES ON THE BOUNDARY

Let $G = (V, E)$ be a simple graph with a set of vertices $V$ and a set of edges $E$; let $L = |V|$ and let $D$ be a positive integer. For every edge $y \in E$, let $\Omega_D^{(y)} = \sum_{\alpha = 1}^{D} |\alpha, \alpha\rangle \in \mathbb{C}^D \otimes \mathbb{C}^D$ be the unnormalized maximally entangled state of dimension $D$. Define $|\Omega_D^{(y)}\rangle = \bigotimes_{y \in E} |\Omega_D^{(y)}\rangle$ and regard it as a vector in $\bigotimes_{v \in V} (\mathbb{C}^D)^{\otimes k_v}$, where $k_v$ is the degree of the vertex $v \in V$. Pictorially, this is the result of placing each $|\Omega_D^{(y)}\rangle$ on the corresponding edge of the graph and regarding the resulting tensor product $|\Omega_D\rangle$ as an unnormalized state on $L$ sites, corresponding to the vertices.

For every $v \in V$, let $A^v : (\mathbb{C}^D)^{\otimes k_v} \to \mathbb{C}^d$ be a linear map. Explicitly, write $A^v$ as a tensor

$$A^v = \sum_{i=1}^{d} \sum_{\alpha_d \in D} A_{i,d} |\alpha\rangle,$$

where $D^{\otimes k_v} = \{|\alpha_1, \ldots, \alpha_{k_v}\rangle : 1 \leq \alpha_j \leq D\}$.

Given a family of linear maps $A = (A^v : v \in V)$, define $\text{tns}^G(A) = (\bigotimes_{v \in V} A^v) |\Omega_D\rangle$. If $|\psi\rangle = \text{tns}^G(A)$ for some choice of linear maps $A$, we say that $(A^v : v \in V)$ is a tensor network state representation for $|\psi\rangle$ with respect to $G$.

Let

$$\text{TNS}_{D,d}^G = \{|\psi\rangle = \text{tns}^G(A) : A = (A^v : v \in V)\}.$$

For $|\psi\rangle \in (\mathbb{C}^d)^{\otimes L}$, the bond dimension of $|\psi\rangle$ (with respect to the graph $G$) is defined as

$$\text{bond}_G(|\psi\rangle) = \min \{D : |\psi\rangle \in \text{TNS}_{D,d}^G\}.$$

We refer, e.g., to [8] for details on this construction. It is straightforward to generalize this construction with different bond dimension $D$ at each edge, but we consider only the uniform case here for simplicity. If $G$ is a ring, a tensor network representation is called a matrix product state representation (MPS) [3] with periodic boundary conditions; if $G$ is a lattice, it is called a projected entangled pair state representation (PEPS) [8]. In addition, if the graph $G$ is regular and we have the same bond dimension at each edge, one can restrict to representations in which we apply the same map at each vertex. We call such tensor network representations translation invariant.

If the graph $G$ contains cycles, then $\text{TNS}_{D,d}^G$ is not closed, unless it coincides with the full $(\mathbb{C}^d)^{\otimes L}$ [14]. Write $\overline{\text{TNS}_{D,d}^G}$ for the closure of $\text{TNS}_{D,d}^G$; the set $\overline{\text{TNS}_{D,d}^G}$ is an algebraic variety and the (closure of the) difference $\overline{\text{TNS}_{D,d}^G} \setminus \text{TNS}_{D,d}^G$ is its algebraic boundary. Following [15], define the border bond dimension of a state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes L}$ as

$$\text{bond}_G^c(|\psi\rangle) = \min \left\{D : |\psi\rangle \in \overline{\text{TNS}_{D,d}^G} \right\}.$$

Clearly $\text{bond}_G(|\psi\rangle) \leq \text{bond}_G^c(|\psi\rangle)$ and $\text{bond}_G^c(-)$ is a lower semicontinuous function.

Following [17, Lemma 3.1.6.2], it is easy to show that a state at the boundary can be approximated along a rational curve; see Fig. 1. In other words, if $|\psi\rangle$ lies on the boundary, then there exists a family of local maps $(A^v : v \in V)$ whose entries are polynomials of degree at most $\delta$ in $\epsilon$, and an integer $a \geq 1$, such that

$$|\psi(\epsilon)\rangle = \text{tns}^G(A(\epsilon)) = e^{a \epsilon} |\psi\rangle + \sum_{j=1}^{c} e^{a \epsilon + j} |\tilde{\psi}_j\rangle;$$

note $|\psi\rangle = \lim_{\epsilon \to 0} e^{-\delta \epsilon} |\psi(\epsilon)\rangle$. Following [15], we say that the state $|\psi\rangle$ is a degeneration of $|\Omega_D\rangle$, and that it admits a border bond dimension $D$ representation; the integers $a$ and $\epsilon$ are called approximation degree and error degree, respectively [18,19].

Note that although $e^{-\delta \epsilon} |\psi(\epsilon)\rangle$ converges, in general there is no sequence of linear maps $(A^v : v \in V)$ such that $(\bigotimes_v A^v) |\Omega_D\rangle$; in this case, we say that the limit exists only nonlocally, and we observe that the terms $\bigotimes_v A^v$ of order lower than $a$ as $\epsilon$ vanish.

For a fixed $a$ and $\delta$, we will define an ansatz class that allows one to optimize over boundary states admitting a border bond representation with approximation degree $a$ and local maps of degree at most $\delta$ in $\epsilon$. We focus on the regime in which $a$ is of constant or linear order in the system size and $\delta$ is constant. To the best of our knowledge, this encompasses all known examples. Moreover, one can always assume $\delta \leq a$, as higher order terms in the local maps do not contribute to the limit.

