Ground-state properties via machine learning quantum constraints

Pei-Lin Zheng*, Si-Jing Du*, and Yi Zhang*

International Center for Quantum Materials, Peking University, Beijing, 100871, China and
School of Physics, Peking University, Beijing, 100871, China

Ground-state properties are central to our understanding of quantum many-body systems. At first glance, it seems natural and essential to obtain the ground state before analyzing its properties; however, its exponentially large Hilbert space has made such studies costly, if not prohibitive, on sufficiently large system sizes. Here, we propose an alternative strategy based upon the expectation values of an ensemble of operators and the elusive yet vital quantum constraints between them, where the search for ground-state properties simply equates to classical constrained minimization. These quantum constraints are generally obtainable via sampling and then machine learning on a large number of systematically consistent quantum many-body states. We showcase our perspective on 1D fermion chains and spin chains for applicability, effectiveness, caveats, and unique advantages, especially for strongly correlated systems, thermodynamic-limit systems, property designs, etc.

Introduciton—The collective behaviors of quantum many-body systems are central to various cutting-edge fields in condensed matter physics and beyond. Despite the nominal simplicity of certain quantum Hamiltonians, e.g., the Hubbard model [1–3], the non-commuting quantum operators squander any advantageous basis, and the exponentially-large Hilbert space renders the solutions and characterizations of ground states costly, limiting the system size and geometry in numerical techniques, e.g., exact diagonalization and density matrix renormalization group (DMRG) [4, 5]. While quantum Monte Carlo methods introduce efficient samplings, they are limited to sign-problem-free cases [6–8]. Also, the ground state solution usually starts from scratch upon slight model modifications, making the systematic study of a complex phase diagram, not uncommon in condensed matter physics [9], even more expensive.

Rather than the abstract quantum many-body ground state, we are usually interested in its properties, such as the ground-state energy and spontaneous-symmetry-breaking order parameters - (linear combinations of) expectation values of target observables. Considering that the minimum energy criteria also concern expectation values, one would be prompt to establish a study based solely on expectation values and cut out the ground state. However, the quantum operators follow nontrivial commutation relations and, as a result, enforce nontrivial quantum constraints upon their expectation values - a role played by the ground state as the mediator. Expectation values violating these quantum constraints do not have an underlying quantum state and may not reflect the true nature of the quantum many-body system. Therefore, such quantum constraints are complicated yet essential for proper expectation-value-based considerations. An example of such quantum constraints is the conformal bootstrap for conformal field theories [10], beyond which, however, the bootstrap reduces to mere bounds and no longer offers a controlled analysis of ground-state properties [11, 12].

On the other hand, recent developments in machine learning [13, 14] have revolutionized data analysis such as image recognition, spam and fraud detection, and autonomous driving [15]. Artificial neural networks (ANNs) can grasp the key yet complex and hidden rules in big datasets and generalize accurately for future scenarios [13–15]. Recently, machine learning has witnessed many explorations at the quantum many-body physics frontier, including quantum state tomography [16, 17], quantum phase recognition [18–25], neural network states [26, 27], experiment interpretations [28–30], etc.

In this letter, we propose studying the ground-state properties of quantum many-body systems within a classical expectation-value framework with quantum constraints over an ensemble of important operators. Then, the ground-state properties amount to constrained minimization. We can generally encode such quantum constraints as ANNs via supervised machine learning on example quantum states. Without loss of generality, we showcase the unique advantages of our strategy on 1D fermion and spin-1/2 models: (1) Compared with the expensive procedure of solving quantum many-body states, evaluations of expectation values are efficient and easily parallelizable among multiple operators and quantum states. (2) Our main effort is to extract and apply the quantum constraints through a large, classical dataset of expectation values, where machine learning techniques excel compatibly and proficiently. (3) Given a sufficiently diverse and representative training set, the obtained quantum constraints work for all Hamiltonians with different parameters, where we iterate the classical constrained minimization with respect to the expectation values of different Hamiltonians. (4) We embed systematic properties such as system size, geometry, and dimensions into the sample quantum many-body states and are rarely limited by them. (5) The quantum constraints also exhibit the competition and symbiosis between the observables, offering recipes for engineering models for desired ground-state properties, even emergent phases.

Algorithm—Our approach consists of steps as follows:
• Start with a large and representative ensemble of quantum many-body states \{\langle \Phi \rangle_\alpha \} systematically consistent with the potential ground state, namely, obeying the expected symmetries and the area law.

• For each \langle \Phi \rangle_\alpha, evaluate the expectation values of a set of operators \{\hat{O}_j\} and contribute a physical data point \langle \hat{O} \rangle_\alpha = \langle \langle \hat{O}_1 \rangle_\alpha, \langle \hat{O}_2 \rangle_\alpha, \ldots \rangle in the \langle \hat{O} \rangle space. Operators with lower orders and spatial extents receive priority due to larger relevance and compatibility with local Hamiltonians. For comparison, unphysical data is obtained by considering deviations from the physical data [31].

• Via supervised machine learning on the training set \{\langle \hat{O} \rangle_\alpha \}, train ANNs \hat{f}(\langle \hat{O} \rangle) to distinguish physical (unphysical) values of \langle \hat{O} \rangle that is allowed (disallowed) by the quantum constraints.

• For the Hamiltonian \hat{H} = \sum_j a_j \hat{O}_j, search the constrained minimum of the energy \hat{E} = \sum_j a_j \langle \hat{O}_j \rangle with the quantum constraints \hat{f}(\langle \hat{O} \rangle). The coordinates \langle \hat{O} \rangle_0 of the resulting minimum offer the expectation values that characterize the ground state.

The first three steps yield the quantum constraints \hat{f}(\langle \hat{O} \rangle) that mark the physical manifold in the classical \langle \hat{O} \rangle space. We expect a relatively smooth and continuous manifold, as the adiabatic theorem ensures that the quantum many-body ground states and the corresponding \langle \hat{O} \rangle evolve continuously in the absence of first-order phase transitions. Importantly, the area law and symmetries vastly reduce the pool of quantum many-body states from the original Hilbert space, and machine learning can summarize and generalize from a limited number of training samples [13, 14], making it feasible to extract the quantum constraints through a polynomial amount of sample states. Since evaluating expectation values is simple and efficient, the key is to obtain a diverse training set representative of the candidate parts of the Hilbert space, e.g., by teaming up multiple quantum many-body ansatzes.

Only the final step that applies the quantum constraints to the model Hamiltonians is repeated throughout a parameter space. Although the classical constrained optimizations are not guaranteed to be fully straightforward, compared with the exponential expense of brute-force quantum algorithms, the overall cost can be much less, especially given the available optimization algorithms and physical intuitions. For example, the solution \langle \hat{O} \rangle_0 for one set of model parameters helps to initialize searches for its neighbors, as \langle \hat{O} \rangle_0 changes continuously in the absence of transitions. For efficiency, we can start with models with exact solutions or controlled approximations [32] and move progressively into other parts of the parameter space, tracking \langle \hat{O} \rangle_0 successively in the process [31].

An important question is the choice of observables \{\langle \hat{O}_j \rangle\}, for which we suggest two criteria: (1) Is the observable likely to appear in target Hamiltonians? (2) Does the observable represent a physical quantity we are interested in? These favor local, low-order operators, and the more, the better, though with added costs. Also, these criteria are soft: irrespective of chosen observables, quantum constraints address the physical realizability of their expectation values in a yes/no fashion: they do lose capacity without certain observables, e.g., tell corresponding degeneracy, but encounter no algorithmic breakdown.

A heuristic example—First, let’s consider a 1D Fermi sea between \(k_L = k_0 - k_F\) and \(k_R = k_0 + k_F\), where we have a simple analytical expression for the quantum constraints. Its expectation values of two-point correlators are:

\[
C_0 = \langle c_x^\dagger c_x \rangle = k_F/\pi \\
C_r = \langle c_{x+r}^\dagger c_x \rangle = \sin(k_Fr)e^{ik_0r}/\pi r, r \neq 0,
\]

irrespective of \(x\) due to the translation symmetry. Expectation values of higher-order operators depend fully on \(C_r\)’s through Wick’s theorem. In particular, the following quantum constraint holds between the most dominant real-valued \(C_0\) and complex-valued \(C_1\):

\[
\pm \pi |C_1| = \sin(\pi C_0),
\]

as illustrated in Fig. 1, which holds as long as there is one Fermi sea and no spontaneous translation symmetry breaking. We note that the physical manifold (black contour) is smooth except for the two endpoints at \(C_0 = 0, 1\), corresponding to Van Hove singularities.

