Time-dependent Neural Galerkin Method for Quantum Dynamics

Alessandro Sinibaldi, Douglas Hendry, Filippo Vicentini, Giuseppe Carleo

Time-dependent Neural Galerkin Method for Quantum Dynamics

Alessandro Sinibaldi,[1,2] Douglas Hendry,[1,2] Filippo Vicentini,[3,4] and Giuseppe Carleo[1,2]

[1] *Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[2] *Center for Quantum Science and Engineering, EPFL, Lausanne, Switzerland*
[3] *CPHT, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France.*
[4] *Collège de France, Université PSL, 11 place Marcelin Berthelot, 75005 Paris, France*

(Dated: December 17, 2024)

We introduce a classical computational method for quantum dynamics that relies on a global-in-time variational principle. Unlike conventional time-stepping approaches, our scheme computes the entire state trajectory over a finite time window by minimizing a physically motivated loss function that enforces the Schrödinger's equation. The variational state is parametrized with a Galerkin-inspired ansatz based on a time-dependent linear combination of time-independent Neural Quantum States. This structure effectively captures the relevant dynamical frequencies in the time evolution and is particularly well-suited for exploring long-time dynamics. We showcase the method's effectiveness by simulating global quantum quenches in the paradigmatic Transverse-Field Ising model in both 1D and 2D. By extracting the asymptotic long-time evolution, we uncover signatures of ergodicity breaking and absence of thermalization in two dimensions. Overall, the method presented here shows competitive performance compared to state-of-the-art time-dependent variational approaches and highlight the potential to explore previously inaccessible dynamical regimes of strongly interacting quantum systems.

Introduction–. The major limitation to exact calculations in quantum many-body physics is the exponential growth of the Hilbert space, which makes systems of more than a handful of particles practically inaccessible to brute-force approaches. Many-body variational methods are a powerful tool to circumvent this issue: instead of manipulating intractably large quantum states, a compressed representation relying on a smaller set of variational parameters can be used. Several classes of variational states have been used to study quantum dynamics, including several incarnations of Tensor Network wave functions [1, 2], as well as Neural Quantum States (NQS) [3].

In the context of time-dependent NQS, conventional approaches to the variational simulation of the dynamics rely on the explicit integration of the Schrödinger's equation to obtain a different variational state at each time-step. This is realized either by means of a stochastic implementation of the time-dependent variational principle (TDVP) [3–5] or by projecting the exactly evolved state at each time in the variational manifold [6–9]. However, those schemes suffer from an accumulation of errors coming from the sequential propagation of the small time-step dynamics.

To circumvent those issues, we investigate a global-in-time variational principle that optimizes the entire time-dependent trajectory at once, departing from the conventional sequential time-stepping paradigm and directly yielding the whole time-dependent solution to the initial value problem. Those global variational principles, or loss functions, minimize the deviations from the Schrödinger's solution at every point in time simultaneously, while the variational ansatz must be able to parametrize the solution at all times. Such approaches have already been explored in the field of Partial Differential Equation integration [10, 11], as a way to circumvent the requirement to construct a discrete mesh, or in the field of Physics-Inspired Neural Networks (PINNs) to solve a complex differential equation [12, 13]. Efforts to extend these concepts to quantum mechanical systems [14] have achieved limited success when compared to conventional approaches. The numerical studies have been constrained to small-scale systems and imaginary-time evolution, falling short of demonstrating practical advantages. While the exact reason remains unclear, we identified three major problems in the literature of PINNs: the loss functions used previously (i) did not fully respect all gauge invariances of the Hilbert space, (ii) underfit the initial condition [15, 16] and the deep neural architectures used before (iii) suffer from a spectral bias, making them incapable of learning high-frequency components [17, 18].

In order to move past such issues, we develop a global-in-time algorithm, combining ideas originating from both Machine Learning literature and physical requirements on the wave functions parameterizations. Our key contributions are twofold. First, we develop a physically-motivated loss function that explicitly preserves the fundamental quantum mechanical requirements of norm and phase invariance, enabling stable and accurate optimization. Second, we introduce a Galerkin-inspired parametrization based on a time-dependent linear combination of a small number of time-independent Neural Quantum States. A Fourier decomposition of the time-dependent coefficients ensures that the relevant dynamical frequencies of the system are well captured, and it generalizes the ansatz chosen for the coarse-grained spectral projection method [19]. This approach is particularly suited for studying fundamental questions in quantum many-body physics, such as thermalization [20–25], many-body localization [22, 26–29], and the emergence of hydrodynamic behavior [30–32], where accurate long-

time dynamics is essential but traditionally difficult to access.

Global-in-time variational principle– Our approach is based on a global dynamical variational principle that directly targets the entire time evolution, rather than evolving the state sequentially by integrating a local-in-time set of differential equations as in t-VMC [3–5] or by performing projections like in p-tVMC [6, 7, 9].

Given a time-dependent quantum state $|\Psi(t)\rangle$ belonging to the Hilbert space $\mathcal{H}$, the Schrödinger's equation

$$\frac{d}{dt}|\Psi(t)\rangle = -iH|\Psi(t)\rangle, \tag{1}$$

determines the evolution of the state under the time-independent Hamiltonian H [33].

We encode the solution to the equation above with a differentiable time-dependent variational state $|\Psi_\theta(t)\rangle$ depending on a set of parameters θ. We stress the difference from *established* time-dependent NQS approaches where the time-dependency is encoded in the parameters, as $|\Psi_{\theta(t)}\rangle$.

The parameters θ that give a valid solution to Eq. (1) can be determined by minimizing some distance between the left-hand side and right-hand side of the equation at all times. We start from the L^2 distance and additionally impose the physical requirements of norm [34] and phase invariance (see Section A for a detailed derivation). We obtain the following time-local loss function measuring the physically relevant deviations from the Schrödinger's dynamics,

$$\mathcal{L}(|\Psi_\theta\rangle) = \left\| P_{\perp|\Psi_\theta\rangle} \frac{|\dot{\Psi}_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} + i\bar{H}\frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right\|^2, \tag{2}$$

where $\bar{H} = H - \frac{\langle\Psi_\theta|H|\Psi_\theta\rangle}{\langle\Psi_\theta|\Psi_\theta\rangle}$ and $P_{\perp|\Psi_\theta\rangle} = 1 - \frac{|\Psi_\theta\rangle\langle\Psi_\theta|}{\langle\Psi_\theta|\Psi_\theta\rangle}$. For brevity, in the previous expressions we omit the explicit time dependence of the variational state.

The solution to the Schrödinger's equation in the interval $[0, T]$ can be obtained by minimizing the integrated loss function

$$L_{[0,T]}(\theta) = \frac{1}{T}\int_0^T dt\, \mathcal{L}(|\Psi_\theta(t)\rangle), \tag{3}$$

assuming that the initial condition $|\Psi_\theta(t=0)\rangle = |\Psi_0\rangle$, where $|\Psi_0\rangle$ is the initial state of the dynamics, is respected.

This loss function is positive semi-definite, $L \geq 0$, and attains the minimum value of $L = 0$ for states exactly satisfying the Schrödinger's equation Eq. (1) at all times in the considered interval $[0, T]$. We also stress that, as detailed in Section A, this global loss function does not require the state $|\Psi_\theta\rangle$ to have a unit norm, which means that, contrary to previous schemes featuring a global loss function [14, 19], it is compatible with generally unnormalized variational parametrizations of the wave function.