III. WEIGHT STATES AND ANSATZ CLASS

One of the working horses of our ansatz class will be the following family of unnormalized states. Fix $a, \delta, L$ and define

$$|\chi_{a,\delta,L}\rangle = \sum_{i_1 + i_2 + \ldots + i_L = a} |i_1i_2, \ldots, i_L\rangle \in (\mathbb{C}^{\delta+1})^{\otimes L}. \quad (1)$$

We say that $|\chi_{a,\delta,L}\rangle$ is the weight state of weight $a$, length $L$, and local dimension $\delta + 1$. These weight states play a role in the differential geometry of homogeneous spaces (see, e.g., [20, chap. 12]) and are of weight zero under certain actions of $SL_2$ (see, e.g., [21]). Note that $L^{-1} |\chi_{1,1,L}\rangle$ is the $W$-state on $L$ sites.

FIG. 1. Schematic representation of a curve converging to the boundary of the variety. The gray surface represents TNS$_{D,d}^G$, the dotted black curve represents a sequence $e^{-\delta \epsilon} |\phi(\epsilon)\rangle$ converging to a state on the boundary, the thick black curve.
FIG. 2. Expansion of the TNS when contracted with the $|\chi_{1,1,3}\rangle$ state in terms of the local tensors corresponding to different degrees for $L=3$ and $a, b = 1$.

Suppose $|\psi\rangle$ is a state satisfying $\text{bond}_G(|\psi\rangle) \leq D$ and the degeneration has approximation degree $a$ and local degree at most $\delta$. Then there are augmented linear maps $B^c:(C^D)^{\otimes k_c} \otimes C^{b+1} \rightarrow C^d$ such that

$$|\psi\rangle = (\bigotimes B^n)[|\Omega_D^G\rangle \otimes |\chi_{a,b,L}\rangle].$$

Indeed, let $(A^v(\epsilon):v \in V)$ be the sequence of local maps realizing the degeneration for $|\psi\rangle$, i.e., $|\psi\rangle = \lim_{\epsilon \rightarrow 0} e^{-\epsilon\text{tns}}(A^v(\epsilon)) = \lim_{\epsilon \rightarrow 0} e^{-\epsilon}(\bigotimes_{v} A^v(\epsilon))|\Omega_D^G\rangle$.

Expanding the local maps in terms of $\epsilon$, one has $A^v(\epsilon) = \sum_{n=0}^{\infty} A^{v,n}\epsilon^n$, where $A^{v,n}:(C^D)^{\otimes k_v} \rightarrow C^d$ are linear maps. Define the augmented $B^c:(C^D)^{\otimes k_c} \otimes C^{b+1} \rightarrow C^d$ by

$$B^c = \sum_{n=0}^{\infty} A^{v,n} \otimes \eta.$$  \tag{2}

Specializing [18, Remark 9] to our setting, we have

$$e^{-\epsilon\text{tns}}(A^v(\epsilon)) = e^{-\epsilon}(\bigotimes_{v} A^v(\epsilon))|\Omega_D^G\rangle = (\bigotimes B^n)[|\Omega_D^G\rangle \otimes |\chi_{a,b,L}\rangle] + |Z(\epsilon)\rangle,$$

where $|Z(\epsilon)\rangle$ converges to 0 as $\epsilon \rightarrow 0$; in particular $|\psi\rangle = (\bigotimes_{v} B^n)|\Omega_D^G\rangle \otimes |\chi_{a,b,L}\rangle$. In other words, this construction results in only the term of $|\psi(\epsilon)\rangle = (\bigotimes_{v} A^v(\epsilon))|\Omega_D^G\rangle$ of degree $a$ in $\epsilon$.

This is illustrated in Fig. 2 for $a=1$ and $L=3$. Note that the pattern of the superpositions is mirrored in the entries of the weight state.

Given a family of linear maps $B = (B^c:(C^D)^{\otimes k_c} \otimes C^{b+1} \rightarrow C^d:v \in V)$, write

$$b\text{TNS}_D^G(B) = (\bigotimes_{v} B^c)[|\Omega_D^G\rangle \otimes |\chi_{a,b,L}\rangle]$$

and define the set

$$b\text{TNS}_D^G(a, b, d) = \{|\psi\rangle = b\text{TNS}_D^G(B) : B = (B^c:v \in V)\}.$$

The parameters $a$ and $d$ should be regarded as additional variational parameters that play a similar role to the bond dimension $D$, in the sense that higher $a$ and $d$ increase the expressive power of the class, but also the complexity of optimizing over it.

The discussion above implies that all states $|\psi\rangle$ satisfying $\text{bond}_G(|\psi\rangle) \leq D$ realized by a degeneration of approximation degree $a$ and local maps of degree at most $d$ are contained in $b\text{TNS}_D^G(a, b, d)$. However, $b\text{TNS}_D^G(a, b, d)$ contains states that do not necessarily arise as a degeneration. Indeed, the family of maps defined in (2) satisfies relations which ensure the lower terms (in $\epsilon$) of the degeneration vanish. On the other hand, $b\text{TNS}_D^G(a, b, d)$ is defined using arbitrary families of local maps ($B^c:v \in V$): as a result, it contains all states that arise as structured superpositions mirroring the entries of the weight states.

