Neural network representation of tensor network and chiral states

Yichen Huang (黄溢辰)

Institute for Quantum Information and Matter, California Institute of Technology
Pasadena, California 91125, USA
ychuang@caltech.edu

Joel E. Moore

Department of Physics, University of California, Berkeley
Materials Sciences Division, Lawrence Berkeley National Laboratory
Berkeley, California 94720, USA
jemoore@berkeley.edu

January 24, 2017

Abstract

We study the representational power of a Boltzmann machine (a type of neural network) in quantum many-body systems. We prove that any (local) tensor network state has a (local) neural network representation. The construction is almost optimal in the sense that the number of parameters in the neural network representation is almost linear in the number of nonzero parameters in the tensor network representation. Despite the difficulty of representing (gapped) chiral topological states with local tensor networks, we construct a quasi-local neural network representation for a chiral p-wave superconductor. This demonstrates the power of Boltzmann machines.

1 Introduction to Boltzmann machines

This section provides a minimum background for those people with no prior knowledge of Boltzmann machines [1]. The goal is to motivate the definition of neural network states in the next section, rather than a general-purpose introduction from the perspective of machine learning.

Formally, a Boltzmann machine is a type of stochastic recurrent neural network. In the language of physicists, it is a classical Ising model on a weighted undirected graph. Each vertex (also known as a unit or a neuron) of the graph carries a classical Ising variable $s_j = \pm 1$ and a local field $h_j \in \mathbf{R}$, where j is the index of the vertex. For a reason that will soon be clear, the set of vertices is divided into the disjoint union of two subsets V and H so that $|V| + |H|$ is the total number of units. Vertices in V are called visible units, and those in H are called hidden units. For notational simplicity, we assume that visible units have small indices $1, 2, \ldots, |V|$, and hidden units have large indices $|V|+1, |V|+2, \ldots, |V|+|H|$. Each edge of the graph carries a weight $w_{jk} \in \mathbf{R}$ that describes the interaction between s_j and s_k. The energy of a configuration is given by

$$E(\{s_j\}) = \sum_j h_j s_j + \sum_{j,k} w_{jk} s_j s_k. \tag{1}$$

A restricted Boltzmann machine [21] is a Boltzmann machine on a bipartite graph, i.e., $w_{jk} \neq 0$ only if the edge (i, k) connects a visible unit with a hidden unit.

At thermal equilibrium, the configurations follow the Boltzmann distribution. Without loss of generality, we fix the temperature $T = 1$. Let

$$Z = \sum_{\{s_j\} \in \{\pm 1\}^{\otimes n}} E(\{s_j\}) \tag{2}$$

be the partition function. The probability of each configuration $\{s_j\}$ is given by $e^{-E(\{s_j\})}/Z$. Furthermore, the probability of each configuration $\{s_{j \leq |V|}\}$ of visible units is the marginal probability obtained by summing over hidden units:

$$P\{s_{j \leq |V|}\} = \frac{1}{Z} \sum_{\{s_{j > |V|}\} \in \{\pm 1\}^{\otimes |H|}} e^{-E(\{s_j\})}. \tag{3}$$

The process of training a Boltzmann machine is specified as follows. The input is a (normalized) probability distribution Q over the configurations of the visible units. (Here we assume, for simplicity, that Q is given. In practice, Q may not be given but we are allowed to sample from Q.) The goal is to adjust the weights w_{jk} such that the probability distribution P (3) best approximates (as measured by Kullback-Leibler divergence or some other distance function) the desired distribution Q. This is a variational minimization problem. We minimize a given objective function of P with respect to the ansatz (3), in which the weights are variational parameters. Note that Monte Carlo techniques are usually used in training a Boltzmann machine.

