Variational subspace methods and application to improving variational Monte Carlo dynamics

Adrien Kahn[1,2,3,4], Luca Gravina[5,6], and Filippo Vicentini[1,2,3,4]

[1]CPHT, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France
[2]Collège de France, Université PSL, 11 place Marcelin Berthelot, 75005 Paris, France
[3]Inria Paris-Saclay, Bâtiment Alan Turing, 1, rue Honoré d'Estienne d'Orves – 91120 Palaiseau
[4]LIX, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France
[5]Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[6]Center for Quantum Science and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

arXiv:2507.08930v2 [quant-ph] 16 Apr 2026

We present a formalism that allows for the direct manipulation and optimization of subspaces, circumventing the need to optimize individual states when using subspace methods. Using the determinant state mapping, we can naturally extend notions such as distance and energy to subspaces, as well as Monte Carlo estimators, recovering the excited states estimation method proposed by Pfau et al. As a practical application, we then introduce Bridge, a method that improves the performance of variational dynamics by extracting linear combinations of variational time-evolved states. We find that Bridge is both computationally inexpensive and capable of significantly mitigating the errors that arise from discretizing the dynamics, and can thus be systematically used as a post-processing tool for variational dynamics.

1 Introduction

The simulation of quantum many-body systems is a major challenge of modern computational physics [1]. Although brute-force diagonalization is incapable of tackling Hilbert spaces with more than a few particles, several families of techniques have been proposed over the years to provide approximate solutions. In particular, variational methods, that rely on the optimization of a parametrized wavefunction, have appeared as efficient tools in approximating solutions of quantum many-body problems, such as dynamics or ground state estimation. Families of variational states that have been studied in recent years include tensor networks [2, 3, 4], variational quantum circuits [5, 6, 7], and neural quantum states [8, 9]. While tensor networks are well suited for low-dimensional problems with limited entanglement, such as ground state estimation for 1D gapped Hamiltonians [10], neural network quantum states have recently appeared as a powerful alternative. They have been successful in approximating the ground state of various spin [11, 12, 13], bosonic [14] and fermionic [15, 16] systems.

In the context of dynamics, generally considered a more challenging problem for variational methods, neural quantum states have also proven useful. Several variational schemes have been proposed to simulate dynamics with neural quantum states, such as the time-dependent variational principle [17, 18, 19], the projected time-dependent variational principle [20, 21], or global-in-time variational principles [22, 23].

Whether it be exact diagonalization or variational Monte Carlo, many methods that operate on quantum states can be improved by subspace methods. One of the most prominent examples of subspace method is the Lanczos algorithm that improves upon the power method by considering linear combinations of the generated vectors [24, 25, 26, 27]. Lanczos or Lanczos-inspired methods have been successfully used in conjunction with variational Monte Carlo, to refine ground state estimations [28, 29, 30] and to simulate dynamics [31] among other applications [32]. Subspace methods in variational

Monte Carlo typically represent the relevant subspace as the span of a given basis of states, each having been obtained by an individual optimization. For constructing the Krylov subspace [28] or for excited states optimization [33] for example, the sequential nature of optimizations causes an accumulation of errors for later basis states. This problem stems from the lack of a formal language to operate directly on subspaces, instead of individual states.

In this paper, we formalise a framework where the fundamental object is the subspace itself, thus providing for a more natural language to derive variational Monte Carlo algorithms operating directly on subspaces. Inspired by a recently proposed method to optimize for several excited states at once [34], we present the *determinant state mapping*, that allows the direct manipulation of subspaces by encoding them into a single quantum state of a larger Hilbert space. We find that this framework not only shines a new light on some previously known results, but also establishes new methods as natural generalization of single quantum state methods.

We then leverage this framework to introduce *Bridge*, a post-processing method for variational dynamics. Using previously generated variational states approximating quantum dynamics, we improve their performance by taking linear combinations of these states [23]. This method is inexpensive compared to the cost of generating the original states, such that it can be used systematically at a negligible cost. We show that in certain circumstances, Bridge can improve the performance of the original variational states by several orders of magnitude. Crucially, we find that this method relies on a specific estimator derived from determinant state framework in order to obtain consistent performances as the number of states increases.

The manuscript is structured as follows. In section 2 we introduce the formalism of the determinant state and several of its applications. In section 3, we present Bridge, and we demonstrate its performance on the square lattice transverse-field Ising model in section 4. Section 2 and the later sections are relatively independent.

2 Subspace methods with the determinant state

In quantum physics, a state is represented by a vector of the Hilbert space $\mathcal{H}$, up to its norm and global phase, that are irrelevant. In other words, quantum states are effectively one-dimensional subspaces of $\mathcal{H}$. In this section, we examine general m-dimensional subspaces of $\mathcal{H}$ and generalise the formalism of pure state quantum mechanics to them. This framework goes beyond purely mathematical considerations, and extends to numerical methods. We will show that it recontextualizes the excited states variational Monte Carlo method proposed by Pfau et al. [34] as the natural generalization of variational Monte Carlo to m-dimensional subspaces, but also allows for the formulation of new methods, as we will illustrate by proposing a method to construct a basis for dynamics.

2.1 Encoding subspaces as determinant states

A natural place to start in order to represent a subspace V is with a basis $\{|\phi_k\rangle\}_k$ of V, such that $V = \text{span}\left(\{|\phi_k\rangle\}_k\right)$. The exponential cost of representing each basis vector can be managed by means of variational encodings (tensor networks, neural networks...) of each basis state $|\phi_k\rangle$.

Regardless of how the basis vectors are encoded, this representation of a subspace relies on an arbitrary choice of basis, such that the mapping

$$V \mapsto \{|\phi_k\rangle\}_k \tag{1}$$

is not well-defined. For example, any permutation of the basis states $|\phi_k\rangle$ encodes the same subspace. More generally, given any complex $m \times m$ invertible matrix B that represents a change of basis, the family $\{|\varphi_k\rangle\}_k$ defined by

$$|\varphi_p\rangle = \sum_k B_{kp} |\phi_k\rangle \tag{2}$$

would parametrize the same subspace, since $\text{span}\left(\{|\phi_k\rangle\}_k\right) = \text{span}\left(\{|\varphi_k\rangle\}_k\right)$. In order to have a well-defined and bijective encoding, we should instead represent a subspace by the set of all its bases. Formally, we define on $\mathcal{H}^m$, the set of families of m vectors of $\mathcal{H}$, the equivalence relation $\sim$ such that two families are equivalent if and only if they span the same subspace. Then, the following map is a well-defined bijective encoding of

a subspace:

$$\begin{cases} \mathrm{Gr}_m(\mathcal{H}) & \to \mathcal{H}^m/\sim \\ V & \mapsto [\{|\phi_k\rangle\}_k] \end{cases}, \tag{3}$$

where $\mathrm{Gr}_m(\mathcal{H})$ is the set of subspaces of $\mathcal{H}$ of dimension m, known as the Grassmannian, and $[\{|\phi_k\rangle\}_k]$ is the equivalence class of $\{|\phi_k\rangle\}_k$.

In practice, however, we can only manipulate a representative of a class, and not the class itself. This implies that that any expression involving a subspace should be independent from the choice of representative.

However, working with representatives in a representative-independent way is well-developed in the context of quantum states. A quantum state is an element of the projective space $\mathrm{Gr}_1(\mathcal{H})$, that is, it is equivalence classes under multiplication by a scalar. We will now show that it is possible to transform equivalence classes of $\mathcal{H}^m$ under change of basis into equivalence classes of $\mathcal{H}^{\otimes m}$ under multiplication by a scalar. This will effectively allow us to map a subspace onto a quantum state.

Inspired by Pfau et al. [34], we propose to encode the subspace $V = \mathrm{span}\left(\{|\phi_k\rangle\}_k\right)$ with the quantum state $|\boldsymbol{\phi_A}\rangle$ given by the antisymmetrization of a given basis,

$$|\boldsymbol{\phi_A}\rangle = \boldsymbol{S}_- |\phi_1\rangle \otimes \ldots \otimes |\phi_m\rangle , \tag{4}$$

where $\boldsymbol{S}_-$ is the antisymmetrization operator (see definition A.3). The central motivation for this encoding is that given two different bases $\{|\phi_k\rangle\}_k$ and $\{|\varphi_k\rangle\}_k$ of the same subspace, related by a change of basis matrix B (see eq. (2)), the antisymmetrization of the two bases will differ only by a complex scalar factor $\det B$, that is,

$$|\boldsymbol{\varphi_A}\rangle = \det B \, |\boldsymbol{\phi_A}\rangle . \tag{5}$$

The proof is reported in proposition A.4. The two corresponding quantum states of $\mathcal{H}^{\otimes m}$ are therefore identical.

For example, it is easy to see that if the basis states $|\phi_k\rangle$ are permuted, their antisymmetrization will be modified by a factor ± 1. The equivalence classes of $\mathcal{H}^m$ under change of basis have thus been turned into equivalence classes of $\mathcal{H}^{\otimes m}$ under multiplication by a scalar, or in other words, into quantum states of $\mathcal{H}^{\otimes m}$.

More specifically, this mapping encodes a subspace by a Slater determinant, such that we call the quantum state $|\boldsymbol{\phi_A}\rangle$ the *determinant state* associated to the subspace $V = \mathrm{span}\left(\{|\phi_k\rangle\}_k\right)$. We emphasize that, when thought of as an encoding of a subspace, both the global phase and the norm of this vector are irrelevant.

Formally, we have established the mapping

$$\begin{cases} \mathrm{Gr}_m(\mathcal{H}) & \to S/(\mathbb{C}, \times) \\ V & \mapsto [\boldsymbol{S}_- |\phi_1\rangle \otimes \ldots \otimes |\phi_m\rangle] \end{cases} \tag{6}$$

where S is the set of Slater determinant vectors, and $S/(\mathbb{C}, \times)$ is thus the set of Slater determinants understood as quantum states, that is, defined up to a complex scalar multiple. We have established that this mapping is well-defined, it is surjective, and we show in the appendix that it is injective (proposition A.5). This mapping that we introduced in the context of encoding subspaces of a Hilbert space is in fact known in mathematics as the *Plücker embedding* [35].

To conclude, we have presented a one-to-one mapping of a subspace of dimension m to a Slater determinant of $\mathcal{H}^{\otimes m}$ constructed from any basis of that subspace. In practice we will still encode the states $|\phi_k\rangle$ themselves, but the representation as a determinant state will serve to ensure that any expression we use is representative-independant. Furthermore, the fact that we can identify a subspace with a quantum state suggests that we can use expressions that are known for quantum states, and apply them on determinant states as guidance to generalize them on subspaces.

2.2 Distance between subspaces from distance between quantum states

Let us now see how the determinant state perspective is helpful for computing distances between subspaces. The space of quantum states, that is, $\mathcal{H}/(\mathbb{C}, \times)$, is endowed with the Fubini-Study metric

$$d\left(|\psi\rangle, |\phi\rangle\right) = \arccos\left(\frac{|\langle\psi|\phi\rangle|}{\|\psi\|\|\phi\|}\right). \tag{7}$$

We emphasize that the Fubini-Study metric is independent of the choice of representative of the quantum states, that is,

$$d\left(\alpha |\psi\rangle, \beta |\phi\rangle\right) = d\left(|\psi\rangle, |\phi\rangle\right), \tag{8}$$

for any $\alpha, \beta \in \mathbb{C}$.

It is then natural to define a Fubini-Study metric between subspaces through their respective determinant states. For any two m-dimensional subspaces U and V of $\mathcal{H}$, with respective arbitrary bases $\{|\psi_k\rangle\}_k$ and $\{|\phi_k\rangle\}_k$, we define

$$d_s(U,V) = d\left(|\boldsymbol{\psi_A}\rangle, |\boldsymbol{\phi_A}\rangle\right). \tag{9}$$

First, we note that thanks to the properties of the determinant state mapping, this function is well-defined, in the sense that it does not depend on the chosen bases. Indeed, if $\{|\tilde{\psi}_k\rangle\}_k$ and $\{|\tilde{\phi}_k\rangle\}_k$ are different bases of U and V respectively, then $d(|\boldsymbol{\psi_A}\rangle, |\boldsymbol{\phi_A}\rangle) = d(|\tilde{\boldsymbol{\psi}}_{\boldsymbol{A}}\rangle, |\tilde{\boldsymbol{\phi}}_{\boldsymbol{A}}\rangle)$, as a consequence of eq. (5) and eq. (8).

Since the Fubini-Study metric is a metric and the determinant state mapping (eq. (6)) is injective, it implies that d_s is itself a metric over $\mathrm{Gr}_m(\mathcal{H})$.[1]

We can examine in more detail the behaviour of this metric. In accordance with the definition of a metric, $d_s(U,V) = 0$ if and only if $U = V$. Conversely, the maximum distance of $\pi/2$ is attained if and only if V contains a non-zero vector that is orthogonal to U (see proposition A.6). More generally, we can relate this distance to the volume of the m-dimensional parallelotope spanned by a orthonormal basis of V orthogonally projected on U (see lemma A.1).

Having seen that the Fubini-Study metric between determinant states defines a metric between subspaces, we can reuse all the tools developed for estimating and optimizing the Fubini-Study metric between states and apply them to subspaces [20, 21, 36].

2.3 Operators acting on subspaces through determinant states

In quantum physics, the two fundamental kinds of operators are unitary and Hermitian. In this section, we examine their natural generalization to operators acting on subspaces.

We consider an operator $\hat{A}$ acting on $\mathcal{H}$ and a subspace V with basis $\{|\phi_k\rangle\}_k$, encoded through its determinant state $|\boldsymbol{\phi_A}\rangle$. The objective of this section is to define an operator $\hat{\boldsymbol{A}}$ acting on $\mathcal{H}^{\otimes m}$ such that $\hat{\boldsymbol{A}}|\boldsymbol{\phi_A}\rangle$ defines the action of $\hat{A}$ on the subspace V in a meaningful way. In this section we will be using the $\hat{\cdot}$ symbol to disambiguate operators $\hat{U}$ from subspaces V.

[1] We remark that, in general, any choice of metric over the space of determinant states would induce a metric over subspaces.

Unitary operators The natural way to define the action of a unitary operator $\hat{U}$ on a subspace V is

$$\hat{U}V = \left\{\hat{U}|\phi\rangle, |\phi\rangle \in V\right\}. \tag{10}$$

The set $\hat{U}V$ is still a subspace, and if $\{|\phi_k\rangle\}_k$ is a basis of V, then $\left\{|\tilde{\phi}_k\rangle\right\}_k = \left\{\hat{U}|\phi_k\rangle\right\}_k$ is a basis of $\hat{U}V$. The determinant state of $\hat{U}V$ can thus be expressed as

$$|\tilde{\boldsymbol{\phi}}_{\boldsymbol{A}}\rangle = \boldsymbol{S}_- \bigotimes_k \hat{U}|\phi_k\rangle = \hat{U}^{\otimes m}|\boldsymbol{\phi_A}\rangle, \tag{11}$$

since $\hat{U}^{\otimes k}$ commutes with $\boldsymbol{S}_-$. We deduce that the natural way to generalize a unitary operator $\hat{U}$ as an operator acting on determinant states is to define $\hat{\boldsymbol{U}} = \hat{U}^{\otimes k}$. We remark that $\hat{\boldsymbol{U}}$ is indeed a unitary operator of $\mathcal{H}^{\otimes m}$.

Hermitian operators Using the previously established generalization of unitaries, we will see that while unitaries are generalized by taking a tensor product, Hermitian operators are generalized by taking a sum. Consider the family of unitary operators generated by the hermitian operator $\hat{H}$

$$\hat{U}_a = e^{-ia\hat{H}},\ a \in \mathbb{R}. \tag{12}$$

In the previous paragraph, we have defined the generalization of $\hat{U}_a$ to the space of determinant states to be $\hat{\boldsymbol{U}}_{\boldsymbol{a}} = \hat{U}_a^{\otimes m}$. Substituting eq. (12) in the definition of the generalized unitary $\hat{\boldsymbol{U}}_{\boldsymbol{a}}$, we obtain,

$$\hat{\boldsymbol{U}}_{\boldsymbol{a}} = e^{-ia\sum_k \hat{\boldsymbol{H}}_{\boldsymbol{k}}}, \tag{13}$$

where $\hat{\boldsymbol{H}}_{\boldsymbol{k}}$ is the operator $\hat{H}$ acting on the k-th copy of the Hilbert space

$$\hat{\boldsymbol{H}}_{\boldsymbol{k}} = \mathbb{1} \otimes \ldots \otimes \underset{\substack{\uparrow \\ k\text{-th position}}}{\hat{H}} \otimes \ldots \otimes \mathbb{1}. \tag{14}$$

The unitary $\hat{\boldsymbol{U}}_{\boldsymbol{a}}$ is therefore generated by the Hermitian operator $\sum_k \hat{\boldsymbol{H}}_{\boldsymbol{k}}$, so we deduce that the natural generalization of $\hat{H}$ as an operator that acts on subspaces through the determinant state is $\hat{\boldsymbol{H}} = \sum_k \hat{\boldsymbol{H}}_{\boldsymbol{k}}$.

One might question why we defined the action of a unitary operator on a subspace as $\hat{U}V$, but not the action of a Hermitian operator as $\hat{H}V$.

There are three reasons. (i) First, unitary operators are operators that are fundamentally meant to be applied on states. The action of a unitary operator on a state defines a transformed state. On the other hand, the action of a Hermitian operator on a state does not fundamentally have a meaning. What is fundamental in the definition of Hermitian operators is their average values, which defines the expected value of a measurement. (ii) The second justification is that unitary operators and Hermitian operators are related by the former being generated by the latter. While our construction explicitly enforced this relationship for the generalized operators, defining the action of $\hat{H}$ on V as $\hat{H}V$ would not satisfy it. In particular, the time evolution operator would no longer be generated by the Hamiltonian. (iii) Lastly, following the same procedure as for unitary operators would have led to the operator $H^{\otimes m}$. While this operator is also Hermitian, dimensional analysis indicates that it is not a proper generalization of $\hat{H}$ since it does not have the dimension of an energy.

2.4 Energy of a subspace and the spectrum of the Rayleigh matrix

In the previous section we examined the generalization of unitary and Hermitian operators for determinant states. Since we have generalized the notion of a Hamiltonian to subspaces, it is now natural to examine the resulting notion of energy.

