Rényi free energy and variational approximations to thermal states

Giacomo Giudice, Aslı Çakan, J. Ignacio Cirac, and Mari Carmen Bañuls

Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany and
Munich Center for Quantum Science and Technology (MCQST), Schellingstraße 4, D-80799 München, Germany
(Dated: December 24, 2020)

We propose the construction of thermodynamic ensembles that minimize the Rényi free energy, as an alternative to Gibbs states. For large systems, the local properties of these Rényi ensembles coincide with those of thermal equilibrium, and they can be used as approximations to thermal states. We provide algorithms to find tensor network approximations to the 2-Rényi ensemble. In particular, a matrix-product-state representation can be found by using gradient-based optimization on Riemannian manifolds, or via a non-linear evolution which yields the desired state as a fixed point. We analyze the performance of the algorithms and the properties of the ensembles on one-dimensional spin chains.

I. INTRODUCTION

From the point of view of thermodynamics, thermal states describe the equilibrium properties of a system. Given a Hamiltonian H, the Gibbs state

$$\rho_G = \frac{1}{\mathcal{Z}_G} e^{-\beta H}, \quad \mathcal{Z}_G = \text{Tr}\, e^{-\beta H} \quad (1)$$

describes the state of the system at a given temperature $1/\beta$. On the other hand, thermal states arise from the principle of maximum entropy [1, 2]: for a given energy, the thermal ensemble is the one that maximizes the von Neumann entropy $S_G(\rho) = -\text{Tr}(\rho \log \rho)$. Equivalently, this can be formulated as the minimization of the free energy

$$F_G(\rho) = \text{Tr}(H\rho) - \frac{1}{\beta} S_G(\rho), \quad (2)$$

so that

$$\rho_G(\beta) = \arg \min_{\substack{\rho \succeq 0 \\ \text{Tr}\,\rho = 1}} F_G(\rho), \quad (3)$$

for some fixed value of β. To keep the notation light, we do not explicitly write the dependence of ρ_G on this parameter. One should keep in mind that the minimum is taken with respect to density operators, i.e. positive-semidefinite operators $\rho \succeq 0$ with a chosen normalization, typically $\text{Tr}(\rho) = 1$. This optimization is not very convenient in practice, since the entropy S_G is often difficult to compute, as it requires information about the entire spectrum of ρ. In a quantum many-body setting, this would require diagonalizing an exponentially large operator, because of the inherent tensor product structure of the Hilbert space.

In the quantum many-body setting, numerical approaches to thermal equilibrium do not try to explicitly solve the optimization above, but resort to different approaches to approximate Eq. (1). Monte Carlo methods use sampling to estimate very efficiently the physical properties from Eq.(1), but they encounter difficulties in scenarios where a sign problem appears, as can happen

for fermionic models or frustrated systems. A different approach is based on tensor networks (TN), where the total state corresponds to the contraction of low-rank tensors and allows for a local description of the physics. This is motivated by the fact that thermal states for a local Hamiltonian obey an area law for the mutual information [3, 4], and hence there is strong theoretical evidence that a tensor network description should be efficient at approximating thermal states [4–8].

In practice, TN are extremely successful for studying thermal equilibrium. In one spatial dimension, matrix product states (MPS) can be used to construct a representation of the (mixed) Gibbs state [9–13] or, combined with sampling, to construct minimally-entangled thermal states (METTS) [14, 15]. Alternatively, the partition function can be represented as a two-dimensional tensor network, and its contraction can be approximated using tensor renormalization group approaches, for instance as originally proposed in [16–18]. The algorithms can also be generalized for two dimensional systems [19–21].

In this work, we study alternative thermodynamic ensembles that, instead of the von Neumann entropy, maximize the α-Rényi entropy [22],

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log \text{Tr}\, \rho^\alpha \quad (4)$$

at a fixed energy. In the limit $\alpha \to 1$, S_α reduces to the von Neumann entropy. By replacing the von Neumann entropy in Eq. (2) by a Rényi entropy, we define a new "free energy"

$$F_\alpha(\rho) = \text{Tr}(H\rho) - \frac{1}{\beta_\alpha} S_\alpha(\rho). \quad (5)$$

We would like to stress that in general the extremizer ρ_α of this function is not the thermal ensemble. However, as we will show in Sec. II, this new ensemble reproduces nonetheless all local expectation values in the thermodynamic limit. The parameter β_α is not, in general, related to the conventional inverse temperature β, but should be treated as a constant for the optimization.

From a TN perspective, the definition in Eq. (5) offers the possibility of directly performing a minimization, since the Rényi entropies in Eq. (4) are efficiently

computable—at least for small integer values of α. In this work we analyze the properties of such ensembles, in particular how they approximate the thermal properties, and present several variational algorithms which can be used to compute them.

For practical purposes, we will often consider the most convenient case $\alpha = 2$, for which Eq. (5) becomes

$$\rho_R := \arg\min_{\rho \succeq 0} F_R, \quad F_R(\rho) = \mathrm{Tr}(H\rho) + \frac{1}{\beta_R} \log \mathrm{Tr}\, \rho^2, \tag{6}$$

where the subscript R represents $\alpha = 2$. In other words, optimizing Eq. (6) is equivalent to finding the most mixed state at a chosen energy. In applied mathematics, the optimization of such a function is known as a non-linear semi-definite programming, and can be tackled with interior-point methods. However, in many-body quantum physics, the dimension of ρ increases exponentially with the system size, making such approaches impractical for large systems.

This paper is organized as follows. In Sec. II, we provide an analytical solution to Eq. (6), expressed in the eigenbasis of the Hamiltonian. Since the eigenbasis of a many-body system is not always accessible, we propose an optimization strategy based on uniform matrix product states, to approximate the purification of ρ_R directly in the thermodynamic limit. This non-linear optimization can be accelerated using state-of-the-art techniques [23] by restricting it to the Grassmann manifold. This is discussed in detail in Sec. III A, and accompanying numerical experiments to benchmark the algorithm are presented. Moreover, we present an alternative technique, based on a non-linear evolution of the density operator in Sec. III B, which flows towards the desired ensemble. To conclude, we discuss possible developments in Sec. IV.

II. THEORETICAL FRAMEWORK

A. Maximal Rényi ensemble

We now show the analytical form of the extremizer of Eq. (5), which has been previously derived for classical distributions [24–26]. We can use this result in the quantum case noticing that the state that minimizes Eq. (5) must be diagonal in the energy eigenbasis $\{|E_k\rangle\}$ and thus its eigenvalues are equivalent to a probability distribution.

In order to find the coefficients $\{p_k\}$ in the density operator $\rho = \sum_k p_k |E_k\rangle \langle E_k|$, $\rho \succeq 0$ which maximizes the Rényi entropy Eq. (4) under the constraints $\mathrm{Tr}\,\rho = 1$ and $\mathrm{Tr}(H\rho) = \bar{E}$, we introduce the Lagrange multipliers

β_α and γ_α. The functional $\mathcal{L}$ is then

$$\mathcal{L}(\rho) = \frac{1}{1-\alpha} \log \sum_k p_k^\alpha$$
$$- \gamma_\alpha \left(\sum_k p_k - 1 \right) - \beta_\alpha \left(\sum_k E_k p_k - \bar{E} \right). \tag{7}$$

At the stationary point, the parameter γ_α can be eliminated [24], and we obtain the *maximal Rényi ensemble* (MRE)

$$\rho_\alpha = \frac{1}{\mathcal{Z}_\alpha} \Pi_{E_\perp} \left(1 - \beta_\alpha \frac{\alpha-1}{\alpha}(H - \bar{E}) \right)^{\frac{1}{\alpha-1}} \Pi_{E_\perp}, \tag{8}$$

where $\mathcal{Z}_\alpha$ is a normalization factor and $\Pi_{E_\perp}$ is a projector onto the eigenvalues below a cutoff energy $E_\perp := \frac{\alpha}{\beta(\alpha-1)} + \bar{E}$ [27]:

$$\Pi_{E_\perp} = \Theta(E_\perp - H),$$
$$\mathcal{Z}_\alpha = \mathrm{Tr} \left[\Pi_{E_\perp} \left(1 - \beta_\alpha \frac{\alpha-1}{\alpha}(H - \bar{E}) \right)^{\frac{1}{\alpha-1}} \right], \tag{9}$$

where $\Theta(\cdot)$ is the Heaviside function.