An interesting question is whether an arbitrary Q can be exactly represented as (3). Due to the positivity of exponential functions, it is easy to see that the answer is no if the probability of some configuration is zero. Nevertheless,

Theorem 1 (Le Roux and Bengio [17]). *Any probability distribution Q can be arbitrarily well approximated by an restricted Boltzmann machine, provided that the number $|H|$ of hidden units is the number of configurations with nonzero probability. In general, $|H| \leq 2^{|V|}$.*

This result has been slightly improved [18]. Indeed, a simple counting argument suggests that an exponential number of hidden units is necessary in general. As a function from $\{\pm 1\}^{\otimes |V|}$ to $[0, 1]$ with one constraint (normalization), the probability distribution Q has $2^{|V|} - 1$ degrees of freedom. Therefore, a sufficiently well approximation of Q requires an exponential number of bits.

2 Neural network states

Carleo and Troyer [6] developed a minimum extension of a Boltzmann machine to the quantum world. In this extension, each vertex still carries a classical spin (bit). This is very different from another extension [2], in which each vertex carries a quantum spin (qubit). It should be clear that the latter extension is more "quantum," for it is operated by quantum computers. In the former extension, we use classical computers to simulate quantum many-body systems.

Expanded in the computational basis $\{\pm 1\}^{\otimes |V|}$, a quantum state ψ of $|V|$ qubits can be viewed as a function from $\{\pm 1\}^{\otimes |V|}$ to $\bar{B}(0, 1)$ with one constraint (normalization), where $\bar{B}(0, 1)$ denotes the region $|z| \leq 1$ in the complex plane. Recall that the distribution Q in the proceeding section is characterized by the probability of each configuration. In comparison, the state ψ is characterized by the probability amplitude of each configuration. This analog between classical probability distributions and quantum states motivates the following ansatz dubbed neural network states.

Consider a graph as in the proceeding section. The visible units correspond to physical qubits, and hidden units are auxiliary degrees of freedom that need to be summed over. The local field

$h_j \in \mathbf{C}$ at each vertex and the weight $w_{jk} \in \mathbf{C}$ carried by each edge are promoted to complex numbers because probability amplitudes are generally complex. A neural network state based on a Boltzmann machine is given by

$$\psi = \frac{1}{Z} \sum_{\{s_1, s_2, \ldots, s_{|V|+|H|}\} \in \{\pm 1\}^{\otimes(|V|+|H|)}} e^{-\sum_j h_j s_j - \sum_{j,k} w_{jk} s_j s_k} \left| \left\{ s_1, s_2, \ldots, s_{|V|} \right\} \right\rangle, \tag{4}$$

where Z is a normalization prefactor. Note that summing over hidden units is very different from taking partial trace. Similarly, a neural network state based on a restricted Boltzmann machine is given by (4) on a bipartite graph as described in the proceeding section.

A very minor modification of the proof of Theorem 1 leads to

Corollary 1. *Any n-qubit quantum state ψ can be arbitrarily well approximated by a neural network state based on a restricted Boltzmann machine with $w_{jk} \in \mathbf{R}, h_j \in \mathbf{C}$, provided that the number $|H|$ of hidden units is the number of configurations with nonzero probability amplitude. In general, $|H| \le 2^n$.*

The formalism of neural network states developed thus far applies to general quantum many-body systems. We now consider the situation that the qubits are arranged on a lattice. The lattice allows us to introduce the notion of locality [15] (with respect to the shortest path metric), which underlies almost all successful methods of simulating quantum lattice systems. Hence, it is desirable to incorporate locality into the neural network. To do this, we define a position for each unit. Let each site of the lattice carry a visible unit and some hidden units. We require that $w_{jk} \ne 0$ only if the units j and k are close to each other. As an example, Deng et al. [8] showed that the ground state of toric code in two dimensions can be exactly represented as a local neural network state based on a restricted Boltzmann machine.

Using variational quantum Monte Carlo techniques ("quantum Monte Carlo" is not a quantum algorithm; rather, it is just a classical Monte Carlo algorithm applied to quantum systems), Carleo and Troyer [6] performed practical calculations for quantum lattice systems in one and two spatial dimensions. For the models they studied, they observed that variational approaches based on neural networks are at least as good as those based on tensor networks. It seems worthwhile to devote more study to the practical performance of neural network states.

3 Tensor network states

This section begins with a very brief introduction to tensor network states following the presentation in Ref. [12], §6.3 or [16], §2.3. Then, it will be clear almost immediately that any (local) tensor network state has a (local) neural network representation.