Optimizing solely over degenerations would entail optimizing over a subset of tensor network states satisfying $O(L^a)$ global polynomial equations of degree at most $a$ which encode the conditions that all terms of order strictly smaller than $a$ must vanish. Even just deciding whether the zeroth-order term vanishes in settings like PEPS on a square lattice is known to be an NP-complete problem [22,23]. Thus $b\text{TNS}_D^G(a, b, d)$ is a superset of $\text{TNS}_D^G(a, b, d)$ which also contains states on the boundary and on which it is possible to optimize with a small overhead when compared with standard tensor network methods and without having to impose global equations, as we will show in the following.

From a numerical point of view, following the standard tensor network methods, one is interested in parametrizing the ansatz class in terms of the local maps. The space of such maps has dimension $d \sum_{v} D^k$ in the standard setting of $\text{TNS}_D^G(a, b, d)$ and has dimension $d(\delta + 1) \sum_{v} D^k$ in the setting of $b\text{TNS}_D^G(a, b, d)$.

IV. PERFORMING COMPUTATIONS IN THE ANSATZ CLASS

An important feature of variational methods in the standard tensor network ansatz class is the possibility of computing expectation values of local observables for tensors in the class.

We evaluate the overhead to compute the expectation value if $|\psi\rangle$ has a representation in $b\text{TNS}_D^G(a, b, d)$ compared to the case where it has a representation in $\text{TNS}_D^G(a, b, d)$.

A standard interpolation argument (see, e.g., [24,25]) shows that if $|\psi\rangle$ is a state in $\text{TNS}_D^G(a, b, d)$ and it is realized via a degeneration of error degree $e$, then $|\psi\rangle \in \text{TNS}_D^G(a, b, d+e)$. This fact is used in [15] to propose a contraction technique for states in $\text{TNS}_D^G(a, b, d+e)$; however, even with constant local degrees $d$, the error degree $e$ depends linearly in the system size $L$, therefore the complexity of this technique grows with $L$.

Here we present two methods, which we call the MPS strategy and the border rank strategy, to perform the same type of computation in $b\text{TNS}_D^G(a, b, d)$ more efficiently. These methods apply to states of $b\text{TNS}_D^G(a, b, d)$ even if they are not elements of $\text{TNS}_D^G(a, b, d)$; the additional information that the state arises as a degeneration provides only a polynomial speed up in [15]. The MPS strategy consists of tensoring the standard tensor network representation of $|\Omega_D^G\rangle$ with an MPS representation of the desired weight state $|\chi_{a,b,L}\rangle$ resulting in a tensor network representation of $|\Omega_D^G\rangle \otimes |\chi_{a,b,L}\rangle$. The border rank strategy relies on the fact that the weight states have low border rank [15,26], a semicontinuous version of tensor rank which is discussed in detail in [27, Sec. I]. Which contraction technique is more advantageous is a subtle question and highly depends on the combinatorics of the underlying graph. The MPS strategy allows for taking the geometry of the graph and the contraction order into account. For relevant cases like PEPS on a two-dimensional lattice, this contraction method provides an overhead bounded polynomially in $a$ (and in particular independent from $L$) in the contraction.
complexity when compared to contracting a PEPS of the same bond dimension. In contrast, the border rank strategy is oblivious to the geometry of graph and the contraction order, but the overhead in the contraction complexity is bounded by $O((a + 1)^2 L)$ when compared to contracting a PEPS of the same bond dimension. We will now discuss the two strategies in more detail.

For the MPS strategy, note that $|\chi_{a, b, L}\rangle$ admits a representation as a matrix product state representation on an (open) chain of bond dimension $a + 1$; see [27, Lemma 3]. Let $P$ be a path on $G$ which visits each vertex of $G$ at least once. One can “lay” the MPS representation of $|\chi_{a, b, L}\rangle$ on the path $P$, resulting in a tensor network representation of $\Omega_D^{(G)} \otimes |\chi_{a, b, L}\rangle$. The bonds of the resulting representation, however, are multiplied by a factor $(a + 1)$ along each edge of $P$; in fact, the factor $(a + 1)$ appears once for each time the corresponding edge appears in $P$; depending on the geometry of the graph, this might significantly increase the bonds of the final tensor network representation. In addition, this procedure is not translation invariant. However, in relevant cases, such as PEPS on a lattice, the path $P$ can be chosen so that it does not involve the same edge more than once: as a result, in this case, $b\text{TN}_{a, b, d} \subseteq \text{TN}_{d(a+1)}^{G}$. In particular, in the case of the square lattice, contracting in $\text{TN}_{d(a+1)}^{G}$ has a complexity which is $(a + 1)^4$ times the complexity of contracting on $\text{TN}_{d}^{G}$. Compared with the interpolation method proposed in [15], the MPS strategy proposed here is more efficient when $a$ is constant in the system size, as the complexity of the interpolation method scales with $L$. The MPS strategy is illustrated in Fig. 3 for a square lattice.