To evaluate Eq. (2) in practice, we use an efficient Monte Carlo estimator (see Section B for a detailed derivation)

$$L_{[0,T]}(\theta) = \frac{1}{T}\int_0^T dt\, \mathbb{E}_{|\Psi_\theta(\sigma,t)|^2}[|\bar{L}_{\text{loc}}(\sigma,t)|^2] \tag{4}$$

where $\bar{L}_{\text{loc}}(\sigma,t) = L_{\text{loc}}(\sigma,t) - \mathbb{E}_{|\Psi_\theta(\sigma,t)|^2}[L_{\text{loc}}(\sigma)]$ and $L_{\text{loc}}(\sigma,t)$ is a statistical estimator. In particular, the estimator corresponds to $L_{\text{loc}}(\sigma,t) = O_t(\sigma,t) + iE_{\text{loc}}(\sigma,t)$, namely the residual between the logarithmic time derivative $O_t(\sigma,t) = \partial_t \log\Psi_\theta(\sigma,t)$ and the local energy $E_{\text{loc}}(\sigma,t) = \langle\sigma|H|\Psi_\theta(t)\rangle/\Psi_\theta(\sigma,t)$. Each term in the integral expression Eq. (3) can be efficiently evaluated using standard Markov Chain Monte Carlo techniques by sampling configurations from $|\Psi_\theta(\sigma,t)|^2$. The time integration can be approximated using any scheme for numerical evaluation of 1D integrals. In our calculations, we rely on the Simpson's 1/3 rule [35] computed over an odd number of equally spaced integration points in $[0, T]$. To find the minimum of Eq. (3), we resort to the standard Adam optimizer [36].

Time-dependent Neural Quantum Galerkin– The global loss function $L_{[0,T]}(\theta)$ allows complete freedom for the choice of the time-dependent variational ansatz $|\Psi_\theta(t)\rangle$. One possible choice is to directly parametrize the wave function $\Psi(\sigma,t)$ as a function of the configurations σ and the time t, as already proposed in [14] by using a Neural Network architecture. However, such *unstructured* approaches are known to lead to poor generalization beyond the time-interval $[0, T]$ considered by the loss function.

In this manuscript we, instead adopt an approach inspired by the Galerkin method [37], taking an ansatz which consists in a linear combination of $M+1$ time-independent basis states $|\phi_i\rangle$ with time-dependent coefficients $c_i(t)$,

$$|\Psi_\theta(t)\rangle = \sum_{i=0}^{M} c_i(t)|\phi_i\rangle. \tag{5}$$

In the previous expression, $|\phi_0\rangle \equiv |\Psi_0\rangle$ is the fixed initial state of the dynamics and $\{|\phi_i\rangle\}_{i=1}^M$ are arbitrary variational states with parameters θ_i, such that $\theta = [\theta_1, \ldots, \theta_M]$. We remark that all the states appearing in Eq. (5) are in general not normalized.

The basis states $\{|\phi_i\rangle\}_{i=1}^M$ are represented as Neural Quantum States (NQS) [3], so we name our method as *time-dependent Neural Quantum Galerkin* (t-NQG). Any possible NQS architecture for standard Variational Monte Carlo can be used for t-NQG, including recurrent neural networks [38–43], convolutional neural networks [44–47] and transformers [48–53]. If the variational basis states enjoy high expressivity, the specific ansatz in Eq. (5) can be made arbitrarily close to the exact solution of the Schrödinger's equation by increasing M. In Section C, it is shown that for a particular choice of the $|\phi_i\rangle$ corresponding to the basis states of the

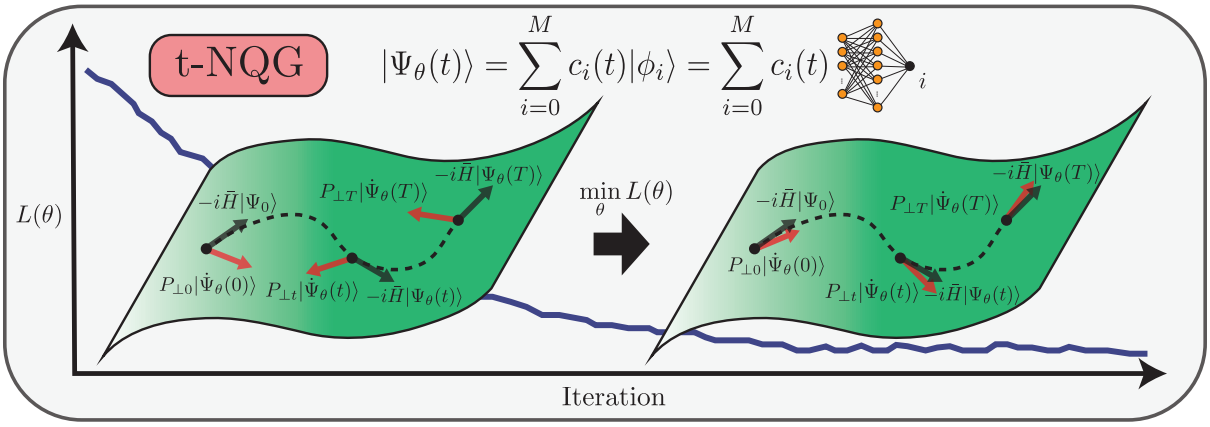

FIG. 1. Sketch of the time-dependent Neural Quantum Galerkin (t-NQG) method for the simulation of quantum dynamics. The approach consists in minimizing the global loss function $L_{[0,T]}(\theta)$ in Eq. (3) matching the Hamiltonian evolution $-i\bar{H}\,|\Psi_\theta(t)\rangle$ and the projected time derivative $P_{\perp t}\,|\dot{\Psi}_\theta(t)\rangle$ at each time t, where $P_{\perp t} = 1 - \frac{|\Psi_\theta(t)\rangle\langle\Psi_\theta(t)|}{\langle\Psi_\theta(t)|\Psi_\theta(t)\rangle}$ is the orthogonal projector to the variational state $|\Psi_\theta(t)\rangle$. The normalization of the state is not indicated for simplicity. The ansatz consists of the linear combination of $M+1$ time-independent basis states $|\phi_i\rangle$ parametrized as Neural Quantum States (NQS) with time-dependent coefficients $c_i(t)$.

coarse-grained approach in Ref. [19], the optimal number of states needs to grow at worst linearly in the total propagation time $M \sim N \times T$. A pictorial sketch of the t-NQG method is also presented in Fig. 1.

Time-dependent linear variational method– For the linear ansatz in Eq. (5) with fixed basis states $\{|\phi_0\rangle \equiv |\Psi_0\rangle, |\phi_1\rangle, \ldots, |\phi_M\rangle\}$, there exist optimal trajectories for the coefficients $c_i(t)$ exactly solving the Schrödinger's equation in the subspace spanned by the basis. These are given by the equations of the time-dependent linear variational method

$$c(t) = \exp\big(-it\,\mathbb{S}^{-1}\mathbb{H}\big)c(0), \tag{6}$$

where $c(0) = [1, 0, \ldots, 0]$ is a $M+1$-dimensional vector setting the initial condition, $\mathbb{S}$ is the overlap/Gram matrix of the basis states, and $\mathbb{H}$ is the reduced Hamiltonian matrix in the subspace of the basis. Their matrix elements are given by

$$\mathbb{S}_{ij} = \langle\phi_i|\phi_j\rangle\,, \qquad \mathbb{H}_{ij} = \langle\phi_i|\,H\,|\phi_j\rangle\,, \tag{7}$$

for $i, j = 0, \ldots, M$. The proof of Eq. (6) is reported in Section E. In this work, we consider a two-step approach. At first, we parametrize the coefficients $c_i(t)$ with a truncated Fourier expansion (see Section D for details) and minimize the loss in Eq. (3) with respect to the parameters of both basis states and coefficients. Then, the dynamics can be further improved by plugging the optimal coefficients coming from the solution of Eq. (6) with the optimized basis states. The matrix elements in Eq. (7) can be efficiently estimated (up to an irrelevant constant) through Monte Carlo sampling as shown in Section F. In Section H we prove that assuming the ansatz Eq. (5) with the optimal coefficients Eq. (6), the loss Eq. (3) is an effective measure of the discrepancy from the exact solution of the dynamics.