Let $[m] = \{1, 2, \ldots, m\}$ be the set of natural numbers from 1 to m. A k-dimensional tensor M is a multivariable function $M : [d_1] \times \cdots \times [d_k] \to \mathbf{C}$, and $D = \max_{j \in [k]} d_j$ is called the bond dimension. It is easy to see that any tensor M can be reshaped to a n-dimensional ($n = \sum_j \lceil \log_2 d_j \rceil$ with $\lceil \cdot \rceil$ the ceiling function) tensor N of bond dimension 2 by representing every input variable in binary. Moreover, there is a straightforward way to identify N with an n-qubit unnormalized state expanded in the computational basis:

$$\psi = \sum_{\{s_j\} \in \{\pm 1\}^{\otimes n}} N \left(\frac{s_1 + 3}{2}, \frac{s_2 + 3}{2}, \ldots, \frac{s_n + 3}{2} \right) |\{s_j\}\rangle. \tag{5}$$

Thus, a neural network state with $|V|$ visible units is a $|V|$-dimensional tensor of bond dimension 2. In particular, each visible unit corresponds to an input variable to the tensor. As a restatement of Corollary 1,

Corollary 2. *Any tensor M can be arbitrarily well approximated by a restricted Boltzmann machine with $w_{jk} \in \mathbf{R}, h_j \in \mathbf{C}$, provided that the number $|H|$ of hidden units is the number of nonzero elements in M. In general, $|H| \leq \prod_{j \in [k]} d_j$.*

Informally, tensor contraction is defined as follows. Suppose we have two 3-dimensional tensors $M(j_1, j_2, j_3), M'(j_1', j_2', j_3')$ of shapes $d_1 \times d_2 \times d_3, d_1' \times d_2 \times d_3'$, respectively. Their contraction on the middle indices is a 4-dimensional tensor

$$\sum_{j \in [d_2]} M(j_1, j, j_3) M'(j_1', j, j_3'). \tag{6}$$

Similarly, suppose we have two Boltzmann machines with units $\{s_j\} = V \cup H, \{s_j'\} = V' \cup H'$, weights w_{jk}, w_{jk}', and local fields h_j, h_j', respectively. Their contraction on the first c visible units is given by (i) identifying s_j with s_j' for $j = 1, 2, \ldots, c$; (2) summing over s_j for $j = 1, 2, \ldots, c$:

$$\sum_{\{s_1, s_2, \ldots, s_{|V|+|H|}\} \in \{\pm 1\}^{\otimes(|V|+|H|)}, \{s_1', s_2', \ldots, s_{|V'|+|H|}'\} \in \{\pm 1\}^{\otimes(|V'|+|H'|)} \text{ s. t. } s_1 = s_1', s_2 = s_2', \ldots, s_c = s_c'} \cdots, \tag{7}$$

where $\cdots$ is given by

$$e^{-\sum_j h_j s_j - \sum_{j,k} w_{jk} s_j s_k - \sum_j h_j' s_j' - \sum_{j,k} w_{jk}' s_j' s_k'} |\{s_{c+1}, s_{c+2}, \ldots, s_{|V|}\}\rangle \otimes |\{s_{c+1}', s_{c+2}', \ldots, s_{|V'|}'\}\rangle. \tag{8}$$

Note that only visible units are allowed to be contracted on, and such visible units become hidden after the contraction. Therefore, the Boltzmann machine after contraction has $|V| + |V'| - 2c$ visible and $|H| + |H'| + c$ hidden units.

A multi-qubit state ψ is a tensor network state if its tensor representation N (5) can be obtained by contracting a network of C-dimensional tensors, where C is a small absolute constant. The bond dimension D of a tensor network is defined as the maximum bond dimension of each constituting tensors in the network.