To illustrate the behavior of Eq. 8, we show in Fig. 1 the distribution of ρ relative to the eigenbasis. The MRE has a distinctive cutoff energy, beyond which the distribution is zero and therefore fairly different from the case of the canonical ensemble. However, in a many-body system, we have to consider that the density of states is not uniform, but becomes increasingly peaked in the middle of the spectrum. Then the distributions, weighted by the density of states, become much more similar.
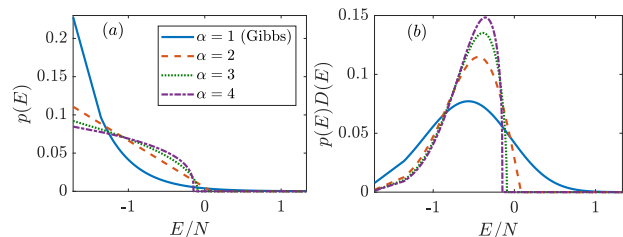

Figure 1. (a) Distribution $p(E)$ of the Maximal Rényi and Gibbs ensembles for different values of α for the Ising model in Eq. (21), with longitudinal and transverse fields respectively $h_x = 1.05$ and $h_z = -0.5$ and system size $N = 10$. The average energy $\bar{E}$ is fixed at $1/4$ of the width of the spectrum. (b) The same distributions weighted with the corresponding density of states $D(E)$, from the approximation in Ref. [28].

Another way of visualizing the relation between the canonical and the Rényi ensembles is to compare their entropies for the same mean energy $\bar{E}$. In Fig. 2 we explicitly show the comparison of von Neumann and 2-Rényi entropies for the ensembles that maximize each of them over the whole energy range for a small system size.
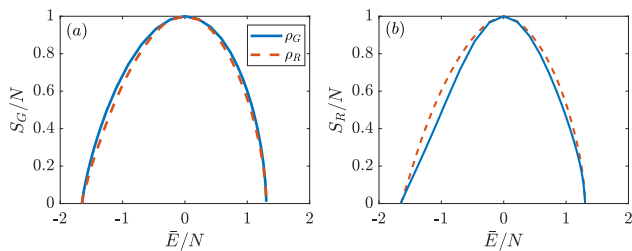
Figure 2. Von Neumann (a) and 2-Rényi (b) entropies for the canonical (solid line) and 2-Rényi (dashed line) ensembles at a given mean energy density, for the same model and parameters as Fig. 1. In both cases, the asymptotic behaviors $\lim_{\beta \to 0} S_G = \lim_{\beta_R \to 0} S_R = 0$ and $\lim_{\beta \to \infty} S_G = \lim_{\beta_R \to \infty} S_R = N$ are recovered.

While the behavior is qualitatively similar, both ensembles only coincide in the limiting cases $\bar{E} = 0$, when the state is maximally mixed (corresponding to the Gibbs ensemble at infinite temperature $\beta = 0$) and $\bar{E} = E_{\min}$ ($E_{\max}$), when the ensemble reduces to the ground (maximally excited) state, corresponding to $\beta \to +\infty$ ($-\infty$).

B. Equivalence of local observables

We now consider a one-dimensional quantum system described by a local Hamiltonian H, an operator in the complex Hilbert space $\mathcal{H}$. This total Hilbert space is formed by the tensor product of N local Hilbert spaces

$$\mathcal{H} = \bigotimes_{n=1}^{N} \mathcal{H}_n \,. \tag{10}$$

The Hamiltonian is restricted to be ℓ-local, i.e. it can then be written in the form

$$H = \sum_{n=1}^{N} h_n, \tag{11}$$

where each h_n acts non-trivially only on sites $n, \ldots, n + \ell - 1$, and has finite operator norm. Additionally, we will assume that almost all local terms satisfy $\|h_n\|_{\mathrm{op}} > 0$, such that the spectrum of H is extensive. For convenience we consider periodic boundary conditions, but open boundary conditions are a special case by choosing $h_N = 0$.

In this setting, it is straightforward to see that the *density of states*

$$D(E) = \delta(H - E) \tag{12}$$

has a variance which scales as $\mathcal{O}(\sqrt{N})$. For specific models, such as strictly 1-local Hamiltonians, it can be shown that $D(E)$ becomes Gaussian in the thermodynamic limit. Under the assumption of a Gaussian density of states, we can then compute the variance of the

energy when we take into account the energy distribution of the ensemble. In the case of the 2-Rényi entropy, it turns out that this can be computed exactly. As described in App. A, in both cases the variances $\langle (\Delta H)^2 \rangle_G$ and $\langle (\Delta H)^2 \rangle_R$ scale as $\mathcal{O}(N)$. Hence, if we think about the normalized energy spectrum, both distributions will be increasingly peaked around the same $\bar{E} = \langle H \rangle$ with a standard deviation $\mathcal{O}(1/\sqrt{N})$ for large N. Hence, the expectation values of local observables become equivalent in the thermodynamic limit. This derives from the correspondence between microcanonical and canonical ensembles [29]. While there exist counterexamples to this correspondence, a sufficient condition for it to hold is that the energy per site converges to a constant [30, 31]. Note that while this argument has been carried out for a Gaussian density of states, we believe that it can be extended to the general case as long as the Hamiltonian is local.

As a final note, we wish to remark that, at least in the case of $\alpha = 2$, we find a correspondence $\beta_R \to \beta$ which holds in the thermodynamic limit. This can be shown in the case of a Gaussian density of states (see App. A), and is observed numerically in both integrable and non-integrable models (see Sec. III A). This is somewhat surprising, since a priori there is no connection between the parameters describing the two different ensembles. From a practical point of view, however, this correspondence is convenient to approximate a thermal ensemble, since we may as well take β_R to be the inverse temperature.

III. VARIATIONAL ALGORITHMS FOR APPROXIMATING THE RÉNYI ENSEMBLE

In this Section, we introduce two different possibilities to obtain numerically the Rényi ensemble in Eq. (8). Although we have a closed form for the exact solution, its use in a many-body setting is impractical because it would require knowledge of the full energy eigenbasis or of the projector in Eq. (8). This motivates the formulation of methods compatible with tensor network techniques. In Sec. III A, we explore how uniform matrix product states can be used to form a purification which represents the density matrix, and its individual tensors can be optimized directly by using techniques from Riemannian optimization. In Sec. III B instead, we propose a non-linear evolution which has Eq. (8) as a fixed point, so that any arbitrary state can be brought to the desired one by simulating this evolution for a sufficient amount of time.

