Neural network representation for minimally entangled typical thermal states

Douglas Hendry, Hongwei Chen, and Adrian Feiguin
Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA

(Received 2 May 2022; revised 23 September 2022; accepted 27 September 2022; published 12 October 2022)

Minimally entangled typical thermal states are a construction that allows one to solve for the imaginary time evolution of quantum many-body systems. By using wave functions that are weakly entangled, one can take advantage of efficient representations in the form of matrix product states. We generalize these ideas to arbitrary variational wave functions and we focus, as an illustration, on the particular case of restricted Boltzmann machines. The imaginary time evolution is carried out using stochastic reconfiguration (natural gradient descent) combined with Monte Carlo sampling. Since the time evolution takes place on the tangent space, deviations between the actual path in the Hilbert space and the trajectory on the variational manifold can be important, depending on the internal structure and expressivity of the variational states. We show how these differences translate into a rescaled temperature and demonstrate the application of the method to quantum spin systems in one and two spatial dimensions.

DOI: 10.1103/PhysRevB.106.165111

I. INTRODUCTION

Quantum Monte Carlo (QMC) is a powerful method that can be applied to problems with hundreds of degrees of freedom in arbitrary dimensions. However, while in principle it can be considered an unbiased numerical technique, in many cases it suffers from some pathological drawbacks. Among these, there is the infamous sign problem, which appears in fermionic and frustrated systems when the complicated nodal structure of the wave function does not guarantee a well-defined positive transition probability [1–9]. Other situations arise when the Monte Carlo updates necessary to make the simulation ergodic are complicated or numerically costly, or when the system is close to a phase transition and global updates are required to fight critical slowing down.

Alternatives to QMC that can overcome such drawbacks are not many and also suffer from limitations. Useful practical approaches that rely on exact diagonalization are limited to small system sizes [10–24]. In quasi-one-dimensional systems, a family of methods based on the density-matrix renormalization group can essentially provide numerically exact results for models with frustration [25–34], and recent advances have put quasi-two-dimensional frustrated lattice problems within reach [35]. At this point, it is important to point out that these approaches rely, in one way or another, on representations of the transfer matrix or the thermal density matrix of the quantum many-body problem in the form of matrix product states (MPSs) or matrix product operators, and, as a consequence, are limited by the entanglement growth in the system as temperature is lowered, implicitly imposing a numerical barrier that is hard to overcome. Recent proposals using entanglement purification with neural networks provide an interesting alternate route [36].

In a thermal state, the expectation value of observables is identical to the value in the canonical ensemble at some temperature T. This idea lies at the foundation of the statistical mechanics and the canonical to microcanonical correspondence, relating the thermodynamic behavior of systems at temperature T to the microstates of a system at some energy $E(T)$. A generic chaotic closed system out of equilibrium is expected to relax to a thermal state after some time. This problem does not require a thermal bath and, in the context of the microcanonical ensemble energy is conserved. However, in order for this to actually occur, certain conditions need to be satisfied: The expectations values of observables within an energy window around $E(T)$ need to vary smoothly, or, rather, to be very "similar." This is the premise behind the eigenstate thermalization hypothesis (ETH) [37,38], and the idea of "typicality." According to this, a thermal state can be represented accurately by a typical pure state in the microcanonical ensemble. This can be exploited to carry out finite-temperature calculations with pure states, which is the foundation behind "minimally entangled typical thermal states" (METTSs) [39,40] and "canonical thermal pure quantum states"(CTPQSs) [16,18,21]. In a nutshell, the recipe is very simple: Start from a random state, such as a linear combination of basis states with random coefficients $|\psi_0\rangle$, and evolve it with the operator $|\psi(\beta)\rangle = \exp(-\beta H/2)|\psi_0\rangle$ (β as customary represents the inverse temperature T). Observables are hence obtained as $\langle\hat{A}\rangle_T = \langle\psi(\beta)|\hat{A}|\psi(\beta)\rangle/\langle\psi(\beta)|\psi(\beta)\rangle$. In the case of METTS, the initial random states are product states, e.g., quantum spins pointing in random directions on the Bloch sphere. Essentially, the algorithm is identical to projector Monte Carlo [41,42], but with the initial state being evolved using a numerically exact method. On the other hand, METTS approaches are based on a *variational* representation of the quantum many-body states in the form of a MPS, and their remarkable accuracy relies on the extraordinary representation power of these wave functions. The fact that entanglement at finite temperatures remains under

control when the initial state is a random product state has enabled some outstanding progress toward understanding the thermodynamic behavior of frustrated magnets [43].

In this work, we take a similar route but use neural network wave functions instead. Although our considerations are general, for illustration purposes we here focus on the simplest form, a restricted Boltzmann machine (RBM). Same as MPS wave functions, RBMs are agnostic to the underlying physics of the problem and hold a remarkable representation power.

This paper is organized as follows: In Sec. II we review how typicality can be used to calculate thermodynamic properties of quantum systems. Section III discusses the practical implementation of these ideas using variational Monte Carlo. Section IV demonstrates the methods with applications to one- and two-dimensional quantum spin systems. Finally, we close with a discussion.

