Supplementary Material
Efficiency of neural quantum states in light of the quantum geometric tensor

Sidhartha Dash, Luca Gravina, Filippo Vicentini, Michel Ferrero, and Antoine Georges

SUPPLEMENTARY NOTE 1: THE AKLT STATE

The AKLT state, also known as a valence-bond solid [1], is the ground state of the spin-1 BLBQ chain for $\theta = \arctan(1/3)$:

$$H_{\text{AKLT}} = \sum_{i=1}^{N-1} J\left[\mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{3}(\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2\right], \tag{1}$$

where $\mathbf{S}_i = (S_{ix}, S_{iy}, S_{iz})$ is the spin-1 operator acting on site i. The AKLT state is exactly known and can be represented by expressing the spin-1 particle by two auxiliary spin-1/2 particles. The spin-1 computational basis can be represented in terms of the triplet states formed with two spin-1/2 particles:

$$|+\rangle = \psi_{11} = |\uparrow\uparrow\rangle, \tag{2}$$

$$|0\rangle = \psi_{12} = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) = \psi_{21}, \tag{3}$$

$$|-\rangle = \psi_{22} = |\downarrow\downarrow\rangle, \tag{4}$$

where $\{|+\rangle, |0\rangle, |-\rangle\}$ is the basis for the spin-1 system, and $\{|\uparrow\rangle, |\downarrow\rangle\}$ is the spin-1/2 basis. Then, the AKLT state is given by

$$\begin{aligned}\left|\Psi_{\text{AKLT}}(\alpha,\beta)\right\rangle = \frac{1}{2^{(N-1)/2}}\, &\psi_{\alpha\beta_1}\epsilon^{\beta_1\alpha_2}\psi_{\alpha_2\beta_2}\epsilon^{\beta_2\alpha_3}\cdots \\ &\psi_{\alpha_i\beta_i}\epsilon^{\beta_i\alpha_{i+1}}\cdots\psi_{\alpha_{N-1}\beta_{N-1}}\epsilon^{\beta_{N-1}\alpha_N}\psi_{\alpha_N\beta},\end{aligned} \tag{5}$$

where ϵ is the Levi-Civita tensor of rank 2, and repeated indices imply a summation. Note that the AKLT state Eq. (5) is written for the case of an open boundary condition, and has two free spin-1/2 variables α, β which correspond to the two outermost spin-1/2s on the chain. This leads to a four-fold degeneracy of the AKLT state. In the total $S_z = 0$ sector, the AKLT state is two-fold degenerate. It is interesting to note in the AKLT state that two adjacent spin-1 variables are never aligned ferromagnetically (both +s or both −s). This can be realized from Eq. (5), given that the Levi-Civita tensor only has off-diagonal terms. A broader consequence of this is that a + (−) can only be followed by a 0 or a − (+), i.e. a state of the form $|+0000000000+\rangle$ is not allowed. Whereas a typical state could take the form $|00+0--+000-0+00-0\rangle$, which essentially has a Néel order when we remove all 0s. This is a consequence of a non-local order, characterized by the *string order parameter*

$$O_{ij} = \left\langle S_{iz} e^{i\pi\sum_{i<k<j} S_{kz}} S_{jz} \right\rangle, \tag{6}$$

in the Haldane phase, which attains the maximum value at the AKLT point.

Furthermore, the AKLT Hamiltonian Eq. (1) can be written as a sum of projection operators into the spin-2 subspace for every two neighboring sites, $\hat{\text{P}}_{S=2}(i, i+1)$. The projection operator can be written by inspecting the eigenvalues of the operator $\hat{\text{X}} = (\mathbf{S}_i + \mathbf{S}_{i+1})^2$, $S(S+1) = 0, 2, 6$ for $S = 0$, $S = 1$, $S = 2$ subspaces respectively. Note that the eigenspaces of the operator $\hat{\text{X}}$, are shared with that of the projection operators. Therefore, we can write