As for the border rank strategy, [27, Lemma 2] shows that the weight states admit expressions of the form $|\chi_{a, b, L}\rangle = \lim_{\epsilon \to 0} |\chi_{a, b, L, \epsilon}\rangle$ where

$$|\chi_{a, b, L, \epsilon}\rangle = e^{-a} \sum_{i=1}^{a+1} |x_i(\epsilon)\rangle^{\otimes d}$$

is a sum of product vectors $|x_i(\epsilon)\rangle^{\otimes d}$ where $|x_i(\epsilon)\rangle$ is an element of $C^{a+1}$ depending linearly in $\epsilon$; in geometric language, this means that $|\chi_{a, b, L}\rangle$ has border rank (at most) $a + 1$. To get some intuition of why this is the case, note that we can also obtain the weight states as

$$|\chi_{a, a, L}\rangle = a! \frac{d^a}{d\epsilon^a} |\Gamma_a(\epsilon)\rangle |_{\epsilon=0},$$

where

$$|\Gamma_a(\epsilon)\rangle = (|0\rangle + \epsilon |1\rangle + \cdots + \epsilon^a |a\rangle)^{\otimes L}.$$ 

Thus, expressing the derivative in (3) as limit of a linear combination of $a + 1$ points on the curve $|\Gamma(\epsilon)\rangle$, one obtains the claim. For example, if $a = d = 1$, $|\chi_{1, 1, L}\rangle$ is the unnormalized $W$-state on $L$ parties, and one has $|\chi_{1, 1, L}\rangle = \lim_{\epsilon \to 0} \left( (|0\rangle + \epsilon |1\rangle)^{\otimes L} - |0\rangle^{\otimes L} \right)$.

This property allows one to compute expectation values using standard methods combined with an interpolation step. To see this, consider a converging sequence of states $|\psi(\epsilon)\rangle$ with $|\psi(\epsilon)\rangle = \lim_{\epsilon \to 0} |\psi(\epsilon)\rangle$ and fix an observable $O$: we provide a technique to compute the expectation value $\langle \phi(\epsilon) | O \phi(\epsilon) \rangle$ assuming that we can evaluate $|\phi(\epsilon)\rangle$ only at nonzero values of $\epsilon$. As observed already in [15], the function $p \epsilon \mapsto p(\epsilon) = e^{-2a} \langle \phi(\epsilon) | O \phi(\epsilon) \rangle$ is a polynomial of degree at most $2(Ld - a)$ in $\epsilon$, and its value at $\epsilon = 0$ coincides with the desired expectation value. One cannot evaluate $p(\epsilon)$ at $\epsilon = 0$ directly as the entries of the involved states diverge. However, via Lagrange interpolation, $p(0)$ is uniquely determined by the value of $p(\epsilon)$ at $2(Ld - a) + 1$ points.

The border rank strategy applies this method to the function $p(\epsilon)$ when $|\psi(\epsilon)\rangle = |\Omega_D^{(G)}\rangle \otimes |\chi_{a, b, L, \epsilon}\rangle$. For instance, for $a = d = 1$, $|\chi_{1, 1, L, \epsilon}\rangle$ is a superposition of two product states. We then have that $|\phi(\epsilon)\rangle$ can be written as the superposition of two tensor network states of bond dimension $D$. As a result, expanding $p(\epsilon)$ one sees that every single evaluation at $\epsilon \neq 0$ can be performed using standard tensor network methods.

We discuss this and the generalization to $a > 1$ in more detail in [27, Sec. III]; we point out here that it is possible to evaluate the expectation value of any observable on $|\psi(\epsilon)\rangle$ by contracting $2(Ld - a) + 1[(a + 1)^2]$ tensor networks states of bond dimension $D$.

V. SEPARATIONS IN THE EFFICIENCY OF REPRESENTATIONS

It is natural to ask to what extent the ansatz class provides more efficient representations of states of interest; in other words, we want to understand how large the gap between border bond dimension and bond dimension can be in the case of states of interest.

In the special case of matrix product states, i.e., when $G$ is a ring of length $L$, [27, Proposition 4] shows that $\text{TN}_{d}^{G} \subseteq \text{TN}_{d}^{G}$; in other words, $\text{bond}^{G}(|\psi\rangle) \leq \text{bond}^{G}(|\psi\rangle) \leq (\text{bond}^{G}(|\psi\rangle))^2$. An analogous result holds for any graph, but the exponent depends on the combinatorics of the graph, and in particular it may depend on $L$, making the upper bound exponential in the system size. For a general graph, as mentioned before, if $\text{bond}^{G}(|\psi\rangle) \leq D$ and the degeneration has error degree $\delta$, then $\text{bond}^{G}(|\psi\rangle) \leq (\delta + 1)D$.

Little is known about lower bounds on the possible separation. The difficulty in obtaining examples of large separations between bond and border bond dimension lies in the fact that essentially all techniques to prove lower bounds for bond dimension give, in fact, a lower bound on the border bond dimension: this is the case for the rank across a cut and
other methods relying on the evaluation of semicontinuous functions.

In the case where \( G \) is the ring with three nodes, consider \(|\text{ghz}_3\rangle\) to be the level three GHZ state on three parties; it has long been known [25] that \( \text{bond}^G(|\text{ghz}_3\rangle) = 2 \); in [15], the authors show \( \text{bond}^G(|\text{ghz}_3\rangle) = 3 \), showing a separation. An additional example is provided [15], where the possibility that the same separation holds also for the RVB state on the kagome lattice is discussed.