Energy of the determinant state Following the generalization of Hermitian operators in section 2.3, the energy of a subspace V spanned by a basis $\{|\phi_k\rangle\}_k$ for a Hamiltonian H is given by

$$\frac{\langle \boldsymbol{\phi_A}|\boldsymbol{H}|\boldsymbol{\phi_A}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle}. \tag{15}$$

Since $\boldsymbol{H}$ commutes with the antisymmetrizer $\boldsymbol{S}_-$, which is the orthogonal projector on the antisymmetric subspace $V_-(\mathcal{H}^{\otimes m})$ (see definition A.2), then $V_-(\mathcal{H}^{\otimes m})$ is an invariant subspace under $\boldsymbol{H}$. Therefore, the restriction of $\boldsymbol{H}$ to the subspace of $\mathcal{H}^{\otimes m}$ spanned by the determinant states is well-defined. If we denote by E_k the ordered eigenvalues of H and by $|E_k\rangle$ their associated eigenstate, then the minimum energy of $\boldsymbol{H}$ restricted to $V_-(\mathcal{H}^{\otimes m})$ is

$$\boldsymbol{E_0} = \sum_{k=0}^{m-1} E_k. \tag{16}$$

The associated eigenstate is the determinant state corresponding to the subspace span $\left(\{|E_k\rangle\}_{0\leq k\leq m-1}\right)$.

Based on this observation, Pfau et al. [34] use eq. (15) as a cost function to optimize for the m lowest excited states. Pfau et al. further show that this energy of a subspace has a simple expression that does not involve determinants,

$$\frac{\langle \boldsymbol{\phi_A}|\boldsymbol{H}|\boldsymbol{\phi_A}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} = \operatorname{tr}\left(G^{-1}G^{(H)}\right), \tag{17}$$

where G is the Gram matrix of the family $\{|\phi_k\rangle\}_k$ and $G^{(H)}$ is the projected Hamiltonian on this family,

$$G_{ij} = \langle \phi_i|\phi_j\rangle, \tag{18}$$

$$G^{(H)}_{ij} = \langle \phi_i|H|\phi_j\rangle. \tag{19}$$

We provide a detailed proof of eq. (17) in proposition A.7.

The appearance of the matrix $G^{-1}G^{(H)}$ within the determinant state energy shows that this framework is in fact a reformulation of subspace methods for eigenvalue problems. In order to better examine this connection, let us first recall the theory of subspace methods for eigenvalue problems, in particular in the context of non-orthonormal bases.

Subspace methods and the Rayleigh matrix In the Lanczos method [24, 26], extremal eigenvalues of a Hermitian matrix A are estimated by orthonormalizing the family of m Krylov vectors $\left\{A^k v\right\}_k$ into the family $\{u_k\}_k$, and taking the eigenvalues of the projected $m\times m$ tridiagonal matrix with coefficients $T_{ij} = u_i^\dagger A u_j$.

The same procedure would also work with different families of vectors ϕ_k, not necessarily orthonormal, as long as they are suitably chosen to have good overlap with the extremal eigenvectors [37, 38]. Given a family of well-chosen linearly independent states $|\phi_k\rangle$, the eigenvalues μ_k of $G^{-1}G^{(H)}$ provide an approximation of the low-lying spectrum of the Hamiltonian H, where

G and $G^{(H)}$ are defined by eqs. (18) and (19).[2]

The eigenvalues μ_k of $G^{-1}G^{(H)}$ only depend on the subspace span $(\{|\phi_k\rangle\}_k)$, and not on the specific choice of basis. These eigenvalues are real.[3]

In particular, these approximations satisfy a generalized variational principle: if we order the eigenvalues μ_k of $G^{-1}G^{(H)}$, then they are upper bounds to the corresponding ordered eigenvalues E_k of the Hamiltonian H,

$$\forall k,\ E_k \leq \mu_k. \tag{20}$$

Furthermore, the eigenvectors $\alpha^{(k)}$ of $G^{-1}G^{(H)}$ define states of the full Hilbert space

$$|\psi^{(k)}\rangle = \sum_p \alpha_p^{(k)} |\phi_p\rangle\,, \tag{21}$$

such that $|\psi^{(k)}\rangle$ has energy μ_k. The proofs of these statements are reported in section B. In section B, we discuss more generally why the eigenvalues μ_k of $G^{-1}G^{(H)}$ can be seen as the natural generalization of the eigenvalues of H in the subspace $V = \text{span}(\{|\phi_{\theta_k}\rangle\}_k)$.

We finally remark that eq. (20) is a generalization of the ground state variational principle

$$\forall\, |\psi\rangle \in \mathcal{H},\ E_0 \leq \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}. \tag{22}$$

In particular, we see that $G^{-1}G^{(H)}$ appears as a generalization of the Rayleigh quotient $\langle\psi|\psi\rangle^{-1}\langle\psi|H|\psi\rangle$. As such, we call $G^{-1}G^{(H)}$ the *Rayleigh matrix*.

From this perspective, a natural variational principle thus consists in minimizing the sum of the eigenvalues μ_k of the Rayleigh matrix, that is, $\text{tr}(G^{-1}G^{(H)})$. This variational principle is known as the Ky Fan variational principle [39].[4]

[2]The matrix $G^{-1}G^{(H)}$ is analogous to the matrix T from the Lanczos method, with the difference that the inverse of the Gram matrix G accounts for the fact that the family $\{|\phi_k\rangle\}_k$ is not necessarily orthonormal.

[3]The matrix $G^{-1}G^{(H)}$ is not Hermitian in general, so it is not obvious that its eigenvalues μ_k should be real. However, this matrix is similar to $G^{-\frac{1}{2}}G^{(H)}G^{-\frac{1}{2}}$ such that they share the same eigenvalues μ_k. And since $G^{-\frac{1}{2}}G^{(H)}G^{-\frac{1}{2}}$ is manifestly Hermitian, we must have $\mu_k \in \mathbb{R}$.

[4]Since the trace of the Rayleigh matrix is independent of the chosen basis, we can choose an orthonormal basis $\{|\psi_k\rangle\}_k$, such that $\text{tr}\left(G^{-1}G^{(H)}\right) = \sum_k \langle\psi_k|H|\psi_k\rangle$, which is the usual form of the Ky Fan variational principle.

Determinant state and Rayleigh matrix We now understand eq. (17) as expressing the equivalence between the determinant state variational principle, and the Ky Fan variational principle.

Furthermore, this connection provides a better understanding of the Rayleigh matrix from the point of view of the determinant state, as the object whose eigendecomposition defines the natural generalization of the spectrum of H restricted to that subspace. More practically, it provides bounds that generalize the ground state variational principle for each approximate energy μ_k found in the subspace (eq. (20)).

In the next section, we explore practical applications of the formalism that we have developed so far in this section.

2.5 Applications of the determinant state to variational Monte Carlo methods

The determinant state is also a convenient tool for practical applications. In a variational context, a subspace V can be encoded by a basis of m variational states $|\phi_{\theta_k}\rangle$ through the determinant state mapping, and we will see in this section that these variational subspace states can also be manipulated efficiently.

Sampling The amplitudes of the determinant state can be easily evaluated. For an element $\boldsymbol{s} = (s_1, s_2, \ldots, s_m)$ of the computational basis of $\mathcal{H}^{\otimes m}$, such that s_k is an element of the computational basis of $\mathcal{H}$, the corresponding amplitude of $|\boldsymbol{\phi_A}\rangle$ can be expressed as

$$\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle = \frac{1}{m!}\det\Phi(\boldsymbol{s}), \tag{23}$$

where the $m \times m$ matrix $\Phi(\boldsymbol{s})$ is defined by $[\Phi(\boldsymbol{s})]_{ij} = \langle s_i|\phi_j\rangle$ (see proof of proposition A.3). As such, even though $|\boldsymbol{\phi_A}\rangle$ is a linear combination of $m!$ states, its amplitudes can be queried efficiently. This also means that the probability distribution $|\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle|^2$ can be approximately sampled from using Markov chain Monte Carlo.[5]

[5]The determinant structure of the amplitudes allows for the use of rank-1 update formulas to update the determinant at a cost of $O(m^2)$ instead of $O(m^3)$ [40]. While this method can increase the speed of sampling, it may lead to an accumulation of numerical errors. Since in our case the cost of evaluating determinant state amplitudes is dominated by the evaluation of $\Phi(\boldsymbol{s})$ and $\Phi^{(H)}(\boldsymbol{s})$ rather

Fubini-Study metric The subspace distance discussed in section 2.2 can be estimated using the decomposition of the Fubini-Study distance between states

$$d_s(U,V) = \arccos\left(\mathbb{E}_{\boldsymbol{s}\sim|\boldsymbol{\psi_A}(\boldsymbol{s})|^2}\left[\frac{\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle}{\langle \boldsymbol{s}|\boldsymbol{\psi_A}\rangle}\right] \mathbb{E}_{\boldsymbol{s}\sim|\boldsymbol{\phi_A}(\boldsymbol{s})|^2}\left[\frac{\langle \boldsymbol{s}|\boldsymbol{\psi_A}\rangle}{\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle}\right]\right), \quad (24)$$

which leads to an efficient estimator with the addition of control variates, as discussed in Ref. [21].

The samples $\boldsymbol{s} = (s_1, \ldots, s_m)$ are configurations of the Hilbert space $\mathcal{H}^{\otimes m}$. They are sampled according to the probability distribution of $|\boldsymbol{\psi_A}\rangle$ and $|\boldsymbol{\phi_A}\rangle$ respectively, namely, $|\det \Psi(\boldsymbol{s})|^2$ and $|\det \Phi(\boldsymbol{s})|^2$.

One possible application of the Fubini-Study distance between subspaces is the search for approximately invariant subspaces, which could be used to simulate dynamics.

Approximately invariant subspace Let $|\psi_0\rangle$ be the initial state for dynamics. If we can identify an m-dimensional subspace V that contains $|\psi_0\rangle$ and that is invariant under the Hamiltonian H, then the dynamics of $|\psi_0\rangle$ under H can be restricted to V without approximation. In general, however, for reasonably small m, we do not expect the existence of an invariant subspace containing $|\psi_0\rangle$.

Alternatively, we can search for an *approximately invariant subspace*, that is, a subspace V which contains $|\psi_0\rangle$ and such that $HV = \{H|\phi\rangle, |\phi\rangle \in V\}$ is close enough to V according to some distance, such as the one defined by eq. (9).[6] To solve this problem, we can variationally encode a subspace $V_\theta = \{|\psi_0\rangle, |\psi_{\theta_1}\rangle, \ldots, |\psi_{\theta_{m-1}}\rangle\}$ and solve the optimization problem

$$\theta^* = \arg\min_\theta \; d_s(V_\theta, HV_\theta). \quad (25)$$

than the computation of the determinant, our numerical results in section 4 were obtained without the rank-1 update method.

[6] As we discussed in section 2.3, applying a Hermitian operator directly on states is not a natural operation. Instead of requiring V_θ to be close to the subspace HV_θ, we could instead consider the subspace $e^{-iH\delta}V_\theta$ for some $\delta \in \mathbb{R}$. While both choices would lead to the optimization problem minimizing at an invariant subspace in the case that it exists, it is not yet clear what the minimum would be in either cases if an invariant subspace does not exist.

The loss function of this optimization problem can be estimated through eq. (24), and estimation of its gradient has also been studied [21].

Once an approximately invariant subspace V^* has been obtained, we simply need to perform the dynamics in that subspace, which is precisely the object of section 3.

Energy estimation It is also straightforward to estimate the energy of a subspace (eq. (15)) using the standard decomposition

$$\frac{\langle \boldsymbol{\phi_A}|\boldsymbol{H}|\boldsymbol{\phi_A}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} = \mathbb{E}_{\boldsymbol{s}\sim|\boldsymbol{\phi_A}(\boldsymbol{s})|^2}\left[\frac{\langle \boldsymbol{s}|\boldsymbol{H}|\boldsymbol{\phi_A}\rangle}{\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle}\right]. \quad (26)$$

As shown in section 2.4, the minimization of this scalar quantity provides the subspace spanned by the m eigenstates of H with lowest energy, but extracting the individual energies and the corresponding states requires computing the entire Rayleigh matrix $G^{-1}G^{(H)}$.

In section 2.4 we argued that the Rayleigh quotient of a single state $\langle\psi|H|\psi\rangle/\langle\psi|\psi\rangle$ can be generalised to the Rayleigh matrix of a subspace $G^{-1}G^{(H)}$. Similarly, the Monte Carlo estimator for the Rayleigh quotient

$$\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} = \mathbb{E}_{s\sim|\psi(s)|^2}\left[\frac{\langle s|H|\psi\rangle}{\langle s|\psi\rangle}\right], \quad (27)$$

is generalised to the following estimator for the Rayleigh matrix (due to Pfau et al. in Ref. [34], see proof in proposition A.8)

$$G^{-1}G^{(H)} = \mathbb{E}_{\boldsymbol{s}\sim|\boldsymbol{\phi_A}(\boldsymbol{s})|^2}\left[\Phi(\boldsymbol{s})^{-1}\Phi^{(H)}(\boldsymbol{s})\right], \quad (28)$$

where the matrices $\Phi(\boldsymbol{s})$ and $\Phi^{(H)}(\boldsymbol{s})$ are defined by

$$[\Phi(\boldsymbol{s})]_{ij} = \langle s_i|\phi_j\rangle, \quad (29)$$

$$\left[\Phi^{(H)}(\boldsymbol{s})\right]_{ij} = \langle s_i|H|\phi_j\rangle. \quad (30)$$

The estimator of the Rayleigh matrix also generalizes the desirable zero-variance property of the estimator of the energy: if the subspace is an invariant subspace under H (and thus contains m proper eigenvectors of H), then the local energy matrix $\Phi(\boldsymbol{s})^{-1}\Phi^{(H)}(\boldsymbol{s})$ does not depend on $\boldsymbol{s}$ (proposition A.9).

Extracting states from subspaces In the excited states estimation method discussed

above, estimating the Rayleigh matrix was required to extract the eigenvalue and eigenstate approximations from the optimized subspace. This process can be extended to any subspace, irrespective of how its basis states were obtained.

In the context of ground state optimization, the solution obtained by variational Monte Carlo can be improved by extracting the linear combination from a basis consisting of the optimization result and other states. Example of states that can be added to the basis include states reached earlier during the optimization, or Krylov states built from the final optimization state [28, 29, 30].

Another possible use of extracting optimal linear combinations out of a subspace from the Rayleigh matrix is the generalization of ground states optimized for given values of the Hamiltonian parameters [41, 42]. Consider a parametric Hamiltonian of the form

$$H(\gamma) = \sum_p \gamma_p H_p. \tag{31}$$

Suppose that we have obtained m states $|\phi_k\rangle = |\phi_{\gamma_k}\rangle$ that approximate the ground state at m different values of the Hamiltonian parameter γ. Exploiting the continuity of eigenstates within a phase, we can approximate the ground state at other values of γ as linear combinations of the states $|\phi_k\rangle$. The Rayleigh matrix at any parameter γ is given by

$$G^{-1}G^{(H(\gamma))} = \sum_p \gamma_p G^{-1}G^{(H_p)}. \tag{32}$$

Therefore, it can be estimated at any value γ with a single estimation of each Rayleigh matrices $G^{-1}G^{(H_p)}$, using either the determinant state estimator (eq. (28)) or the sum of states estimator (eqs. (39) and (40)). Then, for any value γ, we simply need to diagonalize the small $m \times m$ Rayleigh matrix $G^{-1}G^{(H(\gamma))}$ in order to obtain its lowest eigenvalue $\mu^{(\gamma)}$ and the corresponding eigenvector $\alpha^{(\gamma)}$. It defines a state of the full Hilbert space

$$|\psi^{(\gamma)}\rangle = \sum_k \alpha_k^{(\gamma)} |\phi_k\rangle \tag{33}$$

that approximates the ground state of $H(\gamma)$, and has energy $\mu^{(\gamma)}$.

We illustrate this method on a simple example in fig. 1. The Hamiltonian is the transverse-field Ising model defined in eq. (43) on an open chain with 12 sites. We choose the open chain in order to decrease the number of symmetries and ensure that the ground states do not belong to a subspace whose dimension is much smaller than that of the Hilbert space.

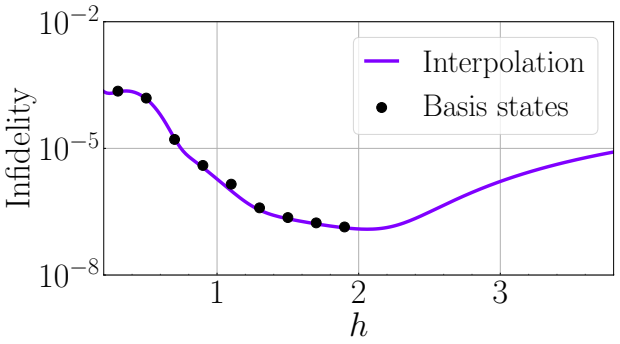

Figure 1: Infidelity between the exact ground state and estimations of the ground state for various values of the transverse-field Ising Hamiltonian parameter h. The parameter J is fixed to $J = 1$. The Hamiltonian is defined on an open chain of 12 sites. The black dots indicate nine ground state approximations obtained for various values of h. These variational states are restricted Boltzmann machines [8] with $\alpha = 4$. The purple line shows the precision of the ground state interpolation obtained by taking linear combination of the nine ground state approximations.

We see that the linear combination states successfully generalize to unseen values of γ, both for interpolation and, to some degree, extrapolation.

The average value of any operator A on these linear combination states can then be conveniently evaluated using the Rayleigh matrix of A. For a given value of γ, let p be the index of the lowest eigenvalue of $G^{-1}G^{(H(\gamma))}$, and let P_γ denote the matrix of eigenvectors of $G^{-1}G^{(H(\gamma))}$. Then, under the assumption that the eigenvalues of $G^{-1}G^{(H(\gamma))}$ are all distinct, the average value of A on the ground state approximation at γ is given by

$$\frac{\langle\psi^{(\gamma)}|A|\psi^{(\gamma)}\rangle}{\langle\psi^{(\gamma)}|\psi^{(\gamma)}\rangle} = \left[P_\gamma^{-1}G^{-1}G^{(A)}P_\gamma\right]_{pp}. \tag{34}$$

A proof is reported with proposition B.3.