$$\begin{aligned}\hat{\text{P}}_{S=2}(i,i+1) &= \frac{1}{24}\hat{\text{X}}(\hat{\text{X}} - 2) \\ &= \frac{1}{2}\mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{6}(\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2 + \frac{1}{3},\end{aligned} \tag{7}$$

where we have used the fact that $\hat{\text{X}} = \mathbf{S}_i^2 + \mathbf{S}_{i+1}^2 + 2\mathbf{S}_i \cdot \mathbf{S}_{i+1} = 4 + 2\mathbf{S}_i \cdot \mathbf{S}_{i+1}$. As a consequence of Eq. (7), we can write the AKLT Hamiltonian Eq. (1) as

$$H_{\text{AKLT}} = 2J \sum_{i=1}^{N-1} \hat{\text{P}}_{S=2}(i, i+1) - \frac{2J}{3}(N-1) \tag{8}$$

An interesting observation from Eq. (8) is that the AKLT state has no two adjacent spin-1s living in the $S = 2$ sector, so as to have the minimum energy. In the picture of the auxiliary spin-1/2 particles, this leads to the formation of singlet states ($S = 0$) between two spin-1/2 particles in the adjacent sites.

Additionally, the AKLT state can be expressed exactly as an MPS with the lowest non-trivial bond dimension $\chi = 2$ [2]:

$$\left|\Psi_{\text{AKLT}}(\alpha,\beta)\right\rangle = \sum_{|\sigma\rangle} A_{\alpha}^{(\sigma_1) i_1} A_{i_1}^{(\sigma_2) i_2} \cdots A_{i_{N-1}}^{(\sigma_N)\beta} |\sigma_1, \sigma_2, \cdots, \sigma_N\rangle, \tag{9}$$

where repeated indices imply a summation, $|\sigma_i\rangle$ denotes the local computational basis for the spin-1 particle at site i, with $\sigma_i = +, 0, -$, and

$$\mathbf{A}^{(+)} = \begin{bmatrix} 0 & \sqrt{\frac{2}{3}} \\ 0 & 0 \end{bmatrix},\ \mathbf{A}^{(0)} = \begin{bmatrix} -\frac{1}{\sqrt{3}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{bmatrix},\ \mathbf{A}^{(-)} = \begin{bmatrix} 0 & 0 \\ -\sqrt{\frac{2}{3}} & 0 \end{bmatrix}. \tag{10}$$

α, β in Eq. (9) are the free indices, corresponding to the two auxiliary spin-1/2s at the boundary for the case of an open-boundary condition. The bond dimension of the required MPS to represent the AKLT state is just one unit higher than that for the product states (which can be represented as MPSs with bond dimension $\chi = 1$), making it the simplest entangled quantum state.

L	Model	NQS spin-flip symmetry	Iterations
8, 10, 12	AFH ($\theta = 0$)	even parity	$15K$ ($L = 8,10$), $20K$ ($L = 12$)
	AKLT ($\theta = \arctan(1/3)$)	even parity	$15K$ ($L = 8,10$), $20K$ ($L = 12$)
	critical ($\theta = \pi/4$)	odd (even for $L = 12$)	$15K$ ($L = 8$), $15-20K$ ($L = 10,12$)
	critical ($\theta = \arctan(2)$)	odd (even for $L = 12$)	$15K$ ($L = 8,10$), $20K$ ($L = 12$)

Supplementary Table I. Details of the infidelity minimization procedure for the spin-1 BLBQ chain (with open boundary conditions) for lengths $L = 8$, 10, and 12. The NQS ansatz is given by the spin-1 RBM (Eq. 11 in the main text). The values shown in the table are used for computations with all values of α shown in figures 2, 3, and 4 in the main text. Note that all computations of the gradient, and the QGT were done by a full summation over the total $S_z = 0$ subspace of the Hilbert space. The learning rate, and the regularization parameter ϵ for the QGT matrix were chosen after a few trials for each case. We used learning rates in the range of 5×10^{-4} to 10^{-2}, and applied a learning rate decay strategy through the optimization process. We used the regularization parameter ϵ in the range of 10^{-8} to 10^{-6} for the Haldane phases, and from 10^{-6} to 10^{-4} for the critical phases, carefully selected to prevent divergence in the updates for each case. The regularization parameter ϵ is also updated in the course of the optimization for improved quality of solutions [3–6].