Theorem 2. *Any tensor network state ψ can be arbitrarily well approximated by a neural network state based on a Boltzmann machine with $w_{jk} \in \mathbf{R}, h_j \in \mathbf{C}$, provided that the number $|H|$ of hidden units is sufficiently large. Let $\emptyset(T_j)$ be the number of nonzero elements in a constituting tensor T_j, and $\#(T_j)$ be the number of elements. It suffices that*

$$|H| = \sum_{j=1}^n (\emptyset(T_j) + \log_2 \#(T_j) + O(1)), \tag{9}$$

where n is the number of constituting tensors. Furthermore, the number of parameters in the neural network representation is upper bounded by

$$\sum_j \emptyset(T_j) \log_2 \#(T_j) + \text{less significant terms.} \tag{10}$$

The neural network is local (translationally invariant) if the tensor network is local (translationally invariant).

4

Proof. We first represent each T_j with a restricted Boltzmann machine as in Corollary 2, and then contract the restricted Boltzmann machines in the same way as T_j's are contracted. Note that the Boltzmann machine after contraction is generically not restricted. The first term on the right-hand side in (9) is the total number of hidden units in all constituting restricted Boltzmann machines, and the other terms are responsible for the production of hidden units in the contraction process. Eq. (10) is the total number of parameters in all constituting restricted Boltzmann machines, and the contraction process does not introduce any new parameters. It is obvious that the construction is locality- and translational-invariance-preserving. $\square$

This result is almost optimal in the sense that the number of parameters in the neural network representation is at most a logarithmic (in the bond dimension) multiple of the number of nonzero parameters in the tensor network representation. In cases that the tensors are sparse (possibly due to the presence of symmetries), this conversion to neural network states has the desirable property of automatically compressing the representation. We close this section by specializing Theorem 2 to matrix product states [19].

Corollary 3. *To leading order, any n-qubit matrix product state of bond dimension D has a neural network representation with $2nD^2$ hidden units and $4nD^2 \log_2 D$ parameters.*

4 Chiral topological states

It seems difficult to obtain a local tensor network representation for gapped chiral topological states. Early attempts did not impose locality [13] or just target at expectation values of local observables rather than the wave function [4]. Recent progress [24, 10] shows that local tensor networks can describe chiral topological states, but the examples there are all gapless. Indeed, Dubail and Read [10] even proved a no-go theorem, which roughly states that for any chiral local free-fermion tensor network state, any local parent Hamiltonian is gapless. Here we construct a Boltzmann machine that approximates the unique ground state of a (gapped) chiral p-wave superconductor. The error of approximation is an inverse polynomial in the system size, and the neural network is quasi-local in the sense that the maximum distance of connections between units is logarithmic in the system size. This example explicitly demonstrates the power of Boltzmann machines.

From now on, we consider fermionic systems, for which some modifications are necessary to account for the exchange statistics of fermions. In particular, each vertex carries a Grassmann number ξ_j rather than an Ising variable s_j, and the sum over Ising variables in hidden units are replaced by the Grassmann integral. We work in the second quantization formalism. Let $c_j, c_j^\dagger$ with $1 \leq j \leq |V|$ be the fermionic annihilation and creation operators. Define $p(\xi_1, \xi_2, \ldots, \xi_{|V|})|0\rangle = p(c_1^\dagger, c_2^\dagger, \ldots, c_{|V|}^\dagger)|0\rangle$ for any polynomial p in $|V|$ variables, where $|0\rangle$ is the vacuum state with no fermions. As an analog of (4),

$$\psi = \frac{1}{Z} \left(\int \mathrm{d}\xi_{|V|+1} \mathrm{d}\xi_{|V|+2} \cdots \mathrm{d}\xi_{|V|+|H|} e^{-\sum_{j,k} w_{jk} \xi_j \xi_k} \right) |0\rangle \tag{11}$$

is a fermionic neural network state. Note that we have set $h_j = 0$. This class of states appeared previously in Ref. [10].