A. Minimization on the MPS manifold

The optimization problem in Eq. (6) can be restricted to the manifold of states described by some class of tensor networks. In particular matrix product states

(MPS) are arguably the most effective ansatz to represent ground states of local, gapped Hamiltonians in one dimension [5, 32–34]. For convenience we consider a uniform MPS, which written in the conventional diagrammatic notation, is

$$|\Psi(A)\rangle = \sum_{\vec{s}} \text{Tr}\left(\ldots A^{s_{n-1}} A^{s_n} A^{s_{n+1}} \ldots\right)|\vec{s}\rangle. \quad (13)$$

Hence, given a local basis $\{\vec{s}\} = \{s_1, \ldots, s_N\}_{s=1,\ldots,d}$, each A is a rank-3 tensor with a physical index of dimension d and two virtual indices contracted with the neighboring tensors, each with dimension D [35–37]. In this Section we focus on uniform MPS for simplicity, but the method can be applied to finite MPS as well.

A natural generalization of MPS for quantum-mechanical operators is a matrix product operator (MPO) [9, 10, 38], which is composed of rank-4 tensors contracted sequentially,

$$O = \quad\quad\quad\quad\quad . \quad (14)$$

The issue with this construction is that it is hard to ensure positivity (if the tensors are over the field $\mathbb{C}$ or $\mathbb{R}$), which is a necessary and physical property for objects like density operators. The problem is that positivity is a global property, which cannot be captured in the local tensors [39–41]. Although an MPO ansatz has been used successfully to approximate the stationary points of dissipative dynamics [42, 43], it is problematic for a variational method since there is no way to vary the local tensors without compromising positivity.

An alternative is to introduce a *locally purified state* [9, 39, 44], which guarantees the positivity of the operator for any local tensor. The construction goes as follows. One considers a pure state, where each site has twice the degrees of freedom, which we call *system* and *ancilla*, so that the local tensor is

$$A = \quad\quad\quad . \quad (15)$$

By tracing out the ancillary degrees of freedom, we obtain a ladder-like tensor network, which represents the density matrix $\rho = \text{Tr}_{\text{anc}} |\Psi\rangle\langle\Psi|$, or graphically

$$\rho = \quad\quad\quad\quad\quad\quad . \quad (16)$$

Shaded boxes represent complex conjugation. It is simple to see that this tensor network is positive semidefinite by

construction. The price to pay is that we have introduced a non-linearity in ρ with respect to the local tensors A^s, so even if the objective function is quadratic in ρ, as in Eq. (6), it will be quartic in the local tensors. Hence we cannot use linear algebra to iteratively optimize the local tensors, as in the case of DMRG [36]. Nonetheless, we can consider the problem in Eq. (6) a non-linear optimization over the tensors of an MPS.

The parametrization of the state in Eq. (13) has an inherent redundancy, since we can perform a gauge transformation on the virtual degrees of freedom of the form $A^s \mapsto X A^s X^{-1}$, for any invertible matrix X. This gauge redundancy of the MPS parametrization allows us to choose the tensors to fulfill the left-gauge condition

$$(17)$$

For the rest of this work, we will often not draw the ancillary degree of freedom, but implicitly assume it is part of the physical leg of each tensor. The tensor ϱ is the (positive-semidefinite) right fixed point of the transfer matrix, which encodes the Schmidt values [36]. Hence, we can view the tensor A as a linear map from the right virtual leg to the left virtual and physical legs which is *isometric*. We will use $W \colon \mathbb{C}^D \mapsto \mathbb{C}^d \times \mathbb{C}^D$ to denote this specific mapping. Alternatively, we can think of W as a matrix, so that we can use the notation $W^\dagger W = \mathbb{1}$ to unambiguously specify the isometricity condition.

Hence we can restrict a generic optimization of an MPS to the optimization of tensors over the Stiefel manifold [45],

$$\text{St}(n,p) = \left\{ W \in \mathbb{C}^{n \times p} | W^\dagger W = \mathbb{1} \right\}. \quad (18)$$

In reality, since there is a unitary freedom remaining in Eq. (17)—namely $A^s \mapsto U^\dagger A^s U$—one can restrict the manifold even further to the Grassmann manifold. The Grassmann manifold should be understood as a quotient manifold, namely all W satisfying the isometricity condition up to a basis rotation, and it is often denoted as $\text{Gr}(n,p) = \text{St}(n,p)/\text{U}(p)$ [45].

In order to optimize a generic function $f(A)\colon \text{Gr}(Dd, D) \to \mathbb{R}$ using any gradient-based optimization, we must be able to compute the gradient with respect to the parameters in A and project it onto the tangent space of the Grassmann manifold. The optimization of differentiable functions on Riemannian manifolds has been the object of extensive studies in mathematics and recently these techniques have been applied to tensor networks [23]. For the self-containedness of this work, we summarize the key ingredients of this optimization in App. B.

For our application, the objective function is given by Eq. (6). For the uniform MPS of Eq. (13), it reduces to

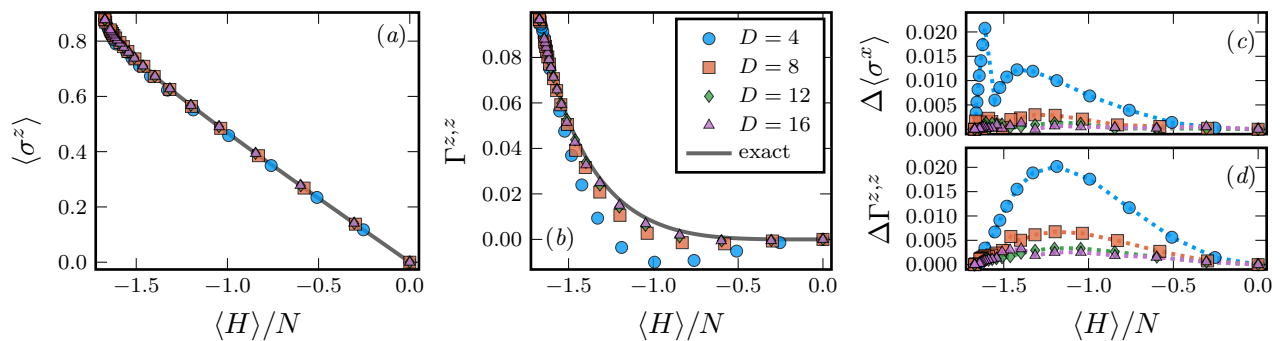$$f_R := \frac{F_R}{N} = \varepsilon + \frac{1}{\beta_R} \log \eta, \quad (19)$$

Figure 3. Magnetization (a) and next-neighbor correlation $\Gamma^{z,z}$ (b) vs. the mean energy density for the Ising model with $h_x = 0$, $h_z = 3/2$, for different bond dimensions D. The dotted line corresponds to the exact results with the same mean energy density. In $(c-d)$, the absolute errors to the exact solution are compared. The $\mathbb{Z}_2$ symmetry is explicitly enforced in the tensors.
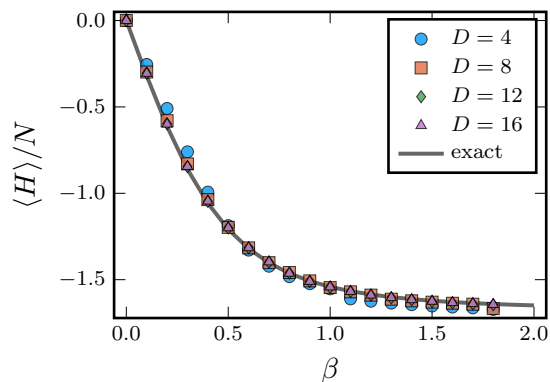

Figure 4. Average energy from Fig. 3 choosing $\beta_R = \beta$.

where $\varepsilon = \text{Tr}(H\rho)/N$ is the energy per site and $\eta = \left(\text{Tr}\,\rho^2\right)^{1/N}$ is the purity per site. Both these terms are computable with standard tensor network routines in polynomial time, for uniform MPS as well as finite MPS. The gradient of Eq. 19 with respect to A is

$$\frac{\partial f_R}{\partial A} = \frac{\partial \varepsilon}{\partial A} + \frac{1}{\beta_R \eta} \frac{\partial \eta}{\partial A}. \qquad (20)$$

As for Eq. (C1), both these quantities $\partial \varepsilon / \partial A$ and $\partial \eta / \partial A$ are simple to obtain, as described in App. C. We thus use this gradient information to perform the optimization on the Riemannian manifold using the l-BFGS algorithm [46, 47]. An open-source implementation of the non-linear optimization in JULIA is available online [48].