II. TYPICAL THERMAL STATES

In this section we follow the reasoning outlined in Ref. [40] to describe thermal averages in terms of typical states. We consider a set of initial states $\{|\phi_0(\xi)\rangle\}_\xi$ with ξ drawn from a probability distribution function $P_0(\xi)$ such that

$$\int d\xi P_0(\xi)|\phi_0(\xi)\rangle\langle\phi_0(\xi)| = 1. \tag{1}$$

We have not explicitly introduced a particular form for these states, yet. The index ξ labels a set initial random product states. Each drawn initial state is evolved in imaginary time $\beta = 1/T$ as

$$|\phi(\beta;\xi)\rangle = e^{-\frac{1}{2}\beta\hat{H}}|\phi_0(\xi)\rangle.$$

Introducing $Z(\beta;\xi) = \langle\phi(\beta;\xi)|\phi(\beta;\xi)\rangle$, the partition function can expressed as $\mathcal{Z}(\beta) = \langle Z(\beta,\xi)\rangle_{P_0}$ and the evolution operator in imaginary time as

$$\frac{1}{\mathcal{Z}(\beta)}e^{-\beta\hat{H}} = \frac{\langle|\phi(\beta;\xi)\rangle\langle\phi(\beta;\xi)|\rangle_{P_0}}{\langle Z(\beta;\xi)\rangle_{P_0}}.$$

Then, for any observable given by operator $\hat{A}$, its thermal average can be expressed as

$$\mathcal{A}(\beta) = \frac{\langle Z(\beta;\xi)A(\beta;\xi)\rangle_{P_0}}{\langle Z(\beta;\xi)\rangle_{P_0}},$$

where

$$A(\beta;\xi) = \left.\frac{\langle\phi|\hat{A}|\phi\rangle}{\langle\phi|\phi\rangle}\right|_{(\beta;\xi)}$$

is the "local" expectation value of the operator in state $|\phi(\beta;\xi)\rangle$. Introducing a finite-temperature distribution $P_\beta(\xi) = [Z(\beta,\xi)/\mathcal{Z}(\beta)]P_0(\xi)$, the thermal average can be expressed compactly as $\mathcal{A}(\beta) = \langle A(\beta;\xi)\rangle_{P_\beta}$.

The importance weights $Z(\beta,\xi)/\mathcal{Z}(\beta)$ can be obtained without explicitly calculating $\langle\phi(\beta;\xi)|\phi(\beta;\xi)\rangle$ for each β. Instead we only need the initial $Z(0;\xi)$ and, for each β, the expectation value of the Hamiltonian:

$$E(\beta;\xi) = \left.\frac{\langle\phi|\hat{H}|\phi\rangle}{\langle\phi|\phi\rangle}\right|_{(\beta;\xi)}.$$

Then, we can exploit that the imaginary time evolution gives

$$\frac{\partial}{\partial\beta}\ln Z(\beta;\xi) = -E(\beta;\xi).$$

Hence,

$$Z(\beta;\xi) = Z(0;\xi)e^{-\int_0^\beta d\beta' E(\beta';\xi)}.$$

Until now, we have not imposed any conditions on the structure of the random initial states. In the METTS algorithm, one chooses them over a Gaussian distribution of random product states. For instance, if the quantum degree of freedom is spins $S = 1/2$ on a lattice $\mathcal{L}$ with N sites, they will be given as

$$|\phi_0(\xi)\rangle = \otimes_{l\in\mathcal{L}}\left(\frac{\xi_{l\uparrow}|\uparrow\rangle_l + \xi_{l\downarrow}|\downarrow\rangle_l}{\sqrt{|\xi_{l\uparrow}|^2 + |\xi_{l\downarrow}|^2}}\right). \tag{2}$$

In this case, the label ξ represents the set of $\xi \in \mathbb{C}^{N\times 2}$ complex numbers that are distributed according to

$$P_0(\xi) = \left(\frac{1}{\pi}\right)^{2N}\exp\left(-\sum_{l\in\mathcal{L}}\sum_{\sigma=\uparrow,\downarrow}|\xi_{l\sigma}|^2\right).$$

This choice of distribution is not the only possible one. However, the average of the outer product has to be equal to the identity, Eq. (1), and a Gaussian distribution is the easiest and most straightforward choice that satisfies this condition. In addition, it also gives us a uniform random distribution over the Bloch sphere for each site [40].

III. METHOD

While the concepts described in the previous section offer a prescription to calculate thermodynamic properties of quantum many-body states, exact calculations can only be carried out in small systems. To scale the computations to large system sizes, we must make some sacrifices: We use a variational representation of the wave functions. For this particular task, the mathematical structure of the wave function has to be flexible enough to be able to represent any quantum state in the spectrum, and not just the ground state. In the original formulation of METTS, matrix product states are used. In our case, we generalize the method to arbitrary variational states and we focus, as an illustration, on the particular case of restricted Boltzmann machines.

In this section we review how to carry out the time evolution of a many-body state on a variational manifold, following an elegant geometrical interpretation presented in Ref. [44]. The time-evolved state will describe a "trajectory" that will be constrained to this manifold and will deviate from the exact trajectory in the full Hilbert space. We find that these deviations can be partially accounted by rescaling the "projected imaginary time" τ such that it corresponds to an actual physical inverse temperature β.

A. Variational imaginary-time evolution

For each random starting state $|\phi_0(\xi)\rangle$, the time-evolved wave function $|\phi(\beta,\xi)\rangle$ is obtained by solving the first-order differential equation $\frac{\partial}{\partial\beta}|\phi(\beta,\xi)\rangle = -(1/2)\hat{H}|\phi(\beta,\xi)\rangle$, with

the constraint that the new wave function has to live on the same variational manifold.

We hereby proceed to summarize projective variational imaginary-time evolution for complex holomorphic variational wave functions. In this discussion we follow the notation and formalism as detailed in Ref. [44]. The states $|\phi(\beta,\xi)\rangle$ are approximated by a class of variational wave functions $\psi_{\mathcal{M}}$ which are holomorphic in terms of parameters $\vec{\theta}\in\mathbb{C}^M$ and define a submanifold in the Hilbert space $\mathcal{H}$, $\mathcal{M}=\psi_{\mathcal{M}}(\mathbb{C}^M)\subset\mathcal{H}$. The projected time evolution results in a series of "equations of motion" for the parameters $\theta_\mu(\tau)$, which are completely equivalent to the well-known "stochastic reconfiguration" (SR) method [45–48]—also known as "natural gradient descent" [49]—used to carry out ground-state calculations. In other words, SR consists of projecting out the ground state by evolving the state in imaginary time on the variational manifold. Therefore, any variational Monte Carlo code that implements SR already contains all the ingredients to evolve any variational state in imaginary time. We refer the reader to a pedagogical description in Ref. [50] for a detailed derivation.