3 Bridge: improving variational dynamics accuracy with subspace methods

Various methods have been proposed for the variational approximation of quantum dynamics, including time-dependent variational Monte Carlo

	Section	Single state	Subspace
Original object	2.1	1-d subspace with basis $\{\|\psi\rangle\}$	m-d subspace with basis $\{\|\phi_k\rangle\}_k$
Quantum state	2.1	$\|\psi\rangle$	$\|\boldsymbol{\phi_A}\rangle = \boldsymbol{S_-} \|\phi_1\rangle \otimes \ldots \otimes \|\phi_m\rangle$
Hilbert space	2.1	$\mathcal{H}$	$\mathcal{H}^{\otimes m}$
Distance	2.2	$d(\|\psi\rangle, \|\phi\rangle) = \arccos\left(\frac{\|\langle\psi\|\phi\rangle\|}{\|\psi\|\|\phi\|}\right)$	$d_s(U,V) = d(\|\boldsymbol{\psi_A}\rangle, \|\boldsymbol{\phi_A}\rangle)$
Unitary operator	2.3	U	$\boldsymbol{U} = U^{\otimes m}$
Hermitian operator	2.3	H	$\boldsymbol{H} = \sum_k \boldsymbol{H_k}$
Energy	2.4	$\frac{\langle\psi\|H\|\psi\rangle}{\langle\psi\|\psi\rangle}$	$\frac{\langle\boldsymbol{\phi_A}\|\boldsymbol{H}\|\boldsymbol{\phi_A}\rangle}{\langle\boldsymbol{\phi_A}\|\boldsymbol{\phi_A}\rangle} = \mathrm{tr}\left(G^{-1}G^{(H)}\right)$
Variational principle	2.4	$E_0 \leq \frac{\langle\psi\|H\|\psi\rangle}{\langle\psi\|\psi\rangle}$	$\forall k,\ E_k \leq \lambda_k\left(G^{-1}G^{(H)}\right)$
Energy estimator	2.5	$\frac{\langle\psi\|H\|\psi\rangle}{\langle\psi\|\psi\rangle} = \mathbb{E}_{s\sim\|\psi(s)\|^2}\left[\frac{\langle s\|H\|\psi\rangle}{\langle s\|\psi\rangle}\right]$	$G^{-1}G^{(H)} = \mathbb{E}_{\boldsymbol{s}\sim\|\boldsymbol{\phi_A}(\boldsymbol{s})\|^2}\left[\Phi(\boldsymbol{s})^{-1}\Phi^{(H)}(\boldsymbol{s})\right]$

Table 1: Summary of the generalisation of quantum states to subspaces, as explored in section 2.

[17, 18], projected-time variational Monte Carlo (p-tVMC) [20, 21], and global-in-time dynamics [22, 23]. In this section, we discuss how the subspace spanned by approximate Arnoldi states obtained by one of the above-mentioned methods can be used to extract linear combinations that improve the precision of the approximate dynamics. In the following, we will refer to this approach as *Bridge*, where the Arnoldi basis states are seen as pillars on which a bridge is built. We remark that the use of linear combinations of Arnoldi states for extrapolating dynamics has already been examined for exact dynamics [27]. A related strategy has recently been discussed in Ref. [23] for global-in-time variational dynamics. We comment on the relation to this work at the end of section 3.2.

We will see that Bridge is a computationally inexpensive method that can significantly mitigate the errors arising from the time discretization of the method that produced the approximate Arnoldi basis states.

Consider a set of m variational states $\{|\phi_{\theta_k}\rangle\}_{k\in\{0,\ldots,m-1\}}$ that approximate the exact dynamics at times $t = k\delta$ of a system evolving according to the Hamiltonian H with initial state $|\phi_{\theta_0}\rangle$. These states may have been produced by t-VMC, p-tVMC, tensor network methods or any other variational approach, such that their amplitudes can be queried efficiently. Since these states define a low-dimensional subspace, we can identify the time-dependent linear combination

$$|\psi_{\alpha(t)}\rangle = \sum_k \alpha_k(t)\,|\phi_{\theta_k}\rangle\,,\ \alpha(t) \in \mathbb{C}^m \tag{35}$$

which solves the Schrödinger equation within the variational space $\mathrm{span}(\{|\phi_{\theta_k}\rangle\}_k)$, and take it as a new and possibly more accurate approximation of the dynamics. Since at time $t = k\delta$ the linear variational state can represent $|\phi_{\theta_k}\rangle$, the optimal state in the subspace $\mathrm{span}(\{|\phi_{\theta_k}\rangle\}_k)$ has at least the accuracy of the original basis states, but it could achieve a better approximation of the dynamics. Although it is not guaranteed, we empirically observe that the linear combinations obtained by this procedure are always close to the optimal linear state, and thus never underperforms compared to the original states $|\phi_{\theta_k}\rangle$.

Importantly, this approach not only improves on previously computed solutions at times $t = k\delta$, but also allows interpolating between times $t = k\delta$ and extrapolating beyond $t = (m-1)\delta$.

In the next section, we will discuss how the parameter $\alpha(t)$ is determined through the time-dependent variational principle. We emphasize that during this optimization, the parameters θ_k remain fixed and only the coefficients α_k are varied. As such, we simplify notations by denoting $|\phi_k\rangle = |\phi_{\theta_k}\rangle$. We assume that the family $\{|\phi_k\rangle\}_k$ is linearly independent, which should always be the case in practice.[7]

3.1 Time-dependent variational principle for a linear state

To determine the time-dependent parameter vector $\alpha(t)$, we use the time-dependent variational principle (TDVP). The linear structure of $|\psi_\alpha\rangle$ greatly simplifies the TDVP, resulting in the equation

$$\dot{\alpha} = -iG^{-1}G^{(H)}\alpha. \tag{36}$$

The direct application of the TDVP commonly used with neural quantum states where the norm and global phase are not constrained would yield a different equation. However, as discussed in section C, it is in fact equivalent to eq. (36).

This equation is an effective Schrödinger equation, where α evolves under the action of an effective Hamiltonian given by the *Rayleigh matrix* $G^{-1}G^{(H)}$, previously discussed in section 2.4. The solution of this equation is given by

$$\alpha(t) = e^{-iG^{-1}G^{(H)}t}\alpha(0). \tag{37}$$

We recall from section 2.4 that the spectrum of the Rayleigh matrix is real. Consequently, the evolution of $\alpha(t)$ is similar to an evolution under a Hermitian Hamiltonian, where the eigencomponents rotate and the dynamics never converges.

Crucially, unlike the general case, the TDVP for a linear state only requires the evaluation of $G^{-1}G^{(H)}$ once, since it does not depend on α. From one evaluation of $G^{-1}G^{(H)}$, we can obtain the solution $\alpha(t)$ from eq. (37) at any times for a negligible cost.

This suggests that Bridge can be used both to improve the precision at times $t = k\delta$, and to interpolate between these times, yielding a truly continuous approximation of the dynamics.

The numerical implementation of Bridge thus consists in two steps. (i) Estimate the Rayleigh matrix. (ii) Obtain $\alpha(t)$ from eq. (37).

The first step requires an estimator for the Rayleigh matrix, which we discuss in the next section.

The second step is simple, since the Rayleigh matrix is a small $m \times m$ matrix that can be exponentiated exactly at a negligible cost. In practice, the basis states $|\phi_k\rangle$ can be close to linear dependence, which causes instability in the exact exponentiation of the Rayleigh matrix estimate. To circumvent this problem, we perform these operations in arbitrary precision.

Time-dependent Hamiltonian We note that while we focus on the case of a time-independent Hamiltonian, Bridge can also be used to tackle time-dependent problems. If the Hamiltonian has a time dependence of the form

$$H(t) = \sum_p \beta_p(t) H_p, \tag{38}$$

then it is enough to estimate the Rayleigh matrix of each term H_p once to obtain the Rayleigh matrix of the Hamiltonian at all times (see eq. (32)).

3.2 Estimation of the Rayleigh matrix

One way of estimating the Rayleigh matrix $G^{-1}G^{(H)}$ is the *determinant state estimator* discussed in the previous section (eq. (28)). Another way, proposed by Sinibaldi et al. in Ref. [23], relies on individually estimating G and $G^{(H)}$ up to a common constant that cancels out in the product $G^{-1}G^{(H)}$. This estimator is based on the decomposition

$$\frac{1}{\|p\|}G_{ij} = \sum_s p(s)\frac{\langle\phi_i|s\rangle\langle s|\phi_j\rangle}{\sum_k |\langle s|\phi_k\rangle|^2}, \tag{39}$$

$$\frac{1}{\|p\|}G^{(H)}_{ij} = \sum_s p(s)\frac{\langle\phi_i|s\rangle\langle s|H|\phi_j\rangle}{\sum_k |\langle s|\phi_k\rangle|^2}, \tag{40}$$

where samples are taken from the distribution

$$p(s) = \frac{1}{\|p\|}\sum_k |\langle s|\phi_k\rangle|^2, \tag{41}$$

and $\|p\| = \sum_s \sum_p |\langle s|\phi_p\rangle|^2$ is a normalization constant. Once $G/\|p\|$ has been obtained, it must

[7] If it is not the case, then the number of states in the family can be reduced until the family is linearly independent. Linear dependence can be immediately noticed when Monte Carlo sampling the Rayleigh matrix with the determinant state estimator (eq. (28)), since the probability of all samples is zero in this case.

	Determinant	Sum
Sampling	$O\left(m\gamma + m^3\right)$	$O\left(m\gamma\right)$
Local term	$O\left(m^2 l\gamma + m^3\right)$	$O\left(ml\gamma + m^2\right)$

Table 2: Cost of the determinant state and sum of states estimators, where γ is the complexity of evaluating the network of one of the basis states, l is the number of connected elements of the Hamiltonian, and m is the number of basis states. *Sampling* refers to the complexity of querying the amplitude of a sample, under the assumption that the new sample proposal only changes the configuration in a single copy of the original Hilbert space $\mathcal{H}$ (but we do not use the rank-1 update determinant formula). *Local term* refers to the complexity of evaluating the local term for a single sample. It is either the local energy matrix $\mathbf{\Phi}(\boldsymbol{s})^{-1}\mathbf{\Phi}^{(H)}(\boldsymbol{s})$ or the local matrix $\frac{\langle\phi_i|s\rangle\langle s|H|\phi_j\rangle}{\sum_k |\langle s|\phi_k\rangle|^2}$.

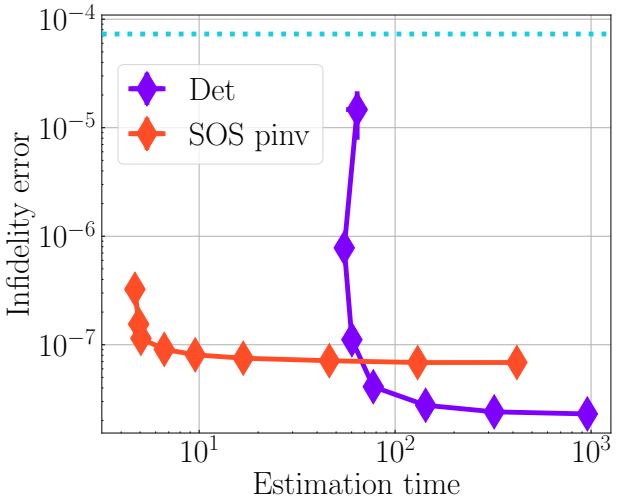

Figure 2: Comparison between the final infidelity error of Bridge performed using two estimators of the Rayleigh matrix as a function of the estimation time of the Rayleigh matrix, corresponding to varying numbers of samples. The two estimators are: in purple, the determinant state estimator of eq. (28), and in red, the sum of states (SOS) estimator of eqs. (39) and (40) where G is inverted using the pseudo-inverse with an appropriately chosen singular value cut-off. The dotted horizontal blue line indicates the performance of the original states $|\phi_k\rangle$. The Hamiltonian is the transverse-field Ising Hamiltonian on a 4×4 lattice with $(h, J) = (1, 0.1)$, and the basis states are 136 CNN states obtained with p-tVMC from time 0 to 1.35 with $\delta = 0.01$.

be inverted in order to obtain the Rayleigh matrix. As G is typically ill-conditioned, we need to use the pseudo-inverse with a well-chosen singular value cut-off. Since this estimator relies on samples drawn from the summed probability distributions of all states $|\phi_k\rangle$, we call it the *sum of states* (SOS) estimator.

In table 2, we report the computational complexity of querying the probability distribution and local estimator for each estimator. While the determinant state estimator is more expensive per sample, numerical experiments reported in fig. 2 indicate that it leads to more accurate results. To evaluate the performance of each estimator, we run Bridge with a given basis of states for a single quench of the 4×4 transverse-field Ising model, which is discussed in more details in section 4. We run Bridge with each Rayleigh matrix estimator on a single A100 GPU with an increasing number of samples. We find that the best performance is obtained with the determinant state estimator, and so we will be using this estimator for the numerical results of section 4. The importance of using the determinant state estimator will be further exemplified with fig. 4, which shows that the performance of the sum of states estimator is inconsistent and highly dependent on the choice of the pseudo-inverse singular value cut-off. In section D, we provide further details regarding the various possible estimators of the Rayleigh matrix and their associated benchmarks, as well as theoretical justification for the superiority of the determinant state estimator. In particular, we justify that the determinant state estimator is more resilient to basis states that are near linear dependence, which is the case for approximate dynamics states such as those obtained by p-tVMC, especially as the size of the basis increases.

Relation to Ref. [23] In Ref. [23], the authors employed a similar method of variational optimization of linear coefficients to post-process global-in-time dynamics, using the sum of states estimator. In their specific case, the basis states are not approximate dynamics states, and therefore do not necessarily suffer from near linear dependence. Additionally, the size of their bases are small, never exceeding 20 states, which might explain why the sum of states estimator was sufficient. However, as detailed in section D, bases of states obtained by the time-dependent variational principle or p-tVMC are typically near linear dependence, which causes the sum of states estimator to fail, such that we instead rely on the more resilient determinant state estimator.

3.3 Estimating observables

Once the parameters $\alpha(t)$ have been obtained, the state $|\psi_{\alpha(t)}\rangle = \sum_k \alpha_k(t)\,|\phi_k\rangle$ can be sampled from and average values of observables can be estimated using standard methods.

Alternatively, the average value of an observable A can be estimated for any number of times from a single estimation. Due to the linearity of the variational state, its average values can be expressed as

$$\frac{\langle\psi_{\alpha(t)}|A|\psi_{\alpha(t)}\rangle}{\langle\psi_{\alpha(t)}|\psi_{\alpha(t)}\rangle} = \frac{\alpha^\dagger(t)G^{(A)}\alpha(t)}{\alpha^\dagger(t)G\alpha(t)}. \tag{42}$$

The matrices G and $G^{(A)}$ can be estimated using the estimator of eqs. (39) and (40). We remark that the determinant state estimator cannot be used here, since it does not provide G and $G^{(A)}$ independently.

While the second method of estimating average values should be more efficient when computing many data points, we did not find a simple method to estimate statistical error bars. As such, for the numerical results, we will restrict ourselves to the first method.[8]

4 Numerical results

We study the performance of Bridge for the dynamics of the transverse-field Ising model

$$H = -J\sum_{\langle i,j\rangle}\sigma_i^{(z)}\sigma_j^{(z)} - h\sum_i \sigma_i^{(x)}, \tag{43}$$

where we examine in particular the time evolution of the average magnetization

$$M_x = \frac{1}{n}\sum_k \sigma_k^{(x)}. \tag{44}$$

We consider square lattices of different sizes, with periodic boundary conditions for $J = 1$ and various values of h/h_c with $h_c = 3.044J$, the value of h at the phase transition. The initial state is the uniform state

$$|\psi_0\rangle = \bigotimes_{k=1}^{n} \frac{|\uparrow\rangle + |\downarrow\rangle}{\sqrt{2}}. \tag{45}$$

[8] We note that the method for estimating average values relying only on Rayleigh matrices detailed in proposition B.3 does not apply here, since it relies on the orthogonality of the Ritz vectors. Here, the states $|\psi_{\alpha(t)}\rangle$ are linear combinations of Ritz vectors, and since the norms of the Ritz vectors are unknown, it is not clear how to complete $|\psi_{\alpha(t)}\rangle$ into an orthogonal basis.

To obtain the original dynamics states $|\phi_k\rangle$, we use p-tVMC [20, 21, 43]. More details are given in section E. The networks used in this section are either convolutional neural networks [18] (for 4×4 lattices) or vision transformers [44] (for 8×8 lattices). Both networks enforce the translational invariance of the dynamics. For more details on the architectures, we refer to appendix H of Ref. [21].

In fig. 3, we consider a 4×4 lattice, where we have access to the exact dynamics, such that we can measure the performance of the various methods by evaluating its infidelity with the exact dynamics. We run Bridge on the states $|\phi_k\rangle$ obtained using p-tVMC with the discretization scheme SLPE2 [21]. In the figure, the light blue line corresponds to the p-tVMC data while the purple line corresponds to Bridge. We examine the two cases $h = 0.1h_c$ (left) and $h = 2h_c$ (right). In the first case, we see that Bridge improves on p-tVMC by one order of magnitude up to time $0.5J$ and then does not improve any more. In the second case, we see that Bridge significantly improves on p-tVMC, for all times, by 3 or 4 orders of magnitudes. In yellow, we plot the optimal infidelity with respect to the exact dynamics that can be achieved within the subspace spanned by the original states $|\phi_k\rangle$, which we observe Bridge is very close to.

To understand when Bridge improves on the states $|\phi_k\rangle$, we detail the two sources of error arising from the p-tVMC method used to obtain them (see section E for more details). (i) The first is *discretization error* emerging from the low-order expansion of the unitary evolution. (ii) The second is *optimization error*, which arises because the optimization problem at each discrete time-step cannot be performed exactly, for example because of finite variational expressivity. In fig. 3, we present in green the exact (non-variational) discretized dynamics, that is, the dynamics when applying exactly the operator resulting from the Taylor expansion of the time evolution operator at each step. When the p-tVMC infidelity is close to the discretized infidelity, this means that the p-tVMC error is dominated by discretization error. Conversely, when the p-tVMC infidelity is above the discretized infidelity, the p-tVMC error is dominated by the optimization error. What we therefore identify in fig. 3 is that Bridge improves the performance of the original states when they

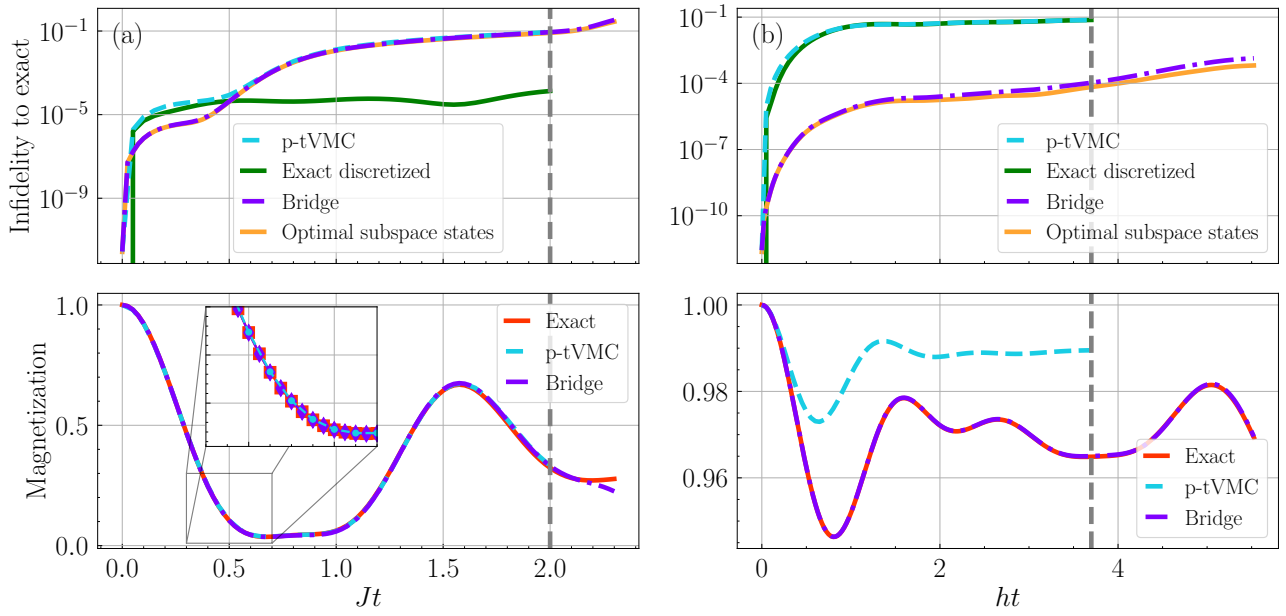

Figure 3: Infidelity to exact dynamics and magnetization on the 4×4 transverse-field Ising model with $h = 0.1h_c$ (a) and $h = 2h_c$ (b). The original states (blue) are convolutional neural networks optimized with p-tVMC using the scheme SLPE2 with time step $J\delta = 0.05$ (a) and $h\delta = 0.05$ (b). For the Bridge states (purple), the Rayleigh matrix is estimated with the determinant state estimator using 3000 samples. Exact discretized (green) refers to the dynamics resulting from the exact application of the discretized time evolution operator. The vertical grey line highlights the final time of the p-tVMC states.

are dominated by discretization error, but not when they are dominated by optimization error.