To conclude, we note that gradient methods cannot guarantee in any way convergence towards the global minimum, but only some local minimum. While Eq. (6) has a unique solution in the cone of the positive operators, the same cannot be said on a uniform MPS manifold of fixed bond dimension.

1. Numerical experiments

For our numerical experiments, we consider the Ising model

$$H = -\sum_k \left(\sigma_k^x \sigma_{k+1}^x + h_z \sigma_k^z + h_x \sigma_k^x \right). \qquad (21)$$

When the parallel field vanishes ($h_x = 0$), the model is integrable, and local observables and correlations have a closed form [49, 50]. We use this model to perform the optimization of Eq. (6) as described in Sec. III A. The parameter β_R is fixed to different values in the interval $\beta_R \in [0, 2]$, and the uniform MPS is optimized until the gradient is sufficiently small [51].

The results of the optimization are shown in Fig. 3, where we plot some local observables such as the magnetization $\langle \sigma_i^z \rangle$ and next-neighbor correlation $\Gamma^{a,b} = \langle \sigma_i^a \sigma_{i+1}^b \rangle - \langle \sigma_i^a \rangle \langle \sigma_{i+1}^b \rangle$ as a function of the mean energy density of the ensemble. By increasing the bond dimension, we increase the number of the free parameters, and the numerical results converge towards the thermal ones. Additionally, the comparison of the thermal observables by setting $\beta_R = \beta$ is shown in Fig. 4. Up to $\beta_R \lesssim 2$, we observe that there is a correspondence between the two ensembles at $\beta_R = \beta$. For $\beta_R \gtrsim 2$ the optimization of Eq. (19) converges to the ground space exactly, especially at small bond dimensions. In order to study the physics of low temperatures, it is therefore more convenient to reexpress the optimization problem in Eq. (6) by introducing a Lagrange multiplier

$$\rho^* = \arg\min_{\rho \succeq 0} \left\{ \text{Tr}\,\rho^2 + \frac{\lambda^2}{2} \left(\text{Tr}(H\rho) - \bar{E} \right)^2 \right\}. \qquad (22)$$

The gradient (see App. C) can be modified accordingly, and the non-linear optimization can be performed in a similar way. This new objective function gets rid of the dependence on β_R, and one can directly choose an energy to target, since $\lim_{\lambda \to \infty} \text{Tr}(H\rho^*) = \bar{E}$. However, if one wishes to explore the behavior of some observable with
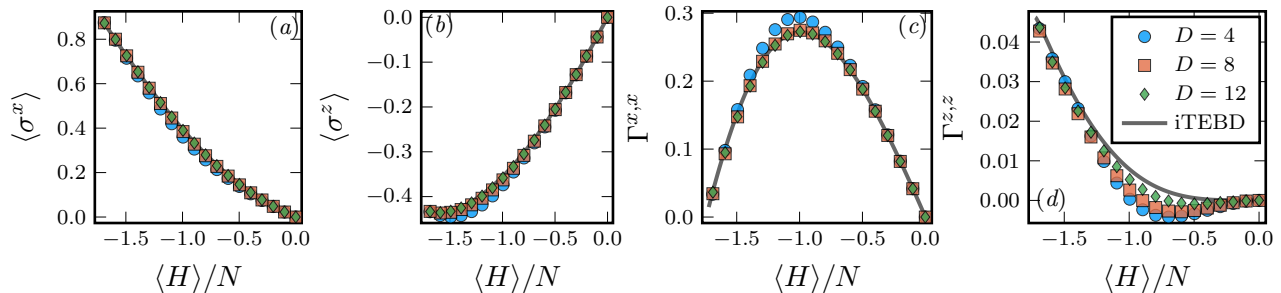
Figure 5. (a–d) Different observables as a function of the mean energy density for the non-integrable case with $h_x = 0.5$, $h_z = -1.05$. The dotted line corresponds to the results given by the iTEBD algorithm.

respect to $\bar{E}$, it is not necessary to perform the extrapolation with $\lambda \to \infty$, but a finite λ is sufficient to obtain an energy in the vicinity of the desired value [52].

We also wish to remark that the method is completely general and does not depend on whether the system is integrable or not. To complete our benchmarks, we present in Fig. 5 a comparison in the case where a parallel field is introduced, making the system non-integrable. In this case exact results are not known, but our results are compared to those of an MPS approximation to the Gibbs state purification obtained with a traditional imaginary time evolution method [9, 53].

B. Non-linear evolution

In Ref. [54], the authors introduced a non-linear evolution for approximating the thermal ensemble with Gaussian states. Here we generalize this idea for the Rényi entropies, which gives rise to an evolution that is efficiently computable with tensor network techniques. We consider a non-linear evolution of a density operator ρ_τ which depends on a real parameter τ

$$\dot{\rho}_\tau := \frac{\partial \rho_\tau}{\partial \tau} = -\frac{1}{2}\left\{\mathcal{I}_\tau - \langle \mathcal{I}_\tau \rangle, \rho_\tau\right\}. \qquad (23)$$

The operator $\mathcal{I}_\tau$ can be chosen such that the fixed point of this evolution gives rise to the MRE. For example, the choice

$$\mathcal{I}_\tau \, \rho_\tau = \beta_R H + \frac{2}{\text{Tr}\,\rho_\tau^2}\rho_\tau \qquad (24)$$

gives rise to the same density operator as Eq. (6). The proof follows similarly from Ref. [54], and it is sufficient to show that the operator $\mathcal{I}_\tau$ in Eq. (24) satisfies the following criteria:

$\text{Tr}\,\rho_\tau = 1, \ \forall \tau \in \mathbb{R}$	Trace conservation	(25a)
$\rho_\tau \succeq 0, \ \forall \tau \in \mathbb{R}$	Positivity conservation	(25b)
$\partial f_R(\rho_\tau)/\partial \tau \leq 0$	Free energy decrease	(25c)

Hence, choosing an appropriate density operator ρ_0 and integrating Eq. (23) over a sufficiently long interval, we

obtain the solution to Eq. (6), since its value can only decrease with time. There is no guarantee of reaching the global minimum—and indeed any eigenstate of H does not evolve under Eq. (23)—but a random choice of the initial state should be sufficient in most cases.
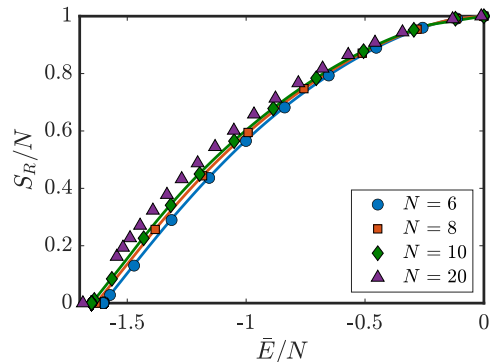

Figure 6. 2-Rényi entropy of the maximal Rényi ensemble obtained with the analytic solution (solid lines) and nonlinear evolution (points). Results are for the Ising model in Eq. (21) with longitudinal and transverse fields respectively $h_x = 0.5$ and $h_z = -1.05$. We also show numerical results for $N = 20$ (triangles) obtained using the non-linear evolution with MPS.