Evolving the state in imaginary time evolution within the manifold requires performing a local projection at each step. The variational parameters that best represent the time-evolved state are obtained by minimizing the projection error

$$\left\|\hat{P}_{T\mathcal{H}}\left(\frac{d}{d\tau}|\psi_{\mathcal{M}}\rangle+\frac{1}{2}\hat{H}|\psi_{\mathcal{M}}\rangle\right)\right\|^2,$$

where use the Fubini-Study metric to measure the "distance" between two wave functions [51,52] with the projector $\hat{P}_{T\mathcal{H}}$ defined as

$$\hat{P}_{T\mathcal{H}}=\left(1-\frac{|\psi_{\mathcal{M}}\rangle\langle\psi_{\mathcal{M}}|}{\langle\psi_{\mathcal{M}}|\psi_{\mathcal{M}}\rangle}\right),$$

which projects out radial dependence, which is given by the $|\psi_{\mathcal{M}}\rangle$ direction.

Algorithm 1 Imaginary-time evolution

Require: $\Delta\tau$ (fixed imaginary time step)
Require: $\xi\sim P_0$ (draw random product state coefficients)
Require: θ_0 (initial VWF coefficients)
Require: $E_0=\langle\hat{H}\rangle(\theta_0)$
Require: $\sigma_0^2=(\langle\hat{H}^2\rangle-\langle\hat{H}\rangle^2)(\theta_0)$
Require: $A_0=\langle\hat{A}\rangle(\theta_0)$
Require: $\tau_0=0,\ \beta_0=0,\ Z_0=1$
while E_i not converged **do**
  $i\leftarrow i+1$
  $\theta_i\leftarrow\theta_{i-1}-\Delta\tau\nabla_\theta\langle\hat{H}\rangle(\theta_{i-1})$
  $E_i\leftarrow\langle\hat{H}\rangle(\theta_i)$
  $\sigma_i^2\leftarrow(\langle\hat{H}^2\rangle-\langle\hat{H}\rangle^2)(\theta_i)$
  $A_i\leftarrow\langle\hat{A}\rangle(\theta_i)$
  $\tau_i\leftarrow\tau_{i-1}+\Delta\tau$
  $\beta_i\leftarrow\beta_{i-1}-(E_i-E_{i-1})/\sigma_i^2$
  $Z_i\leftarrow Z_{i-1}\exp[-(\beta_i-\beta_{i-1})E_i]$
end while

To minimize this cost function we recast the problem in terms of $d\theta^\mu/d\tau\approx\Delta\theta^\mu/\Delta\tau$, which are defined by the tangent space vectors of the variational manifold:

$$|v_\mu\rangle=\hat{P}_{T\mathcal{H}}\frac{\partial}{\partial\theta^\mu}|\psi_{\mathcal{M}}\rangle.$$

The parameter updates $\Delta\vec{\theta}$ are given in terms of the system of equations:

$$g_{\mu\nu}\Delta\theta^\nu=-\frac{\Delta\tau}{2}\frac{\langle v_\mu|\hat{H}|\psi_{\mathcal{M}}\rangle}{\langle\psi_{\mathcal{M}}|\psi_{\mathcal{M}}\rangle},$$

where g is the induced metric tensor:

$$g_{\mu\nu}=\frac{\langle v_\mu|v_\nu\rangle}{\langle\psi_{\mathcal{M}}|\psi_{\mathcal{M}}\rangle}.$$

The resulting evolution of $|\psi_{\mathcal{M}}\rangle$ can be expressed compactly as

$$\hat{P}_{T\mathcal{H}}\frac{d}{d\tau}|\psi_{\mathcal{M}}\rangle=-\frac{1}{2}\hat{P}_{T\mathcal{M}}\hat{H}|\psi_{\mathcal{M}}\rangle,$$

where we introduce the variational tangent space projector:

$$\hat{P}_{T\mathcal{M}}=G^{\mu\nu}\frac{|v_\mu\rangle\langle v_\nu|}{\langle\psi_{\mathcal{M}}|\psi_{\mathcal{M}}\rangle},\quad G^{\mu\nu}g_{\nu\sigma}=\delta^\mu_\sigma,$$

In practice, the procedure to time evolve an initial random state is summarized in the Algorithm 1. As discussed above, each initial trial wave function is drawn from a distribution $P_0(\xi)$ and the parameters are evolved in imaginary time by a small fixed time step $\Delta\tau$. At each time step, expectation values of the energy, variance of the energy, and observables of interest are calculated:

$$E(\vec{\theta}(\tau;\xi))=\left.\frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle}\right|_{\vec{\theta}(\tau;\xi)},\tag{3}$$

$$\sigma^2(\vec{\theta}(\tau;\xi))=\left.\frac{\langle\psi|(\hat{H}-E)^2|\psi\rangle}{\langle\psi|\psi\rangle}\right|_{\vec{\theta}(\tau;\xi)},\tag{4}$$

$$A(\vec{\theta}(\tau;\xi))=\left.\frac{\langle\psi|\hat{A}|\psi\rangle}{\langle\psi|\psi\rangle}\right|_{\vec{\theta}(\tau;\xi)}.\tag{5}$$

These expectation values are then averaged over many initial realizations of ξ.