Now, let’s consider a tight-binding Hamiltonian with nearest-neighbor hopping \(t \in \mathbb{R}\) [33] and Fermi energy \(\mu\):

\[
\hat{H} = \sum_x -t(c_{x+1}^\dagger c_x + c_x^\dagger c_{x+1}) + \mu c_x^\dagger c_x \\
E = (-2t\text{Re}(C_1) + \mu C_0) \times N,
\]

where \(N\) is the system size, and we set \(t = 1\) as our unit of energy hereinafter. As the constraint in Eq. 2 only concerns \(|C_1|\), we set \(\text{Im}(C_1) = 0\) to allow maximal range for Re\((C_1)\). A schematic plot for the solutions of optimal values of (Re\((C_1), C_0\)) is in Fig. 1. More rigorously, we define \(\pi C_0 = y \in (0, \pi)\), and \(\text{Re}(C_1)/C_0 = f(y) = \sin(y)/y \in (0, 1)\) is a single-valued function following the quantum constraint. To minimize \(E \propto |\mu - 2t f(y)| \times y\), the subsequent solutions:

\[
2tf(y) - \mu + 2tf'(y)y = 0 \Rightarrow 2t \cos(y) = \mu \\
C_0 = y/\pi = \arccos(\mu/2t)/\pi \\
\text{Re}(C_1) = yf(y)/\pi = \sin(y)/\pi = \text{sgn}(t)\sqrt{1-\mu^2/4t^2}/\pi
\]

which are consistent with the exact results obtained in the momentum space \(H = \sum_k [-2t \cos(k) + \mu]c_k^\dagger c_k\) for generic values of \(t\) and \(\mu\).
For ground-state properties of Hamiltonians with up to $n^{th}$-nearest-neighbor hopping, we need to employ quantum constraint $C_0 = f(C_1/C_0, C_2/C_0, \cdots)$ on the expectation values $C_i$, $i = 0, 1, 2, \cdots, n_f$, which can be represented by ANNs and trained via supervised machine learning on quantum states with multiple Fermi seas \cite{31}. For general quantum many-body systems, we may not formulate the quantum constraints as a function between the expectation values. It is more convenient to establish a ‘penalty’ function $f(\langle \hat{O} \rangle)$ that measures the extent of $\langle \hat{O} \rangle$’s violations to the quantum constraints \cite{31}, which is also advantageous for allowing more freedom in choices of $\hat{O}$. We will examine such formalism next.

Benchmark examples. We consider 1D fermion insulators with a bipartite unit cell, whose Bloch states take a general form $u(k) = (\cos(\theta_k/2), \sin(\theta_k/2) \exp(i\varphi_k))^T$ with the first (second) component denoting the $A$ ($B$) sublattice. The expectation values of two-point correlators are:

$$
C_0^{AA(BB)} = 0.5 \pm g_0/2,
$$

$$
C_r^{AA(BB)} = \pm g_r/2, r \in \mathbb{Z}^+,
$$

$$
C_r^{AB} = \tilde{g}_{r'}/2, r' \in \mathbb{Z} + 1/2, \quad (5)
$$

the rest obtainable via complex conjugation. $g_r = \int_0^{2\pi} d\theta_k e^{i\theta_k} \cos(\theta_k), \quad \tilde{g}_{r'} = \int_0^{2\pi} d\theta_k e^{i(\theta_k + \varphi)} \sin(\theta_k)$, over which we establish the following quantum constraints:

$$
\sum_r g_r \cdot \tilde{g}_{r+s} + \sum_{r'} \tilde{g}_{r'} \cdot \tilde{g}_{r'+s} = \delta_s. \quad (6)
$$

$g_r$ and $\tilde{g}_{r'}$, related to correlations in insulators, are fast decaying functions of $r$ and $r'$, allowing us to truncate at a finite distance $\Lambda = 20$ unless noted otherwise. We can thus define a positive-definite penalty function:

$$
f(g_r, \tilde{g}_{r'}) = \sum_{s=0}^{\Lambda/2} \left[ \sum_r g_r \cdot \tilde{g}_{r+s} + \sum_{r'} \tilde{g}_{r'} \cdot \tilde{g}_{r'+s} - \delta_s \right]^2 (7)
$$

which yields $\sim 0$ if and only if $\{g_r, \tilde{g}_{r'}\}$ are consistent with the quantum constraints. We note that the derivation of an expression as Eq. 7 is unavailable in generic quantum scenarios. Here for non-interacting fermions, it offers benchmarks to our strategy via machine learning quantum constraints in the following paragraphs.

Starting from random $u(k)$, we obtain $1.92 \times 10^6$ samples of $\{g_r, \tilde{g}_{r'}\}$ consistent with the quantum constraints and no penalty. We also include in the dataset $7.68 \times 10^6$ contrasting samples with small random deviations to $\{g_r, \tilde{g}_{r'}\}$ and corresponding penalties \cite{31}. Besides, we utilize the gauge equivalence to reduce the degrees of freedom \cite{31}. Then, we apply supervised machine learning \cite{13, 14} to train ANNs on the quantum constraints of $\{g_r, \tilde{g}_{r'}\}$ in the neighborhood of small or no violations \cite{31}. In practice, we use the average output of multiple independent ANNs $f^*(g_r, \tilde{g}_{r'})$ as the approximate penalty and their max output as an acceptance threshold to avoid unphysical regions.

To test out these quantum constraints, we study the mean-field solutions of a 1D interacting fermion Hamiltonian at half-filling:

$$
\hat{H} = \sum_x -t(c_x^\dagger c_{x+1} + c_{x+1}^\dagger c_x) + V c_x^\dagger c_{x+1} c_{x+1}^\dagger c_x + \nu \cdot \sum_{r,x} g_r \cdot \tilde{g}_{r+x}. \quad (8)
$$

The underlying assumptions of $f(g_r, \tilde{g}_{r'})$ and $f^*(g_r, \tilde{g}_{r'})$ are that the ground state takes a non-interacting fermion framework, hence the Hartree-Fock approximation, and an emergent bipartite order parameter may spontaneously break the translation symmetry. Likewise, while our strategy straightforwardly applies to any quantum many-body ansatz, e.g., matrix product states \cite{34–36}, neural network states \cite{26, 27}, quantum Monte Carlo methods, ab initio wave functions, even multiple ansatzes at the same time, the resulting quantum constraints will inherit the underlying assumptions and skip lower-energy scenarios beyond such assumptions, if any.

Under these circumstances, the energy expectation value is:

$$
E = \langle \hat{H} \rangle = \left[ -t \left( \text{Re}(\tilde{g}_{1/2}) + \text{Re}(\tilde{g}_{-1/2}) \right) \right] + 0.25V \left( 2 - 2g_0^2 - |g_1|^2 - |g_{-1}|^2 \right) \times N/2. \quad (9)
$$

We look for the constrained minimum $\{g_r, \tilde{g}_{r'}\}$ by minimizing either $L = \tilde{E} + \eta f$ or $L = \tilde{E} + \eta f^*$, where $\tilde{E} = E/(N/2), \eta$ controls the weight of the quantum constraints, and the optimized results approach the physical limit asymptotically when $\eta \to \infty$. In practice, we should balance $\eta$ between too large to allow an efficient search.
acceptance rate and too small to prevent the search from exiting regions represented by the samples. The extrapolation of \( \eta \) may offer a more systematic analysis, and an example is shown in Supplemental Material [31].

Also, we use the expectation values of \( \{g_r, \bar{g}_r\} \) at \( V \) to initialize the search at \( V + \delta V \), and so on so forth. In practice, we start from \( V = 2 \) with an interval of \( \delta V = -0.01 \) [37]. The benchmark results are summarized in Fig. 2 and Ref. [31], and their consistency indicates that given sufficient dataset and training, machine learning can offer a trustworthy path toward quantum constraints. It is worth noting that such soft quantum constraints offer a distinctive and complementary perspective to conventional variational approaches: while the latter bounds the ground states from above, given a search space generally smaller than necessary, our method may approach the ground state from below, where near-physical regions join our consideration yielding a search space larger than permitted.

The quantum-constraint perspective also allows us to design quantum many-body systems like never before. Say we wish to apply certain criteria to expectation values: sometimes it is as simple as the inclusion of the corresponding observables into the Hamiltonian, yet sometimes the criteria do not possess simple interpretations or require a nontrivial origin such as spontaneous symmetry breaking. For instance, to maximize \( \langle \Re(g_1)^\dagger \Re(\tilde{g}_{1/2} + \tilde{g}_{-1/2}), \rangle \in [0.5, 2] \) for 1D fermion insulators on bipartite lattices, we commonly need to consider a variational Hamiltonian, such as:

\[
H_{\text{var}} = \sum_x -t(c^\dagger_{x+1}c_x + \text{h.c.}) + (-1)^x\Delta(c^\dagger_{x+2}c_x + \text{h.c.}) ,
\]

which balances operators favoring \( \Re(\tilde{g}_{1/2} + \tilde{g}_{-1/2}) \) and \( \Re(g_1) \), respectively, and \( \Delta \) is a variational parameter for optimization. More generally, larger variational spaces with additional operators are preferred for thorough searches, and the ground-state solutions may bring additional complications. On the other hand, with the quantum constraints, we can circumvent such difficulties and resort to a constrained maximization. We compare our results in Fig. 3. Further, we can establish the underlying Hamiltonians and quantum states via the strategy in Ref. [38].