One of the simplest examples of a chiral topological phase is the $p + ip$ superconductor [20]. For concreteness, we consider the lattice model in Ref. [5]. Let $c_{\vec{x}}, c_{\vec{x}}^\dagger$ be the fermionic annihilation and creation operators at the site $\vec{x} \in \mathbf{Z}^2$ on a two-dimensional square lattice. Let $\vec{i} = (1, 0)$ and

$\vec{j} = (0,1)$ be the unit vectors in the x and y axes, respectively. The Hamiltonian is

$$H = \sum_{\vec{x} \in \mathbf{Z}^2} \left(c_{\vec{x}+\vec{i}}^\dagger c_{\vec{x}} + c_{\vec{x}+\vec{j}}^\dagger c_{\vec{x}} + c_{\vec{x}+\vec{i}}^\dagger c_{\vec{x}}^\dagger + i c_{\vec{x}+\vec{j}}^\dagger c_{\vec{x}}^\dagger + \text{h.c.} \right) - 2\mu \sum_{\vec{x} \in \mathbf{Z}^2} c_{\vec{x}}^\dagger c_{\vec{x}}, \quad \mu \in \mathbf{R}. \tag{12}$$

Let $\vec{k} = (k_x, k_y)$ be the lattice momentum, and $\int_{\text{BZ}} d\vec{k}$ the integral over the Brillouin zone $(-\pi, \pi]^{\otimes 2}$. To slightly simplify the presentation, we will be sloppy about unimportant overall prefactors in the calculations below. The Fourier transform

$$c_{\vec{x}}^\dagger = \int_{\text{BZ}} d\vec{k} e^{-i\vec{k}\cdot\vec{x}} c_{\vec{k}}^\dagger, \quad c_{\vec{x}} = \int_{\text{BZ}} d\vec{k} e^{i\vec{k}\cdot\vec{x}} c_{\vec{k}} \tag{13}$$

leads to

$$H = \int_{\text{BZ}} d\vec{k} H_{\vec{k}}, \quad H_{\vec{k}} = \Delta_{\vec{k}} c_{\vec{k}}^\dagger c_{-\vec{k}}^\dagger + \Delta_{\vec{k}}^* c_{-\vec{k}} c_{\vec{k}} + 2M_{\vec{k}} c_{\vec{k}}^\dagger c_{\vec{k}} \tag{14}$$

in the momentum space, where

$$\Delta_{\vec{k}} = \sin k_x + i \sin k_y, \quad M_{\vec{k}} = \cos k_x + \cos k_y - \mu. \tag{15}$$

The quasi-particle spectrum is given by

$$E_{\vec{k}} = 2\sqrt{|\Delta_{\vec{k}}|^2 + M_{\vec{k}}^2} = 2\sqrt{\sin^2 k_x + \sin^2 k_y + (\cos k_x + \cos k_y - \mu)^2}. \tag{16}$$

Hence, H is gapless for $\mu = 0, \pm 2$ and gapped otherwise. The unique (unnormalized) ground state of H is

$$\psi = e^{\frac{1}{2} \int_{\text{BZ}} d\vec{k} (v_{\vec{k}}/u_{\vec{k}}) c_{\vec{k}}^\dagger c_{-\vec{k}}^\dagger} |0\rangle, \tag{17}$$

where $|0\rangle$ is the vacuum state, and

$$v_{\vec{k}} = \Delta_{\vec{k}}, \quad u_{\vec{k}} = \sqrt{|\Delta_{\vec{k}}|^2 + M_{\vec{k}}^2} - M_{\vec{k}} \tag{18}$$

so that $v_{\vec{k}} = -v_{-\vec{k}}$, $u_{\vec{k}} = u_{-\vec{k}}$. It is not difficult to see that the model (12) is a trivial superconductor for $|\mu| > 2$. It represents topological superconductors with opposite chirality for $-2 < \mu < 0$ and $0 < \mu < 2$, respectively.