B. Beta correction

Since the approximate imaginary-time evolution corresponds to projecting onto the variational manifold, the time-evolved state will deviate from the exact one. This deviation from the true path can be decomposed locally into two contributions (see Fig. 1):

$$\hat{P}_{T\mathcal{H}}\frac{d}{d\tau}|\psi_{\mathcal{M}}\rangle=-\frac{1}{2}\gamma\hat{P}_{T\mathcal{M}}\hat{H}|\psi_{\mathcal{M}}\rangle+|\eta\rangle,\tag{6}$$

where $|\eta\rangle$ is some error direction orthogonal to both $|\psi_{\mathcal{M}}\rangle$ and $(\hat{H}-E)|\psi_{\mathcal{M}}\rangle$. The factor $\gamma\in[0,1]$ represents the fraction of the distance traveled in the exact imaginary time direction,

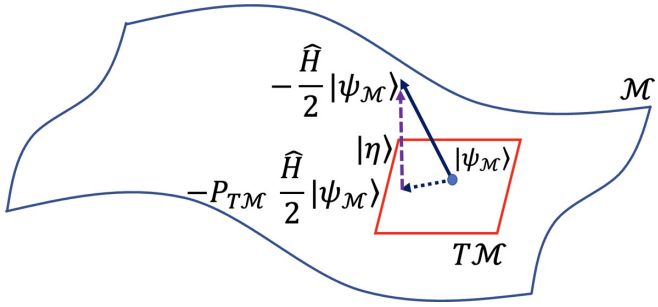

FIG. 1. Cartoon illustrating imaginary-time evolution in the projected variational manifold. The actual time-evolved state differs from the projected one. These difference can be accounted for by a rescaling of the temperature (see text).

which can be explicitly calculated as

$$\gamma = \frac{\langle\psi|(\hat{H}-E)\hat{P}_{T\mathcal{M}}(\hat{H}-E)|\psi\rangle}{\langle\psi|(\hat{H}-E)^2|\psi\rangle} = -\frac{1}{\sigma^2}\frac{dE}{d\tau}. \quad (7)$$

Here we assume that the contribution from the error direction $|\eta\rangle$ is fairly negligible (which is a reasonable assumption for imaginary-time evolution as opposed to real-time evolution). Thus, the remaining error corresponds to the parametrization of the imaginary time. Therefore, we reparametrize τ in terms of β as

$$\frac{d\beta}{d\tau} = \gamma, \quad \beta(\tau;\xi) = -\int_0^\tau d\tau' \frac{1}{\sigma^2}\frac{dE}{d\tau'}\bigg|_{\vec{\theta}(\tau';\xi)}.$$

Following Eq. (7), this parametrization also enforces $dE/d\beta = -\sigma^2$.

C. Restricted Boltzmann machines

Although our considerations are independent of the choice of variational wave function, in the following we focus on a particular example for demonstration purposes: a restricted Boltzmann machine (RBM). A RBM wave function for a system of N spins $S = 1/2$ ($s = \uparrow, \downarrow$) and N_h hidden variables is defined as

$$|\psi(\vec{\theta}\,)\rangle = \sum_{\{\vec{s}\}} \psi(\vec{s};\vec{\theta})|\vec{s}\,\rangle,$$

where $\vec{s} = (s_1, s_2, \ldots, s_N)$ and

$$\psi(\vec{s};\vec{\theta}) = e^{\sum_{l\in\mathcal{L}} a_l s_l} \prod_{i=1}^{N_h} \cosh\left(b_i + \sum_{l\in\mathcal{L}} W_{il}\cdot s_l\right)$$

It is parametrized by a set of complex values $\vec{\theta} = (\vec{a}, \vec{b}, W) \in \mathbb{C}^N \times \mathbb{C}_h^N \times \mathbb{C}^{N_h\times N}$ that are used as variational parameters to minimize some cost function. This cost function is usually a measure of the "distance" between $|\psi(\vec{\theta}\,)\rangle$ and a target wave function.

Besides being proposed as ground-state estimators for variational calculations [50,53–55], their representation power has been instrumental to a number of other applications, such as the calculation of spectral functions [56–58].

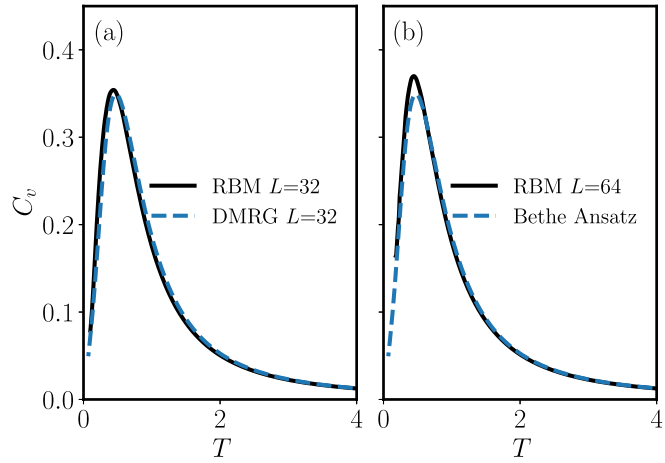

FIG. 2. Specific heat of the Heisenberg chain comparing results obtained with restricted Boltzmann machines and other methods: (a) DMRG for $L = 32$, (b) $L = 64$ and Bethe ansatz. We use $N_h = 3N$ hidden variables and $n = 50$ random initial states.

D. Restricted Boltzmann machine initialization

To implement the procedure with RBMs, it is necessary to first draw a random set of initial product wave functions as described in Sec. II. Fortunately, it is possible to exactly represent (up to an overall constant) a state $|\phi_0(\xi)\rangle$ in Eq. (2) by setting b_l, $W = 0$ and matching the a_l bias term to the corresponding random spin at l:

$$a_l = \tfrac{1}{2}\ln(\xi_{l+}/\xi_{l-}).$$

In practice, a very small Gaussian noise has to be added to W, b such that the derivatives needed in the imaginary-time evolution are not zero. Notice that there is no requirement that the initial state has to be a product state, so the validity of the method is not affected.