Strongly correlated scenarios—Generality is another essential merit of our strategy, which applies straightforwardly to strongly correlated systems and outshines the conventional methods hanging on the mind-boggling quantum many-body ground states themselves. Here, we illustrate the quantum constraints of 1D interacting spin-1/2 chains in the thermodynamic limit. Since the ground states of local Hamiltonians obey the Area Law, we can use tensor network states [4, 5, 39, 40], especially infinite matrix product states [34–36] for infinite system sizes, as our representation of quantum many-body state samples for machine learning quantum constraints. We also emphasize that it is straightforward to generalize and beneficial to include other quantum many-body ansatzes, such as projected states via variational Monte Carlo methods; see Supplemental Materials for examples and results [31].

First, we sample quantum many-body states with random, translation symmetric matrices of dimension \( \chi = 8 \) [41]. Next, we evaluate the expectation values \( \langle O \rangle \) of a series of low-order spin operators \( S^x_\lambda \) and \( S^z_\lambda \) for \( \lambda = x, y, z \) upon a section of length \( l_{\text{max}} = 6 \). Other than \( 6.75 \times 10^4 \) of these quantum-constraints-abiding samples, we also include \( 2.025 \times 10^5 \) contrasting samples with small deviations and corresponding penalties [31]. Then, we perform supervised machine learning on the dataset and
FIG. 3. The maximum of $\langle \text{Re}(g_{1/2}) \rangle$ for 1D fermion insulators on bipartite lattices shows good consistency between constrained optimization via the quantum constraints and searches among the variational Hamiltonian $H_{\text{var}}$ in Eq. 10. Individual expectation values for realizing the maximum, e.g., $g_1$ and $g_{1/2}$, also check out for each $\epsilon$. For the quantum-constraints approaches, we gradually increase $\epsilon$ from $\epsilon = 0.5$ while keeping track of $(g_r, \tilde{g}_r)$. Inset: the ANN $f^*(g_r, \tilde{g}_r)$ and the benchmark $f(g_r, \tilde{g}_r)$ suggest the obtained $(g_r, \tilde{g}_r)$ indeed obey the quantum constraints. $\eta = 1000$.

train ANNs $f^*(\langle O \rangle)$ to recognize how well a target $\langle O \rangle$ aligns with the quantum constraints. We note that the trained ANNs, as well as the previous ANNs $f^*(g_r, \tilde{g}_r)$ and benchmark $f(g_r, \tilde{g}_r)$ for 1D fermion insulators, penalize expectation values’ departure from and thereby enforcing quantum constraints as intended [31].

Subsequently, we use the quantum constraints for ground-state properties of quantum spin Hamiltonians. For instance, we apply our strategy with $f^*(\langle O \rangle)$ to the 1D transverse field ($h = 0$) and the non-integrable longitudinal-transverse field Ising models ($h \neq 0$):

$$H = \sum_j -JS_j^x S_{j+1}^z - g S_j^y - hS_j^z,$$

(11)

where we set $J = 1$ as the unit of energy. The results on the energy expectation value per site $E/N = \langle H \rangle/N = -J(S_0^x S_1^z) - g(S_0^y) - h(S_0^z)$ and beyond in the $N \to \infty$ thermodynamic limit are summarized in Fig. 4 and Supplemental Material [31]. Such quantum constraints are also directly applicable to the spin-1/2 XXZ chains [31]. In summary, we obtain quantitative results on ground-state properties such as energies and short-range correlators yet qualitative trends only on longer-range correlators, making pinpointing phase transitions relatively tricky. An additional or different set of observables addressing critical behaviors may be helpful.

Discussions—We propose to analyze ground-state properties via machine learning quantum constraints on expectation values and complement conventional ground-state-based approaches. Other than the aforementioned advantages, we have yet to establish a controlled quantitative analysis of algorithmic uncertainties, especially for relatively soft degrees of freedom, e.g., the order parameter of a spontaneous symmetry-breaking phase. Qualitative tendencies are observable, and extrapolation of $\eta$, the weight of quantum constraints, offers a partial solution [31]. Also, tensor network states in 2D and beyond may become costly, and other quantum many-body ansatzes may help complement the training data and simultaneously reduce biases originating from respective ansatz. Finally, while systematic presumptions such as the area law and symmetries help narrow the questions and facilitate the calculations, such physics intuitions should sometimes be taken with a grain of salt.

Acknowledgement: We thank Junren Shi, Xiao Yuan, Yang Qi, Zhenduo Wang, and Peng-Cheng Xie for insightful discussions. The authors are supported by the National Key R&D Program of China (No. 2021YFA1401900) and the National Science Foundation of China (No. 12174008). The computation was supported by High-performance Computing Platform of Peking University. Source code is available at https://github.com/PeilinZHENG/MLQC.

* frankzhangyi@gmail.com; # P.-L. Zheng and S.-J. Du are responsible for 1D fermion chains and spin chains respectively and contributed equally.
[1] R. T. Scalettar, E. Y. Loh, J. E. Gubernatis, A. Moreo,
S. R. White, D. J. Scalapino, R. L. Sugar, and E. Dagotto, “Phase diagram of the two-dimensional negative-u hubbard model,” Phys. Rev. Lett. 62, 1407 (1989).

[2] A. Moreo and D. J. Scalapino, “Two-dimensional negative-u hubbard model,” Phys. Rev. Lett. 66, 946 (1991).

[3] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Builik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozlik, X.-W. Liu, A. J. Millis, N. V. Prokof‘ev, M. Qin, G. E. Scuseria, H. Shi, B. V. Ssvistov, L. F. Tocchio, I. S. Tupitsyn, S. R. White, S. Zhang, B.-X. Zheng, Z. Zhu, and E. Gull, “Solutions of the two-dimensional hubbard model: Benchmarks and results from a wide range of numerical algorithms,” Phys. Rev. X 5, 041041 (2015).

[4] S. R. White, “Density matrix formulation for quantum renormalization groups,” Phys. Rev. Lett. 69, 2863 (1992).

[5] U. Schollwöck, “The density-matrix renormalization group,” Rev. Mod. Phys. 77, 259 (2005).

[6] 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).

[7] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, “Quantum monte carlo simulations of solids,” Rev. Mod. Phys. 73, 33 (2001).

[8] M. Troyer and U.-J. Wiese, “Computational complexity and fundamental limitations to fermionic quantum monte carlo simulations,” Phys. Rev. Lett. 94, 170201 (2005).

[9] E. Fradkin, S. A. Kivelson, and J. M. Tranquada, “Colloquium,” Rev. Mod. Phys. 87, 457 (2015).

[10] D. Poland, S. Rychkov, and A. Vichi, “The conformal bootstrap: Theory, numerical techniques, and applications,” Rev. Mod. Phys. 91, 015002 (2019).

[11] X. Han, “Quantum Many-body Bootstrap,” arXiv e-prints, arXiv:2006.06002 (2020).

[12] T. Deutsch, “Exact solution of the many-body problem with a $O(n^6)$ complexity,” (2021).

[13] Michael Nielsen, Neural Networks and Deep Learning (Free Online Book, 2013).

[14] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature 521, 436 (2015).

[15] M. Jordan and T. Mitchell, “Machine learning: Trends, perspectives, and prospects,” Science 349, 255 (2015).

[16] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, “Neural-network quantum state tomography,” Nature Physics 14, 447 (2018).

[17] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, “Reconstructing quantum states with generative models,” Nat. Mach. Intell. 1, 155 (2019).

[18] L. Wang, “Discovering phase transitions with unsupervised learning,” Phys. Rev. B 94, 195105 (2016).

[19] J. Carrasquilla and R. G. Melko, “Machine learning phases of matter,” Nature Physics 13, 431 (2017).

[20] Y. Zhang and E.-A. Kim, “Quantum Loop Topography for Machine Learning,” Phys. Rev. Lett. 118, 216401 (2017).

[21] K. Ch’ng, J. Carrasquilla, R. G. Melko, and E. Khatami, “Machine Learning Phases of Strongly Correlated Fermions,” Phys. Rev. X 7, 031038 (2017).