This behaviour can be understood by looking at the form of each error. At each time step, the discretization error is the remainder of the Taylor expansion formula, which is a linear combination of powers of H applied to the current state.[9] On the other hand, optimization error has no particular structure, which justifies why linear combinations cannot significantly correct it.

In the previous section, we saw that the special structure of the time-dependent variational principle allows the evaluation of the linear parameter α for many times at almost no additional cost, meaning that Bridge can also be used to interpolate between time steps. The insets of fig. 3 show that Bridge is indeed successful at interpolating between the times corresponding to the original states $|\phi_k\rangle$.

Lastly, fig. 3 also examines whether Bridge can extrapolate beyond the times of the $|\phi_k\rangle$. In both cases, we see that while Bridge does have some extrapolating power, its performance predictably wanes over time.

Figure 4 shows the performance of Bridge in the case $(h, J) = (1, 0.1)$ using convolutional neural network states generated with p-tVMC, with varying time steps δ. In this case, Bridge is performed with both the determinant state and the sum of states estimator, for various values of the pseudo-inverse singular value cut-off. We first observe that the performance of the sum of states estimator is inconsistent for different basis sizes, and even for a given basis, it strongly depends on the chosen value of the cut-off. On the other hand, the determinant state estimator displays consistent improvement as the basis size increases, and systematically both outperforms the sum of states estimator and reaches the optimal accuracy within the subspace, further confirming that it is the most adapted estimator of the Rayleigh matrix in the context of Bridge.

Just like in the previously mentioned $h = 2h_c$ case, the dominating source of error is discretization error, meaning that Bridge is expected to significantly improve performance. Indeed, when using the determinant state estimator, we see im-

[9] In the case of SLPE schemes, the Hamiltonian is split between the diagonal part H_0 and the off-diagonal part H_1, such that the remainder would instead be a linear combination of products of H_0 and H_1.

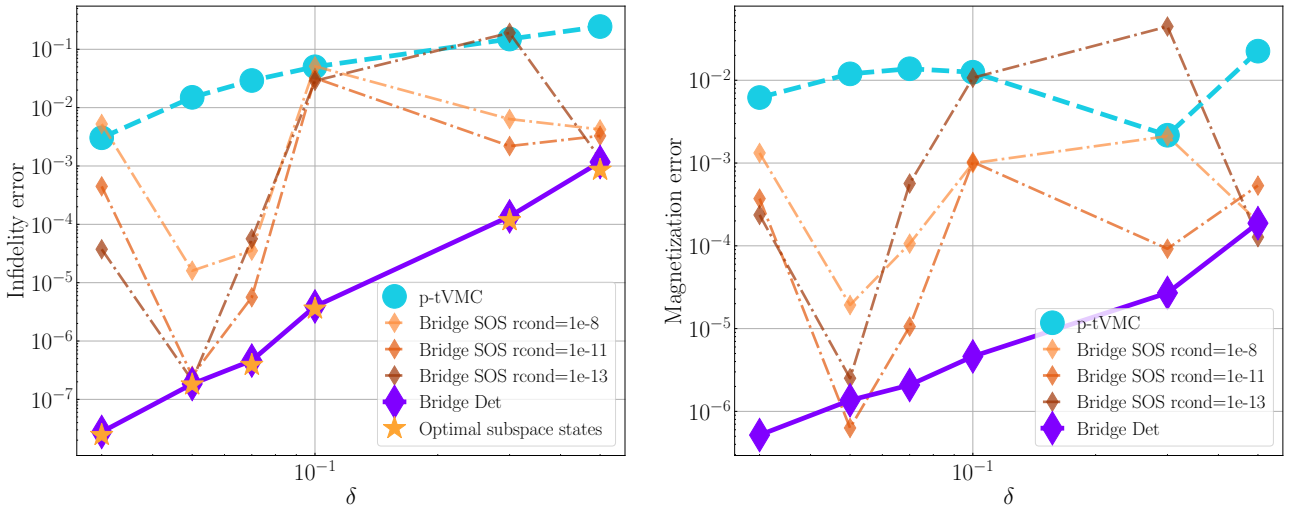

Figure 4: Performance of Bridge with either the determinant state estimator (Det) or sum of states estimator (SOS), using families of p-tVMC states generated with varying time steps δ between time 0 and 1, for the transverse-field Ising Hamiltonian on a 4×4 lattice with $(h, J) = (1, 0.1)$. The performance of Bridge with the sum of states estimator is shown for various values of the pseudo-inverse singular value cut-off `rcond` for inverting G.

provements in the fidelity of the dynamics for all values of δ, ranging from 2 to 5 orders of magnitude. As δ increases, we expect that the quality of the states degrade to the extent that Bridge cannot improve them as much as for lower values of δ.

In fig. 5, we present a practical application of Bridge to a system size of 8×8 lattice sites, beyond the capabilities of exact calculations. Similarly to the results on the 4×4 lattice, we see the most improvement in the case $h = 2h_c$, where we expect to be dominated by discretization error. In this case, the numerical results show that Bridge faithfully reproduces the behaviour of the magnetization, even though the basis states do not. This demonstrates the usefulness of Bridge as a practical method that can improve dynamics at a negligible cost. For that practical application, obtaining the p-tVMC states takes about 12 hours on 4 H100 GPUs for (a) and 17 hours on 12 H100 GPUs for (b), while running Bridge on a single A100 GPU takes 15 minutes for (a) and 60 minutes for (b).

Restarting dynamics from a Bridge state In subplot (b) of fig. 5, going beyond times $ht = 2$ reveals a breakdown of Bridge. The interpretation is that after some time, the original basis states have gone so far away from the exact dynamics that we cannot expect Bridge to be able to make use of them. However, since Bridge successfully improved the dynamics at earlier times, we could run p-tVMC again from time $ht = 2$, starting from the corresponding Bridge state. Doing so, however, requires compressing the linear combination state $\sum_k \alpha_k(t) \, |\phi_{\theta_k}\rangle$ back into a single variational state $|\phi_\theta\rangle$. This operation can be performed with infidelity optimization [21], but it has revealed itself especially challenging in this case. Further investigation into the compression of linear states could allow restarting the dynamics from Bridge solutions whenever the original dynamics strays too far from the exact dynamics.

5 Conclusion

In this paper, we presented the determinant state mapping, which is a natural formalism to treat subspaces of the Hilbert space. This mapping results in both a natural notion of distance between subspaces, and a natural notion of energy which is consistent with the one suggested by subspace methods. The resulting formalism is applicable in the context of variational states, and is compatible with the tools developed for Monte Carlo estimation, while providing new ones. It recontextualizes existing methods and provides a framework where new methods can be developed.

We then examined Bridge, a computationally cheap subspace method relying on tools from this

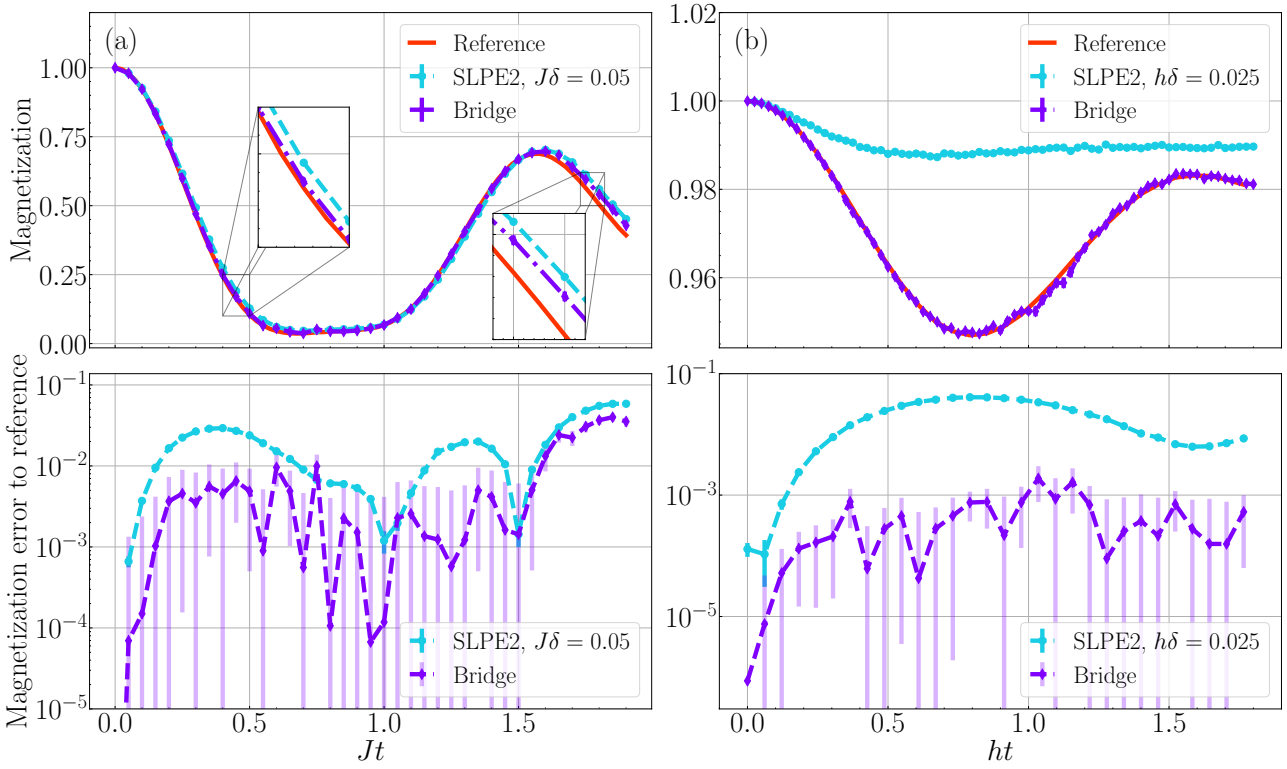

Figure 5: Magnetization on the 8×8 transverse-field Ising model with $h = 0.1h_c$ (a) and $h = 2h_c$ (b). The original states are vision transformers optimized with p-tVMC using the scheme SLPE2 with time step $J\delta = 0.05$ (a) and $h\delta = 0.025$ (b). The reference for (a) are vision transformers optimized with p-tVMC using the higher order scheme SLPE3 with a smaller time step $J\delta = 0.025$. The reference for (b) is convolutional neural network states obtained with t-VMC in Ref. [18], since obtaining good p-tVMC states in this regime would have been too computationally expensive.

framework, which can improve the accuracy of variational dynamics obtained by other established methods. We discussed different estimators for the Rayleigh matrix, and showed that the determinant state estimator is the most suitable estimator in the context of Bridge.

Numerical evidence suggests that Bridge can significantly reduce time step discretization errors, while it does not provide significant improvements for optimization (or variationally-induced) errors. We saw that it is a powerful interpolation tool, while its extrapolation capabilities are more limited.

The low cost and potentially significant improvements provided by Bridge thus make it appear as a method that can be systematically used to post-process states approximating dynamics.

An important point that should be further examined is the compression of linear states, in order to restart the dynamics from Bridge states and further improve the performance of both the original time evolution method and Bridge. Further investigations into Bridge could also examine the impact of enriching the subspace with more states, for examples Krylov states built from any of the $|\phi_k\rangle$.

More importantly, Bridge is thus evidence of the usefulness of subspace methods in the context of variational methods, especially when using the formalism of the determinant state mapping. It therefore invites more investigation into the potential new methods and tools that could be derived from using the determinant state mapping.

Acknowledgments

Simulations were performed with NetKet [45, 46], and at times parallelized with mpi4JAX [47]. This software is built on top of JAX [48], equinox [49] and Flax [50]. We used QuTiP [51, 52] for exact benchmarks and mpmath for arbitrary precision arithmetics [53]. p-tVMC simulations where performed using the repository from [21]. The accompanying code for the manuscript is avail-

able at github.com/NeuralQXLab/variational-subspace-methods.

We acknowledge extensive discussions with A. Shokry and R. Nutakki, as well as insightful discussions with G. Booth, G. Carleo, Z. Denis, A. Sinibaldi and V. Savona. F.V. and A.K. acknowledge support by the French Agence Nationale de la Recherche through the NDQM project, grant ANR-23-CE30-0018. This project was provided with computing HPC and storage resources by GENCI at IDRIS thanks to the grants 2023-AD010514908 and 2024-A0170515698 on the supercomputer Jean Zay's V100/A100 partition.

We also acknowledge a concurrent work by D. Hendry, A. Sinibaldi, and G. Carleo [54], where a similar formalism is discussed to describe subspaces and used to estimate excited states. This work appeared simultaneously on the preprint server.

References

[1] Adam D. Bookatz. "Qma-complete problems". Quantum Information and Computation **14**, 361–383 (2014).

[2] Jacob C Bridgeman and Christopher T Chubb. "Hand-waving and interpretive dance: an introductory course on tensor networks". Journal of Physics A: Mathematical and Theoretical **50**, 223001 (2017).

[3] Román Orús. "A practical introduction to tensor networks: Matrix product states and projected entangled pair states". Annals of Physics **349**, 117–158 (2014).

[4] Ulrich Schollwöck. "The density-matrix renormalization group in the age of matrix product states". Annals of Physics **326**, 96–192 (2011).

[5] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H. Booth, and Jonathan Tennyson. "The variational quantum eigensolver: A review of methods and best practices". Physics Reports **986**, 1–128 (2022).

[6] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. "A variational eigenvalue solver on a photonic quantum processor". Nature Communications**5** (2014).

[7] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". New Journal of Physics **18**, 023023 (2016).

[8] Giuseppe Carleo and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks". Science **355**, 602–606 (2017).

[9] Anna Dawid, Julian Arnold, Borja Requena, Alexander Gresch, Marcin Płodzień, Kaelan Donatella, Kim A. Nicoli, Paolo Stornati, Rouven Koch, Miriam Büttner, Robert Okuła, Gorka Muñoz-Gil, Rodrigo A. Vargas-Hernández, Alba Cervera-Lierta, Juan Carrasquilla, Vedran Dunjko, Marylou Gabrié, Patrick Huembeli, Evert van Nieuwenburg, Filippo Vicentini, Lei Wang, Sebastian J. Wetzel, Giuseppe Carleo, Eliška Greplová, Roman Krems, Florian Marquardt, Michał Tomza, Maciej Lewenstein, and Alexandre Dauphin. "Machine learning in quantum sciences". Cambridge University Press. (2025).

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

[11] Kenny Choo, Titus Neupert, and Giuseppe Carleo. "Two-dimensional frustrated j1-j2 model studied with neural network quantum states". Physical Review B**100** (2019).

[12] Luciano Loris Viteritti, Riccardo Rende, and Federico Becca. "Transformer variational wave functions for frustrated quantum spin systems". Physical Review Letters**130** (2023).

[13] Ao Chen and Markus Heyl. "Empowering deep neural quantum states through efficient optimization". Nature Physics **20**, 1476–1481 (2024).

[14] Zakari Denis and Giuseppe Carleo. "Accurate neural quantum states for interacting lattice bosons". Quantum **9**, 1772 (2025).

[15] Di Luo and Bryan K. Clark. "Backflow transformations via neural networks for

quantum many-body wave functions". Physical Review Letters**122** (2019).

[16] Javier Robledo Moreno, Giuseppe Carleo, Antoine Georges, and James Stokes. "Fermionic wave functions from neural-network constrained hidden states". Proceedings of the National Academy of Sciences**119** (2022).

[17] Giuseppe Carleo, Lorenzo Cevolani, Laurent Sanchez-Palencia, and Markus Holzmann. "Unitary dynamics of strongly interacting bose gases with the time-dependent variational monte-carlo method in continuous space". Physical Review X**7** (2017).

[18] Markus Schmitt and Markus Heyl. "Quantum many-body dynamics in two dimensions with artificial neural networks". Physical Review Letters**125** (2020).

[19] Tiago Mendes-Santos, Markus Schmitt, and Markus Heyl. "Highly resolved spectral functions of two-dimensional systems with neural quantum states". Phys. Rev. Lett. **131**, 046501 (2023).

[20] Alessandro Sinibaldi, Clemens Giuliani, Giuseppe Carleo, and Filippo Vicentini. "Unbiasing time-dependent variational monte carlo by projected quantum evolution". Quantum **7**, 1131 (2023).

[21] Luca Gravina, Vincenzo Savona, and Filippo Vicentini. "Neural projected quantum dynamics: a systematic study". Quantum **9**, 1803 (2025).

[22] Anka Van de Walle, Markus Schmitt, and Annabelle Bohrdt. "Many-body dynamics with explicitly time-dependent neural quantum states". Machine Learning: Science and Technology **6**, 045011 (2025).

[23] Alessandro Sinibaldi, Douglas Hendry, Filippo Vicentini, and Giuseppe Carleo. "Time-dependent neural galerkin method for quantum dynamics". Physical Review Letters**136** (2026).

[24] Gene Golub and Charles Van Loan. "Matrix computations". Johns Hopkins University Press. (2013).

[25] Yousef Saad. "Numerical methods for large eigenvalue problems: Revised edition". Society for Industrial and Applied Mathematics. (2011).

[26] Lloyd N. Trefethen and David Bau. "Numerical linear algebra, twenty-fifth anniversary edition". Society for Industrial and Applied Mathematics. (2022).