IV. NUMERICAL RESULTS

We have implemented the imaginary time evolution for the spin $S = 1/2$ Heisenberg model in one and two dimensions:

$$\hat{H} = \sum_{\langle i,j\rangle} \vec{S}_i \cdot \vec{S}_j,$$

where the sum sums over nearest-neighbor sites on a one-dimensional chain or a square lattice.

Besides the energy (3), we calculate the specific heat as $C_v = -(\beta^2/N)(dE/d\beta) = \beta^2/N\sigma^2$, where σ is the standard deviation of the energy. As an initial benchmark, in Fig. 2 we show the specific heat for chains with $L = 32$ and $L = 64$ sites and periodic boundary conditions obtained with $n = 50$ initial states and $N_h = 3N$ hidden variables, together with the finite-temperature density-matrix renormalization group [31] and Bethe ansatz [59] results for comparison. The curves are barely distinguishable on this scale.

Results for two-dimensional systems (Fig. 3) show good agreement with quantum Monte Carlo data for 6×6 and

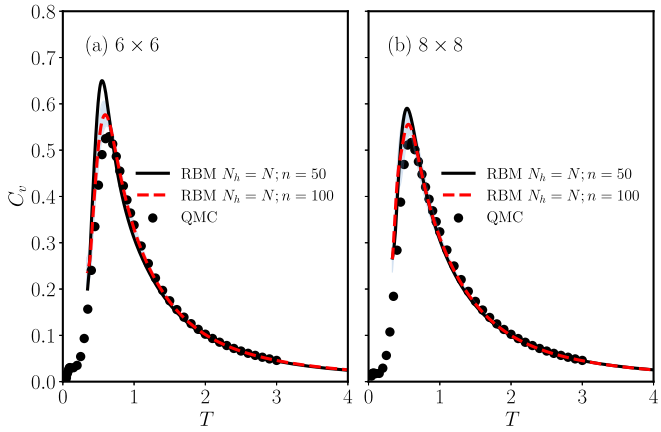

FIG. 3. Specific heat of the two-dimensional Heisenberg model comparing results obtained with restricted Boltzmann machines and quantum Monte Carlo for two system sizes (a) 6×6 and (b) 8×8 with varying number of initial random states n and $N_h = N$ hidden variables. The shaded area represents the error for the dashed curve.

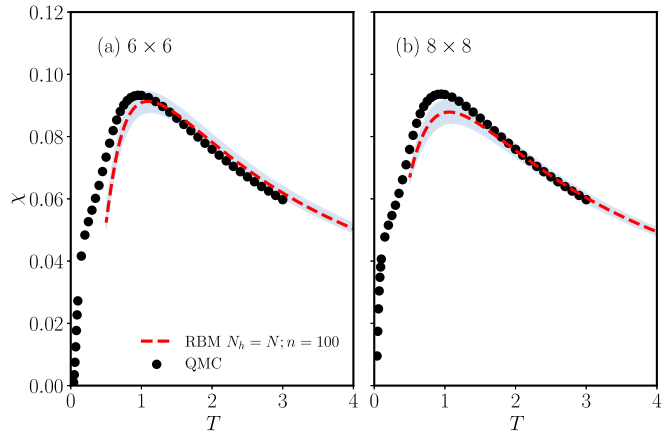

FIG. 5. Magnetic susceptibility of the two-dimensional Heisenberg model comparing results obtained with restricted Boltzmann machines and quantum Monte Carlo for two system sizes (a) 6×6 and (b) 8×8 with $n = 100$ initial random states n and $N_h = N$ hidden variables. The shaded area represents the error for the dashed curve.

8×8 lattices [60–62]. As seen in the figure, the accuracy varies considerably with the number n of initial states and improves noticeably with $n = 100$ in the region around the maximum, $T \approx J = 1$. We have also studied the convergence with the number of hidden variables N_h: even though a larger N_h may help to improve the accuracy, we find that with too many hidden variables, the natural gradient descent may get unstable. In our case we settle for $N_h = N$ since this choice provides a reliable comparison.

The specific heat is directly related to the variance of the energy. While C_v may show some deviation from the exact results, the energy is actually very accurate, as shown in Fig. 4 where the errors are smaller than the symbol size.

On the other hand, the errors for the magnetic susceptibility are considerably larger, as shown in Fig. 5. This is due to the nonlocal nature of the squared magnetization M^2 which involves a summation over all-to-all correlations:

$$\chi = \frac{1}{NT}(\langle M^2\rangle - \langle M\rangle^2) = \frac{1}{NT}\left(\sum_{i,j}\langle S_i^z S_i^z\rangle - \left(\sum_i \langle S_i^z\rangle\right)^2\right). \tag{8}$$

The structure of the correlations is more sensitive and a better measure of the expressivity or representation power of the states. For instance, it has been established that matrix product states can only realize exponentially decaying correlations [63] even though the yield energies with machine precision accuracy in one-dimensional (1D) systems.

We point out that, in these figures, the Monte Carlo sampling error and the error originating from averaging over a finite number of initial random states is plotted as a shaded area centered at the mean. In addition, we account for part of the systematic errors with the beta correction, as described above. Other sources of error originating from the representation power of the RBM wave functions are difficult to estimate. On this scale, the QMC data are essentially exact and the error is smaller than the symbol size.