Long-time quantum dynamics– With the particular form of the ansatz in Eq. (5) decoupling spatial and time degrees of freedom, it is particularly natural to gain access to the long-time dynamics and thermalization of physical quantities. This can be done, for example, by considering the discrete Fourier transform of the optimal coefficients $c_i(t) = \sum_k \tilde{\gamma}_{ik} e^{i\tilde{\omega}_k t}$, where $\tilde{\omega}_k$ are the Fourier frequencies and $\tilde{\gamma}_{ik}$ the corresponding amplitudes. By removing fast oscillating components in the time evolution, the expectation value of an observable O in the long-time limit can be written as

$$\frac{\langle\Psi_\theta(t)|\,O\,|\Psi_\theta(t)\rangle}{\langle\Psi_\theta(t)|\Psi_\theta(t)\rangle} \overset{t\to\infty}{\approx} \frac{\sum_{ijk}\tilde{\gamma}^*_{ik}\mathbb{O}_{ij}\tilde{\gamma}_{jk}}{\sum_{ijk}\tilde{\gamma}^*_{ik}\mathbb{S}_{ij}\tilde{\gamma}_{jk}}, \tag{8}$$

where $\mathbb{O}_{ij} = \langle\phi_i|\,O\,|\phi_j\rangle$. The proof of Eq. (8) is reported in Section G.

Results–. To demonstrate the effectiveness of our approach, we consider the dynamics in the Transverse Field Ising (TFI) model with Hamiltonian

$$H_{\text{TFI}} = -J\sum_{\langle i,j\rangle}\sigma^z_i\sigma^z_j - h\sum_i \sigma^x_i, \tag{9}$$

where $\sigma^{z,x}_i$ are the z, x Pauli matrices acting on site i, J is the coupling strength, h is the transverse magnetic field and $\langle i,j\rangle$ indicates nearest-neighbor sites. Without loss of generality, we set $J = 1$. The TFI model exhibits a quantum phase transition in correspondence of the critical fields $h_c^{\text{1D}} = 1$ for the 1D chain [54] and $h_c^{\text{2D}} = 3.04438(2)$ [55] for the 2D square lattice, separating a ferromagnetic phase for $h < h_c$ from a paramagnetic phase for $h > h_c$. It is a paradigmatic example where the interplay between interactions and the field leads to rich dynamical behaviors.

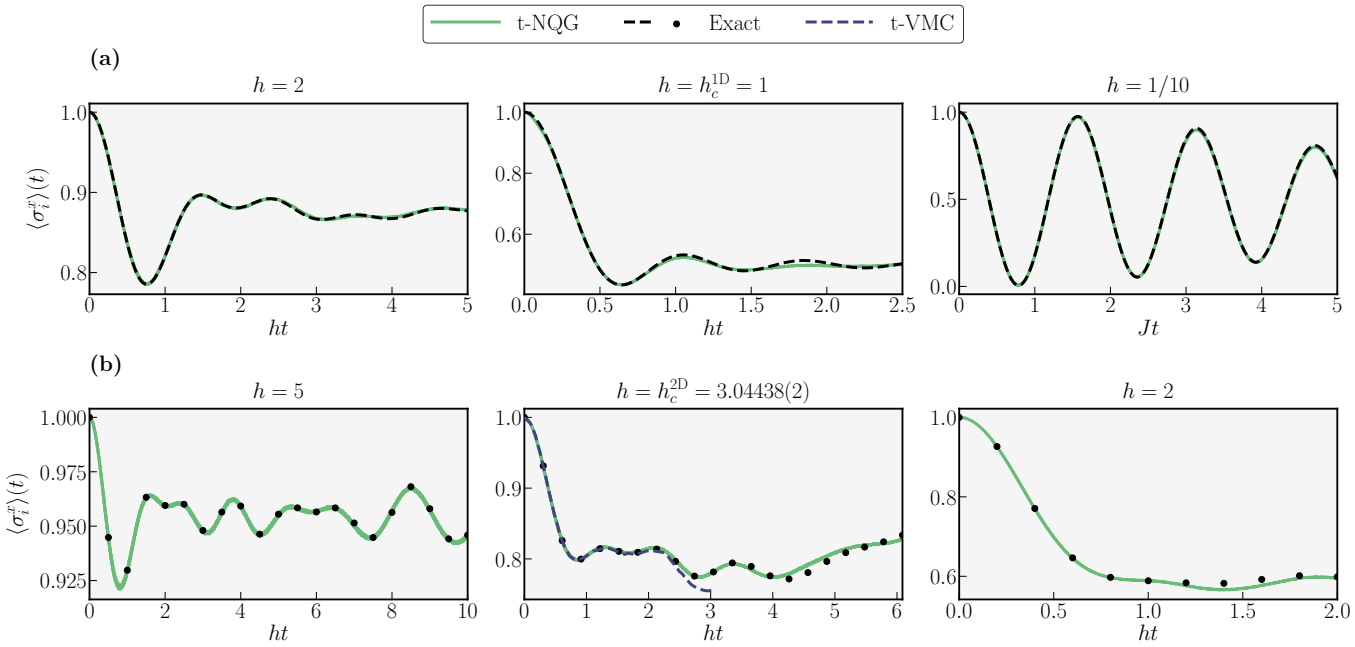

FIG. 2. Time evolution of the transverse magnetization $\langle \sigma_i^x \rangle(t)$ after a global quantum quench in the 1D **(a)** and 2D **(b)** TFI model from the paramagnetically polarized state $|\Psi_0\rangle = \bigotimes_{i=1}^{N} |+\rangle_i$ to the paramagnetic phase, to the critical point and to the ferromagnetic phase. The 1D system consists of $N = 40$ spins while the 2D system is a 6×6 lattice. In the 1D system, we use time sub-intervals of $\Delta T = 0.25$ with $128 + 1$ time integration points for $h = 2, 1$, while sub-intervals of $\Delta T = 0.5$ with $256 + 1$ points for $h = 1/10$. For the 2D, sub-intervals of $\Delta T = 0.2$ with $256 + 1$ integration points are used for all the quenches. We have employed $M = N/2$ RBM basis states. The coefficients are expanded onto $N_b = 64$ Fourier basis functions in 1D, and $N_b = 128$ in 2D. For each integration point the loss is estimated with 512 Monte Carlo samples.

We simulate the dynamics of global quenches both in the 1D and 2D TFI model of N spins with periodic boundary conditions. In particular, we prepare the ground state of H_{TFI} with $h = \infty$, namely the paramagnetically polarized state $|\Psi_0\rangle = \bigotimes_{i=1}^{N} |+\rangle_i$, and evolve it under H_{TFI} for different values of h. This setup has also been used as a benchmark in other variational calculations based on NQS [9, 56].

We have experimentally observed that converging to the minimum of $L_{[0,T]}(\theta)$ becomes harder as the final time T is larger, especially for bigger system sizes, and increasing the number of basis states M does not counter this sufficiently. We therefore partition the loss into several sub-intervals of length ΔT

$$L_{[0,T]}(\theta) = \sum_{i=0}^{[T/\Delta T]-1} L_{[i\Delta T,(i+1)\Delta T]}(\theta_i), \qquad (10)$$

where the parameters θ_i encode the solution in the $i-$th interval. At every interval i, the initial condition is taken to be the wave function at time $i\Delta T$ obtained from the solution of the previous interval. The time-independent basis states are encoded as complex-valued Restricted Boltzmann Machine (RBM) ansätze [3] with different parameters at every interval.