One can determine examples where the separation depends on the system size in the setting of translation invariant (TI) tensor networks. Consider translationally invariant matrix product states with periodic boundary conditions on an odd number of vertices \( L \). Define

\[
|\psi\rangle = \frac{1}{\sqrt{L}} \sum_{k=0}^{L-1} s^k |21010 \ldots 10\rangle \in (\mathbb{C}^3)^{\otimes L},
\]

where \( S \) is the shift operator. Consider the projector \( P \) acting on \( \mathbb{C}^3 \otimes \mathbb{C}^3 \) given by

\[
P = |01\rangle (01) + |10\rangle (10) + |02\rangle (02) + |21\rangle (21).
\]

Define a Hamiltonian \( H \) on a ring of size \( L \) as

\[
H = \sum_{i=0}^{L-1} \left[ (I - P_{i, i+1}) + \frac{1}{2L} |2\rangle \langle 2| \right],
\]

where \( I \) is the identity map, \( P_{i, i+1} \) acts as \( P \) in Eq. (5) on sites \( i, i + 1 \), and addition is taken modulo \( L \). From Proposition 6 in [27] we have that \( |\psi\rangle \) is the unique translationally invariant ground state of this 2-local Hamiltonian on the ring of length \( L \) for odd \( L \).

Moreover, note that \( \text{bond}^{\text{TI}-G}(|\psi\rangle) = 2 \) with a degeneration having \( a = 0 \). Indeed, consider the degeneration defined by the local map \( A(\epsilon): \mathbb{C}^2 \otimes \mathbb{C}^2 \rightarrow \mathbb{C}^3 \) defined as follows: write \( A(\epsilon) = A^0 + \epsilon A^1 \) with

\[
A^0 = |0\rangle \langle 0| + |1\rangle \langle 1|,
\]

\[
A^1 = |0\rangle \langle 1| + |1\rangle \langle 0|.
\]

It is easy to see that \(|\psi\rangle = \lim_{\epsilon \to 0} e^{-\epsilon |A|^{\otimes L} |\Omega^G_{D_j}|} \). In particular, \(|\psi\rangle \in \text{bTNS}^{\text{TI}-G}_{2,1.1.3} \). On the other hand, an adaptation of the results of [3] and [28] shows that \( \text{bond}^{\text{TI}-G}(|\psi\rangle) = \Omega(L^{1/3} / \log(L)) \). This implies system-size-dependent separations asymptotically for large enough \( L \). Furthermore, the results of [3] also imply that \( \text{bond}^{\text{TI}-G}(|\psi\rangle) > 2 \) for all ring sizes.

VI. VARIATIONAL METHODS

In this section, we discuss two widely used methods to find ground states of local Hamiltonians using tensor networks which can be adapted to the ansatz class: gradient descent [29], [10, Sec. 7.1] and imaginary time evolution [10, Sec. 7.2], which is sometimes called decimal block decimation method in this context. As in the case of the computation of expectation values, it is possible to adapt usual tensor network techniques and code to also optimize within the class with minimal effort.

First, we discuss gradient descent methods to find the ground state of a local Hamiltonian. In the standard tensor network setting, given a local Hamiltonian \( H \) and a desired bond dimension, one considers the energy of a tensor network state \( \text{tns}^\epsilon(A) \) as a function depending on the family of local maps \( A = \{ A_v : v \in V \} \). A gradient method computes the gradient of energy function and optimizes with respect to the linear maps. Often it is useful to consider the energy function as a function of only one linear map, optimize with respect to that, and then repeat the procedure alternating among all the linear maps; this method is called alternating gradient descent, and we refer to [10, Sec. 7.2] for more details. In the translation-invariant case, one considers the energy as a function of a single linear map and optimizes with respect to that, as in [29]. In the setting, the same method can be used considering the energy of a state \( \text{tns}^\epsilon(B) \) in \( \text{bTNS}^{\text{G}}_{D, a, d, d} \) as a function of the family of linear maps \( B = \{ B_1, \ldots, V, \ldots, B_v \} \), with \( B_v : \mathbb{C}D^{\otimes k_v} \otimes \mathbb{C}^{d+1} \rightarrow \mathbb{C}^d \).

We give some details to explain how to compute the gradient efficiently in the setting. Suppose we are computing the directional derivative of the function \( \langle \text{tns}^\epsilon(B) | O \text{ tns}^\epsilon(B) \rangle \) for some observable \( O \) in the direction of a vector \( V \) on the \( k \)th component of \( B \). Then

\[
\frac{\partial}{\partial V} \langle \text{tns}^\epsilon(B) | O \text{ tns}^\epsilon(B) \rangle
\]

\[
= 2 \text{Re} \langle \text{tns}^\epsilon(B^1, \ldots, B^{v-1}, V, B^{v+1}, \ldots, B^v) | O \text{ tns}^\epsilon(B) \rangle,
\]

which is the overlap of two states in \( \text{bTNS}^{\text{G}}_{D, a, d, d} \) with respect to the observable \( O \). As explained before, the calculation of the overlap can be done by combining standard contraction methods for tensor network states and one of the contraction strategies outlined before for states in \( \text{bTNS}^{\text{G}}_{D, a, d, d} \). In order to optimize the energy of some local Hamiltonian by alternating gradient descent, one computes the partial derivatives of the function

\[
f(B) = \frac{E(B)}{N(B)},
\]

where \( E(B) = \langle \text{tns}^\epsilon(B) | H \text{ tns}^\epsilon(B) \rangle \) and \( N(B) = \langle \text{tns}^\epsilon(B) | \text{tns}^\epsilon(B) \rangle \). The calculation of the directional derivatives of \( f \) reduces to the valuation of \( E \) and \( N \) and of the their derivatives, which can be done as explained above. Note that this method can be easily generalized to perform gradient descent in order to maximize the overlap with another state.