[22] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, “Machine learning quantum phases of matter beyond the fermion sign problem,” Scientific Reports 7, 8823 (2017).

[23] Y. Zhang, R. G. Melko, and E.-A. Kim, “Machine learning $\mathbb{Z}_2$ quantum spin liquids with quasiparticle statistics,” Phys. Rev. B 96, 245119 (2017).

[24] P. Zhang, H. Shen, and H. Zhai, “Machine learning topological invariants with neural networks,” Phys. Rev. Lett. 120, 066401 (2018).

[25] W. Lian, S.-T. Wang, S. Lu, Y. Huang, F. Wang, X. Yuan, W. Zhang, X. Ouyang, X. Wang, X. Huang, L. He, X. Chang, D.-L. Deng, and L. Duan, “Machine learning topological phases with a solid-state quantum simulator,” Phys. Rev. Lett. 122, 210503 (2019).

[26] G. Carleo and M. Troyer, “Solving the quantum many-body problem with artificial neural networks,” Science 355, 602 (2017).

[27] D.-L. Deng, X. Li, and S. Das Sarma, “Quantum entanglement in neural network states,” Phys. Rev. X 7, 021021 (2017).

[28] A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, “Active learning machine learns to create new quantum experiments,” Proceedings of the National Academy of Sciences 115, 1221 (2018).

[29] Y. Zhang, A. Mesaros, K. Fujita, S. Edkins, M. Hamidian, K. Ch’ng, H. Eisaki, S. Uchida, J. S. Davis, E. Khatami, et al., “Machine learning in electronic-quantum-matter imaging experiments,” Nature 570, 484 (2019).

[30] C. Valagiannopoulos, “Predicting the quantum texture from transmission probabilities,” Journal of Applied Physics 127, 174301 (2020).

[31] Please refer to Supplemental Materials for further details on constrained optimization algorithms, generation of training samples, the settings of our ANNs, and additional examples and results on 1D spin and fermion chains such as multiple Fermi seas, which include Ref. 43–56.

[32] Such scenarios also provide valuable additional training samples.

[33] The arguments can generalize to complex $t$ straightforwardly noting that the minimum energy requires $\arg C_1 = - \arg t$.

[34] 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).

[35] G. Vidal, “Classical simulation of infinite-size quantum lattice systems in one spatial dimension,” Phys. Rev. Lett. 98, 070201 (2007).

[36] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac, “Classical simulation of infinite-size quantum lattice systems in two spatial dimensions,” Phys. Rev. Lett. 101, 250602 (2008).

[37] Although the model at $V = 0$ also offers a good starting point, the truncation in $r$ and $r'$ may cause issues.

[38] J.-B. Wang and Y. Zhang, “Single-shot quantum measurements sketch quantum states,” arXiv e-prints, 2203.01348 (2022).

[39] M. Fannes, B. Nachtergaele, and R. F. Werner, “Finitely correlated states on quantum spin chains,” Communications in Mathematical Physics 144, 443 (1992).

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

[41] We note that random matrix product states have emergent statistics [57–59] against sampling generality at large dimension $\chi$, hence our small $\chi$ choice.

[42] E. Lieb, T. Schultz, and D. Mattis, “Two soluble models of an antiferromagnetic chain,” Annals of Physics 16, 407 (1961).

[43] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in Advances in Neural Information Processing Systems, Vol. 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019).

[44] J. Han and C. Moraga, “The influence of the sigmoid function parameters on the speed of backpropagation learning,” in From Natural to Artificial Neural Computation, edited by J. Mira and F. Sandoval (Springer Berlin Heidelberg, Berlin, Heidelberg, 1995) pp. 195–201.

[45] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).

[46] K. Fukushima, “Visual feature extraction by a multilayered network of analog threshold elements,” IEEE Transactions on Systems Science and Cybernetics 5, 322 (1969).

[47] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, Vol. 15, edited by G. Gordon, D. Dunson, and M. Dudík (PMLR, Fort Lauderdale, FL, USA, 2011) pp. 315–323.

[48] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift.” (2015), arXiv:1502.03167.

[49] J. Hauschild and F. Pollmann, “Efficient numerical simulations with Tensor Networks: Tensor Network Python (TeNPy),” SciPost Phys. Lect. Notes 5 (2018), arXiv:1805.00055.

[50] C. E. Brodley and M. A. Friedl, “Identifying mislabeled training data,” J. Artif. Intell. Res. 11, 131 (1999).

[51] E. Beigman and B. B. Klebanov, “Learning with annotation noise,” in ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computer Linguistics, edited by K. Su, J. Su, and J. Wiebe (The Association for Computer Linguistics, 2009) pp. 280–287.

[52] D. Rolnick, A. Veit, S. J. Belongie, and N. Shavit, “Deep learning is robust to massive label noise,” (2017), arXiv:1705.10694.

[53] Y. Kim, J. Yim, J. Yun, and J. Kim, “NLNL: negative learning for noisy labels,” in 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019 (IEEE, 2019) pp. 101–110.

[54] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, “Deep residual shrinkage networks for fault diagnosis,” IEEE Transactions on Industrial Informatics 16, 4681 (2020).

[55] M. J. D. Powell, “A direct search optimization method that models the objective and constraint functions by linear interpolation,” in Advances in Optimization and Numerical Analysis, edited by S. Gomez and J.-P. Hennart (Springer Netherlands, Dordrecht, 1994) pp. 51–67.

[56] M. J. D. Powell, “Uobyqa: unconstrained optimization by quadratic approximation,” Mathematical Programming 92, 555 (2002).

[57] S. Garnerone, T. R. de Oliveira, and P. Zanardi, “Typicality in random matrix product states,” Phys. Rev. A 81, 022336 (2010).

[58] S. Garnerone, T. R. de Oliveira, S. Haas, and P. Zanardi, “Emergent statistical mechanics from properties of disordered random matrix product states,” PRX Quantum 2, 040308 (2021).
Supplemental Materials: Ground-state properties via machine learning quantum constraints

Pei-Lin Zheng*, Si-Jing Du*, and Yi Zhang*
International Center for Quantum Materials, Peking University, Beijing, 100871, China and
School of Physics, Peking University, Beijing, 100871, China

DETAILS ON SAMPLE GENERATION, SUPERVISED MACHINE LEARNING, AND CONSTRAINED OPTIMIZATION

Our strategy consists of three vital steps: sample generation ($\langle \hat{O} \rangle$ and corresponding penalty), supervised machine learning to train ANNs on quantum constraints $f^*(\langle \hat{O} \rangle)$, and constrained optimization upon target model systems.

Sample generation

In the main text, we propose two different paradigms of quantum constraints $f^*(\langle \hat{O} \rangle)$: first, we can express the quantum constraints as an explicit function between the operator expectation values, e.g., $(\langle \hat{O}_b \rangle = f^*(\langle \hat{O}_1 \rangle, \ldots, \langle \hat{O}_N \rangle)$ for 1D Fermi seas. Here, we only need the physical operator expectation values $(\langle \hat{O} \rangle)$ fully consistent with quantum constraints, which are straightforward to generate by selecting the appropriate ansatz and sampling the corresponding quantum many-body states.

On the other hand, in more general scenarios where relations between $(\langle \hat{O} \rangle)$ are complex and implicit, we have to resort to a penalty function $f^*(\langle \hat{O} \rangle)$ to characterize quantum constraints. Here, in addition to the physical $(\langle \hat{O} \rangle)$ with 0 penalty values, we generate unphysical $(\langle \hat{O} \rangle)'$ by adding random small deviations $\delta(\langle \hat{O} \rangle)$ to physical $(\langle \hat{O} \rangle)$. We leave the discussions of the penalty values for such near-physical samples to the next subsection.

We also exploit the equivalence through gauge transformations, which leave quantum constraints intact, for further simplification. Therefore, it suffices to keep one $(\langle \hat{O} \rangle)$ example for each gauge equivalent class and reduces the degrees of freedom and thus the resulting computational complexity. For instance, for 1D fermion insulators with a bipartite unit cell, we can define the penalty function $f^*(g_r, \tilde{g}_r)$ for real $\tilde{g}_{1/2}$ only, and the penalty function for a generic $\{g_r, \tilde{g}_r\}$ with complex $\tilde{g}_{1/2}$ is obtainable straightforwardly through a gauge transformation.