While results improve notoriously with increasing number n of initial random states, as mentioned earlier there seems to be an apparent horizontal shift of the C_v and χ curves at low temperatures. This can be partially explained in terms of the beta correction. In Fig. 6 we show the actual β as a function of the projected imaginary time τ for two typical runs. While the curves seem to follow an apparent $1:1$ scaling at small β, deviations appear as we approach zero temperature. The flattening of the curves for increasing β (small temperature) indicates that the wave function is essentially converging to the (approximate) ground state. Geometrically, this can be interpreted as the vector $-\hat{H}/2|\psi_{\mathcal{M}}\rangle$ becoming perpendicular to the variational manifold, as one would see in Fig. 1. For that reason, the wave-function parameters cannot vary and the state becomes "stuck" in parameter space. Other sources

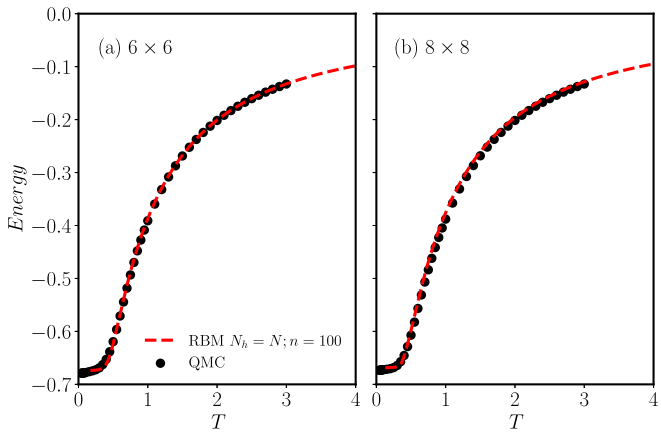

FIG. 4. Energy of the 2D Heisenberg model comparing results obtained with restricted Boltzmann machines and quantum Monte Carlo for two system sizes (a) 6×6 and (b) 8×8 with $n = 100$ initial random states n and $N_h = N$ hidden variables. Errors are smaller than the symbol size.

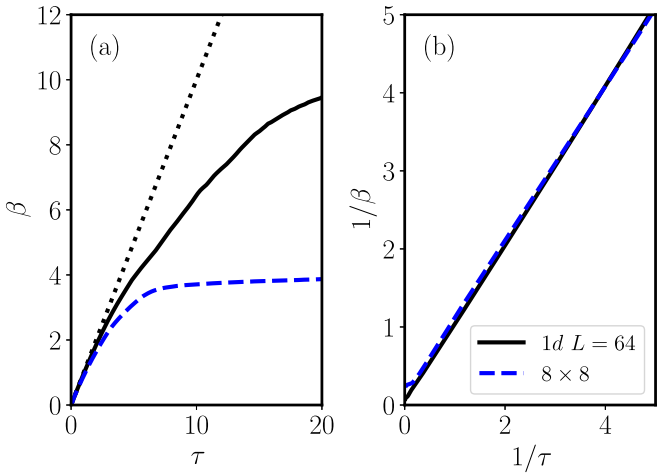

FIG. 6. (a) Actual inverse temperature β versus projected imaginary time τ. We show results for both one- and two-dimensional lattices. The dotted line corresponds to a 1 : 1 perfect scaling. The differences arise from projecting on the variational manifold (see text). (b) Same data as in panel (a) but showing the physical temperature $T = 1/\beta$.

for increase of these deviations could be multiple, but we believe the most significant ones are due to (i) the presence of a finite-size gap in the spectrum, (ii) a small overlap with the actual ground state due to the random initial directions of the spins, or (iii) a poor variational wave function. The most promising route to improve the expressivity of RBMs is by introducing lattice symmetries [64,65], which we have not attempted here.

V. CONCLUSIONS

We have described a method to carry out thermodynamic simulations of quantum many-body models using typicality and variational representations of quantum states. We generalize the idea of METTS to arbitrary variational forms and efficiently carry out the imaginary-time evolution by using natural gradient descent (or stochastic reconfiguration). While in principle one could sample over a random choice of arbitrary initial states as done in the CTPQS approach [16,18,21], the choice of product states simplifies the formulation considerably. In particular, these states can be exactly represented as RBM wave functions. In addition, the low entanglement growth makes the numerical results more controllable. We point out that, unlike CTPQS, our simulations do not account for spatial nor spin symmetries. Incorporating them into the algorithm may result in improved accuracy.

The underlying mathematical structure of the wave functions plays a crucial role in terms of the accuracy of the method. In particular, the wave functions have to be able to represent any state along the imaginary time path. In this work, we pick the particular form of restricted Boltzmann machines as a proof of concept illustration due to their versatility and representation power. We show that the path they follow in the variational manifold differs slightly from the actual imaginary-time evolution for temperatures $T > J/2$. We have found that these deviations can be partially accounted for by correcting the temperature with a rescaling factor that can be easily and systematically calculated at every time step. At lower temperatures, the wave function starts converging to the best variational approximation to the ground state becoming "stuck" in the variational manifold. This sets a limit for the RBM wave functions in terms of their ability to describe the thermodynamic behavior at low temperatures. It is to expect that, by improving the expressivity of the wave function, it might be possible to reach lower temperatures and simultaneously obtain better a ground-state description. The limitations of the wave function can be appreciated in quantities such as the correlation functions and, in particular, magnetic susceptibility, which include all-to-all contributions. The method does not suffer from the sign problem and offers an alternative to matrix product states for studying two-dimensional models with frustration. In addition, with a suitable choice of variational wave function [66], it can readily be extended to fermionic systems.

ACKNOWLEDGMENTS

The authors are grateful to the National Science Foundation for support under Grant No. DMR-2120501. We thank Anders Sandvik for generously sharing his QMC data with us.

[1] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar, Sign problem in the numerical simulation of many-electron systems, Phys. Rev. B **41**, 9301 (1990).