L	Model	NQS spin-flip symmetry	MC samples	Iterations
10	AFH ($\theta = 0$)	even parity	4096	$20K$
	AKLT ($\theta = \arctan(1/3)$)	even parity	4096	$20K$
	critical ($\theta = \pi/4$)	odd parity	4096	$20K$ ($45K$ for $\alpha = 12$)
	critical ($\theta = \arctan(2)$)	odd parity	4096	$20K$

Supplementary Table II. Details of the VMC optimization for the spin-1 BLBQ chain (with open boundary condition) with length $L = 10$. The NQS ansatz is given by the spin-1 RBM (Eq. 11 in the main text). The same values are used for the optimizations with all values of α (unless explicitly mentioned) shown in Supplementary Fig. 3. All computations were done with the symmetry constraint total $S_z = 0$. The learning rates were in the range of 10^{-3} to 10^{-2}, and we also let the learning rate decay through the optimization process. The regularization parameter ϵ was taken to be between 10^{-6} to 10^{-4} for all the phases. Moreover, the regularization parameter ϵ was also updated in the course of the optimization [3–6].

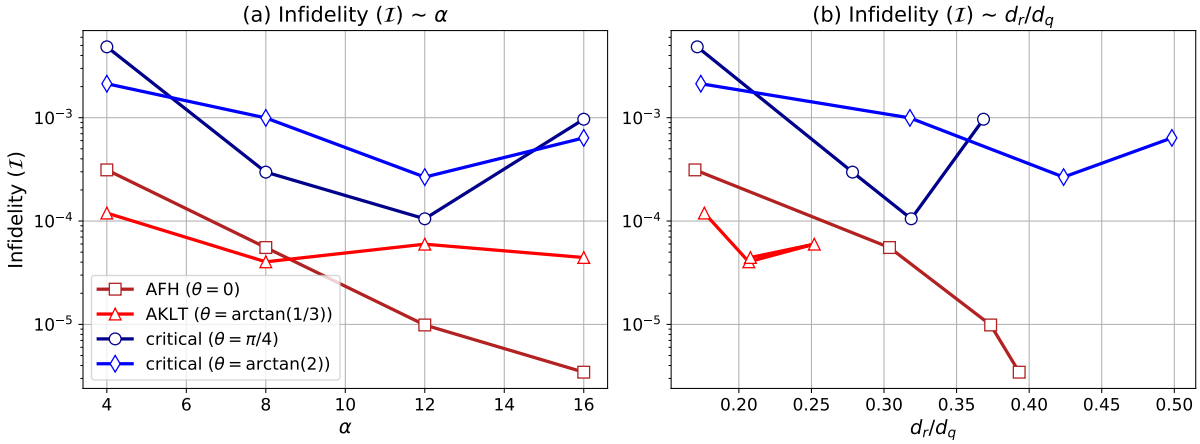
L	α	number of parameters
8	2	288
	4	560
	6	832
	8	1104
	10	1376
10	2	440
	4	860
	6	1280
	8	1700
	10	2120
	12	2540
	14	2960
	16	3380
	18	3800
	20	4220
	22	4640
12	2	624
	4	1224
	6	1824
	8	2424
	10	3024
	12	3624
	14	4224
	16	4824
	18	5424
	20	6024

Supplementary Table III. The number of parameters in the spin-1 RBM ansatz (Eq. 11 in the main text) for different sizes $L = 8,10,12$ of the BLBQ chain, and different hidden layer densities α.