For reference, we show the pseudo-codes for sample generation of 1D fermion insulators with a bipartite unit cell and 1D interacting spin-1/2 chains in Algorithms 1 and 2, respectively.

| Algorithm 1 | Sample generation of 1D fermion insulators |
|-------------|------------------------------------------|
| **Input:**  | number $m$ of physical samples, ratio $p$ of unphysical to physical samples, truncation distance $\Lambda$, random decay rates $\alpha$ |
| **Output:** | dataset $\mathcal{D}$ |
| 1: | $\mathcal{D} = \{\}$; |
| 2: | for $i = 1$ to $m$ do: |
| 3: | Generate a random vector $\mathbf{g}_i' = \{g_{r,i}, \tilde{g}_{r,i}\}$; |
| 4: | $\mathbf{g}_i' = \{ |\tilde{r}| = 0, e^{-\alpha(|\tilde{r}|-1)} | \tilde{r} \neq 0, e^{-\alpha(|\tilde{r}|-0.5)} \}$ for locality; |
| 5: | Fourier transform $\mathbf{g}_i'$ to get $\mathbf{u}_i' = \{u_{k,i}, \tilde{u}_{k,i}\}$; |
| 6: | Normalize $\mathbf{u}_i'$ to get Bloch state $\mathbf{u}_i = \{u_k, \tilde{u}_k\}$; |
| 7: | Inverse Fourier transform $\mathbf{u}_i$ to get $\mathbf{g}_i = \{g_r, \tilde{g}_r\}$; |
| 8: | Gauge transform $\mathbf{g}_i$ to make $\tilde{g}_r_{1/2}$ real; |
| 9: | $\mathcal{D} \leftarrow \{\mathbf{g}_i, \text{penalty value} = 0\}$; |
| 10: | Run Algorithm 3 for basis vectors of physical manifold and return $(\mathbf{g}_i + \delta\mathbf{g}_i, y'_{KLD})$ for $i = 1, \ldots, p$; |
| 11: | for $l = 1$ to $p$ do: |
| 12: | $\mathcal{D} \leftarrow \{\mathbf{g}_i + \delta\mathbf{g}_i, \text{penalty value} = y'_{KLD}\}$; |
| 13: | end for |
| 14: | end for |

| Algorithm 2 | Sample generation of 1D interacting spin-1/2 chains |
|-------------|------------------------------------------------------|
| **Input:**  | number $m$ of physical samples, the ratio $p$ of unphysical to physical samples, truncation distance $l_{max}$, iMPS matrices dimension $\chi$ |
| **Output:** | dataset $\mathcal{D}$ |
| 1: | $\mathcal{D} = \{\}$; |
| 2: | for $i = 1$ to $m$ do: |
| 3: | Generate random $\chi \times \chi$ matrices for iMPS $|\Psi_i\rangle$; |
| 4: | $\langle \hat{O} \rangle_i = \langle \Psi_i | \hat{O} | \Psi_i \rangle = \{S^{(\lambda, \lambda')} i_{\lambda, \lambda'} \}^{|\lambda, \lambda' = z, y, x|1 \leq i \leq \chi;}$ |
| 5: | $\mathcal{D} \leftarrow \{\langle \hat{O} \rangle_i, \text{penalty value} = 0\}$; |
| 6: | Run Algorithm 4 for basis vectors of physical manifold and return $(\langle \hat{O} \rangle_i + \delta\langle \hat{O} \rangle_i, y_{MSE})$ for $i = 1, \ldots, p$; |
| 7: | for $l = 1$ to $p$ do: |
| 8: | $\mathcal{D} \leftarrow \{\langle \hat{O} \rangle_i + \delta\langle \hat{O} \rangle_i, \text{penalty value} = y_{MSE}\}$; |
| 9: | end for |

Penalty values for near-physical samples

We will employ two methods to define and evaluate the penalty value of near-physical (unphysical) samples $(\langle \hat{O} \rangle)' = (\langle \hat{O} \rangle) + \delta(\langle \hat{O} \rangle) = \{\hat{O}_1, \hat{O}_2, \ldots\}$. Using the (weighted) norm of $\delta(\langle \hat{O} \rangle)$ is an overestimate to the penalty value,
as $\delta(\hat{O})$ is not fully normal to the physical manifold in general, and we need to separate its projection along the physical manifold for a more sound estimation.

Our first method starts from a Kullback-Leibler (KL) divergence perspective \cite{1, 2}. For such, we re-express the observables $\{\hat{O}_j\}$ as a set of two-level projection operators $\{P_j\}$, and correspondingly, the expectation values in terms of $\langle \hat{P} \rangle = \{f_1, 1 - f_1, f_2, 1 - f_2, \ldots \}$ for a physical state, $f_\gamma = \langle \Phi_0 | P_j | \Phi_0 \rangle$, and $\langle \hat{P}' \rangle = \{f'_1, 1 - f'_1, f'_2, 1 - f'_2, \ldots \}$ for an unphysical state, $f'_\gamma = f_\gamma + \delta f_\gamma$. Then, we can define the penalty value $y_{KLD}$ as the minimum total KL divergence between the given probabilities $\{f'_\gamma\} \in [0, 1]$ and the quantum probabilities $\langle \Phi | P_j | \Phi \rangle$ among all states:

$$y_{KLD} = \min_{\{\Phi\}} \sum_\gamma f'_\gamma \log \left( \frac{f'_\gamma}{\langle \Phi | P_j | \Phi \rangle} \right) + (1 - f'_\gamma) \log \left( \frac{1 - f'_\gamma}{\langle \Phi | 1 - P_j | \Phi \rangle} \right)$$

$$\approx \min_{\{\Phi\}} \sum_\gamma \frac{(\delta f_\gamma - \langle \Phi_0 | \hat{P}_j | \delta \Phi \rangle - \langle \delta \Phi | \hat{P}_j | \Phi_0 \rangle)^2}{2 f_\gamma (1 - f_\gamma)},$$

where we have assumed $\delta f_\gamma$ is small and $|\Phi\rangle$ is not far from $|\Phi_0\rangle$, and performed a second-order Taylor expansion to the last step:

$$|\Phi\rangle = \frac{|\Phi_0\rangle + |\delta \Phi\rangle}{\sqrt{(|\Phi_0\rangle + |\delta \Phi\rangle)(|\Phi_0\rangle + |\delta \Phi\rangle)}}. \quad (2)$$

$y_{KLD}$ is non-negative definite. $y_{KLD} = 0$ if and only if $\{f'_\gamma\}$ is physical with an existing $|\Phi\rangle$ so that $\{f'_\gamma\} = \langle \Phi_0 | \hat{P}_j | \Phi \rangle$; otherwise, $y_{KLD} > 0$ and describes the extent to which $\{f'_\gamma\}$ deviates from the physical manifold, measured from the closest quantum state $|\Phi\rangle$ towards $\{f'_\gamma\}$ with the best effort.

In practice, we first sample $|\delta \Phi\rangle$ and perform a principal component analysis (PCA) on $\{\Phi_0 | \hat{P} \delta \Phi \} + \langle \delta \Phi | \hat{P} | \Phi_0 \rangle$ - the basis vectors in the physical manifold; then, we project out these directions from $\{\delta f_\gamma\}$, and evaluate the penalty value $y_{KLD}$ following Eq. 1. We summarize the pseudo-code in Algorithm 3. In Fig. 1, we show a comparison between $y_{KLD}$ and the benchmark $f(g_r, \tilde{g}_r')$ following Eq. 7 in the main text. We observe a positive connection between the two penalty functions, suggesting that $y_{KLD}$ is a generally available and reasonably consistent indicator for quantum constraints.

Our second method is based upon the mean squared error (MSE). It defines the penalty value for expectation values $\langle \hat{O} \rangle'$ as the following expression:

$$y_{MSE} = \min_{\{\Phi\}} \sum_j \left( \langle \hat{O}_j \rangle' - \langle \Phi | \hat{O}_j | \Phi \rangle \right)^2$$

$$= \min_{\{\Phi\}} \sum_j \left( \delta(\hat{O}_j') - \left( \langle \Phi | \hat{O}_j | \Phi \rangle - \langle \hat{O}_j \rangle \right) \right)^2. \quad (3)$$

Similar to $y_{KLD}$, for small deviations $\delta(\hat{O}_j)$ and $|\Phi\rangle$ similar to Eq. 2, we can sample $\langle \Phi | \hat{O} | \Phi \rangle - \langle \hat{O} \rangle$ for basis vectors in the physical manifold, and evaluate $y_{MSE}$ following Eq. 3. We summarize the corresponding pseudo-code in Algorithm 4.