The second variational method that we consider is imaginary time evolution. Imaginary time evolution relies on the fact that, given a local Hamiltonian \( H \) and a state \(|\psi\rangle\) with nonzero overlap with the ground state, the state \( e^{-\beta H} |\psi\rangle \) converges to the ground state of \( H \) as \( \beta \) diverges to infinity. The method approximates the map \( e^{-\beta H} \) by a sequence of local maps through Trotterization. This sequence of local maps is applied to the current state, and it is easy to see that the resulting state has a larger overlap with the ground state than the initial one. This procedure can be easily implemented in \( \text{bTNS}^{\text{G}}_{D, a, d, d} \) by applying the local map to the physical indices of the underlying tensors.

More precisely, let \( e^{-\beta H_{v_1,v_2}} : (\mathbb{C}^d)^{\otimes 2} \rightarrow (\mathbb{C}^d)^{\otimes 2} \) be a two-local imaginary time evolution term acting on nodes \( v_1 \) and \( v_2 \) connected by an edge \( e_1 \). Consider a Schmidt decomposition
of the operator $e^{-\beta H_{12}}$:

$$e^{-\beta H_{12}} = \sum_{\ell=1}^{d^2} X_\ell \otimes Y_\ell.$$ 

Then the vector $e^{-\beta H_{12}} \text{bTNS}_G(B)$ can be obtained directly by enlarging the bond dimension across the edge $e_1$ by $d^2$ and updating the local maps $B^{v_1}$ and $B^{v_2}$. Thus, after applying one two-local map on each edge of the network, a state in $\text{bTNS}_G^{D,a,b,d}$ is mapped to a state in $\text{bTNS}_G^{D,a,b,d'}$. Hence, the bond dimension will (potentially) increase exponentially with the number of applied steps.

In the standard tensor network setting, this is addressed by truncating the Schmidt decomposition of $e^{-\beta H_{12}} \text{bTNS}_G(A)$. In the following, we show that the ansatz class supports a suitable truncation of the bond dimension after a certain number of iterations; in other words, we provide a method of finding an approximation of a state $|\psi_1\rangle \in \text{bTNS}_G^{D,a,b,d}$ by a state $|\psi_2\rangle \in \text{bTNS}_G^{D,a,b,d}$ with $D_2 < D_1$. The gradient descent methods discussed before can be used to this end, as we can optimize the overlap of the state $|\psi_1\rangle$ with respect to states in $\text{bTNS}_G^{D,a,b,d}$. However, a method to perform this truncation by solely considering the local maps $B^i$ is desirable, as in every iteration of gradient descent the whole state has to be contracted.

In the case of MPS with open boundary conditions, this problem is solved by first contracting two subsequent local maps $A^{v_1}, A^{v_2} : C^D \otimes C^D \rightarrow C^d$ along their common edge. Let $A^{v_12}$ be the map we obtain this way. Seeing it as matrix $A^{v_12} : C^D \otimes C^D \rightarrow C^d \otimes C^D$, we then perform a singular value decomposition and discard all singular values below a certain threshold. After the truncation, we obtain local maps $A^{v_1}, A^{v_2}$ with a smaller bond dimension on that edge, as desired. A Schmidt decomposition shows that this form of truncation is indeed optimal. However, in the case of tensor networks with cycles, the optimal truncation strategy is a subtle issue [30]. This is primarily due to the fact that edges on a cycle do not induce a bipartition of the state and, thus, a Schmidt decomposition of the state. Nevertheless, truncations based on a purely local SVD truncation, the simple-update algorithm [31], perform well in practice.

In principle, SVD-based truncation techniques can be readily applied to the $\text{bTNS}_G^{D,a,b,d}$ class. We can simply again contract $B^{v_1}$ and $B^{v_2}$ along the shared edge, obtaining a map $B^{v_12}$. We then perform a SVD and subsequent truncation of the matrix $B^{v_12} : C^D \otimes C^D \otimes (C^D)^{\otimes (k_1-1)} \rightarrow C^d \otimes C^D \otimes (C^D)^{\otimes (k_2-1)}$.

This would allow us to truncate the bond dimension of states in $\text{bTNS}_G^{D,a,b,d}$ locally. However, note that this truncation strategy did not take the special structure of the states in $\text{bTNS}_G^{D,a,b,d}$ into account, which may lead to suboptimal truncations. To illustrate this more concretely, take $A^{v_1}$, which we now set to

$$A^1 = |0\rangle \langle 0| \otimes (2) + |2\rangle \langle 2| \otimes (2) + 2|3\rangle \langle 3| \otimes (2).$$

Note that adding this extra subspace did not change the resulting state. The matrix $B^{v_12}$ we obtain for this representation is

$$B^{v_12} = \begin{pmatrix} 0 & |0\rangle \langle 0| & 0 & 0 & 0 & 0 \\ |1\rangle \langle 1| & 0 & 0 & 0 & |1\rangle \langle 0| \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A^1_1 \otimes A^1_2 \end{pmatrix} , \quad (8)$$

where each entry corresponds to a $4 \times 4$ matrix. Note that the upper left $3 \times 3$ block submatrix corresponds to the crossing of degree 0 terms, while the upper right and lower left correspond to degree 0 and 1 terms. Finally, the lower right corresponds to the crossing of degree 1 with 1. Observe that the lower right submatrix does not contribute to the resulting state when we contract with the state $|\chi_{1,1,1,1}\rangle$. This is because it corresponds to a term of degree 2. Thus, it is important to take this into account when performing a truncation.

The matrix in (8) is an extreme example of the issue that not all parts of the submatrix $C$ contribute equally to the state. Suppose that we wish to truncate this bond from three to two by performing a SVD of (8) and discarding the four smallest singular values. This truncation would lead us to discard the subspace spanned by $|0\rangle, |1\rangle$. But this choice of truncation would then result in the 0 state, as it preserved only the submatrix corresponding to degree 2 and set the other to 0.