In Fig. 2 we compare the time evolution simulated with t-NQG against exact benchmarks both in the 1D and 2D system. We observe that t-NQG is able to accurately reproduce several quench dynamics in different phases up to fairly large times. For the more challenging quench at $h = h_c^{2D}$, we also compare with the state-of-the-art t-VMC simulation [56], demonstrating that t-NQG is capable to reach higher precision and longer times. This improvement stems from the fact that t-VMC accumulates errors during the dynamics due to the local-in-time integration, while our method maintains high accuracy

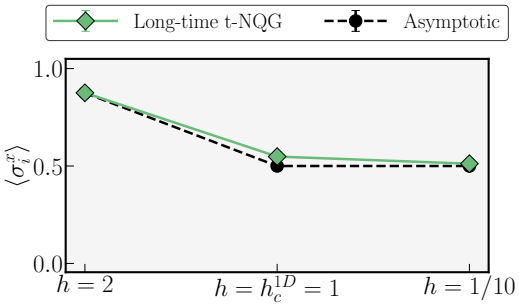

FIG. 3. Long-time value of the transverse magnetization $\langle \sigma_i^x \rangle$ predicted from t-NQG compared with the asymptotic exact value in the 1D spin chain with $L = 40$ sites for the three quenches. The error bars on the long-time t-NQG are assigned by repeating the calculation for 10 independent realizations. The lines are guides for the eye.

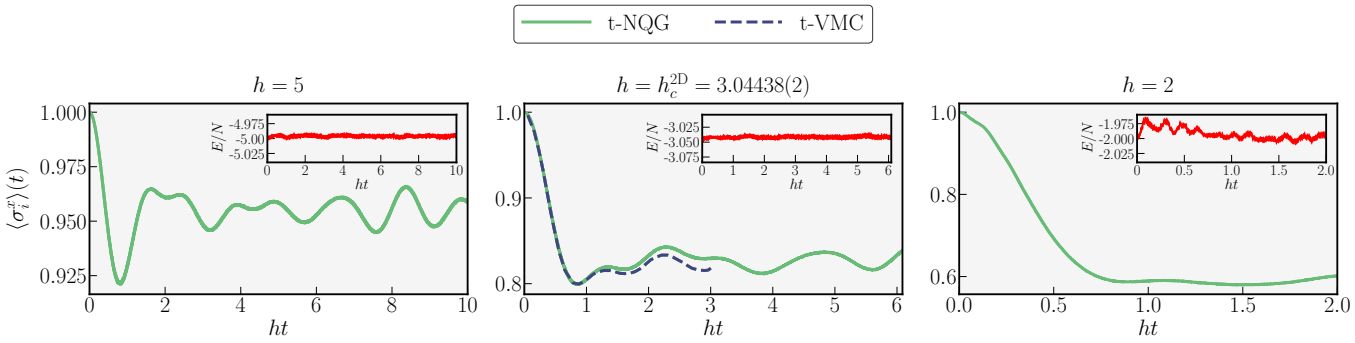

FIG. 4. Time evolution of the transverse magnetization $\langle \sigma_i^x \rangle(t)$ after a global quantum quench in the 2D TFI model from the paramagnetically polarized state $|\Psi_0\rangle = \bigotimes_{i=1}^{N} |+\rangle_i$ to the paramagnetic phase, to the critical point and to the ferromagnetic phase. The system is a 8×8 lattice. Time sub-intervals of $\Delta T = 0.2$ with $256 + 1$ integration points are used for all the quenches. We have employed $M = N/2$ RBM basis state and $N_b = 128$ Fourier basis functions. For each integration point the loss is estimated with 512 Monte Carlo samples. The inset displays the energy density E/N during the time evolution.

even at long times by directly targeting the entire trajectory. While t-VMC can only access the information at the current time-step, t-NQG takes a global perspective on the full trajectory and can adjust the early dynamics to better accommodate later evolution. The $h = 2$ dynamics in 2D reveals to be arduous due to the more complicated optimizations, probably coming from the difficulty in learning ferromagnetic basis states. In Fig. 3 we compare the predicted long-time value of the observable with the asymptotic value of the exact dynamics, showing that t-NQG is able to faithfully extrapolate to the infinite time limit only by accessing a portion of the finite-time dynamics.

After benchmarking on the exactly solvable systems, we investigate the time evolutions in a 8×8 lattice where the exact dynamics is not accessible. The results are displayed in Fig. 4 and are consistent with the trajectories obtained for the smaller system size. Similarly to Fig. 2, we observe that t-VMC [56] undershoots with respect to t-NQG for the critical quench. The insets of Fig. 4 show the value of the energy density during the time evolutions. The energy is well conserved for $h = 5, h_c^{2D}$, while it exhibits wider oscillations for $h = 2$. This supports the accuracy of the calculations for the larger lattice and is consistent with the greater complexity of simulating the $h = 2$ dynamics already observed in the smaller system.

For the 8×8 lattice, we also compute the long-time dynamics from t-NQG and we compare it with the thermal expectation value $\langle O \rangle_{\text{therm}} = \text{Tr}[e^{-\beta_{\text{eff}} H} O]/\text{Tr}[e^{-\beta_{\text{eff}} H}]$ to test the thermalization hypothesis [57, 58]. The effective inverse temperature β_{eff} is fixed by the conservation of energy condition, namely by solving the equation $\langle H \rangle_{\text{therm}} = \langle \Psi_0 | H | \Psi_0 \rangle / \langle \Psi_0 | \Psi_0 \rangle$. In the presence of ergodic dynamical behavior, the effective thermal average should coincide with the long-time dynamics, according to the Boltzmann prescription. The thermal expectation values are computed by Quantum Monte Carlo (QMC) simulations based on the loop algorithm [59–61]. Fig. 5 shows that the long-time value of the observable matches

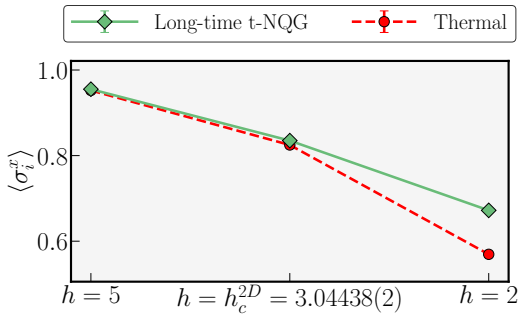

FIG. 5. Long-time value of the transverse magnetization $\langle \sigma_i^x \rangle$ predicted from t-NQG compared with the thermal value computed from Quantum Monte Carlo in the 2D 8×8 lattice for the three quenches. The error bars on the long-time t-NQG are assigned by repeating the calculation for 10 independent realizations. The lines are guides for the eye.

the effective thermal average for $h = 5, h_c^{2D}$, meaning that these quench dynamics are ergodic and thermalize in the long time limit. For $h = 2$, the t-NQG predictions deviate significantly from QMC calculations. This behavior suggests a potential breakdown of ergodicity and thermalization, where the system becomes trapped in long-lived metastable states when driven far from equilibrium—a phenomenon reminiscent of the behavior observed in interacting lattice bosons [62]. However, we note that this deviation could also arise from limitations in our finite-time window, as the t-NQG fit may not extend far enough to accurately capture the true long-time behavior. A definitive assessment of this intriguing possibility of non-thermalization in the quenched 2D TFI model would require additional systematic investigations of longer time scales and different system sizes.