In case we have not exhausted the basis vectors of the physical manifold, the evaluated penalty values may have discrepancies with the definitions in Eqs. 1 and 3. In
Algorithm 4 Evaluation of penalty value $y_{MSE}$

Input:
the set of observables $\{\hat{O}_i\}$, number $p$ of unphysical state per physical state $\{\hat{\Phi}_0\}$, the corresponding physical sample $\{\hat{\mathbf{O}}\}$, number $n$ of reference quantum-state samples, truncation threshold $\epsilon$, hyper-parameter $\delta$

Output:
$\{\langle \hat{\mathbf{O}} + \delta(\hat{\mathbf{O}})^i \rangle, y_{MSE} \}_{i=1, \ldots, p}$

1: $\mathbf{X} = []$
2: for $i = 1$ to $n$ do:
3: Randomly generate $|\Psi_i\rangle$:
4: Compute $|\Phi\rangle = \frac{|\Phi_0 + \delta(\Psi_i)\rangle}{\sqrt{\langle (|\Phi_0 + \delta(\Psi_i)\rangle (|\Phi_0 + \delta(\Psi_i)\rangle)}}$;
5: Compute $\mathbf{x}_i = \left[\left(\Phi(\hat{O}_j|\Phi) - \hat{O}_j\right)_{j=1,2,\ldots}^p\right]$;
6: Normalize $\mathbf{x}_i$;
7: $\mathbf{X} \leftarrow \mathbf{x}_i$;
8: end for
9: Compute covariance matrix $\mathbf{K}_{XX} = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X}^T - \bar{\mathbf{X}})$;
10: Diagonalize $\mathbf{K}_{XX} = \mathbf{V} \Lambda \mathbf{V}^T$;
11: Drop eigenvectors with eigenvalues less than $\epsilon$ from $\mathbf{V}$;
12: Get the projector to physical manifold $\mathbf{W} = \mathbf{V} \Lambda \mathbf{V}^T$;
13: for $l = 1$ to $p$ do:
14: Generate a random vector $\delta(\hat{\mathbf{O}})^l = \{\delta(\hat{O}_j)^l\}$;
15: Calculate $(I - \mathbf{W})\delta(\hat{\mathbf{O}})^l$ then $y_{MSE}^{l}$ following Eq. 3;
16: end for

FIG. 2. The architecture of our ANNs used as the $f^*$ function for the quantum constraints consists of a couple of fully connected (fc) layers and four residual blocks with potential shortcuts. Our choices for the ANN layer width (number of neurons in a layer) vary with the target systems: we set embedding size and hidden size as 500 and 100 for the 1D fermion chains and 600 and 144 for the 1D spin chains, respectively.

Parallel for 1000 epochs to guarantee good convergence. For each ANN, we keep the best performing one on a separate validation set during its entire training process, which takes less than 12 hrs on one NVIDIA RTX 3090 GPU.

For 1D spin-1/2 chains, we stretch the label $y$ with the function $\tanh$ for better regression results. We use the mean-square error as the cost function and apply the stochastic gradient decent method with the Adam optimizer with an initial learning rate of $10^{-3}$ and L2 regularization of $10^{-5}$, and a batch size of 360. We also employ the StepLR scheduler with a multiplier of 0.5 to the learning rate for every 50 epochs. We train multiple ANNs in parallel for at least 500 epochs to ensure good convergence and keep the ANN with the minimum loss on a validation set separated from the training set. The entire training process takes less than 12 hrs on one NVIDIA RTX 3090 GPU.

Constrained optimization

With the quantum constraints $f$ or $f^*$, we employ an adiabatic-algorithm-based constrained optimization algorithm for the ground-state properties of target Hamiltonians or design targets, which we summarize in Algorithm 5. Our choices of sample datasets concentrate on

Artificial neural networks and supervised machine learning

We apply supervised machine learning on the generated datasets to train our ANNs, each of which consists of two hidden layers and four residual blocks [8] with ReLU activation functions [9, 10] and batch normalization [11]; see Fig. 2 for an illustration of our ANN architecture. The output neuron has a sigmoid activation function [12] and (its average over multiple ANNs) stands for the $f^*$ function on the quantum constraints.

For 1D fermion chains, we first apply the function $\tanh(500y)$ to stretch the label $y$ for better compatibility with the range $[0, 1]$ of the sigmoid function of the ANN output neuron. We use the mean-square error as our cost function and apply the stochastic gradient descent method with the Adam optimizer with a learning rate of $10^{-3}$, L2 regularization of $10^{-6}$, a batch size of 4096, and the StepLR scheduler with a learning rate decay ratio of 0.5 for every 200 epochs. We train multiple ANNs in parallel for 1000 epochs to guarantee good convergence. For each ANN, we keep the best performing one on a separate validation set during its entire training process, which takes less than 12 hrs on one NVIDIA RTX 3090 GPU.

For 1D spin-1/2 chains, we stretch the label $y$ with the function $\tanh$ for better regression results. We use the mean-square error as the cost function and apply the stochastic gradient decent method with the Adam optimizer with an initial learning rate of $10^{-3}$ and L2 regularization of $10^{-5}$, and a batch size of 360. We also employ the StepLR scheduler with a multiplier of 0.5 to the learning rate for every 50 epochs. We train multiple ANNs in parallel for at least 500 epochs to ensure good convergence and keep the ANN with the minimum loss on a validation set separated from the training set. The entire training process takes less than 12 hrs on one NVIDIA RTX 3090 GPU.

Constrained optimization
FIG. 3. The penalty functions with respect to selected expectation values’ departure from the quantum constraints illustrate their tendency to restore the expectation values to physically allowed values. We note that the strengths of the quantum constraints may vary from case to case, e.g., the ANN $f^*$ and the benchmark $f$ for departure of $g_0$ and $\hat{g}_{1/2}$ in 1D fermion insulators, and the ANN $f^*$ for departure of $\langle S_i^z \rangle$ and $\langle S_i^z S_j^z \rangle$ in 1D spin chains.

scenarios consistent or nearly consistent with the quantum constraints, so we need to avoid deeply unphysical scenarios where our ANNs have little or no experience during their supervised machine learning and do not respond meaningfully. Therefore, we recommend starting with initial conditions $\langle \hat{O} \rangle_i$ that are known to be consistent with the quantum constraints, either via exact solutions or controlled approximation, and then change the optimization targets gradually, using each step’s optimal $\langle \hat{O} \rangle_0$ as the initial condition for the next, and so on so forth. For example, as we tune the parameter of a Hamiltonian in small steps, the ground-state properties $\langle \hat{O} \rangle$ of one step should not be too far from its neighboring steps, except for the case of an abrupt first-order phase transition. Such a scheme keeps our searches in the ‘river bed’ of physical or near physical $\langle \hat{O} \rangle$ that the ANNs are accustomed to.

The rough landscape of an ANN may cause additional complications, and trap the optimization in local minima, or mislabel unphysical $\langle \hat{O} \rangle$ as physical and cause a diversion or even a mistake. To counter these issues, we employ the idea of ensemble learning and use the average output of multiple ANNs as the value of $f^*(\langle \hat{O} \rangle)$. We also set a threshold that the maximum outputs of multiple ANNs should not exceed to eliminate the likelihood of false ‘physical,’ and keep multiple searches in parallel and multiple $\langle \hat{O} \rangle$ outcomes as the seeds for the next step to increase the acceptance rate. In addition, we include small, random deviations from each seed before optimization to avoid being trapped in local minima. We note that these tricks are no longer necessary when we use the well-behaved quantum-constraint expressions such as $f$, which only requires a few $(O(1))$ initial points for the search to avoid cases of falling into a local minimum. Intuitively, the analytical function $f$ is much smoother than $f^*$, as illustrated in Fig. 3. At the same time, the target physical region in the $\langle \hat{O} \rangle$ space itself should not cause too much roughness concern as the expectation values evolve gradually with the underlying quantum many-body states. This suggests there is certainly room for improvement on the approximation quality and the landscape roughness of the ANNs.

In practice, we parallelize multiple searches from different seeds (initial conditions) and use the Adam optimizer with a step size of $10^{-3}$ and the StepLR scheduler with a slow-down ratio of 0.5 for every 200 steps for an overall 800 steps on target minimization. We note that Algorithm 5 uses only gradient information, and in principle, adding the information of the Hessian matrix can yield better results, which we leave to future studies.

In addition, we can also use a classification ANN $f^*$ that labels physical (unphysical) $\langle \hat{O} \rangle$ samples as 0 (1). As the landscape of such $f^*$ will have a large number of barren plateaus and steep cliffs, simple gradient-based optimizations are no longer suitable, and non-derivative optimization methods, such as genetic algorithm, optimization by quadratic approximation [13, 14], may be considered in the future. However, we note that the non-convex optimization problem is NP-hard, and there is no guarantee that the result obtained must be a global minimum irrespective of the optimization method.