We see we should not care if a truncation changes the lower right matrix substantially, while approximately preserving the other blocks. Furthermore, the upper right and lower left matrices contribute only once to the resulting state, while the submatrices corresponding to constant order degrees appear $L - 1$ times.

In light of this, let us now discuss an heuristic algorithm to perform truncations taking the special structure of the states in $\text{bTNS}_G^{D,a,b,d}$ into account. The first step is to recall the variational formulation of the truncation of singular values. For a matrix $C$ with SVD $C = UDV$ and some truncation rank $r$, denote by $\sqrt{D}$ a $n \times r$ matrix containing the square root of the largest $r$ singular values of $C$ on the diagonal. It is well known that $A = U \sqrt{D} V$ minimize

$$(A_r, B_r) \mapsto \|C - A_r B_r\|_F , \quad (9)$$

where $A_r$ and $B_r$ are rank $r$ matrices and $\| \cdot \|_F$ is the Frobenius norm. This is sometimes referred to as the Eckart-Young theorem. Our approach will be based on picking a different norm to perform this optimization depending on a parameter $0 < p \leq 1$. This parameter $p$ encodes by how much we want to suppress submatrices that correspond to higher orders. For a map $B^{v_12} : C^D \otimes C^D \otimes (C^D)^{\otimes k_1} \rightarrow C^D \otimes C^D \otimes (C^D)^{\otimes k_2}$, consider the weighted Frobenius norm:

$$\|B^{v_12}\|_{F,p}^2 = \sum_{\eta_1, \eta_2=0}^{d} p^{2(\eta_1 + \eta_2)} \sum_{i,j,\alpha,\beta} |B^{v_12}_{(\eta_1, i, \alpha, \beta), (\eta_2, j, \beta)}|^2 .$$

We see that submatrices that correspond to higher degrees contribute less to this norm. Indeed, if we denote by $B^{v_12}_\eta$ the submatrix for indices such that $\eta_1 + \eta_2 = \eta$ we have that

$$\|B^{v_12} - B^{v_12}_\eta\|_{F,p}^2 = \sum_{\eta=0}^{d} p^{2\eta} \|B^{v_12}_\eta - \tilde{B}^{v_12}_\eta\|_{F}^2 .$$
for any two matrices $B^{v_1, v_2}_q, \tilde{B}^{v_1, v_2}_q$. In the previous example with $\alpha, \beta = 1$, the Frobenius norm of the upper left matrix is multiplied by 1, the upper right and lower left by $p$, and the lower right by $p^2$. We see that this norm is less sensitive to the Frobenius distance of the submatrices corresponding to higher degree terms. Thus, truncating the bond dimension w.r.t. to this norm, truncation errors in higher degree terms will contribute less, as desired, and the parameter $p$ controls by how much. For example, in the case of $B^{v_1, v_2}$ in Eq. (8), we see that picking $p < 1/2$ is enough to ensure that we truncate the subspace spanned by $T$. [3] is discarded, as desired. Unfortunately, it is not clear at this point how to pick the parameter $p$ in an optimal fashion. One possibility is to perform the truncation for different values of $p$ and compare the resulting overlap with the original state.

Moreover, performing the truncation w.r.t. this norm can be easily implemented through standard SVD techniques combined with a rescaling step. Defining the map $\tilde{B}^{v_1, v_2}$ with entries

$$
\tilde{B}^{v_1, v_2}_{(i_1, i_2), (j_1, j_2)} = p^{|i_1 + j_2|} B^{v_1, v_2}_{(i_1, i_2), (j_1, j_2)},
$$

it is easy to see that $\| \tilde{B}^{v_1, v_2} \|_F = \| B^{v_1, v_2} \|_{F, p}$.

Let $A^{v_1}_r$ and $A^{v_2}_r$ be the matrices obtained by performing a truncated SVD of $\tilde{B}^{v_1, v_2}$ to rank $r$, where $r$ is the desired truncated bond dimension. By the Eckart-Young theorem, we have that $A^{v_1}_r, A^{v_2}_r$ minimize $\| \tilde{B}^{v_1, v_2} - A^{v_1}_r A^{v_2}_r \|_F$ among all matrices of rank at most $r$. It is then easy to see that by defining $A^{v_1}_r(i_1, i_2) = p^{-|i_1|} A^{v_1}_r(i_1, i_2)$ and $A^{v_2}_r(i_1, i_2)$ analogously, we have that $A^{v_1}_r, A^{v_2}_r$ minimize $\| A^{v_1, v_2} - A^{v_1}_r A^{v_2}_r \|_{F, p}$ in the spirit of Eq. (9).

Thus, by suitably rescaling the initial tensor, performing the usual truncation methods through an SVD, and then scaling back, it is possible to perform the truncation taking into account the contribution of each subspace and performing only local operations. Moreover, the computational complexity of performing all of these steps is comparable with that of locally truncating the bond dimension of a state in the tensor network setting, as they differ only by the rescaling steps.