Conclusion– In this work, we introduce a classical variational method for simulating the dynamics of many-body quantum systems. The scheme optimizes the entire quantum trajectory at once avoiding the accumula-

tion of errors typical of time-stepping algorithms. Our approach relies on a global-in-time variational principle, in the form of a physically motivated loss function enforcing the Schrödinger's equation at each time, and employs a Galerkin-inspired ansatz based on Neural Quantum States (NQS). We name our method *time-dependent Neural Quantum Galerkin* (t-NQG). We demonstrate the capabilities of t-NQG by simulating global quench dynamics in the 1D and 2D Transverse Field Ising model, achieving competitive performance with established time-dependent variational schemes. Thanks to the space-time factorized form of the ansatz, we are able to investigate the long-time limit of the dynamics and thermalization of the system. This work paves the way for leveraging NQS to study unexplored out-of-equilibrium phenomena in strongly-correlated quantum systems. Many extensions and applications could be envisaged. We mention here that it is particularly natural to provide an extension featuring more expressive, deep neural networks as basis states, beyond the simple RBM adopted in this work. Moreover, applications to benchmark the dynamics of noisy quantum computers are especially natural, extending the capabilities of other classical simulation approaches, typically limited to short time scales or one-dimensional geometries.

Data availability– The numerical simulations with the t-NQG method are based on NetKet [63, 64]. The code will be made public in a later revision of the manuscript. The exact benchmarks are realized using QuSpin [65, 66] and the finite-temperature Quantum Monte Carlo calculations using the ALPS library[60, 61].

Acknowledgments– We thank R. Martinazzo, Z. Denis, L. L. Viteritti, and L. Fioroni for insightful discussions and M. Bukov for helping with the exact simulations. A. S. is supported by SEFRI under Grant No. MB22.00051 (NEQS - Neural Quantum). F.V. acknowledges support by the French Agence Nationale de la Recherche through the NDQM project, grant ANR-23-CE30-0018. We acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LEONARDO, hosted by CINECA (Italy) and the LEONARDO consortium through the EuroHPC Development Access call EHPC-DEV-2024D10-055.

Note: during the preparation of this manuscript, we became aware of a related work that has been carried on in parallel by A. Van de Walle, M. Schmitt and, A. Bohrdt, which will appear simultaneously on the preprint server.

APPENDIX

A. Loss function

Here we provide the complete derivation of the loss function used in the main text. In general, to satisfy Eq. (1) with a variational ansatz $|\Psi_\theta\rangle$ one can minimize the L^2 loss between the time derivative and the Hamiltonian evolution

$$\| \, |\dot{\Psi}_\theta\rangle + iH \, |\Psi_\theta\rangle \, \|^2. \tag{11}$$

For physical applications, however, the loss function above is unsatisfactory, since it does not incorporate the geometry of quantum states. Specifically, a physically robust loss function must be invariant under two fundamental transformations: arbitrary (possibly time-dependent) changes in the normalization and global phase rotations of the state $|\Psi_\theta\rangle$. To make the loss invariant under norm changes, it is enough to consider the distance between the normalized states, namely

$$\mathcal{L}'(|\Psi_\theta\rangle) = \left\| \frac{d}{dt}\left(\frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right) + iH \frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right\|^2. \tag{12}$$

To guarantee the invariance under the phase variation, we consider how Eq. (12) is modified after the transformation $|\Psi_\theta\rangle \to e^{i\phi} |\Psi_\theta\rangle$, where $\phi = \phi(t)$. We obtain

$$\begin{aligned} \mathcal{L}'(e^{i\phi}|\Psi_\theta\rangle) &= \left\| \frac{d}{dt}\left(\frac{e^{i\phi}|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right) + ie^{i\phi} H \frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right\|^2 = \left\| \frac{d}{dt}\left(\frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right) + iH \frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} + i\dot{\phi} \frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right\|^2 = \\ &= \mathcal{L}(|\Psi_\theta\rangle) + 2\,\mathrm{Re}\left(i \frac{\langle\epsilon(\Psi_\theta)|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right)\dot{\phi} + \dot{\phi}^2, \end{aligned} \tag{13}$$

where we denote $|\epsilon(\Psi_\theta)\rangle \equiv \frac{d}{dt}\left(\frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} \right) + iH \frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}}$. We note that Eq. (13) depends only on $\dot{\phi}$ and not on ϕ. We can impose that the loss is invariant under phase variations by choosing the $\dot{\phi}$ which minimizes $\mathcal{L}'(e^{i\phi}|\Psi\rangle)$, in

the same way as done in the time-dependent variational principle [67]. Thus, by putting $\partial_{\dot{\phi}}\mathcal{L}'(e^{i\phi}|\Psi_\theta\rangle) = 0$ we get the optimal phase velocity

$$\dot{\phi} = \mathrm{Im}\left(\frac{\langle\dot{\Psi}_\theta|\Psi_\theta\rangle}{\langle\Psi_\theta|\Psi_\theta\rangle}\right) - \langle H\rangle, \tag{14}$$

where $\langle H\rangle = \langle\Psi_\theta| H |\Psi_\theta\rangle / \langle\Psi_\theta|\Psi_\theta\rangle$. Plugging Eq. (14) into Eq. (13) and rewriting yields the fully norm and phase invariant expression used in the main text

$$\mathcal{L}(|\Psi_\theta\rangle) = \left\| \left(1 - \frac{|\Psi_\theta\rangle\langle\Psi_\theta|}{\langle\Psi_\theta|\Psi_\theta\rangle}\right)\frac{|\dot{\Psi}_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}} + i(H - \langle H\rangle)\frac{|\Psi_\theta\rangle}{\sqrt{\langle\Psi_\theta|\Psi_\theta\rangle}}\right\|^2. \tag{15}$$

We note that imposing the invariance under norm and phase results in including the projector onto the state $|\Psi_\theta\rangle$ and by shifting the Hamiltonian with the energy of the state.

B. Monte Carlo evaluation of the loss function

The loss function Eq. (2) can be efficiently evaluated using Monte Carlo sampling. Here we provide the complete derivation of the stochastic estimators. We start by introducing the operator L through its action on $|\Psi_\theta\rangle$

$$L|\Psi_\theta\rangle = |\dot{\Psi}_\theta\rangle + iH|\Psi_\theta\rangle. \tag{16}$$

We remark that L is not Hermitian in general. The loss $\mathcal{L}(|\Psi_\theta\rangle)$ can be compactly written as

$$\mathcal{L}(|\Psi_\theta\rangle) = \frac{\langle\Psi_\theta|L^\dagger L|\Psi_\theta\rangle}{\langle\Psi_\theta|\Psi_\theta\rangle} - \frac{\langle\Psi_\theta|L^\dagger|\Psi_\theta\rangle\langle\Psi_\theta|L|\Psi_\theta\rangle}{\langle\Psi_\theta|\Psi_\theta\rangle\langle\Psi_\theta|\Psi_\theta\rangle}. \tag{17}$$

The previous expression corresponds to the quantum variance of the operator L. By introducing the completeness relation of a basis $\{|\sigma\rangle\}$ of the Hilbert space, Eq. (17) can be evaluated as the statistical variance of the local estimator of L

$$\mathcal{L}(|\Psi_\theta\rangle) = \mathbb{E}_{|\Psi_\theta(\sigma)|^2}[|L_{\mathrm{loc}}(\sigma)|^2] - |\mathbb{E}_{|\Psi_\theta(\sigma)|^2}[L_{\mathrm{loc}}(\sigma)]|^2, \tag{18}$$

where the local estimator is

$$L_{\mathrm{loc}}(\sigma) = \frac{\langle\sigma| L |\Psi_\theta\rangle}{\langle\sigma|\Psi_\theta\rangle} = O_t(\sigma) + iE_{\mathrm{loc}}(\sigma), \tag{19}$$

where $O_t(\sigma) = \partial_t \log\Psi_\theta(\sigma) = \dot{\Psi}_\theta(\sigma)/\Psi_\theta(\sigma)$ is the logarithmic time derivative and $E_{\mathrm{loc}}(\sigma) = \langle\sigma|H|\Psi_\theta\rangle/\Psi_\theta(\sigma)$ is the local energy.