Algorithm 5 Constrained Optimization

**Input:**
- tuning parameter $t$, target function $h((\langle \hat{O} \rangle); t)$, quantum constraints $f^*(\langle \hat{O} \rangle)$, initial $\langle \hat{O} \rangle_i$, penalty weight $\eta$

**Output:**
- $h_0(t) = \min_{(\hat{O})} [h(t)]$, $O_0(t) = \arg \min_{(\hat{O})} [h(t)]$

1: $h_0(t) = \emptyset$, $O_0(t) = \emptyset$
2: for $t$ in $t$ do:
3: $L(\langle \hat{O} \rangle) = h((\langle \hat{O} \rangle); t) + \eta f^*(\langle \hat{O} \rangle)$;
4: Minimize $L$ by gradient-based algorithms with $\langle \hat{O} \rangle_i$, as
5: the initial point for $h_0$ and $\langle \hat{O} \rangle_0$;
6: $O_0(t) \leftarrow \langle \hat{O} \rangle_0 = \arg \min_{(\hat{O})} \in S^+ [h(t)]$;
7: $h_0(t) \leftarrow h_0 = h((\langle \hat{O} \rangle_0); t)$;
8: $\langle \hat{O} \rangle_i = \langle \hat{O} \rangle_0$;
9: end for
MACHINE LEARNING QUANTUM CONSTRAINTS FOR 1D FERMI SEAS

Quantum constraints for multiple Fermi seas

Generalizing the 1D Fermi sea scenarios to ground states with multiple Fermi seas, we can identify a ground state with the set of Fermi points \( \{k_n, n = 1, \ldots, 2n_{fs}\} \), arranged from small to large, where \( n_{fs} \) is an upper bound of the number of Fermi seas [15]. The vital expectation values for the ground state are its two-point correlators \( C_r = \langle c_{x+r}c_x \rangle \):

\[
C_0 = \frac{\sum_{n=1}^{2n_{fs}} (-1)^n k_n}{2\pi} \quad \text{and} \quad C_r = \frac{\sum_{n=1}^{2n_{fs}} (-1)^n \exp(ik_n r)}{2\pi ir}, \quad r \geq 1,
\]

which are complex-value decaying functions of the correlation distance \( r \). By sampling \( \{k_n, n = 1, \ldots, 2n_{fs}\} \), we obtain our expectation-value dataset following Eq. 4. Then, with \( \{C_0/C_0, \ldots, C_{n_{fs}}/C_0\} \) as the inputs and \( C_0 \) as the intended output, we train our ANNs on a quantum constraint in the form of

\[
C_0 = f^* \left( \frac{C_1}{C_0}, \ldots, \frac{C_{n_{fs}}}{C_0} \right)
\]

as the quantum constraints of 1, 2, or 3 Fermi seas. We also cross-check training sets of 5 \( \times \) 10\(^5\) samples with exactly 1, 2, or 3 Fermi seas on validation sets of 1 \( \times \) 10\(^5\) samples. Unsurprisingly, the ANNs trained with more Fermi seas generalizes well to the cases with fewer Fermi seas, but not vice versa. We use the average output of two ANNs trained on operator expectation values from 1, 2, and 3 Fermi seas and two ANNs trained on exactly 3 Fermi seas as the quantum constraints \( f^* \) for the subsequent constrained optimization processes, whose details are discussed in the previous section.

We can apply the obtained \( f^* \) towards the ground-state properties of the following Hamiltonians:

\[
\hat{H} = \sum_x \left[ -\sum_{r=1}^{r_{cut}} \left( t_r c_{x+r}^\dagger c_x + \text{h.c.} \right) - \mu c_x^\dagger c_x \right],
\]

where \( t_r \) is the complex \( r\text{-th} \)-nearest-neighbor hopping amplitude, \( 1 \leq r \leq r_{cut} \), and \( \mu \) is the Fermi energy. We can represent the average energy per site as the operator expectation values in Eq. 4:

\[
\bar{E} = \langle \hat{H} \rangle \big/ N = -\mu C_0 - \sum_{r=1}^{r_{cut}} t_r C_r + \text{c.c.}
\]

\[
= -\left[ \mu + \sum_{r=1}^{r_{cut}} (t_r C_r/C_0 + \text{c.c.}) \right] f^* \left( \frac{C_1}{C_0}, \frac{C_2}{C_0}, \ldots, \frac{C_{n_{fs}}}{C_0} \right),
\]

where the introduction of \( f^* \) in the second line requires \( n_{fs} \leq r_{cut} \). In hindsight, the \( r_{cut}\text{-th} \)-nearest-neighbor hopping can potentially give rise to \( r_{cut} \) Fermi seas in the ground states, see Fig. 4 inset; therefore, the quantum-state samples should at least incorporate such scenarios for adequate diversity and generalizability.

Consequently, we transform the problem into a conventional optimization for the minimum of \( \bar{E} \) with respect to the parameters \( \{\text{Re}(C_1), \text{Im}(C_1), \ldots, \text{Re}(C_{n_{fs}}), \text{Im}(C_{n_{fs}})\} \). Since there is no unphysical expectation-value settings to worry about during the optimization in this formalism, it suffices to start from a random initial condition consistent with \( n_{fs} = 3 \), and minimize \( \bar{E} \) in Eq. 6 with Adam optimizer and StepLR scheduler for a given Hamiltonian in Eq. 5. For instance, we summarize the results of the obtained expectation values of ground-state energy \( \bar{E} \) and electron density \( C_0 \) for \( t_1 = 1 - i, t_2 = 0.5 - 0.9i, t_3 = 1 - i \) and \( \mu \in [-2, 2] \) in Fig. 4.

On the other hand, we can solve the ground states of
Eq. 5 in the momentum space $\hat{H} = \sum_k (\epsilon_k - \mu) c^\dagger_k c_k$,
\[
\epsilon_k = -2 \sum_{r=1}^{\text{ext}} \text{Re}(t_r) \cos(kr) - \text{Im}(t_r) \sin(kr),
\] (7)
see Fig. 4 inset. For a given $\mu$, the ground state fill the Fermi sea $c^\dagger_k c_k = 1$ for $\epsilon_k < \mu$ and $c^\dagger_k c_k = 0$ otherwise, and we have up to $n_{fs} = 3$ Fermi seas for $\mu \in [-2, 2]$. With these ground states, we can calculate the exact values of $\hat{E}$ and $\hat{C}_0$ and use them as benchmarks in the Fig. 4. In most cases, our approach gets satisfactorily accurate results, where the average error of $\hat{E}$ and $\hat{C}_0$ are $1.3 \times 10^{-2}$ and $2.8 \times 10^{-2}$, respectively.

**Additional details on machine learning quantum constraints for 1D fermion insulators**

In this subsection, we provide further details and results on our quantum-constraints strategy on 1D fermion insulators with a bipartite unit cell.

First, we note the following transformations between correlators and projection operators:
\[
\hat{P}^{AA(BB)}_0 = \frac{\hat{c}^\dagger \hat{c}}{2}, \quad r \in \mathbb{Z}^+,
\]
\[
\hat{P}^{AA(BB)}_r = \frac{\hat{c}^\dagger \hat{c} + \hat{c} \hat{c}^\dagger}{2}, \quad r \in \mathbb{Z}^+,
\]
\[
\hat{P}^{AA(BB)}_r' = \frac{\hat{c}^\dagger \hat{c} + \hat{c} \hat{c}^\dagger}{2}, \quad r' \in \mathbb{Z}^+ + \frac{1}{2},
\]
\[
\hat{P}^{AB}_{r'} = \frac{\hat{c}^\dagger \hat{c} + \hat{c} \hat{c}^\dagger}{2}, \quad r' \in \mathbb{Z}^+ + \frac{1}{2},
\]
and the corresponding expectation values are:
\[
\langle \hat{P}^{AA} \rangle = C^{AA}_0 = \frac{1 + g_0}{2},
\]
\[
\langle \hat{P}^{BB} \rangle = C^{BB}_0 = \frac{1 - g_0}{2},
\]
\[
\langle \hat{P}^{AA} \rangle = \text{Re}(C^{AA}_r) + C^{AA}_0 = \frac{\text{Re}(g_r) + 1 + g_0}{2},
\]
\[
\langle \hat{P}^{BB} \rangle = \text{Re}(C^{BB}_r) + C^{BB}_0 = \frac{-\text{Re}(g_r) + 1 - g_0}{2},
\]
\[
\langle \hat{P}^{AA} \rangle = \text{Im}(C^{AA}_r) + C^{AA}_0 = \frac{\text{Im}(g_r) + 1 + g_0}{2},
\]
\[
\langle \hat{P}^{BB} \rangle = \text{Im}(C^{BB}_r) + C^{BB}_0 = \frac{-\text{Im}(g_r) + 1 - g_0}{2},
\]
\[
\langle \hat{P}^{AB} \rangle = \text{Re}(C^{AB}_r) + C^{AB}_0 = \frac{\text{Re}(\tilde{g}_r) + 1}{2},
\]
\[
\langle \hat{P}^{AB} \rangle = \text{Im}(C^{AB}_r) + C^{AB}_0 = \frac{\text{Im}(\tilde{g}_r) + 1}{2}.
\]