We present some numerical experiments on small system sizes in Fig. 6, where the energy eigenbasis is available. In all cases the numerically integrated density operator converges to the ensemble in Eq. (8). The evolution is discretized by expanding Eq. (23) to first order,

$$\rho_{\tau+\delta\tau} \approx e^{-\frac{\delta\tau}{2}(\mathcal{I}_\tau - \langle \mathcal{I}_\tau \rangle)}\rho_\tau e^{-\frac{\delta\tau}{2}(\mathcal{I}_\tau - \langle \mathcal{I}_\tau \rangle)}. \qquad (26)$$

If the time step is chosen to be sufficiently small, then this evolution will converge to the desired fixed point. This is witnessed by the fact that the Rényi entropy reaches the theoretical maximum for each mean energy, as shown in Fig. 6. As a proof of concept, we also perform the integration using MPS, in particular using the TDVP scheme [55, 56] to update the state at each time step. In practice however, we observe that the time step required to obtain accurate results

scales unfavourably with the system size, and we have yet to fully understand if the evolution becomes ill-conditioned for large system sizes. Notwithstanding, it is possible that different integration schemes allow for large time steps without compromising the stability of the evolution. We leave this as a venue for future work.

IV. OUTLOOK

In this paper, we have introduced a new approach to compute thermal expectation values. Instead of attempting to approximate the minimum of the free energy, we construct a novel ensemble that maximizes the 2-Rényi entropy for the same mean energy, and—in the thermodynamic limit—reproduces local observables of the corresponding Gibbs ensemble.

We have shown that this ensemble can be efficiently approximated using tensor networks and have presented variational algorithms to obtain such an approximation. It is possible to work directly in the thermodynamic limit and use an MPS representation of the ensemble, which optimizes the objective function in Eq. (6). Despite the simple form of this function, the optimization is non-linear and must be tackled with gradient-based methods. The fundamental reason is that the positivity constraint in tensor networks is highly non-local, and one way of enforcing it is via a purification. The convergence can be accelerated with techniques from manifold optimization, but a fundamental limitation is the high contraction cost. Indeed, for a purification of bond dimension D, the time complexity involved in computing the purity (see App. C)

is $\mathcal{O}(D^5)$, which is significantly higher than the typical $\mathcal{O}(D^3)$ for other popular MPS algorithms, such as time evolution or ground state search. Coincidentally, the former is the same leading cost of the original formulation of DMRG with periodic boundary conditions [57]. Although the time-complexity is higher, we observe that a moderate bond dimension captures well the ensemble and its local properties, both in integrable and non-integrable models.

As an alternative to gradient-based optimization, we also propose an alternative method based on a non-linear evolution of the density operator. Under this evolution, the objective function in Eq. (6) is monotonically decreasing, and hence flows to the maximal Rényi ensemble.

Despite these limitations, we believe more efficient cost functions could be devised. Additionally, the ideas outlined here could be applied to other wave-function ansätze. For example, in recent works [58–61] neural networks have been optimized with variational Monte Carlo to describe the steady state of dissipative dynamics. Such techniques could be adapted to perform the optimization described in this work.

[1] E. T. Jaynes, Information Theory and Statistical Mechanics. II, Physical Review **108**, 171 (1957).

[2] E. T. Jaynes, Information Theory and Statistical Mechanics, Physical Review **106**, 620 (1957).

[3] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac, Area Laws in Quantum Systems: Mutual Information and Correlations, Physical Review Letters **100**, 070502 (2008).

[4] T. Kuwahara, Á. M. Alhambra, and A. Anshu, Improved thermal area law and quasi-linear time algorithm for quantum Gibbs states, arXiv:2007.11174 (2020).

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

[6] M. B. Hastings and T. Koma, Spectral Gap and Exponential Decay of Correlations, Communications in Mathematical Physics **265**, 781 (2006).

[7] A. Molnar, N. Schuch, F. Verstraete, and J. I. Cirac, Approximating Gibbs states of local Hamiltonians efficiently with projected entangled pair states, Physical Review B **91**, 045138 (2015).

[8] J. Guth Jarkovský, A. Molnár, N. Schuch, and J. I. Cirac, Efficient Description of Many-Body Systems with Matrix Product Density Operators, PRX Quantum **1**, 010304 (2020).

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

[10] M. Zwolak and G. Vidal, Mixed-state dynamics in one-dimensional quantum lattice systems: A time-dependent superoperator renormalization algorithm, Physical Review Letters **93**, 207205 (2004).

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

[12] B.-B. Chen, Y.-J. Liu, Z. Chen, and W. Li, Series-expansion thermal tensor network approach for quantum lattice models, Physical Review B **95**, 161104 (2017).

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

[14] S. R. White, Minimally Entangled Typical Quantum States at Finite Temperature, Physical Review Letters **102**, 190601 (2009).

[15] E. M. Stoudenmire and S. R. White, Minimally entangled typical thermal state algorithms, New Journal of Physics **12**, 055026 (2010).

[16] T. Nishino, Density Matrix Renormalization Group Method for 2D Classical Models, Journal of the Physical Society of Japan **64**, 3598 (1995).

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

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

[19] P. Czarnik, L. Cincio, and J. Dziarmaga, Projected Entangled Pair States at Finite Temperature: Imaginary Time Evolution with Ancillas, Physical Review B **86**, 245101 (2012).

[20] P. Czarnik and J. Dziarmaga, Fermionic projected entangled pair states at finite temperature, Physical Review B **90**, 035144 (2014).

[21] A. Kshetrimayum, M. Rizzi, J. Eisert, and R. Orús, Tensor network annealing algorithm for two-dimensional thermal states, Physical Review Letters **122**, 070502 (2019).

[22] A. Rényi, On measures of entropy and information, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (The Regents of the University of California, 1961).

[23] M. Hauru, M. Van Damme, and J. Haegeman, Riemannian optimization of isometric tensor networks, arXiv:2007.03638 (2020).

[24] A. Bashkirov, Maximum Rényi entropy principle for systems with power-law Hamiltonians, Physical Review Letters **93**, 130601 (2004).

[25] D. C. Brody, I. R. Buckley, and I. C. Constantinou, Option price Calibration from Rényi entropy, Physics Letters A **366**, 298 (2007).

[26] C. Bunte and A. Lapidoth, Maximizing Rényi entropy rate, in *2014 IEEE 28th Convention of Electrical & Electronics Engineers in Israel (IEEEI)* (IEEE, 2014) pp. 1–4.

[27] Symmetrically, there is also a solution with a projector onto energies above the cutoff energy: $E > E_\perp$. For simplicity we ignore this solution, as it is the Rényi-equivalent of negative temperatures.

[28] Y. Y. Atas and E. Bogomolny, Spectral density of a one-dimensional quantum Ising model: Gaussian and multi-Gaussian approximations, Journal of Physics A: Mathematical and Theoretical **47**, 335201 (2014).