[2] H. J. M. van Bemmel, D. F. B. ten Haaf, W. van Saarloos, J. M. J. van Leeuwen, and G. An, Fixed-Node Quantum Monte Carlo Method for Lattice Fermions, Phys. Rev. Lett. **72**, 2442 (1994).

[3] M. Troyer and U.-J. Wiese, Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations, Phys. Rev. Lett. **94**, 170201 (2005).

[4] R. K. Kaul, Marshall-positive SU(n) quantum spin systems and classical loop models: A practical strategy to design sign-problem-free spin Hamiltonians, Phys. Rev. B **91**, 054413 (2015).

[5] J. Yoo, S. Chandrasekharan, R. K. Kaul, D. Ullmo, and H. U. Baranger, On the sign problem in the Hirsch–Fye algorithm for impurity problems, J. Phys. A: Math. Gen. **38**, 10307 (2005).

[6] Z.-X. Li, Y.-F. Jiang, and H. Yao, Solving the fermion sign problem in quantum Monte Carlo simulations by Majorana representation, Phys. Rev. B **91**, 241117(R) (2015).

[7] Z.-X. Li, Y.-F. Jiang, and H. Yao, Majorana-Time-Reversal Symmetries: A Fundamental Principle for Sign-Problem-Free Quantum Monte Carlo Simulations, Phys. Rev. Lett. **117**, 267002 (2016).

[8] S. Karakuzu, K. Seki, and S. Sorella, Solution of the sign problem for the half-filled Hubbard-Holstein model, Phys. Rev. B **98**, 201108(R) (2018).

[9] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, Phys. Rev. Res. **2**, 033075 (2020).

[10] D. A. Drabold and O. F. Sankey, Maximum Entropy Approach for Linear Scaling in the Electronic Structure Problem, Phys. Rev. Lett. **70**, 3631 (1993).

[11] J. Jaklič and P. Prelošek, Lanczos method for the calculation of finite-temperature quantities in correlated systems, Phys. Rev. B **49**, 5065 (1994).

[12] M. Aichhorn, M. Daghofer, H. G. Evertz, and W. von der Linden, Low-temperature Lanczos method for strongly correlated systems, Phys. Rev. B **67**, 161103(R) (2003).

[13] M. W. Long, P. Prelovšek, S. El Shawish, J. Karadamoglou, and X. Zotos, Finite-temperature dynamical correlations using the microcanonical ensemble and the Lanczos algorithm, Phys. Rev. B **68**, 235106 (2003).

[14] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, The kernel polynomial method, Rev. Mod. Phys. **78**, 275 (2006).

[15] H. Avron and S. Toledo, Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix, J. Assoc. Comput. Mach. **58**, 1 (2011).

[16] S. Sugiura and A. Shimizu, Canonical Thermal Pure Quantum State, Phys. Rev. Lett. **111**, 010401 (2013).

[17] O. Hanebaum and J. Schnack, Advanced finite-temperature Lanczos method for anisotropic spin systems, Eur. Phys. J. B **87**, 194 (2014).

[18] M. Hyuga, S. Sugiura, K. Sakai, and A. Shimizu, Thermal pure quantum states of many-particle systems, Phys. Rev. B **90**, 121110(R) (2014).

[19] F. Roosta-Khorasani and U. Ascher, Improved bounds on sample size for implicit matrix trace estimators, Found. Comput. Math. **15**, 1187 (2015).

[20] A. K. Saibaba, A. Alexanderian, and I. C. F. Ipsen, Randomized matrix-free trace and log-determinant estimators, Numer. Math. **137**, 353 (2017).

[21] S. Sugiura, Canonical Thermal Pure Quantum State, in *Formulation of Statistical Mechanics Based on Thermal Pure Quantum States* (Springer Singapore, Singapore, 2017), pp. 15–30.

[22] S. Okamoto, G. Alvarez, E. Dagotto, and T. Tohyama, Accuracy of the microcanonical Lanczos method to compute real-frequency dynamical spectral functions of quantum models at finite temperatures, Phys. Rev. E **97**, 043308 (2018).

[23] J. Schnack, J. Richter, and R. Steinigeweg, Accuracy of the finite-temperature Lanczos method compared to simple typicality-based estimates, Phys. Rev. Res. **2**, 013186 (2020).

[24] P. Weinberg, Enhanced convergence of quantum typicality using a randomized low-rank approximation, arXiv:2102.02293.

[25] T. Nishino and K. Okunishi, Product wave function renormalization group, J. Phys. Soc. Jpn. **64**, 4084 (1995).

[26] R. J. Bursill, T. Xiang, and G. A. Gehring, The density matrix renormalization group for a quantum spin chain at non-zero temperature, J. Phys.: Condens. Matter **8**, L583 (1996).

[27] X. Q. Wang and T. Xiang, Transfer-matrix density-matrix renormalization-group theory for thermodynamics of one-dimensional quantum systems, Phys. Rev. B **56**, 5061 (1997).

[28] K. Maisinger and U. Schollwöck, Thermodynamics of Frustrated Quantum Spin Chains, Phys. Rev. Lett. **81**, 445 (1998).

[29] T. Nishino and N. Shibata, Efficiency of asymmetric targeting for finite-*T* DMRG, J. Phys. Soc. Jpn. **68**, 3501 (1999).

[30] F. Verstraete and J. I. Cirac, Renormalization algorithms for quantum-many body systems in two and higher dimensions, arXiv:cond-mat/0407066.

[31] A. E. Feiguin and S. R. White, Finite-temperature density matrix renormalization using an enlarged Hilbert space, Phys. Rev. B **72**, 220401(R) (2005).

[32] Y.-K. Huang, P. Chen, and Y.-J. Kao, Accurate computation of low-temperature thermodynamics for quantum spin chains, Phys. Rev. B **86**, 235102 (2012).