C. Optimal number of basis states for accurate t-NQG dynamics

Here we present a proof showing that with the ansatz in Eq. (5) it is possible to keep the error of the variational dynamics below a small finite value by increasing the number of basis states M polynomially with the final time T and the system size N. For the specific purpose of the calculation, we consider the basis states to be the ones of the coarse-grained (CG) approach to quantum dynamics [19], but these considerations are general and can be extended to any enough expressive set of basis states such as Lanczos or Chebyshev vectors [68].

In the CG approach the approximate time-evolved state is constructed as $|\Psi^{\mathrm{CG}}(t)\rangle = \sum_i e^{-it\mu_i}|w_i\rangle$, where $\mu_0, \ldots, \mu_{M-1} \in [E_{\min}, E_{\max}]$ are M energies uniformly distributed in the spectrum of H and the corresponding states $|w_i\rangle$ capture the components of the initial state $|\Psi_0\rangle$ on the energy eigenstates with energies closest to μ_i. The CG states are obtained by minimizing

$$\sum_i \langle w_i|(H-\mu_i)^2|w_i\rangle - \langle\lambda|w_i\rangle + \langle w_i|\lambda\rangle, \tag{20}$$

where $|\lambda\rangle$ is the Lagrange multiplier for the constraint on the initial condition $|\Psi_0\rangle = \sum_i |w_i\rangle$. The minimal solution gives $|w_i\rangle = W_i|\Psi_0\rangle$ with the weighting operator W_i of the energy μ_i defined as

$$W_i = \frac{(H-\mu_i)^{-2}}{\sum_j (H-\mu_j)^{-2}}. \tag{21}$$

The CG evolved state can also be written as $|\Psi^{\mathrm{CG}}(t)\rangle = U^{\mathrm{CG}}(t)|\Psi_0\rangle$ with the CG approximation of the time evolution operator corresponding to $U^{\mathrm{CG}}(t) = \sum_i e^{-it\mu_i} W_i$. The operator $U^{\mathrm{CG}}(t)$ approximates the exact propagator $U(t) = e^{-itH}$ if the weighting operators satisfy $W_i \approx \delta(H - \mu_i)$. The error of the CG dynamics can thus be written as $\| |\Psi(t)\rangle - |\Psi^{\mathrm{CG}}(t)\rangle \| = \| U(t)\,|\Psi_0\rangle - U^{\mathrm{CG}}(t)\,|\Psi_0\rangle\,\| = \|R(t)|\Psi_0\rangle\|$ where the error operator $R(t)$ corresponds to

$$R(t) = U(t) - U^{\mathrm{CG}}(t) = \frac{\sum_i (H-\mu_i)^{-2}(e^{-itH} - e^{-it\mu_i})}{\sum_j (H-\mu_j)^{-2}}. \tag{22}$$

Choosing the CG energies so that they are evenly distributed on the interval $[E_{\min}, E_{\max}]$ as $\mu_i = E_{\min} + ((2l+1)/2M)(E_{\max} - E_{\min})$ for $l = 0, \ldots, M-1$ bounds the operator 2-norm of the error operator linearly in time as $\|R(t)\|_2 \leq t(E_{\max} - E_{\min})/2M$. Finally, the error of the CG evolved state, considering $\langle\Psi_0|\Psi_0\rangle = 1$ for simplicity, is bounded in terms of the final time T and the number of CG states M as

$$\| |\Psi(t)\rangle - |\Psi^{\mathrm{CG}}(t)\rangle \| \leq \frac{1}{2}(E_{\max} - E_{\min})\frac{T}{M} \leq O\left(\frac{NT}{M}\right), \quad \forall\, t \in [0, T] \tag{23}$$

in the case of a physical Hamiltonian with extensive energy spectrum.

D. Truncated Fourier series for the coefficients

The time-dependent coefficients $c_i(t)$ of Eq. (5) must satisfy the initial conditions $c_i(t=0) = \delta_{i,0}$. To achieve this, we set $c_0(t) = 1\ \forall t$ and we expand the other coefficients in a truncated Fourier basis respecting the initial condition

$$c_i(t) = \sum_{k=1}^{N_b} \gamma_{ik}(e^{i\omega_k t} - 1), \quad i > 0 \tag{24}$$

where N_b is the number of basis functions, γ_{ik} are variational parameters, and the frequencies ω_k are initialized to energies evenly spaced in the spectrum of H. The minimum and maximum energies of H are estimated by standard Variational Monte Carlo. The choice of the ω_k ensures proper coverage of the relevant dynamical time scales and is motivated also by the coarse-grained dynamics of [19]. To enhance the expressivity of the ansatz, we let however the frequencies be variational, such that the set of parameters θ includes γ_{ik} and ω_k as well.

E. Equations of the time-dependent linear variational method

To derive Eq. (6), we first write the Schrödinger's equation Eq. (1) for a linear ansatz Eq. (5) obtaining

$$\sum_j \dot{c}_j(t)\,|\phi_j\rangle = -i \sum_j c_j(t) H\,|\phi_j\rangle\,. \tag{25}$$

Then, we can search for a solution of Eq. (25) in the subspace spanned by $\{|\phi_i\rangle\}$ by projecting it onto each basis state, leading to the set of equations

$$\sum_j \mathbb{S}_{ij}\dot{c}_j(t) = -i \sum_j \mathbb{H}_{ij} c_j(t) \quad \forall\ i, \tag{26}$$

where the elements of $\mathbb{S}$ and $\mathbb{H}$ are defined as in Eq. (7). The solution of the system of ordinary differential equations Eq. (26) is known and corresponds to the exponential form in Eq. (6).

F. Monte Carlo estimation of the matrix elements

For all the computations in the subspace spanned by the basis states such as Eq. (6), it is enough to know the overlap matrix $\mathbb{S}$ and the matrix representation of any observable O in the basis, say $\mathbb{O}$, up to a common constant k. This observation is essential to be able to estimate their entries with Monte Carlo sampling, since in general the

states $|\phi_i\rangle$ are not normalized. Therefore, we can introduce an arbitrary probability distribution $\Pi(\sigma)$ that can be used to estimate $\mathbb{S}_{ij}/k$ and $\mathbb{O}_{ij}/k$ where k is the normalization of Π. Indeed, we can write

$$\begin{aligned}
\frac{\mathbb{S}_{ij}}{\sum_{\sigma'}\Pi(\sigma')} &= \frac{\sum_\sigma \Pi(\sigma)}{\sum_{\sigma'}\Pi(\sigma')}\left[\frac{\phi_i^*(\sigma)\phi_j(\sigma)}{\Pi(\sigma)}\right] = \mathbb{E}_{\Pi(\sigma)}\left[\frac{\phi_i^*(\sigma)\phi_j(\sigma)}{\Pi(\sigma)}\right], \\
\frac{\mathbb{O}_{ij}}{\sum_{\sigma'}\Pi(\sigma')} &= \frac{1}{2}\frac{\sum_\sigma \Pi(\sigma)}{\sum_{\sigma'}\Pi(\sigma')}\left[\frac{\phi_i^*(\sigma)\langle\sigma|O|\phi_j\rangle}{\Pi(\sigma)} + \frac{\langle\phi_i|O|\sigma\rangle\phi_j(\sigma)}{\Pi(\sigma)}\right] = \frac{1}{2}\mathbb{E}_{\Pi(\sigma)}\left[\frac{\phi_i^*(\sigma)\langle\sigma|O|\phi_j\rangle}{\Pi(\sigma)} + \frac{\langle\phi_i|O|\sigma\rangle\phi_j(\sigma)}{\Pi(\sigma)}\right].
\end{aligned} \tag{27}$$

We note that for the observable matrix we employ a symmetrized estimator. Since we need a distribution Π with support over all the basis states to accurately estimate the expectation values in Eq. (27), we consider $\Pi(\sigma) = \sum_{i=0}^{M}|\phi_i(\sigma)|^2$.