**VII. NUMERICAL RESULTS**

In this section we compare numerical methods in the standard tensor network setting with the analogous methods in the ansatz class. Let us start by illustrating how we expect numerics to behave when trying to approximate a state on the boundary through states in the interior in an example. To this end, consider the state $|T\rangle \in (C^9)^{\otimes 3}$ given by

$$
|T\rangle = \frac{1}{\sqrt{17}} (|005\rangle + |016\rangle + |040\rangle + |126\rangle + |160\rangle + |227\rangle + |251\rangle + |262\rangle + |338\rangle + |373\rangle + |384\rangle + |430\rangle + |501\rangle + |632\rangle + |703\rangle + |714\rangle + |824\rangle).
$$

(10)

In [27, Sec. II B], we show that $|T\rangle$ satisfies

$$
3 = \text{bond}^C(|T\rangle) < \text{bond}^C(T),
$$

where $C_1$ is the ring on three vertices. Moreover, the approximation degree of the border bond dimension representation is $a = 1$.

Figure 4 records the results of performing alternating gradient descent to approximate the state $|T\rangle$ in different ansatz classes. More precisely, we are performing alternating gradient descent by optimizing the overlap with the bond dimension 17 representation of the state $|T\rangle$ given by the decomposition in Eq. (10). We do so because we are unaware of a better representation of $|T\rangle$, although [27, Lemma 5] guarantees that $\text{bond}^C(|T\rangle) \leq 3$. Although the state $|T\rangle$ can be approximated arbitrarily well with states of bond dimension 3, we observe that the convergence is slow. Although by [27, Lemma 5] the maximal bond dimension is $D = 9$, we already observe a fast convergence with $D = 5$. On the other hand, since $\text{bond}^C(|T\rangle) = 3$ with a degeneration having

**FIG. 4.** Logarithm of the distance to $|T\rangle$ after number of iterations for alternating gradient descent and different bond dimensions on the ring.

**FIG. 5.** Energy achieved for a TI-TNS ansatz compared to bTNS for the Hamiltonian defined in Eq. (6). The ground-state energy is normalized to 0, and the ring is of size 11. Note that the time taken for one iteration of bTNS for $D = 2$ is roughly half that of one in TNS with $D = 5$. We picked the best energy value achieved over 40 random starting points for each curve. We note a faster convergence for bTNS.
the isotropic Heisenberg model for a translation invariant ansatz and a faster convergence for bTNS.

FIG. 6. Energy achieved with the imaginary time evolution for the isotropic Heisenberg model for a translation invariant ansatz and $L = 12$. We normalized the ground-state energy to be 0 and picked $p = 0.9$ for the truncation algorithm of the degeneration. We observe a faster convergence for bTNS.

approximation degree $1$, we have $|T \rangle \in \text{bTNS}^{\text{CI},1,1,0}_3$; indeed, performing gradient descent in this class, we already observe fast convergence, which can be observed only for bond dimension $5$ in the standard tensor network setting.

Figure 5 records the results in the translation invariant setting of the Hamiltonian described in (6). Fix a ring of length $L = 9$. The ground state $|\psi \rangle$ of (4) satisfies $\text{bond}^{\text{CI},T}_1(|\psi \rangle) = 2$ with approximation degree $a = 1$, whereas $\text{bond}^{\text{CI},T}_1(|\psi \rangle) > 2$. Gradient descent on bTNS$^{\text{CI},T}_1,1,3$ outperforms the standard matrix product state ansatz for $D = 2, 4$, in the sense that it obtains lower energies, giving an indication of the numerical viability of our method. Moreover, an iteration of gradient descent with $D = 4$ in TNS takes roughly double the time of one in bTNS with $D = 2$ and $a = 1$.

In order to benchmark the ansatz class and the algorithms for imaginary time evolution, we performed the method on the isotropic Heisenberg model on a ring of size $L$, comparing the results achieved by the standard matrix product state method with the ones obtained in the ansatz class. The Hamiltonian of the isotropic Heisenberg model on the ring $C_L$ is given by

$$H = \sum_{k=1}^{L} (\sigma_k^x \sigma_{k+1}^x + \sigma_k^y \sigma_{k+1}^y + \sigma_k^z \sigma_{k+1}^z),$$

where $\sigma_k^j$ are the Pauli matrices acting on site $k$. The number of local parameters for a state in our ansatz class with bond dimension $D$ is $(a + 1)D^2$, so we compare states in our class to MPS with bond dimension $\lceil \sqrt{a + TD} \rceil$. We used imaginary time evolution methods and a translationally invariant ansatz to find the ground state and picked the initial tensor at random. We see in Fig. 6 that states in our class converge faster. Although we do not have a provable separation in the required bond dimension for this model, these results indicate the potential of our method for models of physical interest.

VIII. CONCLUSION

We presented numerical evidence showing that degenerations of tensor network states are a valuable tool for both the numerical and analytical study of the tensor network ansatz class. Many directions remain to be explored in future work, from both the analytical and numerical point of view. On the numerical side, it will be interesting to go beyond one-dimensional systems and see how our enlarged ansatz class performs for higher dimensional lattices. In particular, we believe that larger separations in complexity can be observed for PEPS, even when not restricting to translationally invariant systems. Moreover, we believe that our ansatz class provides a natural framework to study excited states. From the analytical and algorithmic point of view, our work raises many questions. Just to name a few, it is natural to ask about normal forms, the scaling of correlations, and how degenerations behave in the thermodynamical limit.

ACKNOWLEDGMENTS

This work was supported by VILLUM FONDEN via the QMath Centre of Excellence under Grant No. 10059 and the European Research Council (Grant agreement No. 818761). A.H.W. thanks the VILLUM FONDEN for its support with a Villum Young Investigator Grant (Grant No. 25452).

optimization at the boundary of the tensor network states, Quantum Info. Comput. 12, 346 (2012).