[33] B. Bruognolo, J. von Delft, and A. Weichselbaum, Symmetric minimally entangled typical thermal states, Phys. Rev. B **92**, 115105 (2015).

[34] B.-B. Chen, L. Chen, Z. Chen, W. Li, and A. Weichselbaum, Exponential Thermal Tensor Network Approach for Quantum Lattice Models, Phys. Rev. X **8**, 031082 (2018).

[35] L. Chen, D.-W. Qu, H. Li, B.-B. Chen, S.-S. Gong, J. von Delft, A. Weichselbaum, and W. Li, Two-temperature scales in the triangular-lattice Heisenberg antiferromagnet, Phys. Rev. B **99**, 140404(R) (2019).

[36] Y. Nomura, N. Yoshioka, and F. Nori, Purifying Deep Boltzmann Machines for Thermal Quantum States, Phys. Rev. Lett. **127**, 060601 (2021).

[37] M. Srednicki, Chaos and quantum thermalization, Phys. Rev. E **50**, 888 (1994).

[38] J. M. Deutsch, Eigenstate thermalization hypothesis, Rep. Prog. Phys. **81**, 082001 (2018).

[39] E. M. Stoudenmire and S. R. White, Minimally entangled typical thermal state algorithms, New J. Phys. **12**, 055026 (2010).

[40] S. R. White, Minimally Entangled Typical Quantum States at Finite Temperature, Phys. Rev. Lett. **102**, 190601 (2009).

[41] R. Blankenbecler and R. L. Sugar, Projector Monte Carlo method, Phys. Rev. D: Part. Fields **27**, 1304 (1983).

[42] N. Trivedi and D. M. Ceperley, Green-function Monte Carlo study of quantum antiferromagnets, Phys. Rev. B **40**, 2737 (1989).

[43] A. Wietek, R. Rossi, F. Šimkovic, M. Klett, P. Hansmann, M. Ferrero, E. M. Stoudenmire, T. Schäfer, and A. Georges, Mott Insulating States with Competing Orders in the Triangular Lattice Hubbard Model, Phys. Rev. X **11**, 041013 (2021).

[44] L. Hackl, T. Guaita, T. Shi, J. Haegeman, E. Demler, and J. I. Cirac, Geometry of variational methods: Dynamics of closed quantum systems, SciPost Phys. **9**, 048 (2020).

[45] S. Sorella, Green Function Monte Carlo with Stochastic Reconfiguration, Phys. Rev. Lett. **80**, 4558 (1998).

[46] S. Sorella and L. Capriotti, Green function Monte Carlo with stochastic reconfiguration: An effective remedy for the sign problem, Phys. Rev. B **61**, 2599 (2000).

[47] S. Sorella, Wave function optimization in the variational Monte Carlo method, Phys. Rev. B **71**, 241103(R) (2005).

[48] E. Neuscamman, C. J. Umrigar, and G. K.-L. Chan, Optimizing large parameter sets in variational quantum Monte Carlo, Phys. Rev. B **85**, 045103 (2012).

[49] S.-i. Amari, Natural gradient works efficiently in learning, Neural Comput. **10**, 251 (1998).

[50] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Neural-Network Quantum States, String-Bond

States, and Chiral Topological States, Phys. Rev. X **8**, 011006 (2018).
[51] J. P. Provost and G. Vallee, Riemannian structure on manifolds of quantum states, Commun. Math. Phys. **76**, 289 (1980).
[52] D. C. Brody and L. P. Hughston, Geometric quantum mechanics, J. Geom. Phys. **38**, 19 (2001).
[53] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).
[54] H. Saito, Solving the Bose-Hubbard model with machine learning, J. Phys. Soc. Jpn. **86**, 093001 (2017).
[55] Z. Cai and J. Liu, Approximating quantum many-body wave functions using artificial neural networks, Phys. Rev. B **97**, 035116 (2018).
[56] D. Hendry and A. E. Feiguin, Machine learning approach to dynamical properties of quantum many-body systems, Phys. Rev. B **100**, 245123 (2019).
[57] D. Hendry, H. Chen, P. Weinberg, and A. E. Feiguin, Chebyshev expansion of spectral functions using restricted Boltzmann machines, Phys. Rev. B **104**, 205130 (2021).
[58] R. Koch and J. L. Lado, Neural network enhanced hybrid quantum many-body dynamical distributions, Phys. Rev. Res. **3**, 033102 (2021).
[59] A. Klümper and D. C. Johnston, Thermodynamics of the Spin-1/2 Antiferromagnetic Uniform Heisenberg Chain, Phys. Rev. Lett. **84**, 4701 (2000).
[60] A. W. Sandvik, Stochastic series expansion method with operator-loop update, Phys. Rev. B **59**, R14157 (1999).
[61] K. H. Höglund and A. W. Sandvik, Impurity effects at finite temperature in the two-dimensional $S = 1/2$ Heisenberg antiferromagnet, Phys. Rev. B **70**, 024406 (2004).
[62] A. Sandvik (private communication).
[63] F. Pollmann, S. Mukerjee, A. M. Turner, and J. E. Moore, Theory of Finite-Entanglement Scaling at One-Dimensional Quantum Critical Points, Phys. Rev. Lett. **102**, 255701 (2009).
[64] Y. Nomura, Helping restricted Boltzmann machines with quantum-state representation by restoring symmetry, J. Phys.: Condens. Matter **33**, 174003 (2021).
[65] H. Chen, D. Hendry, P. Weinberg, and A. E. Feiguin, Systematic improvement of neural network quantum states using a Lanczos recursion, arXiv:2206.14307.
[66] D. Pfau, J. Spencer, A. de G. Matthews, and W. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).