G. Long-time dynamics of observables

By computing the expectation value of a generic observable O on the ansatz Eq. (5) where the coefficients $c_i(t)$ are expanded in a Fourier series $c_i(t) = \sum_k \tilde{\gamma}_{ik}e^{i\tilde{\omega}_k t}$, we obtain

$$\frac{\langle\Psi_\theta(t)|O|\Psi_\theta(t)\rangle}{\langle\Psi_\theta(t)|\Psi_\theta(t)\rangle} = \frac{\sum_{ijkl}\tilde{\gamma}_{ik}^*\tilde{\gamma}_{jl}e^{i(\omega_l-\omega_k)t}\langle\phi_i|O|\phi_j\rangle}{\sum_{ijkl}\tilde{\gamma}_{ik}^*\tilde{\gamma}_{jl}e^{i(\omega_l-\omega_k)t}\langle\phi_i|\phi_j\rangle} \tag{28}$$

Now, in the limit of $t\to\infty$ all the oscillating factors in the numerator and the denominator have a negligible contribution with respect to the non-oscillating terms, so they can be neglected leading to

$$\frac{\langle\Psi_\theta(t)|O|\Psi_\theta(t)\rangle}{\langle\Psi_\theta(t)|\Psi_\theta(t)\rangle} \overset{t\to\infty}{\approx} \frac{\sum_{ijk}\tilde{\gamma}_{ik}^*\tilde{\gamma}_{jk}\langle\phi_i|O|\phi_j\rangle}{\sum_{ijk}\tilde{\gamma}_{ik}^*\tilde{\gamma}_{jk}\langle\phi_i|\phi_j\rangle}, \tag{29}$$

which corresponds to Eq. (8).

H. Bound on the exact dynamics

The dynamics of the ansatz Eq. (5) with fixed basis states and the optimal coefficients Eq. (6) can be written in terms of the projected Hamiltonian $H_Q = QHQ$ as $|\Psi_\theta(t)\rangle = e^{-itH_Q}|\Psi_0\rangle$, where $Q = \sum_{i,j=0}^{M}(\mathbb{S}^{-1})_{ij}|\phi_i\rangle\langle\phi_j|$ is the orthogonal projector onto the subspace spanned by the basis. For simplicity, we consider the initial state $|\Psi_0\rangle$ to be normalized, such that consequently $|\Psi_\theta(t)\rangle$ keeps a unit norm, but the calculation is completely general. Under this condition, our local-in-time loss function becomes

$$\mathcal{L}(t) = \|(1-|\Psi_\theta(t)\rangle\langle\Psi_\theta(t)|)|\dot{\Psi}_\theta(t)\rangle + i\bar{H}|\Psi_\theta(t)\rangle\|^2 = \|(1-|\Psi_\theta(t)\rangle\langle\Psi_\theta(t)|)(|\dot{\Psi}_\theta(t)\rangle + iH|\Psi_\theta(t)\rangle)\|^2 = \||r(t)\rangle\|^2, \tag{30}$$

where the residual state is defined as $|r(t)\rangle = |\dot{\Psi}_\theta(t)\rangle + iH|\Psi_\theta(t)\rangle = i(1-Q)H|\Psi_\theta(t)\rangle$. The last equality in Eq. (30) is non trivial and comes from the fact that $(1-|\Psi_\theta(t)\rangle\langle\Psi_\theta(t)|)|r(t)\rangle = |r(t)\rangle$ since $(1-|\Psi_\theta(t)\rangle\langle\Psi_\theta(t)|)(1-Q) = 1-Q$. We remark that the loss can also be written in terms of the matrices in the basis subspace as $\mathcal{L}(t) = \||r(t)\rangle\|^2 = c^\dagger(t)\Sigma c(t)$, where $c(t)$ corresponds to Eq. (6) and $\Sigma = \mathbb{H}^{(2)} - \mathbb{H}\mathbb{S}^{-1}\mathbb{H}$ with $\mathbb{H}^{(2)}_{ij} = \langle\phi_i|H^2|\phi_j\rangle$.

The integrated error with respect to the exact evolution $\Delta^2(T) = \int_0^T dt\,\||\epsilon(t)\rangle\|^2$ where $|\epsilon(t)\rangle = |\Psi(t)\rangle - |\Psi_\theta(t)\rangle$ with $|\Psi(t)\rangle = e^{-itH}|\Psi_0\rangle$ can be bounded by the global loss L of Eq. (3). Indeed, the time derivative of the error can be expanded as

$$\begin{aligned}
\frac{d}{dt}\||\epsilon(t)\rangle\|^2 &= \frac{d}{dt}\Big(\langle\Psi(t)|\Psi(t)\rangle + \langle\Psi_\theta(t)|\Psi_\theta(t)\rangle - \langle\Psi_\theta(t)|\Psi(t)\rangle - \langle\Psi(t)|\Psi_\theta(t)\rangle\Big) \\
&= -2\,\mathrm{Re}[\langle\dot{\Psi}(t)|\Psi_\theta(t)\rangle + \langle\Psi(t)|\dot{\Psi}_\theta(t)\rangle] = -2\,\mathrm{Re}[i\langle\Psi(t)|H|\Psi_\theta(t)\rangle - i\langle\Psi(t)|QH|\Psi_\theta(t)\rangle] \\
&= -2\,\mathrm{Re}[\langle\Psi(t)|r(t)\rangle] \le 2|\langle\Psi(t)|r(t)\rangle| \le 2\||\Psi(t)\rangle\|\cdot\||r(t)\rangle\| = 2\sqrt{\mathcal{L}(t)}.
\end{aligned} \tag{31}$$

Therefore, we can write

$$\Delta^2(T) \le 2\int_0^T dt\,\sqrt{\mathcal{L}(t)} \le 2T\max_t \mathcal{L}(t) \le 2T\sqrt{L}, \tag{32}$$

where the last inequality follows from $\left(\int_0^T dt\ \sqrt{\mathcal{L}(t)}\right)^2 \leq \left(\int_0^T dt'\right)\left(\int_0^T dt\ \mathcal{L}(t)\right) = T^2 \cdot L$. Therefore, the error with the exact dynamics can be bounded as

$$\Delta(T) \leq \sqrt{2T \max_t \mathcal{L}(t)} \leq \sqrt{2T} L^{1/4}. \tag{33}$$

From this, it is also possible to bound the error on the expectation value of any observable O at each time as:

$$|\delta O(t)| = |\langle O\rangle_{|\Psi(t)\rangle} - \langle O\rangle_{|\Psi_\theta(t)\rangle}| = |\langle \epsilon(t)|O|\Psi(t)\rangle + \langle \Psi(t)|O|\epsilon(t)\rangle - \langle \epsilon(t)|O|\epsilon(t)\rangle| \leq \tag{34}$$

$$\leq \|O\|_2 (2\| \,|\epsilon(t)\rangle\,\| + \| \,|\epsilon(t)\rangle\,\|^2) \leq \|O\|_2 (2\Delta(t) + \Delta^2(t)) \leq 2\|O\|_2 (\sqrt{2t} L^{1/4} + t\sqrt{L}), \tag{35}$$

where $\|O\|_2$ indicates the operator 2-norm of O.