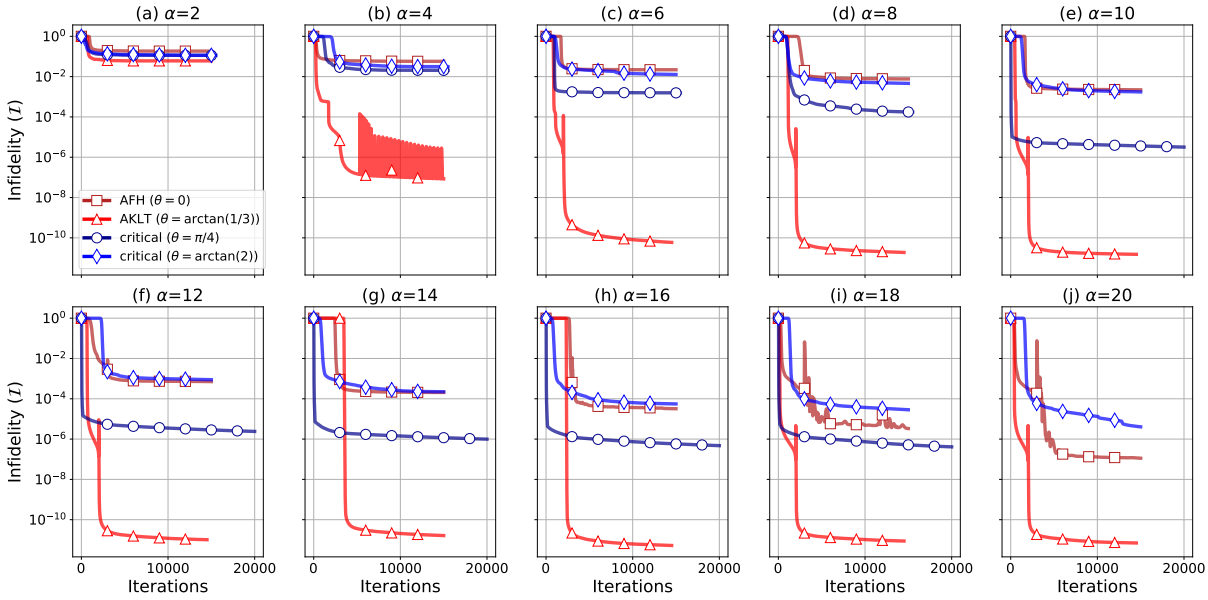
SUPPLEMENTARY NOTE 2: INFIDELITY OF THE NQS, AND QGT, FOR VMC COMPUTATIONS ON THE SPIN-1 BLBQ CHAIN WITH $L = 10$

We also perform VMC optimizations [7] for the spin-1 bilinear-biquadratic (BLBQ) chain of length $L = 10$ (with open boundary conditions), to supplement our results from the infidelity minimizations in the main text. The gradients and the QGT are estimated approximately by the Monte Carlo sampling procedure at each optimization step. However, we estimate the infidelity of the NQS w.r.t. the exact ground state (Eq. 12 in the main text), and the QGT at the end of the VMC optimization exactly by a full summation over the basis states. In Supplementary Fig. 1(a), we plot the NQS infidelity, after convergence of the VMC procedure, as a function of α. We observe that the infidelity of the NQS decreases as α increases, for smaller values of α. It continues to do so up to $\alpha = 16$ for the AFH phase, up to $\alpha = 8$ for the AKLT phase, and up to $\alpha = 12$ for the two critical phases. The infidelity for the AKLT phase saturates beyond $\alpha = 8$ with small oscillations, while the NQS approximation becomes worse for the two critical phases beyond $\alpha = 12$.