[27] Fabrizio Minganti and Dolf Huybrechts. "Arnoldi-lindblad time evolution: Faster-than-the-clock algorithm for the spectrum of time-independent and floquet open quantum systems". Quantum **6**, 649 (2022).

[28] Jia-Qi Wang, Rong-Qiang He, and Zhong-Yi Lu. "Generalized lanczos method for systematic optimization of neural-network quantum states". Physical Review B**113** (2026).

[29] Hongwei Chen, Adrian Feiguin, Douglas Hendry, and Phillip Weinberg. "Systematic improvement of neural network quantum states using lanczos". In Advances in Neural Information Processing Systems 35. Page 7490–7503. NeurIPS 2022. Neural Information Processing Systems Foundation, Inc. (NeurIPS) (2022).

[30] Sandro Sorella. "Generalized lanczos algorithm for variational quantum monte carlo". Physical Review B**64** (2001).

[31] Sebastian Paeckel, Thomas Köhler, Andreas Swoboda, Salvatore R. Manmana, Ulrich Schollwöck, and Claudius Hubig. "Time-evolution methods for matrix-product states". Annals of Physics **411**, 167998 (2019).

[32] André Melo, Gaspard Beugnot, and Fabrizio Minganti. "Variational perturbation theory in open quantum systems for efficient steady state computation" (2025).

[33] Kenny Choo, Giuseppe Carleo, Nicolas Regnault, and Titus Neupert. "Symmetries and many-body excitations with neural-network quantum states". Physical Review Letters**121** (2018).

[34] David Pfau, Simon Axelrod, Halvard Sutterud, Ingrid von Glehn, and James S. Spencer. "Accurate computation of quantum excited states with neural networks". Science**385** (2024).

[35] Phillip Griffiths and Joseph Harris. "Principles of algebraic geometry". Wiley. (1994).

[36] Vojtech Havlicek. "Amplitude Ratios and Neural Network Quantum States". Quantum **7**, 938 (2023).

[37] M. P. Nightingale and Vilen Melik-Alaverdian. "Optimization of ground- and excited-state wave functions and van der waals clusters". Physical Review Letters**87** (2001).

[38] Shi-Xin Zhang and Lei Wang. "A unified variational framework for quantum excited states" (2025).

[39] Lexin Ding, Cheng-Lin Hong, and Christian Schilling. "Ground and excited states from ensemble variational principles". Quantum **8**, 1525 (2024).

[40] Phani K. V. V. Nukala and P. R. C. Kent. "A fast and efficient algorithm for slater determinant updates in quantum monte carlo simulations". The Journal of Chemical Physics**130** (2009).

[41] Thomas Duguet, Andreas Ekström, Richard J. Furnstahl, Sebastian König, and Dean Lee. "Colloquium: Eigenvector continuation and projection-based emulators". Reviews of Modern Physics**96** (2024).

[42] Yannic Rath and George H. Booth. "Interpolating numerically exact many-body wave functions for accelerated molecular dynamics". Nature Communications**16** (2025).

[43] Jannes Nys, Gabriel Pescia, Alessandro Sinibaldi, and Giuseppe Carleo. "Ab-initio variational wave functions for the time-dependent many-electron schrödinger equation". Nature Communications**15** (2024).

[44] Luciano Loris Viteritti, Riccardo Rende, and Federico Becca. "Transformer variational wave functions for frustrated quantum spin systems". Phys. Rev. Lett. **130**, 236401 (2023).

[45] Giuseppe Carleo, Kenny Choo, Damian Hofmann, James E.T. Smith, Tom Westerhout, Fabien Alet, Emily J. Davis, Stavros Efthymiou, Ivan Glasser, Sheng-Hsuan Lin, Marta Mauri, Guglielmo Mazzola, Christian B. Mendl, Evert van Nieuwenburg, Ossian O'Reilly, Hugo Théveniaut, Giacomo Torlai, Filippo Vicentini, and Alexander Wietek. "Netket: A machine learning toolkit for many-body quantum systems". SoftwareX **10**, 100311 (2019).

[46] Filippo Vicentini, Damian Hofmann, Attila Szabó, Dian Wu, Christopher Roth, Clemens Giuliani, Gabriel Pescia, Jannes Nys, Vladimir Vargas-Calderón, Nikita Astrakhantsev, and Giuseppe Carleo. "Netket 3: Machine learning toolbox for many-body quantum systems". SciPost Physics Codebases (2022).

[47] Dion Häfner and Filippo Vicentini. "mpi4jax: Zero-copy mpi communication of jax arrays". Journal of Open Source Software **6**, 3419 (2021).

[48] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. "JAX: composable transformations of Python+NumPy programs" (2018).

[49] Patrick Kidger and Cristian Garcia. "Equinox: neural networks in jax via callable pytrees and filtered transformations" (2021).

[50] Jonathan Heek, Anselm Levskaya, Avital Oliver, Marvin Ritter, Bertrand Rondepierre, Andreas Steiner, and Marc van Zee. "Flax: A neural network library and ecosystem for JAX" (2024).

[51] J.R. Johansson, P.D. Nation, and Franco Nori. "Qutip: An open-source python framework for the dynamics of open quantum systems". Computer Physics Communications **183**, 1760–1772 (2012).

[52] J.R. Johansson, P.D. Nation, and Franco Nori. "Qutip 2: A python framework for the dynamics of open quantum systems". Computer Physics Communications **184**, 1234–1240 (2013).

[53] The mpmath development team. "mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 1.3.0)". `https://mpmath.org/` (2023).

[54] Douglas Hendry, Alessandro Sinibaldi, and Giuseppe Carleo. "Grassmann variational monte carlo with neural wave functions" (2025).

[55] Matija Medvidović and Giuseppe Carleo. "Classical variational simulation of the quantum approximate optimization algorithm". npj Quantum Information **7** (2021).

[56] Kaelan Donatella, Zakari Denis, Alexandre Le Boité, and Cristiano Ciuti. "Dynamics with autoregressive neural quantum states: Application to critical quench dynamics". Phys. Rev. A **108**, 022210 (2023).

A Determinant state formalism

A.1 Antisymmetric subspace and antisymmetrizer

We recall the definitions of the antisymmetric subspace and the antisymmetrizer operator.

Definition A.1 (Permutation operator)**.** *The permutation operator $\boldsymbol{P_\sigma}$ over the Hilbert space $\mathcal{H}^{\otimes m}$ corresponding to a permutation $\sigma \in S_m$ is defined on product states by*

$$\boldsymbol{P_\sigma} \left|\phi_1\right\rangle \otimes \left|\phi_2\right\rangle \otimes \ldots \otimes \left|\phi_m\right\rangle = \left|\phi_{\sigma^{-1}(1)}\right\rangle \otimes \left|\phi_{\sigma^{-1}(2)}\right\rangle \otimes \ldots \otimes \left|\phi_{\sigma^{-1}(m)}\right\rangle . \tag{46}$$

The reason why the states are permuted with σ^{-1} instead of σ is because only this definition produces a representation of S_m on $\mathcal{H}^{\otimes m}$, that is, it is the only one satisfying $\boldsymbol{P_{\sigma\sigma'}} = \boldsymbol{P_\sigma P_{\sigma'}}$.

Definition A.2 (Antisymmetric subspace)**.** *The antisymmetric subspace $V_-(\mathcal{H}^{\otimes m})$ is defined by*

$$V_-(\mathcal{H}^{\otimes m}) = \left\{\left|\psi\right\rangle \in \mathcal{H}^{\otimes m} \text{ such that } \forall \sigma \in S_m,\ \boldsymbol{P_\sigma} \left|\psi\right\rangle = \mathrm{sgn}(\sigma) \left|\psi\right\rangle\right\} . \tag{47}$$

Definition A.3 (Antisymmetrizer operator)**.** *The antisymmetrizer $\boldsymbol{S_-}$ is the operator of $\mathcal{H}^{\otimes m}$ defined by*

$$\boldsymbol{S_-} = \frac{1}{m!} \sum_\sigma \mathrm{sgn}(\sigma) \boldsymbol{P_\sigma}, \tag{48}$$

where $\boldsymbol{P_\sigma}$ is the permutation operator that permutes the states according to the permutation σ.

We also recall several simple properties of the antisymmetrizer that will be useful throughout the appendix.

Proposition A.1. *The antisymmetrizer $\boldsymbol{S_-}$ is the orthogonal projector on the antisymmetric subspace $V_-(\mathcal{H}^{\otimes m})$. In other words, it satisfies:*

- $\boldsymbol{S_-^2} = \boldsymbol{S_-}$,
- $\boldsymbol{S_-^\dagger} = \boldsymbol{S_-}$,
- $im\,(\boldsymbol{S_-}) = V_-(\mathcal{H}^{\otimes m})$.

Proposition A.2. *The antisymmetrizer satisfies*

$$\forall \sigma,\ \boldsymbol{S_- P_\sigma} = \boldsymbol{P_\sigma S_-} = \mathrm{sgn}(\sigma) \boldsymbol{S_-}. \tag{49}$$

Finally, before getting to prove proper results, we show a simple equality that we will extensively use throughout this section.

Proposition A.3. *Let $\{\left|\phi_k\right\rangle\}_k$ and $\{\left|\varphi_k\right\rangle\}_k$ be two families of m states of $\mathcal{H}$. If we define by $\left|\boldsymbol{\phi}\right\rangle$ and $\left|\boldsymbol{\varphi}\right\rangle$ the product states of each family*

$$\left|\boldsymbol{\phi}\right\rangle = \left|\phi_1\right\rangle \otimes \left|\phi_2\right\rangle \otimes \ldots \otimes \left|\phi_m\right\rangle , \tag{50}$$

$$\left|\boldsymbol{\varphi}\right\rangle = \left|\varphi_1\right\rangle \otimes \left|\varphi_2\right\rangle \otimes \ldots \otimes \left|\varphi_m\right\rangle , \tag{51}$$

and by $\left|\boldsymbol{\phi_A}\right\rangle = \boldsymbol{S_-} \left|\boldsymbol{\phi}\right\rangle$ and $\left|\boldsymbol{\varphi_A}\right\rangle = \boldsymbol{S_-} \left|\boldsymbol{\varphi}\right\rangle$ their antisymmetrization, then we have

$$\left\langle\boldsymbol{\phi_A}\middle|\boldsymbol{\varphi_A}\right\rangle = \left\langle\boldsymbol{\phi}\middle|\boldsymbol{S_-}\middle|\boldsymbol{\varphi}\right\rangle = \frac{1}{m!} \det S, \tag{52}$$

where S is the $m \times m$ overlap matrix defined by $S_{ij} = \langle\phi_i|\varphi_j\rangle$.

Proof. We have

$$\langle \boldsymbol{\phi_A} | \boldsymbol{\varphi_A} \rangle = \langle \boldsymbol{\phi} | \boldsymbol{S}_-^\dagger \boldsymbol{S}_- | \boldsymbol{\varphi} \rangle \tag{53}$$

$$= \langle \boldsymbol{\phi} | \boldsymbol{S}_- | \boldsymbol{\varphi} \rangle \,, \tag{54}$$

where we used proposition A.1. We now expand the antisymmetrizer

$$\langle \boldsymbol{\phi} | \boldsymbol{S}_- | \boldsymbol{\varphi} \rangle = \frac{1}{m!} \sum_\sigma \operatorname{sgn}(\sigma) \langle \phi_1 | \varphi_{\sigma(1)} \rangle \langle \phi_2 | \varphi_{\sigma(2)} \rangle \dots \langle \phi_m | \varphi_{\sigma(m)} \rangle \tag{55}$$

$$= \frac{1}{m!} \det S. \tag{56}$$

□

A.2 Determinant state mapping

We now prove the property that justifies the definition of the determinant state mapping (eq. (6)).

Proposition A.4 (Slater determinant under a change of basis)**.** *Let $\{|\phi_k\rangle\}_k$ be a family of m states of $\mathcal{H}$, and let $\{|\varphi_k\rangle\}_k$ be the family of m states defined by the linear transformation*

$$|\varphi_p\rangle = \sum_{k=1}^{m} B_{kp} |\phi_k\rangle \,, \tag{57}$$

where B is an invertible $m \times m$ matrix. Then, the antisymmetrization of both product states are colinear, and we have

$$\boldsymbol{S}_- |\varphi_1\rangle \otimes \dots \otimes |\varphi_m\rangle = \det(B) \boldsymbol{S}_- |\phi_1\rangle \otimes \dots \otimes |\phi_m\rangle \,. \tag{58}$$

Proof. We start by expanding the $|\varphi_p\rangle$,

$$\boldsymbol{S}_- |\varphi_1\rangle \otimes \dots \otimes |\varphi_m\rangle = \sum_{k_1,\dots,k_m} B_{k_1,1} \dots B_{k_m,m} \boldsymbol{S}_- |\phi_{k_1}\rangle \otimes \dots \otimes |\phi_{k_m}\rangle \,. \tag{59}$$

From proposition A.3, the amplitude of the antisymmetrized state $\boldsymbol{S}_- |\phi_{k_1}\rangle \otimes \dots \otimes |\phi_{k_m}\rangle$ corresponding the basis element $\boldsymbol{s} = (s_1, \dots, s_m)$ is given by $\det \Phi(\boldsymbol{s})$, where $[\Phi(\boldsymbol{s})]_{ij} = \langle s_i | \phi_{k_j} \rangle$. Therefore, if the m-uplet $(k_1, \dots, k_m)$ has at least one repeated value, then $\Phi(\boldsymbol{s})$ has repeated columns. As such, for any basis state $\boldsymbol{s}$, we have $\det \Phi(\boldsymbol{s}) = 0$, and so we conclude that $\boldsymbol{S}_- |\phi_{k_1}\rangle \otimes \dots \otimes |\phi_{k_m}\rangle = 0$.

We can thus restrict the sum in eq. (59) to only m-uplet with no repeated values. An m-uplet of values $k_i \in \{1, \dots, m\}$ such that no two k_i are identical is effectively a permutation $\sigma : i \mapsto k_i$. We can thus rewrite this expression as a sum over permutations,

$$\boldsymbol{S}_- |\varphi_1\rangle \otimes \dots \otimes |\varphi_m\rangle = \sum_\sigma B_{\sigma(1),1} \dots B_{\sigma(m),m} \boldsymbol{S}_- |\phi_{\sigma(1)}\rangle \otimes \dots \otimes |\phi_{\sigma(m)}\rangle \tag{60}$$

$$= \sum_\sigma B_{\sigma(1),1} \dots B_{\sigma(m),m} \boldsymbol{S}_- \boldsymbol{P}_{\boldsymbol{\sigma}^{-1}} |\phi_1\rangle \otimes \dots \otimes |\phi_m\rangle \tag{61}$$

$$= \left(\sum_\sigma \operatorname{sgn}(\sigma) B_{\sigma(1),1} \dots B_{\sigma(m),m} \right) \boldsymbol{S}_- |\phi_1\rangle \otimes \dots \otimes |\phi_m\rangle \tag{62}$$

$$= \det(B) \boldsymbol{S}_- |\phi_1\rangle \otimes \dots \otimes |\phi_m\rangle \,. \tag{63}$$

where we used proposition A.2. □

We have now proven eq. (5), which ensures that the determinant state mapping eq. (6) is well-defined.

We now show that the determinant state mapping is also injective. We first prove a lemma that is also useful in interpreting the distance between determinant states defined by eq. (9).

Lemma A.1 (Fidelity between determinant states)**.** *Let U and V be m-dimensional subspaces of $\mathcal{H}$ with respective* ***orthonormal*** *bases $\{|\psi_k\rangle\}_k$ and $\{|\phi_k\rangle\}_k$. Then, the fidelity between the corresponding determinant states can be expressed as*

$$\frac{|\langle \boldsymbol{\psi_A}|\boldsymbol{\phi_A}\rangle|^2}{\|\boldsymbol{\psi_A}\|^2\|\boldsymbol{\phi_A}\|^2} = \det G_{\tilde{\phi}}, \tag{64}$$

where $G_{\tilde{\phi}}$ is the Gram matrix of the image of the family $\{|\phi_k\rangle\}_k$ by P_U the orthogonal projector on U,

$$\left[G_{\tilde{\phi}}\right]_{ij} = \langle \phi_i|P_U^\dagger P_U|\phi_j\rangle\,. \tag{65}$$

Proof. First, we emphasize that the left-hand term does not depend on the chosen bases. It is only the right-hand term that requires $G_{\tilde{\phi}}$ to be the Gram matrix of the projection of an orthonormal basis of V.

Using proposition A.3, we can rewrite the left-hand term as

$$\frac{|\langle \boldsymbol{\psi_A}|\boldsymbol{\phi_A}\rangle|^2}{\|\boldsymbol{\psi_A}\|^2\|\boldsymbol{\phi_A}\|^2} = \frac{|\det S|^2}{\det G_\psi \det G_\phi}, \tag{66}$$

where the matrix S is defined by $S_{ij} = \langle\psi_i|\phi_j\rangle$ and the matrices G_ψ and G_ϕ are the Gram matrices of the families $\{|\psi_k\rangle\}_k$ and $\{|\phi_k\rangle\}_k$ respectively. Since we have chosen the two families to be orthonormal, it follows that these two Gram matrices are the identity.

Thus, only the numerator is left. It can be rewritten as

$$\begin{aligned}\frac{|\langle \boldsymbol{\psi_A}|\boldsymbol{\phi_A}\rangle|^2}{\|\boldsymbol{\psi_A}\|^2\|\boldsymbol{\phi_A}\|^2} &= |\det S|^2 &(67)\\ &= (\det S)^*(\det S) &(68)\\ &= \det\left(S^\dagger S\right). &(69)\end{aligned}$$

The matrix elements of $S^\dagger S$ are given by

$$\left[S^\dagger S\right]_{ij} = \sum_k \langle\phi_i|\psi_k\rangle\,\langle\psi_k|\phi_j\rangle\,. \tag{70}$$

We identify $P_U = \sum_k |\psi_k\rangle\,\langle\psi_k|$ the orthogonal projector on U, such that we can rewrite the elements of $S^\dagger S$ as

$$\begin{aligned}\left[S^\dagger S\right]_{ij} &= \langle\phi_i|P_U|\phi_j\rangle &(71)\\ &= \langle\phi_i|P_U^2|\phi_j\rangle &(72)\\ &= \langle\phi_i|P_U^\dagger P_U|\phi_j\rangle &(73)\\ &= \langle\tilde{\phi}_i|\tilde{\phi}_j\rangle\,. &(74)\end{aligned}$$

where we introduced $|\tilde{\phi}_k\rangle = P_U\,|\phi_k\rangle$. We also introduce $G_{\tilde{\phi}}$, the Gram matrix of the projected family $\left\{|\tilde{\phi}_k\rangle\right\}_k$, defined by

$$\left[G_{\tilde{\phi}}\right]_{ij} = \langle\tilde{\phi}_i|\tilde{\phi}_j\rangle\,, \tag{75}$$

such that we have $S^\dagger S = G_{\tilde{\phi}}$. We indeed conclude that

$$\frac{|\langle \boldsymbol{\psi_A}|\boldsymbol{\phi_A}\rangle|^2}{\|\boldsymbol{\psi_A}\|^2\|\boldsymbol{\phi_A}\|^2} = \det G_{\tilde{\phi}}. \tag{76}$$

□

This lemma allow us to easily prove the injectivity of the determinant state mapping.