[29] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part 1*, 3rd ed., Course of Theoretical Physics No. Vol. 5 (Butterworth Heinemann, Oxford, 1991).

[30] H. Touchette, *Equivalence and Nonequivalence of the Microcanonical and Canonical Ensembles: A Large Deviations Study*, PhD Thesis (2003).

[31] R. S. Ellis, H. Touchette, and B. Turkington, Thermodynamic versus statistical nonequivalence of ensembles for the mean-field Blume–Emery–Griffiths model, Physica A: Statistical Mechanics and its Applications **335**, 518 (2004).

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

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

[34] Y. Huang, Computing energy density in one dimension, arXiv:1505.00772 (2015).

[35] F. Verstraete, V. Murg, and J. Cirac, Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems, Advances in Physics **57**, 143 (2008).

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

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

[38] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, Matrix product operator representations, New Journal of Physics **12**, 025012 (2010).

[39] G. De Las Cuevas, N. Schuch, D. Pérez-García, and J. Ignacio Cirac, Purifications of multipartite states: Limitations and constructive methods, New Journal of Physics **15**, 123021 (2013).

[40] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert, Locality of Temperature, Physical Review X **4**, 031019 (2014).

[41] G. De las Cuevas, T. S. Cubitt, J. I. Cirac, M. M. Wolf, and D. Pérez-García, Fundamental limitations in the purifications of tensor networks, Journal of Mathematical Physics **57**, 071902 (2016).

[42] E. Mascarenhas, H. Flayac, and V. Savona, Matrix-product-operator approach to the nonequilibrium steady state of driven-dissipative quantum arrays, Physical Review A **92**, 022116 (2015).

[43] J. Cui, J. I. Cirac, and M. C. Bañuls, Variational Matrix Product Operators for the Steady State of Dissipative Quantum Systems, Physical Review Letters **114**, 220601 (2015).

[44] A. H. Werner, D. Jaschke, P. Silvi, M. Kliesch, T. Calarco, J. Eisert, and S. Montangero, Positive Tensor Network Approach for Simulating Open Quantum Many-Body Systems, Physical Review Letters **116**, 237201 (2016).

[45] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds* (Princeton University Press, Princeton, N.J. ; Woodstock, 2008).

[46] D. C. Liu and J. Nocedal, On the limited memory BFGS method for large scale optimization, Mathematical Programming **45**, 503 (1989).

[47] J. Nocedal and S. J. Wright, eds., *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering (Springer-Verlag, New York, 1999).

[48] https://github.com/giacomogiudice/RenyiOptimization.jl.

[49] S. Katsura, Statistical Mechanics of the Anisotropic Linear Heisenberg Model, Physical Review **127**, 1508 (1962).

[50] T. Niemeijer, Some exact calculations on a chain of spins, Physica **36**, 377 (1967).

[51] The optimization halts after the norm of the gradient vector in tangent space is smaller than 10^{-6}.

[52] In our simulations, we set $\lambda = 10$. Since the purity per site $0.5 \leq \eta \leq 1$ is order 1, we expect deviations in energy density around $\mathcal{O}(1/\lambda^2)$.

[53] G. Vidal, Efficient Classical Simulation of Slightly Entan-

gled Quantum Computations, Physical Review Letters **91**, 147902 (2003).

[54] T. Shi, E. Demler, and J. I. Cirac, Variational Approach for Many-Body Systems at Finite Temperature, Physical Review Letters **125**, 180602 (2020).

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

[56] L. Vanderstraeten, J. Haegeman, and F. Verstraete, Tangent-space methods for uniform matrix product states, SciPost Physics Lecture Notes , 7 (2019).

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

[58] N. Yoshioka and R. Hamazaki, Constructing neural stationary states for open quantum many-body systems, Physical Review B **99**, 214306 (2019).

[59] M. J. Hartmann and G. Carleo, Neural-Network Approach to Dissipative Quantum Many-Body Dynamics, Physical Review Letters **122**, 250502 (2019).

[60] A. Nagy and V. Savona, Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems, Physical Review Letters **122**, 250501 (2019).

[61] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Variational Neural-Network Ansatz for Steady States in Open Quantum Systems, Physical Review Letters **122**, 250503 (2019).

[62] M. Hartmann and O. Hess, Gaussian quantum fluctuations in interacting many particle systems, Letters in Mathematical Physics **68**, 103 (2004).

[63] M. Hartmann, G. Mahler, and O. Hess, Spectral Densities and Partition Functions of Modular Quantum Systems as Derived from a Central Limit Theorem, Journal of statistical physics **119**, 1139 (2005).

[64] J. Keating, N. Linden, and H. Wells, Spectra and eigenstates of spin chain Hamiltonians, Communications in Mathematical Physics **338**, 81 (2015).

[65] A. Edelman, T. A. Arias, and S. T. Smith, The Geometry of Algorithms with Orthogonality Constraints, SIAM Journal on Matrix Analysis and Applications **20**, 303 (1998).

[66] V. Zauner-Stauber, L. Vanderstraeten, M. T. Fishman, F. Verstraete, and J. Haegeman, Variational optimization algorithms for uniform matrix product states, Physical Review B **97**, 045145 (2018).

Appendix A: Calculation of the variance for a Gaussian density of states

In this Section we compute the mean energy $\langle H \rangle$ and variance $\langle (\Delta H)^2 \rangle = \langle H^2 \rangle - \langle H \rangle^2$ for the Gibbs and 2-Rényi ensembles, assuming the density of states is a Gaussian of the form

$$D(E) = \exp\left(-\frac{(E - E_{\mathrm{mid}})^2}{2\sigma^2 N}\right), \qquad (A1)$$

where N is the system size and σ is a constant independent of N. Additionally, without loss of generality, let us assume it is centered at $E_{\mathrm{mid}} = 0$.

For local Hamiltonians as Eq. (11), it was shown that the density of states weakly converges to a Gaussian in the thermodynamic limit, as a consequence of Lyapunov's central limit theorem [62–64]. As an alternative proof, once can take a an ancillary copy of the system, and consider the state $|\Xi\rangle$ which is the tensor product of maximally entangled pairs between system and ancilla:

$$|\Xi\rangle = \bigotimes_n \frac{1}{\sqrt{|\mathcal{H}_n|}} \sum_{i=1}^{|\mathcal{H}_n|} |i\rangle_{\mathrm{sys}} |i\rangle_{\mathrm{anc}}. \qquad (A2)$$

In the doubled system, the state $|\Xi\rangle$ is a product state, and one can apply directly the Theorem in Ref. [62] to obtain the desired result.

However, the rate of convergence to the central limit theorem is larger than $\mathcal{O}(1/\sqrt{N})$, and one should take into account the finite-size corrections when computing expectation values. Hence, we can think of Eq. (A1) as an toy model of actual local Hamiltonians, and derive results under this assumption.

For the Gibbs ensemble, we have that the partition function is

$$\mathcal{Z}_G = \int_{-\infty}^{+\infty} e^{-\beta E} D(E)\, \mathrm{d}E. \qquad (A3)$$

This leads to

$$\langle H \rangle_G = -\frac{\partial}{\partial \beta} \log \mathcal{Z}_G = -\beta \sigma^2 N, \qquad (A4a)$$

$$\langle (\Delta H)^2 \rangle_G = \frac{\partial^2}{\partial \beta^2} \log \mathcal{Z}_G = \sigma^2 N. \qquad (A4b)$$

Naturally, these results hold only in the region around the peak of the Gaussian, and break down when one tries to take the limit of $\beta \to \infty$.