[1] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[2] A. J. Daley, C. Kollath, U. Schollwock, and G. Vidal, Journal of Statistical Mechanics-Theory and Experiment , P04005 (2004).
[3] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[4] G. Carleo, F. Becca, M. Schiro, and M. Fabrizio, Scientific Reports **2**, 243 (2012).
[5] G. Carleo, F. Becca, L. Sanchez-Palencia, S. Sorella, and M. Fabrizio, Phys. Rev. A **89**, 031602 (2014).
[6] K. Donatella, Z. Denis, A. Le Boité, and C. Ciuti, Phys. Rev. A **108**, 022210 (2023).
[7] A. Sinibaldi, C. Giuliani, G. Carleo, and F. Vicentini, Quantum **7**, 1131 (2023).
[8] J. Nys, G. Pescia, A. Sinibaldi, and G. Carleo, Nature Communications **15**, 9404 (2024).
[9] L. Gravina, V. Savona, and F. Vicentini, arXiv preprint (2024), arXiv:2410.10720.
[10] I. Lagaris, A. Likas, and D. Fotiadis, IEEE Transactions on Neural Networks **9**, 987–1000 (1998).
[11] J. Sirignano and K. Spiliopoulos, Journal of Computational Physics **375**, 1339–1364 (2018).
[12] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman, arXiv preprint (2020), arXiv:2001.04385.
[13] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, Acta Mechanica Sinica **37**, 1727–1738 (2021).
[14] J. Wang, Z. Chen, D. Luo, Z. Zhao, V. M. Hur, and B. K. Clark, arXiv preprint (2021), arXiv:2108.02200.
[15] S. Wang, Y. Teng, and P. Perdikaris, SIAM Journal on Scientific Computing **43**, A3055–A3081 (2021).
[16] A. Krishnapriyan, A. Gholami, S. Zhe, R. Kirby, and M. W. Mahoney, in *Advances in Neural Information Processing Systems*, Vol. 34, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan (Curran Associates, Inc., 2021) pp. 26548–26560.
[17] S. Wang, X. Yu, and P. Perdikaris, Journal of Computational Physics **449**, 110768 (2022).
[18] S. Wang, H. Wang, and P. Perdikaris, Computer Methods in Applied Mechanics and Engineering **384**, 113938 (2021).
[19] P. Xie and W. E, Phys. Rev. B **103**, 024304 (2021).
[20] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Science **353**, 794 (2016).
[21] P. Reimann, Nature communications **7**, 10821 (2016).
[22] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Rev. Mod. Phys. **91**, 021001 (2019).
[23] T. Saha, P. Ghosal, P. Bej, A. Banerjee, and P. Deb, Physics Letters A **509**, 129501 (2024).
[24] I. A. Maceira and A. M. Läuchli, arXiv preprint (2024), arXiv:2409.18863.
[25] T. I. Andersen, N. Astrakhantsev, A. H. Karamlou, J. Berndtsson, J. Motruk, A. Szasz, J. A. Gross, A. Schuckert, T. Westerhout, Y. Zhang, *et al.*, arXiv preprint (2024), arXiv:2405.17385.
[26] R. Nandkishore and D. A. Huse, Annu. Rev. Condens. Matter Phys. **6**, 15 (2015).
[27] J. Smith, A. Lee, P. Richerme, B. Neyenhuis, P. W. Hess, P. Hauke, M. Heyl, D. A. Huse, and C. Monroe, Nature Physics **12**, 907 (2016).
[28] J.-y. Choi, S. Hild, J. Zeiher, P. Schauß, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, Science **352**, 1547 (2016).
[29] P. Sierant, M. Lewenstein, A. Scardicchio, L. Vidmar, and J. Zakrzewski, Reports on Progress in Physics 10.1088/1361-6633/ad9756 (2024).
[30] Á. S. Sanz, S. Miret-Artés, Á. S. Sanz, and S. Miret-Artés, A Trajectory Description of Quantum Processes. II. Applications: A Bohmian Perspective , 271 (2014).
[31] O. A. Castro-Alvaredo, B. Doyon, and T. Yoshimura, Physical Review X **6**, 041065 (2016).
[32] T. Banks and A. Lucas, Physical Review E **99**, 022105 (2019).
[33] While we only discuss the time-independent case, our derivation would be equivalent in the case of time-dependent Hamiltonians.
[34] By working with normalized states, such as autoregressive NQS, we could forego this requirement. However, imposing those invariances comes at a negligible computational cost.
[35] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, USA, 2007).

[36] D. P. Kingma and J. Ba, arXiv preprint (2014), arXiv:1412.6980.
[37] L. C. Evans, *Partial differential equations*, Vol. 19 (American Mathematical Society, 2022).
[38] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Physical Review Research **2**, 023358 (2020).
[39] C. Roth, arXiv preprint 2003.06228 (2020).
[40] M. Hibat-Allah, E. M. Inack, R. Wiersema, R. G. Melko, and J. Carrasquilla, Nature Machine Intelligence **3**, 952 (2021).
[41] M. Hibat-Allah, R. G. Melko, and J. Carrasquilla, arXiv preprint (2022), arXiv:2207.14314.
[42] D. Wu, R. Rossi, F. Vicentini, and G. Carleo, Phys. Rev. Res. **5**, L032001 (2023).
[43] E. Ibarra-García-Padilla, H. Lange, R. G. Melko, R. T. Scalettar, J. Carrasquilla, A. Bohrdt, and E. Khatami, arXiv preprint (2024), arXiv:2411.07144.
[44] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, Physical Review B **98**, 104426 (2018).
[45] C. Roth and A. H. MacDonald, arXiv preprint (2021), arXiv:2104.05085.
[46] C. Roth, A. Szabó, and A. H. MacDonald, Phys. Rev. B **108**, 054410 (2023).
[47] C. Fu, X. Zhang, H. Zhang, H. Ling, S. Xu, and S. Ji, in *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)* (SIAM, 2024) pp. 490–498.
[48] L. L. Viteritti, R. Rende, A. Parola, S. Goldt, and F. Becca, arXiv preprint (2023), arXiv:2311.16889.
[49] Y.-H. Zhang and M. Di Ventra, Physical Review B **107**, 075147 (2023).
[50] K. Sprague and S. Czischek, Communications Physics **7**, 90 (2024).
[51] H. Lange, G. Bornet, G. Emperauger, C. Chen, T. Lahaye, S. Kienle, A. Browaeys, and A. Bohrdt, arXiv preprint arXiv:2406.00091 (2024).
[52] R. Rende, F. Gerace, A. Laio, and S. Goldt, Physical Review Research **6**, 023057 (2024).
[53] R. Rende, L. L. Viteritti, L. Bardone, F. Becca, and S. Goldt, Communications Physics **7**, 260 (2024).
[54] G. B. Mbeng, A. Russomanno, and G. E. Santoro, SciPost Physics Lecture Notes , 082 (2024).
[55] H. W. Blöte and Y. Deng, Physical Review E **66**, 066110 (2002).
[56] M. Schmitt and M. Heyl, Phys. Rev. Lett. **125**, 100503 (2020).
[57] M. Srednicki, Physical review e **50**, 888 (1994).
[58] B. Blaß and H. Rieger, Scientific Reports **6**, 38185 (2016), number: 1 Publisher: Nature Publishing Group.
[59] S. Todo and K. Kato, Phys. Rev. Lett. **87**, 047203 (2001).
[60] A. F. Albuquerque *et al.*, Journal of Magnetism and Magnetic Materials **310**, 1187 (2007), wOS:000247618700217.
[61] B. Bauer *et al.*, Journal of Statistical Mechanics: Theory and Experiment **2011**, P05001 (2011).
[62] G. Carleo, F. Becca, M. Schiró, and M. Fabrizio, Scientific reports **2**, 243 (2012).
[63] G. Carleo *et al.*, SoftwareX , 100311 (2019).
[64] F. Vicentini *et al.*, SciPost Phys. Codebases , 7 (2022).
[65] P. Weinberg and M. Bukov, SciPost Physics **2**, 003 (2017).
[66] P. Weinberg and M. Bukov, SciPost Physics **7**, 020 (2019).
[67] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Quantum **3**, 191 (2019).
[68] M. Motta, W. Kirby, I. Liepuoniute, K. J. Sung, J. Cohn, A. Mezzacapo, K. Klymko, N. Nguyen, N. Yoshioka, and J. E. Rice, Electronic Structure **6**, 013001 (2024).