Proposition A.5 (Injectivity of the determinant state mapping)**.** *The mapping*

$$\begin{cases} \mathrm{Gr}_m(\mathcal{H}) & \to S/(\mathbb{C},\times) \\ V & \mapsto [S_- \left|\phi_1\right\rangle \otimes \ldots \otimes \left|\phi_m\right\rangle] \end{cases} \tag{77}$$

is injective.

Proof. Let U and V be m-dimensional subspaces of $\mathcal{H}$ with respective bases $\{\left|\psi_k\right\rangle\}_k$ and $\{\left|\phi_k\right\rangle\}_k$, such that $\left|\boldsymbol{\psi_A}\right\rangle$ and $\left|\boldsymbol{\phi_A}\right\rangle$ are the same quantum state. We want to show that $U = V$.

Since $\left|\boldsymbol{\psi_A}\right\rangle$ and $\left|\boldsymbol{\phi_A}\right\rangle$ are the same quantum state, their fidelity is 1,

$$\frac{|\langle \boldsymbol{\psi_A}|\boldsymbol{\phi_A}\rangle|^2}{\|\boldsymbol{\psi_A}\|^2\|\boldsymbol{\phi_A}\|^2} = 1. \tag{78}$$

Without loss of generality, $\{\left|\psi_k\right\rangle\}_k$ and $\{\left|\phi_k\right\rangle\}_k$ can be chosen to be orthonormal bases. From lemma A.1, it follows that the Gram matrix $G_{\tilde{\phi}}$ has determinant 1. Using the Hadamard inequality on the Gram determinant, we deduce that

$$\prod_k \|P_U \left|\phi_k\right\rangle\|^2 \geq 1. \tag{79}$$

However, since the vectors $\left|\phi_k\right\rangle$ are normalized, we also have

$$\forall k,\ \|P_U \left|\phi_k\right\rangle\|^2 \leq 1. \tag{80}$$

From both inequalities, we deduce that

$$\forall k,\ \|P_U \left|\phi_k\right\rangle\|^2 = 1. \tag{81}$$

This implies that each state $\left|\phi_k\right\rangle$ is in U. Since they form a basis of V, it follows that $V \subset U$, and since they have the same dimension, we conclude that $U = V$. We have thus shown that if two subspaces have the same determinant state, then they are the same subspace. In other words, the determinant state mapping is injective. □

A.3 Distance between subspaces

In section 2.2, we introduced the distance between subspaces U and V with respective bases $\{\left|\psi_k\right\rangle\}_k$ and $\{\left|\phi_k\right\rangle\}_k$ defined by

$$d_s\left(U,V\right) = d\left(\left|\boldsymbol{\psi_A}\right\rangle, \left|\boldsymbol{\phi_A}\right\rangle\right). \tag{82}$$

To help better understand this distance, the next proposition characterizes maximally distant subspaces under that metric.

Proposition A.6 (Maximally distant subspaces)**.** *Two m-dimensional subspaces U and V of $\mathcal{H}$ have the maximum distance of $\pi/2$ for d_s, if and only if V contains a non-zero vector orthogonal to U.*

Proof. We use the characterization of d_s given by lemma A.1. If V contains a non-zero vector $\left|\phi_1\right\rangle$ orthogonal to U, then this vector can be completed into an orthogonal basis $\{\left|\phi_k\right\rangle\}_k$ of V. The orthogonal projection of this family onto U is not full rank since the first vector is 0, and the resulting Gram matrix $G_{\tilde{\phi}}$ is thus singular. The fidelity between the subspaces is 0, and so their distance is $\pi/2$.

Conversely, suppose that $d_s(U,V) = \pi/2$. Again, using lemma A.1, this implies that for any orthogonal basis $\{\left|\phi_k\right\rangle\}_k$ of V, the Gram matrix $G_{\tilde{\phi}}$ of the orthogonal projection of these states on U satisfies

$$\det G_{\tilde{\phi}} = \cos d_s(U,V) = 0. \tag{83}$$

We conclude that the states $P_U |\phi_k\rangle$ are linearly dependent and that there exists $\alpha \in \mathbb{C}^m,\ \alpha \neq 0$ such that

$$\sum_k \alpha_k P_U |\phi_k\rangle = 0 \tag{84}$$

$$\iff P_U \sum_k \alpha_k |\phi_k\rangle = 0. \tag{85}$$

Since the states $|\phi_k\rangle$ are linearly independent, we conclude that $\sum_k \alpha_k |\phi_k\rangle$ is a non-zero state of V that is orthogonal to U. □

A.4 Energy of a subspace

Proposition A.7 (Energy of the determinant state [34])**.** *Let H be a Hamiltonian acting on $\mathcal{H}$, and let V be an m-dimensional subspace of $\mathcal{H}$ spanned by a basis $\{|\phi_k\rangle\}_k$. The energy of the subspace defined through the determinant state is the trace of its Rayleigh matrix,*

$$\frac{\langle \boldsymbol{\phi_A}|\boldsymbol{H}|\boldsymbol{\phi_A}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} = \operatorname{tr}\left(G^{-1}G^{(H)}\right). \tag{86}$$

where G is the Gram matrix and $G^{(H)}$ is the projected Hamiltonian

$$G_{ij} = \langle \phi_i|\phi_j\rangle, \tag{87}$$

$$G^{(H)}_{ij} = \langle \phi_i|H|\phi_j\rangle. \tag{88}$$

Proof. For convenience, we denote by $|\boldsymbol{\phi}\rangle = |\phi_1\rangle \otimes \ldots \otimes |\phi_m\rangle$ the product state formed from the basis $\{|\phi_k\rangle\}_k$.

We start by rewriting the left-hand term using proposition A.1 and the fact that $\boldsymbol{H}$ commutes with $\boldsymbol{S_-}$,[10]

$$\frac{\langle \boldsymbol{\phi_A}|\boldsymbol{H}|\boldsymbol{\phi_A}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} = \frac{\langle \boldsymbol{\phi}|\boldsymbol{S_-^\dagger H S_-}|\boldsymbol{\phi}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} \tag{89}$$

$$= \frac{\langle \boldsymbol{\phi}|\boldsymbol{S_-^2 H}|\boldsymbol{\phi}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} \tag{90}$$

$$= \frac{\langle \boldsymbol{\phi}|\boldsymbol{S_- H}|\boldsymbol{\phi}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} \tag{91}$$

$$= \frac{\langle \boldsymbol{\phi_A}|\boldsymbol{H}|\boldsymbol{\phi}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} \tag{92}$$

$$= \sum_p \frac{\langle \boldsymbol{\phi_A}|\boldsymbol{H_p}|\boldsymbol{\phi}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle}. \tag{93}$$

□

According to proposition A.3, the denominator can be expressed as $\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle = \frac{1}{m!}\det G$.

We now consider the numerator $\langle \boldsymbol{\phi_A}|\boldsymbol{H_p}|\boldsymbol{\phi}\rangle$. If we define the family $\left\{|\tilde{\phi}_k\rangle\right\}_k$ by

$$|\tilde{\phi}_k\rangle = \begin{cases} |\phi_k\rangle & \text{if } k \neq p \\ H|\phi_k\rangle & \text{if } k = p \end{cases}, \tag{94}$$

then the numerator is

$$\langle \boldsymbol{\phi_A}|\boldsymbol{H_p}|\boldsymbol{\phi}\rangle = \langle \boldsymbol{\phi_A}|\tilde{\boldsymbol{\phi}}\rangle, \tag{95}$$

$$= \frac{1}{m!}\det G^{(H)}_{p\to p}, \tag{96}$$

[10] The Hamiltonian $\boldsymbol{H}$ commutes with each of the permutation operators $\boldsymbol{P_\sigma}$, so it also commutes with $\boldsymbol{S_-}$.

where we used proposition A.3, and we defined $G^{(H)}_{p\to p}$, the matrix G where the p-th column was replaced by the p-th column of $G^{(H)}$,

$$\left[G^{(H)}_{p\to p}\right]_{ij} = \begin{cases} \langle \phi_i | \phi_j \rangle & \text{if } j \neq p \\ \langle \phi_i | H | \phi_p \rangle & \text{if } j = p \end{cases}. \tag{97}$$

We have thus obtained

$$\frac{\langle \boldsymbol{\phi_A} | \boldsymbol{H} | \boldsymbol{\phi_A} \rangle}{\langle \boldsymbol{\phi_A} | \boldsymbol{\phi_A} \rangle} = \sum_p \frac{\det G^{(H)}_{p\to p}}{\det G}. \tag{98}$$

As we pointed out, the matrix $G^{(H)}_{p\to p}$ is nothing but the matrix G where the p-th column was replaced by the p-th column of $G^{(H)}$. We can therefore use Cramer's rule to rewrite the ratio of determinants. If we denote the p-th column of $G^{(H)}$ by $\left[G^{(H)}\right]_{:,p}$, we can rewrite the energy as

$$\begin{aligned} \frac{\langle \boldsymbol{\phi_A} | \boldsymbol{H} | \boldsymbol{\phi_A} \rangle}{\langle \boldsymbol{\phi_A} | \boldsymbol{\phi_A} \rangle} &= \sum_p \left[G^{-1} \left[G^{(H)}\right]_{:,p} \right]_p && (99) \\ &= \sum_p \left[G^{-1} G^{(H)} \right]_{pp} && (100) \\ &= \operatorname{tr}\left(G^{-1} G^{(H)} \right). && (101) \end{aligned}$$

Proposition A.8 (Estimator of the Rayleigh matrix [34])**.** *The Rayleigh matrix can be decomposed as*

$$G^{-1} G^{(H)} = \mathbb{E}_{\boldsymbol{s}\sim|\boldsymbol{\phi_A}(\boldsymbol{s})|^2} \left[\Phi(\boldsymbol{s})^{-1} \Phi^{(H)}(\boldsymbol{s}) \right], \tag{102}$$

with

$$\begin{aligned} \left[\Phi(\boldsymbol{s})\right]_{ij} &= \langle s_i | \phi_j \rangle, && (103) \\ \left[\Phi^{(H)}(\boldsymbol{s})\right]_{ij} &= \langle s_i | H | \phi_j \rangle. && (104) \end{aligned}$$

Proof. We consider the matrix element (k, p) of the Rayleigh matrix. Using Cramer's rule again, we can reexpress it as a ratio of determinants,

$$\begin{aligned} \left[G^{-1} G^{(H)} \right]_{kp} &= \left[G^{-1} G^{(H)}_{:,p} \right]_k && (105) \\ &= \frac{\det G^{(H)}_{k\to p}}{\det G}. && (106) \end{aligned}$$

where $G^{(H)}_{k\to p}$, the matrix G where the k-th column was replaced by the p-th column of $G^{(H)}$.

If we define the family $\left\{|\tilde{\phi}_i\rangle\right\}_i$ by

$$|\tilde{\phi}_i\rangle = \begin{cases} |\phi_i\rangle & \text{if } i \neq k \\ H\,|\phi_p\rangle & \text{if } i = k \end{cases}, \tag{107}$$

then the matrix $G^{(H)}_{k\to p}$ can be written as $\left[G^{(H)}_{k\to p}\right]_{ij} = \langle \phi_i | \tilde{\phi}_j \rangle$. Using proposition A.3, we can therefore rewrite the matrix element as

$$\left[G^{-1} G^{(H)} \right]_{kp} = \frac{\langle \boldsymbol{\phi_A} | \tilde{\boldsymbol{\phi}}_{\boldsymbol{A}} \rangle}{\langle \boldsymbol{\phi_A} | \boldsymbol{\phi_A} \rangle}. \tag{108}$$

From there, we can follow the usual steps to provide a decomposition as an expectation value,

$$\left[G^{-1}G^{(H)}\right]_{kp} = \sum_{\boldsymbol{s}} \frac{\langle \boldsymbol{\phi_A}|\boldsymbol{s}\rangle \langle \boldsymbol{s}|\tilde{\boldsymbol{\phi}}_{\boldsymbol{A}}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} \tag{109}$$

$$= \sum_{\substack{\boldsymbol{s} \\ \langle \boldsymbol{\phi_A}|\boldsymbol{s}\rangle \neq 0}} \frac{\langle \boldsymbol{\phi_A}|\boldsymbol{s}\rangle}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} \langle \boldsymbol{s}|\tilde{\boldsymbol{\phi}}_{\boldsymbol{A}}\rangle \tag{110}$$

$$= \sum_{\substack{\boldsymbol{s} \\ \langle \boldsymbol{\phi_A}|\boldsymbol{s}\rangle \neq 0}} \frac{|\langle \boldsymbol{\phi_A}|\boldsymbol{s}\rangle|^2}{\langle \boldsymbol{\phi_A}|\boldsymbol{\phi_A}\rangle} \frac{\langle \boldsymbol{s}|\tilde{\boldsymbol{\phi}}_{\boldsymbol{A}}\rangle}{\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle} \tag{111}$$

$$= \mathbb{E}_{\boldsymbol{s}\sim|\boldsymbol{\phi_A}(\boldsymbol{s})|^2} \left[\frac{\langle \boldsymbol{s}|\tilde{\boldsymbol{\phi}}_{\boldsymbol{A}}\rangle}{\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle}\right]. \tag{112}$$

Let us now examine the object inside the expectation value. Applying proposition A.3 yet again, we can reexpress both the numerator and denominator as determinants,

$$\langle \boldsymbol{s}|\tilde{\boldsymbol{\phi}}_{\boldsymbol{A}}\rangle = \frac{1}{m!} \det \Phi^{(H)}_{k\to p}(\boldsymbol{s}), \tag{113}$$

$$\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle = \frac{1}{m!} \det \Phi(\boldsymbol{s}), \tag{114}$$

where we defined $[\Phi(\boldsymbol{s})]_{ij} = \langle s_i|\phi_j\rangle$ and

$$\left[\Phi^{(H)}_{k\to p}(\boldsymbol{s})\right]_{ij} = \langle s_i|\tilde{\phi}_j\rangle \tag{115}$$

$$= \begin{cases} \langle s_i|\phi_j\rangle & \text{if } j \neq k \\ \langle s_i|H|\phi_p\rangle & \text{if } j = k \end{cases}. \tag{116}$$

In other words, if we define the matrix $\Phi^{(H)}(\boldsymbol{s})$ by $\left[\Phi^{(H)}(\boldsymbol{s})\right]_{ij} = \langle s_i|H|\phi_j\rangle$, then the matrix $\Phi^{(H)}_{k\to p}(\boldsymbol{s})$ is $\Phi(\boldsymbol{s})$ where the k-th column was replaced by the p-th column of $\Phi^{(H)}(\boldsymbol{s})$. Using Cramer's rule again, we can therefore reexpress

$$\frac{\langle \boldsymbol{s}|\tilde{\boldsymbol{\phi}}_{\boldsymbol{A}}\rangle}{\langle \boldsymbol{s}|\boldsymbol{\phi_A}\rangle} = \frac{\det \Phi^{(H)}_{k\to p}(\boldsymbol{s})}{\det \Phi(\boldsymbol{s})} \tag{117}$$

$$= \left[\Phi(\boldsymbol{s})^{-1}\Phi^{(H)}(\boldsymbol{s})\right]_{kp}. \tag{118}$$

Inserting this expression in eq. (112) for every matrix element, we obtain the final result,

$$G^{-1}G^{(H)} = \mathbb{E}_{\boldsymbol{s}\sim|\boldsymbol{\phi_A}(\boldsymbol{s})|^2} \left[\Phi(\boldsymbol{s})^{-1}\Phi^{(H)}(\boldsymbol{s})\right]. \tag{119}$$

□

Proposition A.9 (Zero-variance property of the Rayleigh matrix estimator)**.** *If the states $|\phi_k\rangle$ span an invariant subspace of H, then the local energy matrix $\Phi(\boldsymbol{s})^{-1}\Phi^{(H)}(\boldsymbol{s})$ is a constant function of $\boldsymbol{s}$.*

Proof. If the states $|\phi_k\rangle$ span an invariant subspace, then there exists an (m, m) matrix B such that

$$\forall p,\ H\,|\phi_p\rangle = \sum_k B_{kp}\,|\phi_k\rangle\,. \tag{120}$$

The matrix $\Phi^{(H)}(\boldsymbol{s})$ can therefore be rewritten as

$$\left[\Phi^{(H)}(\boldsymbol{s})\right]_{ij} = \langle s_i|H|\phi_j\rangle \tag{121}$$

$$= \sum_k B_{kj}\,\langle s_i|\phi_k\rangle \tag{122}$$

$$= [\Phi(\boldsymbol{s})B]_{ij}\,. \tag{123}$$

We can then insert this expression in the Rayleigh matrix to obtain

$$\Phi(\boldsymbol{s})^{-1}\Phi^{(H)}(\boldsymbol{s}) = B, \tag{124}$$

such that the local energy matrix is independent of the value of the sample $\boldsymbol{s}$. We conclude that the variance of the estimator of the Rayleigh matrix induced by proposition A.8 is therefore 0 if the states $|\phi_k\rangle$ span an invariant subspace. □

B Eigenvalues of the Rayleigh matrix as the generalization of eigenvalues in a subspace

In the main text, we introduced the eigenvalues μ_k of the Rayleigh matrix $G^{-1}G^{(H)}$. In this section, we justify why the μ_k, called the *Ritz values*, can be understood as the natural generalization of the eigenvalues of an operator when restricted to a subspace that is not necessarily an invariant subspace. In particular, we will prove the generalized variational principle of eq. (20).

We start by recalling two important results of linear algebra: the min-max theorem and the Cauchy interlacing theorem. While these theorems are expressed in the context of an orthonormal basis, we will see that their results can be easily extended to an arbitrary basis.

Theorem B.1 (Min-max theorem [24])**.** *Let H be an $n \times n$ Hermitian operator acting on the Hilbert space $\mathcal{H}$, and let E_k, $k \in \{0, \dots, n-1\}$ be its eigenvalues indexed in increasing order. The eigenvalue E_k can be expressed as a maximum over subspaces of $\mathcal{H}$ of dimension $k+1$,*

$$E_k = \min_{W \in \mathrm{Gr}_{k+1}(\mathcal{H})} \max_{\substack{|\psi\rangle \in W \\ |\psi\rangle \neq 0}} \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}, \tag{125}$$

where we recall that $\mathrm{Gr}_{k+1}(\mathcal{H})$ is the set of subspaces of $\mathcal{H}$ with dimension $k+1$.