For the 2-Rényi distribution, we cannot use the trick of deriving the partition function with respect to β, since we cannot interpret it as a generating function. We can however express everything in terms of the truncated moments

$$\Phi_m := \int_{-\infty}^{E_\perp} E^m D(E)\, \mathrm{d}E. \qquad (A5)$$

The upper integration limit is related to the mean energy and β_R as $E_\perp = \bar{E} + \frac{2}{\beta_R}$. These moments enjoy a recurrence relation of the form $\Phi_{m+2} = \sigma^2 \partial \Phi_m / \partial \sigma$. Additionally, Φ_1 is analytical because the integrand is the derivative of a Gaussian. This allows us to establish the identities

$$\Phi_2 = \sigma^2 N \Phi_0 + E_\perp \Phi_1, \qquad (A6a)$$

$$\Phi_3 = \left(2\sigma^2 N + E_\perp^2\right) \Phi_1. \qquad (A6b)$$

By dividing the partition function by $\beta/2$, we can then compute the mean energy for this ensemble as

$$\langle H \rangle_R = \frac{E_\perp \Phi_1 - \Phi_2}{E_\perp \Phi_0 - \Phi_1} = -\sigma^2 N \frac{\Phi_0}{E_\perp \Phi_0 - \Phi_1}. \qquad (A7)$$
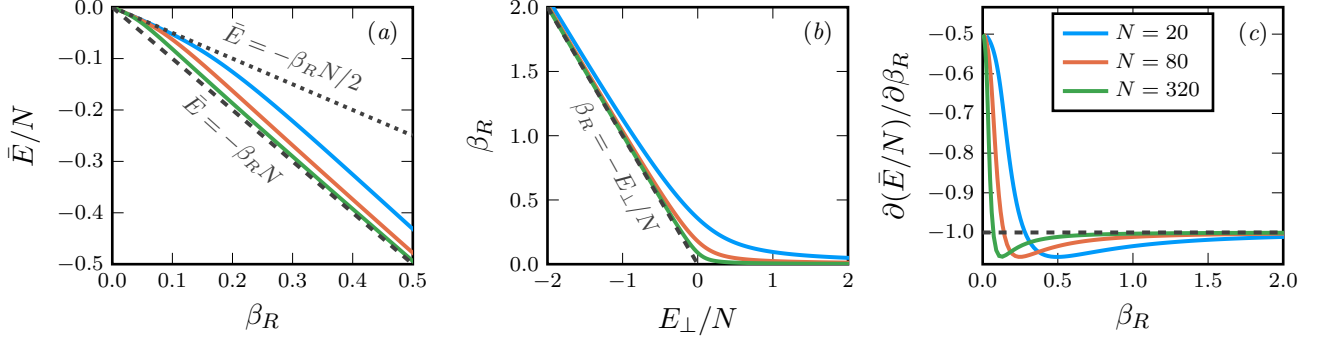
Figure 7. (a) The energy density of the MRE as a function of the parameter β_R, for increasing system sizes. For clarity, in all plots we set $\sigma^2 = 1$. In the high-temperature limit ($\bar{E} \to 0$), the behavior $\bar{E} = \beta N/2$ (dotted line) indicates that $\beta_R \approx 2\beta$. However, as the system size is increased, at non-zero energy density the equation of state approaches $\bar{E} = \beta N$, which is analogous to the Gibbs ensemble. Hence, in the limit of $N \to \infty$ we can identify $\beta_R \approx \beta$. This can be understood in (b), for which $\beta_R \approx \max(E_\perp/N, 0) + \mathcal{O}(1/\sqrt{N})$. (c) Taking the derivative $\partial(\bar{E}/N)/\partial\beta_R$, we identify a fast-varying regime, corresponding to $E_\perp < 0$, and a second regime at $E_\perp > 0$ where the derivative is close to -1.

Equating this result to $\bar{E}$ allows us to express Φ_0 in terms of Φ_1

$$\Phi_0 = \frac{\bar{E}}{\bar{E}E_\perp + \sigma^2 N}\Phi_1. \tag{A8}$$

Using this last relation, we can write the variance as

$$\langle(\Delta H)^2\rangle_R = \bar{E}E_\perp + 2\sigma^2 N - \bar{E}^2 = 2\sigma^2 N + \frac{2\bar{E}}{\beta_R}. \tag{A9}$$

For the Gibbs ensemble, notice that β and $\bar{E}$ are co-linear, $\bar{E} = -\beta\sigma^2 N$. At infinite temperature ($\bar{E} = 0$), we have trivially $\beta = \beta_R = 0$. Expanding ρ_G around $\beta = 0$, we obtain $\rho_G \approx 1 - \beta H$. Comparing this with the form of the MRE, we can easily conclude that in this limit $\beta_R \approx 2\beta$. However, one should take into account the thermodynamic limit. Indeed, as shown in Fig. 7, at non-zero $\bar{E}$, increasing the system size leads to an equation of state which asymptotically approaches $\bar{E} = -\beta_R\sigma^2 N$. This is due to the fact that the cutoff $E_\perp$ becomes proportional to β_R. Indeed, at $E_\perp = 0$, one has that $\beta_R(E_\perp = 0) = \mathcal{O}(1/\sqrt{N})$. The point $E_\perp = 0$ also corresponds to a stationary point of $\partial\bar{E}/\partial\beta_R$. Taking derivatives, one obtains a relation between $\bar{E}$ and β only

$$\frac{\partial\bar{E}}{\partial\beta_R} = \frac{\partial\bar{E}}{\partial E_\perp}\frac{\partial E_\perp}{\partial\beta_R} = \frac{\sigma^2 N}{\beta_R^2}\frac{1 + 2\frac{\bar{E}}{\beta_R\sigma^2 N}}{1 + \frac{\bar{E}}{\beta_R\sigma^2 N}}. \tag{A10}$$

As the system size is increased, the derivative converges towards a constant, as shown in Fig. 7. This allows us to conclude that, for a Gaussian density of states and $\beta_R \gg 1/\sqrt{N}$, we have $\beta_R \approx \beta$.

Appendix B: Technical details on Grassmann manifolds

The gradient of a function on a Riemannian manifold belongs to the tangent space of the manifold itself. A generic tangent vector to a uniform MPS is a linear combination of the partial derivative with respect to the single tensor. This can be seen as a vector embedded in Hilbert space, composed of an (infinite) sum of MPS vectors

$$|\Delta(B)\rangle = B\frac{\partial}{\partial A}|\Psi(A)\rangle$$

$$\tag{B1}$$

where the sum runs over all physical sites.