Furthermore, we plot the ratio of the rank of the QGT matrix d_r (defined by a cutoff 10^{-5}) w.r.t. the *effective*



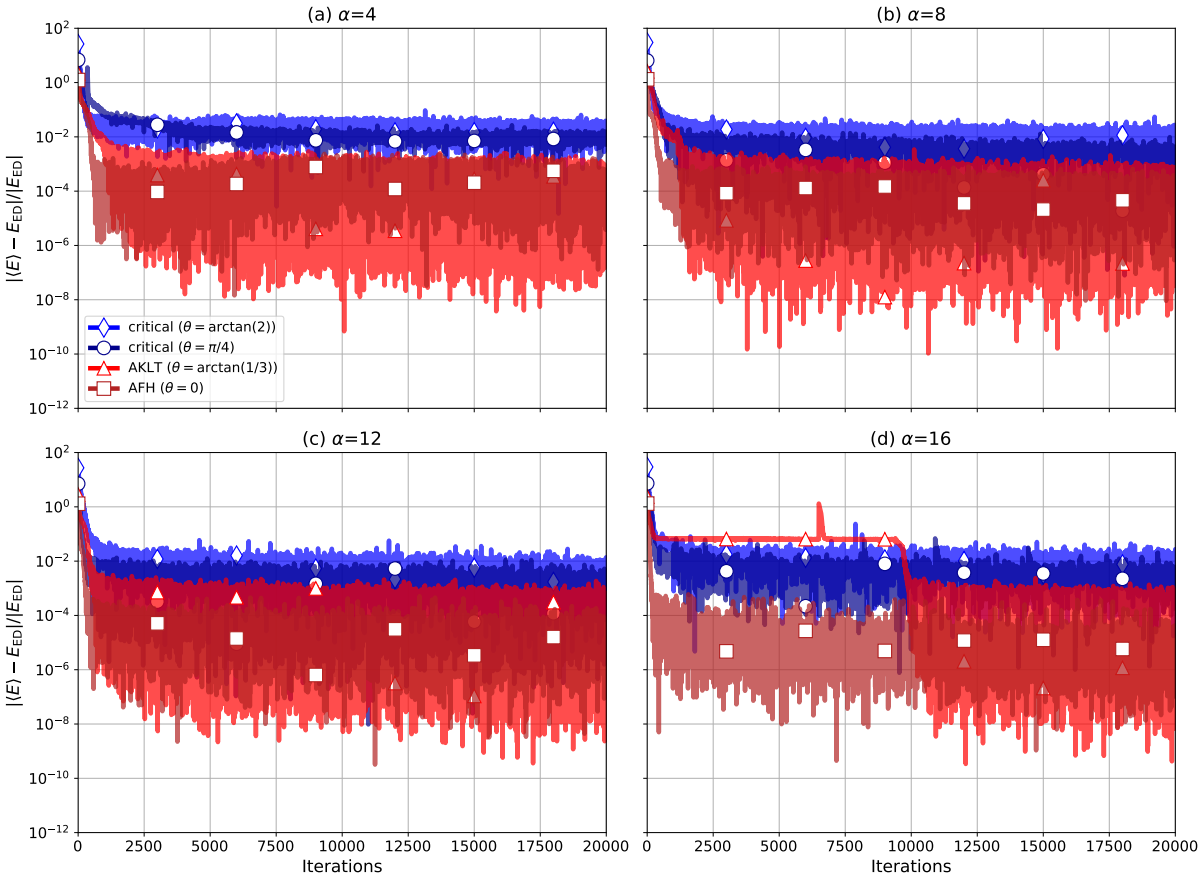
Supplementary Figure 1. Infidelity and QGT for VMC on the spin-1 BLBQ chain. The infidelity (I), after convergence of the VMC procedure for the spin-1 BLBQ chain with length $L = 10$, is shown in panel (a) as a function of α, and in panel (b) as a function of the ratio of the QGT rank d_r (defined by a cutoff 10^{-5}) w.r.t. d_q. Results are shown for four phases of the spin-1 BLBQ chain (corresponding to the four marked points in Fig. 1 in the main text). Note that the infidelity and the QGT are measured exactly at the end of the VMC optimization.