Indeed, a state of the form (17) has an exact local neural network representation (11) [10] if $u_{\vec{k}}, v_{\vec{k}}$ are constant-degree trigonometric polynomials, i.e., polynomials in $e^{\pm i k_x}, e^{\pm i k_y}$. In particular, we have each site $\vec{x}$ carry a visible and a hidden unit. To simplify the notation, the Grassmann number in the visible unit is identified with $c_{\vec{x}}^\dagger$, and that in the hidden unit is denoted by $\xi_{\vec{x}}$. The Fourier transform of Grassmann numbers is defined as

$$\xi_{\vec{x}} = \int_{\text{BZ}} d\vec{k} e^{-i\vec{k}\cdot\vec{x}} \xi_{\vec{k}}. \tag{19}$$

In the momentum space, it is easy to see that the state (17) can be represented as

$$\psi = \int [d\xi_{\vec{k}}] e^{\int_{\text{BZ}} d\vec{k} \left(v_{\vec{k}} \xi_{-\vec{k}} c_{\vec{k}}^\dagger - \frac{1}{2} \xi_{-\vec{k}} v_{-\vec{k}} u_{\vec{k}} \xi_{\vec{k}} \right)} |0\rangle, \tag{20}$$

where $\int [d\xi_{\vec{k}}]$ denotes the integral over all Grassmann numbers in the momentum space with a proper measure. Transforming to the real space, $[d\xi_{\vec{k}}]$ becomes $\prod_{\vec{x} \in \mathbf{Z}^2} \xi_{\vec{x}}$, and the exponent in parentheses is local because $u_{\vec{k}}, v_{\vec{k}}$ are trigonometric polynomials (the maximum distance of connection between units is proportional to the degree of the trigonometric polynomials). Thus, (20) reduces to (11)

with additional properties: (i) the neural network representation is translationally invariant; (ii) there are no connections between visible units.

Generically, exact representations are too much to ask for; hence approximate representations are acceptable. Furthermore, we usually have to work with a finite system size in order to rigorously quantify the error of the approximation and the succinctness of the representation. We now justify these statements with an example. Matrix product states constitute an excellent ansatz for ground states in one-dimensional gapped systems [14]. A generic gapped ground state in a chain of n spins (i) cannot be exactly written as a matrix product state of bond dimension $e^{o(n)}$; (ii) is not expected to be approximated (in the sense of 99% fidelity) by a matrix product state with bond dimension $O(1)$ because errors may accumulate while we truncate the Schmidt coefficients across every cut [23]; (iii) can be approximated by a matrix product state with bond dimension $n^{o(1)}$ [3].

The goal is to construct an as-local-as-possible neural network representation (11) for the ground state (17) of the model (12). For concreteness, we consider $0 < \mu < 2$ so that the only common zero of $u_{\vec{k}}$ and $v_{\vec{k}}$ is at $\vec{k} = \vec{0}$. It should be clear the same result applies to other values of μ provided that the system is gapped. The main idea is to approximate $u_{\vec{k}}$ (18) with a trigonometric polynomial. The approximation error is exponentially small in the degree of the polynomial because $u_{\vec{k}}$ is a real analytic function of $\vec{k}$. The locality of the neural network is due to the smallness of the degree.

We now provide the details of the construction. Consider the model (12) on a finite square lattice of size $n \times n$ with periodic boundary conditions. Assume n is odd. This slightly simplifies the analysis, but it should be clear that a minor modification of the analysis leads to the same result for even n. We abuse the notation by letting BZ denote the set of viable points $\{0, \pm 2\pi/n, \dots, \pm(n-1)\pi/n\}^{\otimes 2}$ for lattice momenta in the Brillouin zone, and $\frac{1}{2}\mathrm{BZ} = \{\vec{k} \in \mathrm{BZ} | k_x > 0 \vee (k_x = 0 \wedge k_y > 0)\}$ be the "right half" of BZ. The Fourier transform (13) becomes discrete:

$$c_{\vec{x}}^{\dagger} = \frac{1}{n} \sum_{\vec{k} \in \mathrm{BZ}} e^{-i\vec{k} \cdot \vec{x}} c_{\vec{k}}^{\dagger}, \quad c_{\vec{x}} = \frac{1}{n} \sum_{\vec{k} \in \mathrm{BZ}} e^{i\vec{k} \cdot \vec{x}} c_{\vec{k}}, \tag{21}$$

and the Hamiltonian in the momentum space is given by

$$H = 2M_{\vec{k}=\vec{0}} c_{\vec{k}=\vec{0}}^{\dagger} c_{\vec{k}=\vec{0}} + 2 \sum_{\vec{k} \in \frac{1}{2}\mathrm{BZ}} \left(M_{\vec{k}} c_{\vec{k}}^{\dagger} c_{\vec{k}} + M_{\vec{k}} c_{-\vec{k}}^{\dagger} c_{-\vec{k}} + \Delta_{\vec{k}} c_{\vec{k}}^{\dagger} c_{-\vec{k}}^{\dagger} + \Delta_{\vec{k}}^{*} c_{-\vec{k}} c_{\vec{k}} \right). \tag{22}$$