A tangent vector parametrized by a tensor B has an inherent gauge freedom to it. The explicit transformation that leaves the vector invariant is $B^s \mapsto B^s + XA^s - A^sX$, for any $D \times D$ matrix X. Indeed, the set of derivatives $|\partial_\mu\Psi(A)\rangle$ form an overcomplete basis. Hence, by introducing the orthogonal complement of A [23], such that
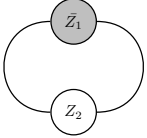

$$\tag{B2}$$

we can parametrize the tangent vectors as
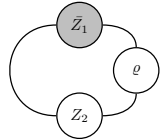
$$\tag{B3}$$

where Z is a $D(d-1) \times D$ matrix. This parametrization arises quite naturally if one considers the tangent vectors to be embedded in the original Hilbert space. In this case, the choice of Eq. (B3) corresponds to imposing orthogonality of the tangent space, $\langle \Psi(A)|\Delta(B)\rangle = 0$. We remark that Eq. (B3) is exactly the parametrization of the tangent space for Grassmann manifolds, as derived traditionally [45].
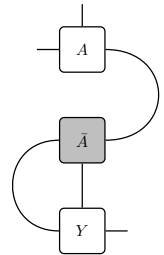
Hence we can consider the problem Eq. (6) as an optimization of a tensor A over the Grassmann manifold. One particularity arises from the choice of metric in the tangent space. In Riemannian manifold optimization, one usually chooses the Euclidean metric [45]

$$\langle \Delta_1(B_1)|\Delta_2(B_2)\rangle_{\text{Eucl}} = \text{Re}\,\text{Tr}(B_1^\dagger B_2)$$

$$= \text{Re} \qquad . \quad \text{(B4)}$$

However this is not the most natural choice in this setting, since the underlying physical Hilbert space prescribes the metric

$$\langle \Delta_1(B_1)|\Delta_2(B_2)\rangle_{\text{Hilb}} \propto \text{Re}\,\text{Tr}(B_1^\dagger B_2 \varrho)$$

$$= \text{Re} \qquad . \quad \text{(B5)}$$

Note that, as opposed to Eq. (B4), this metric depends on the current "point" $|\Psi(A)\rangle$ of the manifold. In practice we notice that the choice of the metric is not very important for the optimization, and the Euclidean metric poses the advantage of not having to invert a potentially ill-conditioned fixed point ϱ when projecting onto the manifold. Additionally, the use of the Euclidean metric is not necessarily deleterious since, compared to Eq. (B5), it will magnify the importance of small Schmidt values of the state $|\Psi(A)\rangle$. In Ref. [23], the authors have proposed a non-linear preconditioner that acts as a compromise between these two metrics. Regardless of our choice, the metric allows us to project arbitrary Hilbert space vectors onto the tangent space. The projection operator for the Euclidean metric in Eq. (B4) reads [65]

$$\mathcal{P}_A(Y) = \qquad - \qquad . \quad \text{(B6)}$$

For the optimization of a generic function $f(A)$: $\text{Gr}(Dd, D) \to \mathbb{R}$, we can compute the gradient $\nabla f(A)$ without taking into account the isometricity condition and then projecting onto the tangent space [45]. This projected gradient $\text{grad}\,f(A) := \nabla f(A)$, can be considered the direction of steepest ascent on the manifold, while its magnitude can be used as a convergence criterion.

The last ingredient necessary for a gradient descent algorithm is defining a *retraction*. Loosely speaking, we need to define a curve on the manifold such that we can move in a direction specified by a tangent vector by a "step size" α. Hence a retraction $\mathcal{R}_A(\alpha\Delta)$ can be any smooth curve such that the it (i) starts at A, $\mathcal{R}_A(0) = A$ and (ii) is consistent with $\text{d}\,\mathcal{R}(\alpha\Delta)/\text{d}\alpha = \Delta$. Different choices of retraction exist, but the most natural choice is a retraction that follows the manifold geodesics, i.e. the shortest path that connects two points on the manifold. Remarkably, the geodesic retraction on a Grassmannian manifold relative to the Euclidean metric is known and is relatively efficient to compute [65]: given some point A, the retraction of some tangent vector $\Delta = A^\perp Z$ (see Eq. (B3)) is

$$\mathcal{R}_A(\alpha\Delta) = e^{\alpha Q}A, \quad \text{(B7)}$$

where

$$Q = \begin{pmatrix} A & A^\perp \end{pmatrix} \begin{pmatrix} 0 & -Z^\dagger \\ Z & 0 \end{pmatrix} \begin{pmatrix} A^\dagger \\ A^{\perp\dagger} \end{pmatrix}. \quad \text{(B8)}$$

This consistutes the bare minimum to define a gradient descent algorithm on the Grassmann manifold. In practice, the convergence of gradient descent can be very slow, and, in Euclidean space, several methods that just use first-order information. For example, conjugate gradient adjusts the gradient with the previous search direction, and quasi-Newton methods—notably l-BFGS [46, 47]—uses the previous iterations to create a low-rank approximation of the inverse Hessian. In order to adapt these methods to optimization on manifolds it is sufficient to define a *vector transport*, a way of transporting a tangent vector at a previous point of the manifold to the current one. In other words, for a retraction $A' = \mathcal{R}_A(\Delta)$, a vector transport $\mathcal{T}_\Delta(\Omega)$ maps a tangent vector Ω at A to a tangent vector at A'. A typical way of defining transport is via differentiated retraction, i.e.

$$\mathcal{T}_\Delta(\Omega) = \left.\frac{\text{d}}{\text{d}\alpha}\mathcal{R}_A(\Delta + \alpha\Omega)\right|_{\alpha=0}. \quad \text{(B9)}$$

Using Eq. (B7), we obtain

$$\mathcal{T}_{\alpha\Delta}(\Omega) = e^{\alpha Q}\Omega. \quad \text{(B10)}$$

Appendix C: Calculation of the gradient

For simplicity, we assume that the Hamiltonian is 2-local and $h_n = h \,\forall N$, but the algorithm can be readily generalized to a non-trivial unit cell and any Hamiltonian which has an MPO form [66]. Once the leading fixed point ϱ is computed, the energy density reduces to the following network:

$$\varepsilon = \quad \text{[diagram]} \quad . \tag{C1}$$

In order to compute the derivative $\partial\varepsilon/\partial\bar{A}$ of Eq. (C1) it is useful to define the left and right environments corresponding to the geometric sum of the terms in Hamiltonian over each half-infinite chain

$$H_L = \quad \text{[diagram]} \quad , \qquad H_R = \quad \text{[diagram]} \quad , \qquad \mathbb{E} = \quad \text{[diagram]} \quad . \tag{C2}$$

The notation $(1-\mathbb{E})^{\mathrm{P}} = (1-\mathbb{E}+|\varrho\rangle\!\rangle\langle\!\langle\mathbb{1}|)^{-1}$ is used to denote the geometric where the divergent part—corresponding to the leading eigenpair—is subtracted [66]. Hence the gradient $\partial\varepsilon/\partial\bar{A}$, without accounting for the constraint is

$$\frac{\partial\varepsilon}{\partial\bar{A}} = \quad \text{[diagram]} \quad + \quad \text{[diagram]} \quad + \quad \text{[diagram]} \quad + \quad \text{[diagram]} \quad . \tag{C3}$$

When computing the purity, we can retain the leading eigenvalues and eigenvectors of the transfer element of ρ^2

$$\Sigma_L = \eta \quad \Sigma_L \quad \text{[diagram]} \quad , \qquad \Sigma_R = \eta \quad \text{[diagram]} \quad \Sigma_R \quad , \qquad \Sigma_L \quad \Sigma_R = 1. \tag{C4}$$

This is by far the costliest computional step relative to the bond dimension D, since it scales as $\mathcal{O}(D^5)$ as opposed to the other steps which are all $\mathcal{O}(D^3)$. The gradient $\partial\eta/\partial\bar{A}$ becomes

$$\frac{\partial\eta}{\partial\bar{A}} = \quad \Sigma_L \quad \text{[diagram]} \quad \Sigma_R \quad + \quad \Sigma_L \quad \text{[diagram]} \quad \Sigma_R \quad . \tag{C5}$$

We can then put together Eq. (C3) and Eq. (C5) to obtain the gradient $\partial f_R/\partial\bar{A}$ in Eq. (20). In order to compute the gradient on the Grassmann manifold, we must then project the unconstrained gradient using Eq. (B6).