Moreover, we notice that while measurements of most expectation values such as the ground-state energies are precise and straightforward, the estimation for the order parameters of the spontaneous symmetry breaking phase is relatively crude and dependent on the quality of the quantum constraints and details of the constrained optimization, see Fig. 5 for illustrations. We attribute such difficulties to two main factors: the order parameter is relatively soft, and small deviations of it do not yield a meaningful rise in energy, making the constrained optimization relatively insensitive to its value; on the other hand, the ANNs $f^*$ tend to have lower penalties for deviations in the order parameter thus less restoring power to enforce consistency with the quantum constraints, see Fig. 3. Adding more specific training samples to emphasize the quantum constraints on the order parameter may improve the ANNs in such aspects, which we leave to future work. In comparison, we note that the order parameter from the high-quality quantum constraints $f$ is able to achieve satisfactory consistency. Its deviations at small $V$ from Hartree-Fock results are likely due to the impact of the cut-off $\Lambda$ as the mean-field gap diminishes.
Also, we summarize in Fig. 6 our results on designing quantum many-body systems to maximize \((\text{Re}(\tilde{g}_1))^2 + \text{Re}(\tilde{g}_{1/2}) + \tilde{y}_{1/2})\) in addition to Fig. 3 in the main text. As we expect, larger (smaller) \(\eta\) enforces stricter (looser) quantum constraints and tends to underestimate the optimal outcomes less (more). As the quantum constraints are essentially rigid \(\eta \to \infty\) for truly physical states, we may scale the target properties with respect to \(\eta^{-1}\) for extrapolation, e.g., Fig. 6 inset. One caveat, however, is that larger \(\eta\) also reduces the manifold of expectation values we deem ‘physical,’ thus the overall acceptance rate, leading to possible underestimations due to incomplete searches within the time and step limits, as is the case for ANN quantum constraints \(f^*\) at large \(\eta\) in Fig. 6. Therefore, throughout and careful studies are necessary to avoid misleading extrapolations and interpretations.

**MACHINE LEARNING QUANTUM CONSTRAINTS FOR 1D INTERACTING SPIN-1/2 CHAINS**

**Additional details on machine learning quantum constraints for 1D spin-1/2 chains**

We use the infinite matrix product state representation to represent and sample our quantum states for 1D spin-1/2 systems in the thermodynamic limit:

\[
|\Psi\rangle = \sum_{\cdots, s_{n-1}, s_n, s_{n+1}, \cdots} \cdots M^{[n-1]} s_{n-1} M^{[n]} s_n M^{[n+1]} \cdots \times |\cdot \cdots, s_{n-1}, s_n, s_{n+1}, \cdots\rangle,
\]

where \(s_n = \uparrow, \downarrow\) is the spin on the \(n^{th}\) site. For a translation symmetric state, we can further simplify the expression by identify \(M^{[n]}_{\uparrow\downarrow} \to M_{\uparrow\downarrow}\), which are matrices with dimension \(\chi\). The expectation values of operators defined on a subsystem of size \(l_{\text{max}}\) can be evaluated straightforwardly by contract all configurations to the left and to the right of the subsystem. For our demonstrations, we set \(\chi = 8\) and \(l_{\text{max}} = 6\), which is sufficient for most scenarios away from criticality, and sample the \(M_{\uparrow\downarrow}\) matrices randomly for quantum states that we evaluate expectation values for our sample dataset.

The subsequent supervised machine learning and constrained optimization processes are similar to those of the 1D fermion insulator case, and detailed in previous sections, except that we no longer have a theoretical counterpart \(f\) for benchmarks. Despite the strongly-correlated physics present in such 1D spin chains and the inclusion of higher-order operators beyond the Wick’s Theorem, the data size and computational cost for obtaining \(f^*\) and the subsequent constrained optimization are actually smaller than those of the 1D fermion insulators under current settings, since we introduce a shorter truncation length \(l_{\text{max}} \ll \Lambda\) that reduces the enlisted operators. Also, importantly, the quantum states of the 1D spin chains are assumed to remain fully translation symmetric, while the quantum states in 1D fermion insulators are assumed to be translationally symmetric.
have a bipartite unit cell that essentially doubles the active degrees of freedom. Consequently, it suffices for us to employ a similar ANN architecture and even fewer training samples for the quantum constraints \( f^* \) on 1D spin chains.

With the quantum constraints \( f^* \), we can evaluate the ground-state properties of 1D spin Hamiltonian in Eq. 11 in the main text via constrained optimization. Similar to Fig. 4 in the main text, we show in Fig. 7 additional data on the ground-state energy with different weights \( \eta \) for the penalty.

On the other hand, the measurements of expectation values of certain degrees of freedom are once again relatively more difficult for possibly similar reasons discussed in the previous section, see Fig. 8. Correspondingly, we indeed observe flatter penalty responses versus the departure in the expectation values corresponding to such operators, especially \( S_x^g \), see Fig. 3. Still, our results on their expectation values follow the correct qualitative tendencies and even display some faint anomalous signatures at \( g \approx 0.5 \), the critical point of the transverse field Ising model \( (h = 0) \).

Machine learning quantum constraints for 1D spin-\(1/2\) chains with projected wave-function samples and Variational Monte Carlo method

In the main text, we establish quantum constraints based on Hartree Fock state and matrix product state samples, which can also generalize to and crossover various other quantum many-body ansatzes straightforwardly. In addition to the benefit of a more extensive training set, such generalizations also help reduce the bias associated with a particular ansatz and offer a more comprehensive sampling of (the relevant parts of) the Hilbert space.

Here, we consider the following projected wave-functions for 1D spin chains:

\[
|\Psi\rangle = \tilde{P}|\Psi_s\rangle|\Psi_\uparrow\rangle,
\]

where \(|\Psi_s\rangle\), \( s = \uparrow, \downarrow \) are free-fermion states and \( \tilde{P}\) is a projection operator onto the one-fermion-per-site Hilbert space: \( \tilde{P} = \prod_i(n_{\uparrow i} + n_{\downarrow i})(2 - n_{\uparrow i} - n_{\downarrow i}) \). In practice, we sample the parent free-fermion states \(|\Psi_s\rangle\) and evaluate the expectation values \( \langle \hat{O} \rangle \) of \(|\Psi\rangle\) through variational Monte Carlo method:

\[
\langle \hat{O}_j \rangle = \sum_\alpha \langle \Phi|\alpha\rangle\langle\alpha|\Phi\rangle \sum_\beta \langle\alpha|\hat{O}_j|\beta\rangle \frac{\langle\beta|\Phi\rangle}{\langle\alpha|\Phi\rangle},
\]

where \( \alpha \) and \( \beta \) are basis in the one-fermion-per-site Hilbert space. After generation of samples \( \langle \hat{O} \rangle \), we can carry out the subsequent supervised machine learning and constrained optimization as before.

We summarize the results in Fig. 9. While the overall performance is acceptable, we do not observe improvements over the performance in Fig. 4 in the main text, which only concerns samples from matrix product states. Possibly, this is because the matrix product representation is more suitable for the ground states of the transverse field Ising model; therefore, the inclusion of projected wave-function samples, though increasing the size and diversity of the training set, brings limited extra
value to this particular model application. Also, the projected wave-function ansatz is real-space-based and performs on finite system sizes, e.g., $L = 50$ in the current case; therefore, its incompatibility with the infinite matrix product states, though small, may confuse the ANNs upon slightly contradictory quantum constraints.

Quantum constraints for spin-1/2 XXZ chains

We can straightforwardly carry over our quantum constraints for 1D interacting spin-1/2 chains to analyze the ground-state properties of 1D XXZ model with ferromagnetic interactions ($J_z = 1$ and $1 \leq J_{xx} \leq 2$):

$$H = \sum_i \left[-J_{xx}(S_i^x S_{i+1}^x + S_i^y S_{i+1}^y) + J_z S_i^z S_{i+1}^z\right]. \quad (13)$$

In the absence of spontaneous translation symmetry breaking, the expectation value of energy per site is

$$E/N = -J_{xx}\langle S_0^x S_1^x \rangle - J_{xx}\langle S_0^y S_1^y \rangle - J_z\langle S_0^z S_1^z \rangle, \quad (14)$$

and its ground-state value in the $N \rightarrow \infty$ thermodynamic limit is obtained via constrained optimization and summarized in Fig. 10.
[16] E. Lieb, T. Schultz, and D. Mattis, “Two soluble models of an antiferromagnetic chain,” Annals of Physics 16, 407 (1961).