Supplementary Figure 2. Evolution of the infidelity (I) of the NQS ansatz w.r.t. the exact ground state during the infidelity minimization procedure. The infidelity (I) is plotted (in $\log_{10}$ scale) against the minimization steps for the AFH, AKLT and the two critical phases of the spin-1 BLBQ chain with open boundary conditions for $L = 10$. The panels (a)-(j) show the optimization curves for different hidden layer densities $\alpha = 2, 4, 6, \cdots, 20$ of the spin-1 RBM (Eq. 11 in the main text). The markers on the curves are shown every 3000 iterations to improve readability.

quantum dimension d_q in Supplementary Fig. 1(b). We observe that an initial increase in α leads to an increase in the ratio d_r/d_q, while the infidelity of the NQS decreases. This indicates a correlation between d_r/d_q and the quality of the NQS approximation for smaller values of α, as also observed for the case of infidelity minimizations (see Fig. 3 in the main text). While, for larger values of α, a larger value of d_r/d_q does not necessarily lead to a better solution. This suggests that the NQS optimization is not able to utilize the additional *relevant directions*, that come with increasing α, to improve the accuracy of the NQS for larger values of α.

We also plot the spectra of the QGT (Eq. 6 in the main text), as histograms in Supplementary Fig. 5, at the end of the VMC procedure for the four phases of the spin-1 BLBQ chain with length $L = 10$, along with the cumula-



Supplementary Figure 3. Evolution of the energy calculated with the NQS during the VMC procedure. The relative error in the energy calculated with the NQS w.r.t. the exact diagonalization result, i.e. $|\langle E\rangle - E_{\rm ED}|/|E_{\rm ED}|$, is plotted (in $\log_{10}$ scale) as a function of the VMC steps, for the AFH, AKLT and the two critical phases of the spin-1 BLBQ chain (with open boundary conditions) for L=10. Panels (a), (b), (c), and (d) correspond to the hidden layer densities $\alpha = 4$, 8, 12, and 16 respectively.

tive distributions in the respective insets. The histograms indicate that increasing α leads to an increase in the number of QGT eigenvalues with moderate magnitudes. This is a bit different from the QGT spectra extracted from the converged infidelity minimizations (see Fig. 4 in the main text), where we see a significant increase in the number of large eigenvalues with an increase in α. This may explain why the improvement of the NQS approximation eventually stops as α increases for the VMC optimizations, even when the ratio d_r/d_q increases.

SUPPLEMENTARY NOTE 3: DETAILS OF THE INFIDELITY OPTIMIZATION AND VMC

We show the infidelity minimization curves in Supplementary Fig. 2, for various phases of the spin-1 BLBQ chain with length $L = 10$. The hyperparameters of the optimization are given in Supplementary table I.

We impose global spin-flip symmetries on the NQS ansätze, as prescribed in ref. [8]. For the AFH ($\theta = 0$) and the AKLT ($\theta = \arctan(1/3)$) states, we impose a global spin-flip symmetry with an even parity,