Proof. For simplicity, we will denote

$$E_{\max}(W) = \max_{\substack{|\psi\rangle \in W \\ |\psi\rangle \neq 0}} \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}. \tag{126}$$

In order to prove that $E_k = \min_{W \in \mathrm{Gr}_{k+1}(\mathcal{H})} E_{\max}(W)$, we need to show:

$$\textbf{(i)} \quad \forall W \in \mathrm{Gr}_{k+1}(\mathcal{H}), E_{\max}(W) \geq E_k \tag{127}$$

$$\textbf{(ii)} \quad \exists W \in \mathrm{Gr}_{k+1}(\mathcal{H}),\ E_{\max}(W) = E_k. \tag{128}$$

We denote by $|E_k\rangle$ the k-th eigenvector of H, corresponding to the eigenvalue E_k.

(i) Consider the $(n-k)$-dimensional subspace $V = \mathrm{span}\left(|E_k\rangle, |E_{k+1}\rangle, \dots, |E_{n-1}\rangle\right)$. For any state $|\psi_\alpha\rangle = \sum_{p=k}^{n-1} \alpha_p |E_p\rangle \in V$, the Rayleigh quotient satisfies

$$\frac{\langle\psi_\alpha|H|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle} = \frac{\sum_{p=k}^{n-1} E_p|\alpha_p|^2}{\sum_{p=k}^{n-1}|\alpha_p|^2} \geq E_k. \tag{129}$$

Since any subspace $W \in \mathrm{Gr}_{k+1}(\mathcal{H})$ has dimension $k+1$, its intersection with the $(n-k)$-dimensional subspace V is of dimension greater or equal to 1. Therefore, there exists a non-zero vector of W that also belongs to V. By eq. (129), we deduce that $E_{\max}(W) \leq E_k$.

(ii) We consider the subspace $W^* = \text{span}\left(|E_0\rangle, |E_1\rangle, \ldots, |E_k\rangle\right)$. For a state $|\psi_\alpha\rangle = \sum_{p=0}^{k} \alpha_p |E_p\rangle \in W^*$, the Rayleigh quotient is given by

$$\frac{\langle\psi_\alpha|H|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle} = \frac{\sum_{p=0}^{k} E_p|\alpha_p|^2}{\sum_{p=0}^{k}|\alpha_p|^2}. \tag{130}$$

Its maximum is thus $E_{\max}(W^*) = E_k$.

□

Based on the characterization of eigenvalues given by the min-max theorem, the following theorem justifies how the Ritz values μ_k are the natural extension of eigenvalues when restricted to a subspace. From a variational perspective, they are the smallest values that can be obtained within V.

Theorem B.2 (Cauchy interlacing theorem)**.** *Let H be an $n \times n$ Hermitian operator. Let V be an m-dimensional subspace with **orthonormal** basis $\{|\phi_k\rangle\}_k$. We define the matrix $G^{(H)}$ by $G^{(H)}_{ij} = \langle\phi_i|H|\phi_j\rangle$. If we denote by E_k and μ_k the respective eigenvalues of H and $G^{(H)}$, indexed in increasing order, then we have*

$$\mu_k = \min_{\substack{W \in \text{Gr}_{k+1}(\mathcal{H}) \\ W \subset V}} \max_{\substack{|\psi\rangle \in W \\ |\psi\rangle \neq 0}} \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}, \tag{131}$$

and

$$\forall k \leq m,\ E_k \leq \mu_k. \tag{132}$$

Proof. Applying the min-max theorem to μ_k, we get

$$\mu_k = \min_{W \in \text{Gr}_{k+1}(V)} \max_{\substack{\alpha \in W \\ \alpha \neq 0}} \frac{\alpha^\dagger G^{(H)} \alpha}{\alpha^\dagger \alpha}. \tag{133}$$

We can rewrite

$$\begin{aligned}
\alpha^\dagger G^{(H)} \alpha &= \sum_{kp} \alpha_k^* \langle\phi_k|H|\phi_p\rangle \alpha_p && (134)\\
&= \left(\sum_k \alpha_k^* \langle\phi_k|\right) H \left(\sum_p \alpha_p |\phi_p\rangle\right) && (135)\\
&= \langle\psi_\alpha|H|\psi_\alpha\rangle, && (136)
\end{aligned}$$

where we defined $|\psi_\alpha\rangle = \sum_k \alpha_k |\phi_k\rangle$.

The denominator can also be reexpressed as

$$\begin{aligned}
\alpha^\dagger \alpha &= \sum_k \alpha_k^* \alpha_k && (137)\\
&= \sum_{kp} \alpha_k^* \delta_{kp} \alpha_p && (138)\\
&= \sum_{kp} \alpha_k^* \langle\phi_k|\phi_p\rangle \alpha_p && (139)\\
&= \langle\psi_\alpha|\psi_\alpha\rangle, && (140)
\end{aligned}$$

where we used the orthogonality of the family $\{|\phi_k\rangle\}_k$.

We can thus rewrite the min-max theorem as

$$\mu_k = \min_{W \in \text{Gr}_{k+1}(V)} \max_{\substack{\alpha \in W \\ \alpha \neq 0}} \frac{\langle\psi_\alpha|H|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle}. \tag{141}$$

The set $W_{\mathcal{H}} = \{|\psi_\alpha\rangle, \alpha \in W\}$ defines a $(k+1)$-dimensional subspace of $\mathcal{H}$ which is included in V. Conversely, any $(k+1)$-dimensional subspace of $\mathcal{H}$ which is included in V can be written as $W_{\mathcal{H}}$ for some subspace W of V. We can therefore rewrite the min-max theorem on μ_k once again,

$$\mu_k = \min_{W \in \mathrm{Gr}_{k+1}(V)} \max_{\substack{|\psi\rangle \in W_{\mathcal{H}} \\ |\psi\rangle \neq 0}} \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} \tag{142}$$

$$= \min_{\substack{W \in \mathrm{Gr}_{k+1}(\mathcal{H}) \\ W \subset V}} \max_{\substack{|\psi\rangle \in W \\ |\psi\rangle \neq 0}} \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}. \tag{143}$$

It then follows immediately from the min-max theorem on E_k that

$$\forall k \leq m,\ E_k \leq \mu_k. \tag{144}$$

□

In the Cauchy interlacing theorem, the Ritz values μ_k are defined as the eigenvalues of the matrix $G^{(H)}$ constructed from an orthonormal basis of the subspace V. We note that for a given Hermitian operator H, they are independent of the chosen orthonormal basis, and only depend on the subspace V. The following result indicates how to obtain the Ritz values from matrices constructed from any basis of V.

Proposition B.1 (Ritz values from non-orthonormal families)**.** *Let H be an $n \times n$ Hermitian operator. Let V be an m-dimensional subspace. The corresponding Ritz values μ_k can be obtained from any basis $\{|\phi_k\rangle\}_k$ of V as the eigenvalues of the Rayleigh matrix $G^{-1}G^{(H)}$, where $G_{ij} = \langle\phi_i|H|\phi_j\rangle$ and $G^{(H)}_{ij} = \langle\phi_i|H|\phi_j\rangle$.*

Proof. Let $\left\{|\tilde{\phi}_k\rangle\right\}_k$ be an orthogonalization of the family $\{|\phi_k\rangle\}_k$ (for example through the Gram-Schmidt process), such that the matrices of columns of each families are related by $\tilde{\Phi} = \Phi A$ where $\Phi_{sk} = \langle s|\phi_k\rangle$ and $\tilde{\Phi}_{sk} = \langle s|\tilde{\phi}_k\rangle$, and A is an invertible $m \times m$ matrix that defines the orthogonalization process.

The Ritz values μ_k are the eigenvalues of the matrix $\tilde{G}^{(H)} = \tilde{\Phi}^\dagger H \tilde{\Phi}$. Expressed in terms of Φ, we have $\tilde{G}^{(H)} = A^\dagger \Phi^\dagger H \Phi A = A^\dagger G^{(H)} A$. The eigenvalues are unchanged upon conjugation by an invertible matrix such as A. Therefore, the μ_k are also the eigenvalues of $A\tilde{G}^{(H)}A^{-1} = AA^\dagger G^{(H)}$. On the other hand, we have

$$G = \Phi^\dagger \Phi \tag{145}$$

$$= (A^{-1})^\dagger \tilde{\Phi}^\dagger \tilde{\Phi} A^{-1} \tag{146}$$

$$= (A^{-1})^\dagger A^{-1} \tag{147}$$

$$= (AA^\dagger)^{-1}. \tag{148}$$

where we used the orthonormality of the family $\left\{|\tilde{\phi}_k\rangle\right\}_k$. We can thus substitute $A\tilde{G}^{(H)}A^{-1} = G^{-1}G^{(H)}$, such that the Ritz values μ_k are indeed the eigenvalues of the Rayleigh matrix. □

The results we have given so far on the Ritz values can be illustrated by the special case where V is an invariant subspace of H. Then, the restriction of H to V is well-defined, and it can be shown to be the operator $G^{-1}G^{(H)}$. Therefore, in this special case, the Ritz values are actual eigenvalues of H.

Since each Ritz value μ is an eigenvalue of the Rayleigh matrix, it has an associated eigenvector α. From this eigenvector, we can define a state $|\psi_\alpha\rangle = \sum_p \alpha_p |\phi_p\rangle$ of the full Hilbert space, called *Ritz vector*. In the next proposition, we show that the Ritz vector $|\psi_\alpha\rangle$ has energy μ.

Proposition B.2 (Energy of Ritz vectors)**.** *The eigenvectors $\alpha^{(k)}$ of $G^{-1}G^{(H)}$ corresponding to the eigenvalues μ_k define states of the full Hilbert space $|\psi_{\alpha^{(k)}}\rangle = \sum_p \alpha_p^{(k)} |\phi_p\rangle$ such that the average value of $|\psi_{\alpha^{(k)}}\rangle$ on H is μ_k.*

Proof. Consider α, an eigenvector of $G^{-1}G^{(H)}$ corresponding to the eigenvalue μ. Then, the energy of the associated state $|\psi_\alpha\rangle = \sum_p \alpha_p |\phi_p\rangle$ is given by

$$\frac{\langle\psi_\alpha|H|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle} = \frac{\sum_{p,p'} \alpha_p^* \langle\phi_p|H|\phi_{p'}\rangle \alpha_{p'}}{\sum_{p,p'} \alpha_p^* \langle\phi_p|\phi_{p'}\rangle \alpha_{p'}} \tag{149}$$

$$= \frac{\alpha^\dagger G^{(H)} \alpha}{\alpha^\dagger G \alpha} \tag{150}$$

$$= \mu \frac{\alpha^\dagger G \alpha}{\alpha^\dagger G \alpha} \tag{151}$$

$$= \mu, \tag{152}$$

where we used $G^{(H)}\alpha = \mu G \alpha$. □

This result means that the Ritz values μ_k are not just abstract quantities. They are associated with states of the full Hilbert space that have energy μ_k, and that can be easily constructed.

We have thus seen how the energy of the Ritz vectors $|\psi_{\alpha^{(k)}}\rangle$ can be trivially obtained. We will now see how the average value of other observables can be obtained. While there exist other methods of computing observables of the Ritz vectors as detailed in section 3.3, the following method has the advantage of only involving Rayleigh matrices.

Proposition B.3 (Average values of Ritz vectors)**.** *Let A be an operator. If the Ritz values are all distinct, then the average value of the k-th Ritz vector $|\psi_{\alpha^{(k)}}\rangle$ is given by*

$$\frac{\langle\psi_{\alpha^{(k)}}|A|\psi_{\alpha^{(k)}}\rangle}{\langle\psi_{\alpha^{(k)}}|\psi_{\alpha^{(k)}}\rangle} = \left[P^{-1}G^{-1}G^{(A)}P\right]_{kk}, \tag{153}$$

where P is the matrix of eigenvectors of $G^{-1}G^{(H)}$.

Proof. We first need to prove that under the assumption that the Ritz values are all distinct, the Ritz vectors are orthogonal.

The eigendecomposition of the Rayleigh matrix $G^{-1}G^{(H)} = PDP^{-1}$ defines a family of Ritz vectors whose matrix of columns is $U = \Phi P$, where $\Phi_{sk} = \langle s|\phi_k\rangle$. If we define $Q = \Phi G^{-1}\Phi^\dagger$ the orthogonal projector on the subspace spanned by the states $|\phi_k\rangle$ and $H_Q = QHQ$ the Hamiltonian projected in this subspace, we notice that

$$H_Q U = \Phi G^{-1}\Phi^\dagger H \Phi G^{-1}\Phi^\dagger \Phi P \tag{154}$$

$$= \Phi G^{-1} G^{(H)} P \tag{155}$$

$$= \Phi P D \tag{156}$$

$$= UD, \tag{157}$$

such that the Ritz vectors are also eigenvectors of the projected Hamiltonian H_Q, with the Ritz values as the corresponding eigenvalues.

Under the assumption that the Ritz values are all different, the Ritz vectors are thus eigenvectors of a Hermitian operator for distinct eigenvalues and are therefore orthogonal (but not necessarily orthonormal).

We denote by $|\psi_{\alpha^{(k)}}\rangle$ the Ritz vectors, that is, the columns of the matrix U. Let $\{|\psi_p\rangle\}_p$ be an orthogonal family of $n-m$ vectors orthogonal to all states $|\psi_{\alpha^{(k)}}\rangle$, that completes $\{|\psi_{\alpha^{(k)}}\rangle\}_k$ into an orthogonal basis $\mathcal{B}$. We denote by Ψ the matrix of columns of the family $\{|\psi_p\rangle\}_p$.

Let R be the matrix of the basis $\mathcal{B}$ in the canonical basis, that is,

$$R = \begin{pmatrix} \Phi P & \Psi \end{pmatrix}. \tag{158}$$

Then, the matrix of A in the basis $\mathcal{B}$ can be expressed as[11]

$$\begin{aligned} A_{\mathcal{B}} &= R^{-1} A R && (159)\\ &= R^{-1} \left(R^\dagger\right)^{-1} R^\dagger A R && (160)\\ &= \left(R^\dagger R\right)^{-1} R^\dagger A R. && (161) \end{aligned}$$

Let us examine the block structure of $R^\dagger R$,

$$\begin{aligned} R^\dagger R &= \begin{pmatrix} P^\dagger \Phi^\dagger \\ \Psi^\dagger \end{pmatrix} \begin{pmatrix} \Phi P & \Psi \end{pmatrix} && (162)\\ &= \begin{pmatrix} P^\dagger \Phi^\dagger \Phi P & P^\dagger \Phi^\dagger \Psi \\ \Psi^\dagger \Phi P & \Psi^\dagger \Psi \end{pmatrix} && (163)\\ &= \begin{pmatrix} P^\dagger G_\phi P & 0 \\ 0 & G_\psi \end{pmatrix}, && (164) \end{aligned}$$

where the off-diagonal blocks are zero since the states $|\psi_{\alpha^{(k)}}\rangle$ and $|\psi_p\rangle$ are orthogonal. We use the indices of G to distinguish between the Gram matrices of the states $|\phi_k\rangle$ and $|\psi_p\rangle$.

Let us then write down the blocks of $R^\dagger A R$,

$$R^\dagger A R = \begin{pmatrix} P^\dagger G_\phi^{(A)} P & P^\dagger G_{\phi\psi}^{(A)} \\ G_{\psi\phi}^{(A)} P & G_\psi^{(A)} \end{pmatrix}. \tag{165}$$

Combining both expressions, we finally get

$$A_{\mathcal{B}} = \begin{pmatrix} P^{-1} G_\phi^{-1} G_\phi^{(A)} P & P^{-1} G_\phi^{-1} G_{\phi\psi}^{(A)} \\ G_\psi^{-1} G_{\psi\phi}^{(A)} P & G_\psi^{-1} G_\psi^{(A)} \end{pmatrix}. \tag{166}$$

Since $A_{\mathcal{B}}$ is the matrix of A in an orthogonal basis, we know that for $i, j \leq m$,

$$\left[P^{-1} G_\phi^{-1} G_\phi^{(A)} P\right]_{ij} = \frac{\langle \psi_{\alpha^{(i)}} | A | \psi_{\alpha^{(j)}} \rangle}{\langle \psi_{\alpha^{(i)}} | \psi_{\alpha^{(i)}} \rangle}. \tag{167}$$

In particular, average values on a Ritz vector can be obtained as

$$\left[P^{-1} G_\phi^{-1} G_\phi^{(A)} P\right]_{ii} = \frac{\langle \psi_{\alpha^{(i)}} | A | \psi_{\alpha^{(i)}} \rangle}{\langle \psi_{\alpha^{(i)}} | \psi_{\alpha^{(i)}} \rangle}. \tag{168}$$

□

[11] We make the usual abuse of notation of denoting both the operator and its matrix in the canonical basis by the same symbol A.

C Variational principle for linear variational states

We derive the (normalized) variational principle for a linear variational state, and show that it is in fact equivalent to the unnormalized variational principal in this case. This form is more simple and convenient than the direct form of the normalized variational principle.

Proposition C.1. *Let $\{|\phi_k\rangle\}_k$ be a family of linearly independent states, and let $|\psi_\alpha\rangle$ be the linear variational state defined on this basis,*

$$|\psi_\alpha\rangle = \sum_k \alpha_k |\phi_k\rangle , \ \alpha \in \mathbb{C}^m. \tag{169}$$

Then, the time-dependent variational principle applied to $|\psi_\alpha\rangle$ for a Hamiltonian H results in the variational dynamics

$$\dot{\alpha} = -iG^{-1}G^{(H)}\alpha. \tag{170}$$

Proof. We recall that for a given parameter α, the time-dependent variational principle gives the variation of the parameter as a solution of the equation

$$S\dot{\alpha} = -iF, \tag{171}$$

where the quantum geometric tensor S and forces vector F are given in general by

$$S_{pp'} = \frac{\partial_{\theta_p^*} \langle\psi_\alpha| \, \partial_{\theta_{p'}} |\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle} - \frac{(\partial_{\theta_p^*} \langle\psi_\alpha|) \, |\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle} \frac{\langle\psi_\alpha| \, \partial_{\theta_{p'}} |\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle}, \tag{172}$$

$$F_p = \frac{(\partial_{\theta_p^*} \langle\psi_\alpha|) H \, |\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle} - \frac{\langle\psi_\alpha|H|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle} \frac{(\partial_{\theta_p^*} \langle\psi_\alpha|) \, |\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle}. \tag{173}$$

In the case of the linear variational state $|\psi_\alpha\rangle$, the derivatives are easily obtained as

$$\partial_{\theta_p} |\psi_\alpha\rangle = |\phi_p\rangle . \tag{174}$$

Plugging these derivatives into S and F, we get

$$S = \frac{1}{\alpha^\dagger G \alpha} \left[G - \frac{G\alpha\alpha^\dagger G}{\alpha^\dagger G \alpha} \right], \tag{175}$$

$$F = \frac{1}{\alpha^\dagger G \alpha} \left[G^{(H)}\alpha - \frac{\alpha^\dagger G^{(H)} \alpha}{\alpha^\dagger G \alpha} G\alpha \right], \tag{176}$$

where $G_{ij} = \langle\phi_i|\phi_j\rangle$ and $G^{(H)}_{ij} = \langle\phi_i|H|\phi_j\rangle$ are respectively the Gram matrix and projected Hamiltonian corresponding to the family $\{|\phi_k\rangle\}_k$.