The ground state is

$$\psi = e^{\sum_{\vec{k} \in \frac{1}{2}\mathrm{BZ}} \left(v_{\vec{k}}/u_{\vec{k}}\right) c_{\vec{k}}^{\dagger} c_{-\vec{k}}^{\dagger}} |0\rangle = \bigotimes_{\vec{k} \in \frac{1}{2}\mathrm{BZ}} \psi_{\vec{k}}, \quad \psi_{\vec{k}} := e^{\left(v_{\vec{k}}/u_{\vec{k}}\right) c_{\vec{k}}^{\dagger} c_{-\vec{k}}^{\dagger}} |0\rangle. \tag{23}$$

Consider the state

$$\phi^{(m)} = e^{\sum_{\vec{k} \in \frac{1}{2}\mathrm{BZ}} \left(v_{\vec{k}}/u_{\vec{k}}^{(m)}\right) c_{\vec{k}}^{\dagger} c_{-\vec{k}}^{\dagger}} |0\rangle = \bigotimes_{\vec{k} \in \frac{1}{2}\mathrm{BZ}} \phi_{\vec{k}}^{(m)}, \quad \phi_{\vec{k}}^{(m)} := e^{\left(v_{\vec{k}}/u_{\vec{k}}^{(m)}\right) c_{\vec{k}}^{\dagger} c_{-\vec{k}}^{\dagger}} |0\rangle, \tag{24}$$

where $u_{\vec{k}}^{(m)}$ is a degree-m trigonometric polynomial obtained by expanding $u_{\vec{k}}$ in Fourier series and computing the partial sum up to order m. Similar to (20), $\phi^{(m)}$ has an exact neural network state representation (11) such that the maximum distance of connections between units is $O(m)$.

As $u_{\vec{k}}$ is a real analytic function of $\vec{k}$, its Fourier coefficients of order m decay exponentially as e^{-am} for some constant $a > 0$. Note that a is a function of μ and can be solved analytically in the regime the energy gap is small; see, e.g., Ref. [22], §2. Therefore,

$$\left| u_{\vec{k}}^{(m)} - u_{\vec{k}} \right| = O(e^{-am}), \quad \forall \vec{k} \in \frac{1}{2}\mathrm{BZ}. \tag{25}$$

7

Furthermore, the absolute values of $u_{\vec{k}}, v_{\vec{k}}$ are bounded away from 0:

$$\left|u_{\vec{k}}\right| \geq \Omega(1/n^2), \quad \left|v_{\vec{k}}\right| \geq \Omega(1/n), \quad \forall \vec{k} \in \frac{1}{2}\text{BZ}. \tag{26}$$

Let $\mathcal{N}$ be the operator that normalizes a state. Eqs. (25), (26) imply

$$f_{\vec{k}} := \left|\left\langle \mathcal{N}\phi_{\vec{k}}^{(m)} | \mathcal{N}\psi_{\vec{k}} \right\rangle\right| \geq 1 - 1/n^3, \quad \forall \vec{k} \in \frac{1}{2}\text{BZ}, \tag{27}$$

provided that $m = O(\log n)$ with a sufficiently large constant prefactor hided in the big-O notation. As ψ and $\phi^{(m)}$ are product states in the momentum space, the fidelity is given by

$$\left|\left\langle \mathcal{N}\phi^{(m)} | \mathcal{N}\psi \right\rangle\right| = \prod_{\vec{k} \in \frac{1}{2}\text{BZ}} f_{\vec{k}} \geq (1 - 1/n^3)^{O(n^2)} = 1 - O(1/n). \tag{28}$$

Acknowledgments and notes

Part of this work was presented on November 27, 2014 (Thanksgiving day!) at the Perimeter Institute for Theoretical Physics. Very recently, we became aware of some related papers [7, 9, 11], which studied the relationship between neural and tensor network states using different approaches. In particular, Theorem 2 and Corollary 3 are stronger than Theorem 3 in Ref. [11].