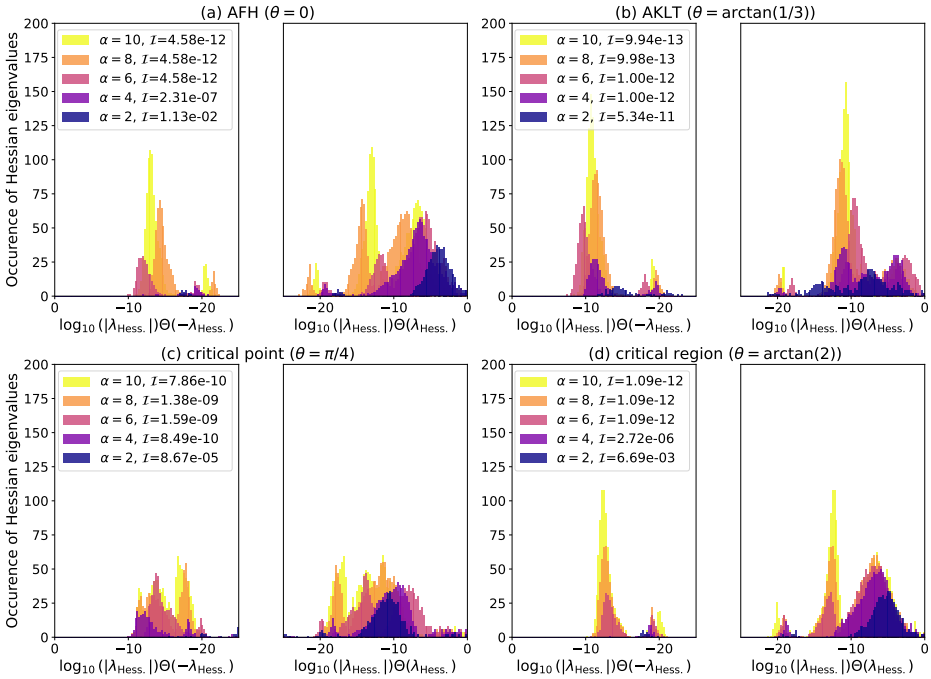
$$\psi_\theta^{(+)}(\sigma) = \frac{1}{2}\left[\psi_\theta(\sigma) + \psi_\theta(-\sigma)\right], \tag{11}$$

while for the two critical phases, we impose the global spin-flip symmetry with an odd parity,

$$\psi_\theta^{(-)}(\sigma) = \frac{1}{2}\left[\psi_\theta(\sigma) - \psi_\theta(-\sigma)\right], \tag{12}$$

for lengths $L = 8$, and 10. Note that we use a global spin-flip symmetry with an even parity for all four phases for $L = 12$. The correct symmetry sector for a given phase and chain length was determined by examining the exact diagonalization (ED) solution. The above symmetry prescription was used for both infidelity minimization and VMC procedures.

Additionally, we also show the optimization curves for the VMC procedure for various phases of the spin-1 BLBQ chain with $L = 10$ in Supplementary Fig. 3(a)-(d). The hyperparameters of the computations are given in Supplementary table II. Furthermore, the number of parameters in the NQS ansätze for different values of L and α are listed in Supplementary table III.



Supplementary Figure 4. Spectra of the Hessian for infidelity minimization on the spin-1 BLBQ chain. The distribution of the normalized eigenvalues (w.r.t. the maximum eigenvalue) of the Hessian is shown after convergence of the infidelity minimization on the spin-1 BLBQ chain (with open boundary conditions) for $L = 8$. The eigenvalues are shown in $\log_{10}$ scale. Results are shown for the AFH, AKLT, and the two critical phases of the spin-1 BLBQ chain in subplots (a), (b), (c), and (d) respectively. The left panel of each subplot shows the distribution of the negative eigenvalues (magnitudes), and the right panel of each subplot shows the distribution of the positive eigenvalues of the Hessian. In each subplot, the x-axis of the right panel increases from left to right, and that of the left panel increases from right to left.

To compute the parameter update $\delta\theta$ predicted by natural gradient descent (NGD), we need to solve the linear inverse problem $\mathbf{G}\delta\theta = \boldsymbol{F}$. The most popular method for estimating $\delta\theta$ is the least squares method, according to which

$$\delta\theta = \underset{x}{\operatorname{argmin}} \|\mathbf{G}x - \boldsymbol{F}\|^2. \tag{13}$$