We notice that we always have $S\alpha = 0$, such that S is not invertible. Thus, we cannot simply take $\dot{\alpha} = -iS^{-1}F$ as our variation of the parameter. The linear system has either no solution or has a solution space which is an affine space of dimension larger than 1. It is easy to see that one particular solution is $-iG^{-1}G^{(H)}\alpha$, and that the set of all solutions is thus given by

$$-iG^{-1}G^{(H)}\alpha + \ker S. \tag{177}$$

We now need to identify the kernel of S. Since G has full rank m and $G\alpha\alpha^\dagger G = G\alpha(G\alpha)^\dagger$ is a rank 1 matrix, then S can only have rank m or $m-1$. But we know that S is singular, so it must have rank $m-1$. As such, we have $\ker S = \mathrm{span}(\alpha)$ and the set of solutions to eq. (171) is

$$\left\{ -iG^{-1}G^{(H)}\alpha + \lambda\alpha, \ \lambda \in \mathbb{C} \right\} \tag{178}$$

for any $\lambda \in \mathbb{C}$.

Let us assume that we pick an arbitrary solution at all times, in such a way that $\lambda(t)$ is an integrable function with antiderivative $\Lambda(t)$. Then the resulting dynamics is given by

$$\alpha(t) = e^{-i\left(G^{-1}G^{(H)}+\Lambda(t)\right)}\alpha_0. \tag{179}$$

Since $\Lambda(t)$ is a scalar function, it doesn't change the resulting quantum state. The function

$$\alpha(t) = e^{-iG^{-1}G^{(H)}}\alpha_0 \tag{180}$$

describes the same quantum state as eq. (179), meaning that the choice of λ is irrelevant. We can thus always choose the solution with $\lambda = 0$. We conclude that in this case, the TDVP equation is equivalent to

$$\dot{\alpha} = -iG^{-1}G^{(H)}\alpha. \tag{181}$$

□

D Estimators of the Rayleigh matrix

In this section we explore in more details the various possible estimators of the Rayleigh matrix $G^{-1}G^{(H)}$. The two classes of estimators that we consider are the determinant state estimator from eq. (28) and the sum of states estimator from eqs. (39) and (40). One remark we can make concerning the sum of states estimator is that as the number of states m increases, the probability distribution it relies on becomes increasingly unadapted to the specific matrix elements it estimates. This might lead to an increase in the variance of this estimator. Compared to the determinant state estimator, however, it does have the advantage of being much cheaper to compute for a given number of samples (see table 2). We also note that both estimators reduce to the standard estimator of the energy when $m = 1$.

We now discuss various details about these two estimators that lead to different possible implementations, each of which we benchmark in fig. 6.

Keep or discard imaginary parts When estimating the energy of a single state using the local energy, the resulting estimate is complex and only becomes real in the limit of infinite samples. As such, the real part of the resulting estimate is taken. Similarly, the Rayleigh matrix is known to have real eigenvalues, but the estimators using a finite number of samples will produce estimates that have eigenvalues with a non-zero imaginary part. As such, we can transform the resulting estimate by discarding the imaginary part of its eigenvalues. In fig. 6, for all estimators, we examine both keeping and discarding the imaginary part of the eigenvalue in order to identify which variant is more efficient.

Numerical stability The sum of states estimator produces G and $G^{(H)}$ individually, such that they still need to be combined into $G^{-1}G^{(H)}$. The inversion of G can be especially ill-conditioned as the states $|\phi_k\rangle$ get close to being linearly dependent. This means that the Monte Carlo error incurred by estimating G stochastically is significantly amplified when taking its inverse. We therefore examine two methods to construct $G^{-1}G^{(H)}$. The first method consists in taking the inverse in high precision. The second method consists in taking the pseudo-inverse for a given singular value cut-off. We compare those two methods in fig. 6.

We note that the determinant state estimator suffers much less from the ill-conditionedness of G than the sum of states estimator. In the sum of states estimator, the matrix G that we invert is known up to Monte Carlo error ϵ_{MC}. If we denote the condition number of G by κ, the resulting error on G^{-1} is bounded by $\kappa\epsilon_{\text{MC}}$. However, in the case of the determinant state estimator, we instead invert $\Phi(\boldsymbol{s})$, which is known up to numerical precision $\epsilon_{\text{NP}} \sim 10^{-15}$. Assuming that the condition numbers of the matrices $\Phi(\boldsymbol{s})$ for the sampled configurations $\boldsymbol{s}$ are of similar magnitude to κ, the resulting error is instead bounded by $\kappa\epsilon_{\text{NP}}$. Since typically $\epsilon_{\text{NP}} \ll \epsilon_{\text{MC}}$, we see that the determinant state estimator is expected to be significantly more stable to large κ than the sum of states estimator.

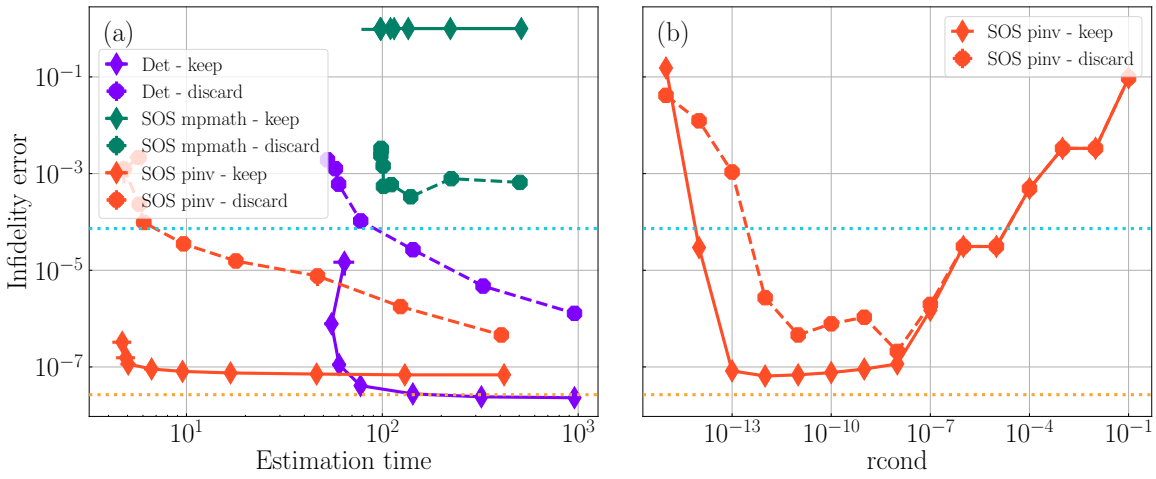

Figure 6: (a) Comparison between the final infidelity error of Bridge performed using various estimators of the Rayleigh matrix as a function of the estimation time of the Rayleigh matrix. The Hamiltonian is the transverse-field Ising Hamiltonian on a 4×4 lattice with $(h, J) = (1, 0.1)$, and the basis states are 136 CNN states obtained with p-tVMC from time 0 to 1.35 with $\delta = 0.01$. The Rayleigh matrix is estimated either with the determinant state (Det) estimator of eq. (28) or with the sum of states (SOS) estimator of eqs. (39) and (40) for various numbers of samples. For the sum of states, the inverse of G was either taken exactly with 400 digits of precision (mpmath) or using the pseudo-inverse (pinv) with an singular value cut-off of 10^{-11}. For each of the three methods, full lines show the performance when keeping the imaginary part of the eigenvalues of the Rayleigh matrix, and dashed lines show the performance when discarding them. The dotted horizontal orange line indicated the optimal performance that can be achieved in the subspace. In theory, it should be below all other points, but it is subject to numerical error. The dotted horizontal blue line indicates the performance of the original states $|\phi_k\rangle$. (b) Performance of Bridge when estimating the Rayleigh matrix with the sum of states estimator and inverting G with the pseudo-inverse for various choices of the singular value cut-off rcond. The estimation is performed with 10^5 samples which corresponds to the rightmost point of the corresponding curve in plot (a).

Numerical comparison In order to compare the performance of the two estimators and their variants, we could compare the resulting estimates of the Rayleigh matrix to the exact Rayleigh matrix. However, when the states $|\phi_k\rangle$ start being almost linearly dependent, the exact calculation of the Rayleigh matrix starts being numerically unstable. While the computation of $G^{-1}G^{(H)}$ from G and $G^{(H)}$ can be performed in high precision, the exact computation of G and $G^{(H)}$ alone must be performed in double precision, since the overhead associated to using unoptimized high-precision libraries like mpmath makes it unpractical to use in this case. The small numerical error accumulated in the computation of G and $G^{(H)}$ is then amplified when taking the ill-conditioned inverse of G.

Since the exact calculation of the Rayleigh matrix cannot be trusted, we instead use the performance of Bridge as a benchmark, where the exact dynamics used as the reference can be computed without significant numerical error. In figure 6, we show the performance of Bridge using the determinant state estimator, the sum of states estimator with high precision inversion of G and the sum of states estimator with inversion of G using the pseudo-inverse at various singular value cut-offs. For each, we either keep or discard the imaginary part of the eigenvalues of the resulting estimate of the Rayleigh matrix. For more details on the setting of the dynamics, see section 4. We conclude from this figure that the optimal estimator is either the sum of states estimator that keeps the imaginary part with an appropriately chosen singular value cut-off for a good performance and a very fast estimation (less than 10 seconds), or the determinant state estimator that keeps the imaginary part for the optimal performance at an increased cost (about 5 minutes).

In the context of Bridge, irrespective of the chosen method, the calculation time will be dominated by the generation of the original basis states $|\phi_k\rangle$. Therefore, we use the determinant state estimator that keeps the imaginary part of the eigenvalues, since it achieves the optimal final performance and does not require tuning the singular value cut-off.

We further show that the determinant state estimator is more suitable for Bridge than the pseudo-

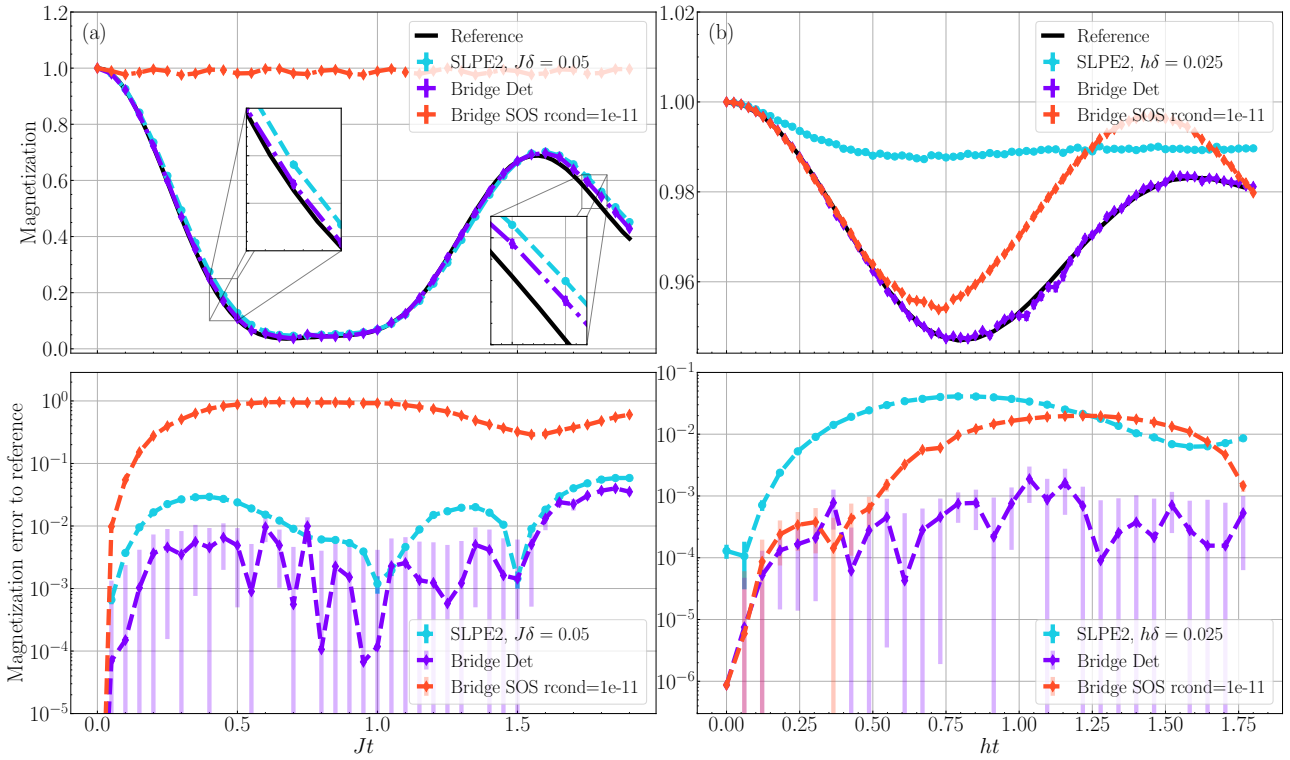

Figure 7: Magnetization on the 8×8 transverse-field Ising model with $h = 0.1h_c$ (a) and $h = 2h_c$ (b). The original states are vision transformers optimized with p-tVMC using the scheme SLPE2 with time step $J\delta = 0.05$ (a) and $h\delta = 0.025$ (b). The reference for (a) are vision transformers optimized with p-tVMC using the higher order scheme SLPE3 with a smaller time step $J\delta = 0.025$. The reference for (b) is convolutional neural network states obtained with t-VMC in Ref. [18], since obtaining good p-tVMC states in this regime would have been too computationally expensive. In both cases, Bridge is performed with either the determinant state estimator (Det) or sum of states estimator (SOS). For the sum of states estimator, G is inverted with the pseudo-inverse, with a singular value cut-off $\texttt{rcond} = 10^{-11}$.

inverse sum of states estimator with fig. 7 which is fig. 5 with an additional curve depicting Bridge performed with the pseudo-inverse sum of states estimator for a given value of the pseudo-inverse singular value cut-off. It confirms that the determinant state estimator also outperforms the sum of states estimator for larger system sizes. In particular, Bridge with the sum of states estimator is even less accurate than the original basis states themselves in the case (a). This is further aggravated by the fact that in both cases, the sum of states estimator ran for significantly longer. In all cases, Bridge was run on a single A100 GPU. With the determinant state estimator, it took 15 minutes for (a) and 60 minutes for (b). With the sum of states estimator, it took 150 minutes for both (a) and (b).

E Projected-time variational Monte Carlo

In this section, we summarize the core features of the projected-time variational Monte Carlo (p-tVMC) method, referring the reader to [20, 21] for a more comprehensive treatment. The main goal of p-tVMC is to numerically solve the time-dependent Schrödinger equation $|\psi(t+\delta)\rangle = e^{\Lambda\delta}|\psi(t)\rangle$ on a discrete time mesh with step size δ. We assume that the exact dynamics is well approximated by a variational state, meaning that there exist parameters $\theta(t)$ such that $|\psi(t)\rangle \approx |\psi_{\theta(t)}\rangle$. Using the McLachlan variational principle, the evolution problem is mapped onto an optimization problem in parameter space:

$$\theta(t+\delta) = \arg\max_{\eta} \mathcal{F}\left(|\psi_\eta\rangle, e^{-iH\delta}|\psi_{\theta(t)}\rangle\right), \tag{182}$$

with the fidelity

$$\mathcal{F}(|\psi\rangle, |\phi\rangle) = \frac{\langle\psi|\phi\rangle \langle\phi|\psi\rangle}{\langle\psi|\psi\rangle \langle\phi|\phi\rangle} = \mathbb{E}_{x\sim|\psi(x)|^2 \; y\sim|\phi(y)|^2} \left[\frac{\phi(x)}{\psi(x)} \frac{\psi(y)}{\phi(y)}\right] \tag{183}$$

measuring the discrepancy between two quantum states.

The p-tVMC approach leverages an expansion of the exponential map $e^{\Lambda\delta}$ to finite order in δ, yielding discretizations specifically suited for the optimization in eq. (182) [20, 21, 43, 55, 56]. These expansions typically assume product forms, such as

$$e^{\Lambda\delta} = \prod_{k=1}^{s} (\mathbb{1} - ia_k H\delta) + O\left(\delta^{s+1}\right), \quad \text{or} \quad e^{\Lambda\delta} = \prod_{k=1}^{s} (\mathbb{1} - ia_k H_1\delta)\, e^{-ib_k H_0\delta} + O\left(\delta^{o(s)+1}\right), \tag{184}$$

where H_0 is diagonal in the computational basis, and H_1 captures the off-diagonal part, with $H = H_0 + H_1$. Importantly, the number of substeps s scales polynomially in the order $o = o(s)$ of the approximation. These discretizations are known as the *Linear Product Expansion* (LPE) and the *Splitted Linear Product Expansion* (SLPE) respectively. The values of the coefficients a_k, b_k are determined semi-analytically and are tabulated in [21]. A key advantage of these expansions is their ability to achieve high-order temporal accuracy by approximately decomposing the single optimization step [eq. (182)] into a sequence of s smaller and efficiently computable optimization problems. Introducing the intermediate parameters $\theta^{(k)}$, with initial condition $\theta^{(0)} \equiv \theta(t)$ and final result $\theta^{(s)} \equiv \theta(t+\delta)$, the procedure becomes:

$$\theta^{(k)} = \arg\max_{\eta} \mathcal{F}\left(|\psi_\eta\rangle, U_k |\psi_{\theta^{(k-1)}}\rangle\right), \quad k \in \{1, \ldots, s\}, \tag{185}$$

where each operator U_k corresponds to a factor in the chosen product expansion [21, 43].

The error associated with the p-tVMC method arises from two distinct sources. Firstly, a finite *discretization error* remains due to truncation of the expansions of the exponential map [eq. (184)]. It can be systematically controlled and made arbitrarily small as the time step $\delta \to 0$, at the expense of an increased computational cost. Secondly, each optimization step introduces an *optimization error*, arising from finite sampling errors, limited expressivity of the variational ansatz, and challenging optimization landscapes. This error accumulates over successive steps and can significantly affect the overall accuracy. The optimization problems at each substep are typically solved using natural gradient descent combined with Monte Carlo approximations of the loss function and its gradient, as detailed in [20, 21].