References

[1] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(1):147 – 169, 1985.

[2] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko. Quantum Boltzmann machine. arXiv:1601.02036, 2016.

[3] I. Arad, A. Kitaev, Z. Landau, and U. Vazirani. An area law and sub-exponential algorithm for 1d systems. arXiv:1301.1162, 2013.

[4] B. Béri and N. R. Cooper. Local tensor network for strongly correlated projective states. *Physical Review Letters*, 106(15):156401, 2011.

[5] A. Bernevig and T. Neupert. Topological superconductors and category theory. arXiv:1506.05805, 2015.

[6] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. arXiv:1606.02318, 2016.

[7] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang. On the equivalence of restricted Boltzmann machines and tensor network states. arXiv:1701.04831, 2017.

[8] D.-L. Deng, X. Li, and S. Das Sarma. Exact machine learning topological states. arXiv:1609.09060, 2016.

[9] D.-L. Deng, X. Li, and S. Das Sarma. Quantum entanglement in neural network states. arXiv:1701.04844, 2017.

[10] J. Dubail and N. Read. Tensor network trial states for chiral topological phases in two dimensions and a no-go theorem in any dimension. *Physical Review B*, 92(20):205307, 2015.

[11] X. Gao and L.-M. Duan. Efficient representation of quantum many-body states with deep neural networks. arXiv:1701.05039, 2017.

[12] S. Gharibian, Y. Huang, Z. Landau, and S. W. Shin. Quantum hamiltonian complexity. *Foundations and Trends in Theoretical Computer Science*, 10(3):159–282, 2015.

[13] Z.-C. Gu, F. Verstraete, and X.-G. Wen. Grassmann tensor network states and its renormalization for strongly correlated fermionic and bosonic states. arXiv:1004.2563, 2010.

[14] M. B. Hastings. An area law for one-dimensional quantum systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(08):P08024.

[15] M. B. Hastings. Locality in quantum systems. In J. Frohlich, M. Salmhofer, W. De Roeck, V. Mastropietro, and L. F. Cugliandolo, editors, *Quantum Theory from Small to Large Scales*, volume 95 of *Lecture Notes of the Les Houches Summer School*, chapter 3, pages 171–212. Oxford University Press, Oxford, UK, 2012.

[16] Y. Huang. *Classical simulation of quantum many-body systems*. PhD thesis, University of California, Berkeley, 2015.

[17] N. Le Roux and Y. Bengio. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Computation*, 20(6):1631–1649, 2008.

[18] G. Montufar and N. Ay. Refinements of universal approximation results for deep belief networks and restricted boltzmann machines. *Neural Computation*, 23(5), 2011.

[19] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix product state representations. *Quantum Information and Computation*, 7(5-6):401–430, 2007.

[20] N. Read and D. Green. Paired states of fermions in two dimensions with breaking of parity and time-reversal symmetries and the fractional quantum Hall effect. *Physical Review B*, 61(15):10267–10297, 2000.

[21] P. Smolensky. Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, and PDP Research Group, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, chapter 6, pages 194–281. MIT Press, Cambridge, MA, USA, 1986.

[22] L. N. Trefethen. *Finite Difference and Spectral Methods for Ordinary and Partial Differential Equations*. unpublished text, 1996, available at http://people.maths.ox.ac.uk/trefethen/pdetext.html.

[23] F. Verstraete and J. I. Cirac. Matrix product states represent ground states faithfully. *Physical Review B*, 73(9):094423, 2006.

[24] T. B. Wahl, H.-H. Tu, N. Schuch, and J. I. Cirac. Projected entangled-pair states can describe chiral topological states. *Physical Review Letters*, 111(23):236805, 2013.