The unbiased solution to this problem reads $\delta\theta = \mathbf{G}^{-1}\boldsymbol{F}$. It is often the case, however, that the matrix $\mathbf{G}$ is ill-conditioned, rank-deficient, or of ill-determined rank. The computation of a meaningful approximate solution of the linear system, therefore, in general requires that the system be replaced by a nearby system that is less sensitive to perturbations. This replacement is referred to as *regularization* [9]. Tikhonov regularization is one of the oldest and most popular regularization methods. The standard formulation of Tikhonov regularization [10] includes an L_2 regularization into Eq. (13) which reads

$$\delta\theta_\epsilon = \underset{x}{\operatorname{argmin}} \|\mathbf{G}x - \boldsymbol{F}\|^2 + \epsilon\|x\|^2. \tag{14}$$

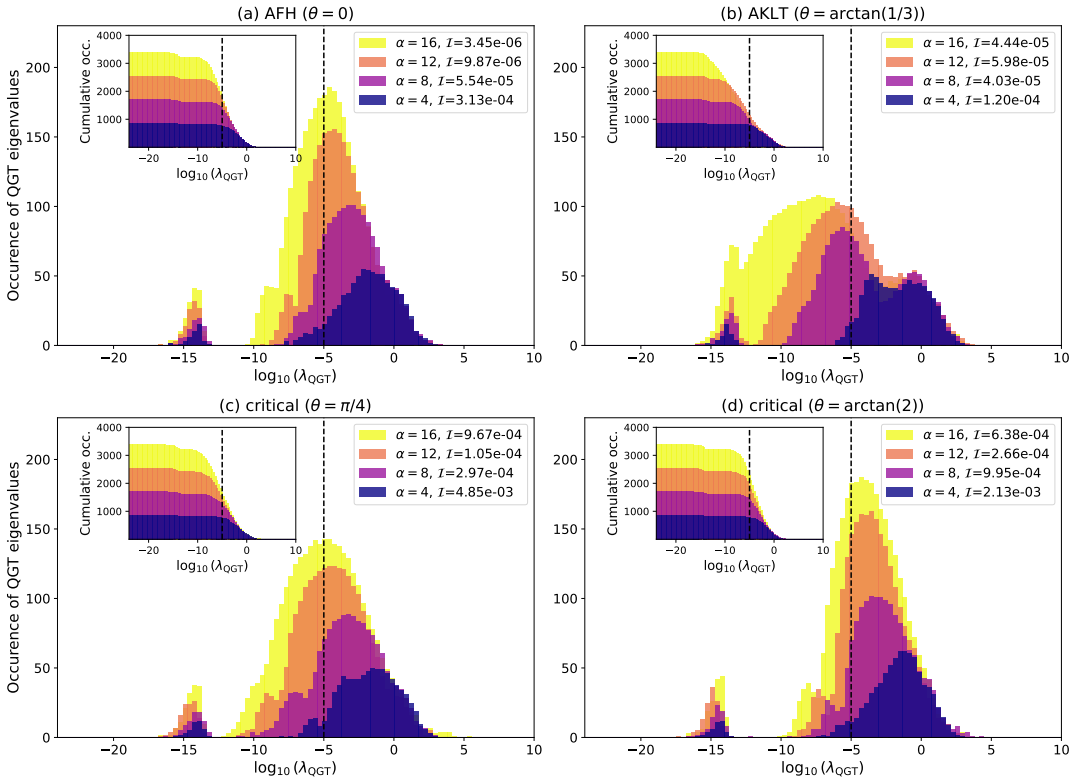
This equation admits the closed-form solution

$$\delta\theta_\epsilon = (\mathbf{G} + \epsilon\mathbb{1})^{-1}\boldsymbol{F}. \tag{15}$$

Note that not all elements of this one-parameter family of ridge estimates $\delta\theta_\epsilon$ are viable solutions to the ubiased problem in Eq. (13). The problem of selecting the optimal regularization parameter ϵ has been studied for decades, with no universal approach having been identified [11]. Classical methods such as the Morozov discrepancy principle, the L-curve method [3–6], and generalized cross-validation [9] define somewhat heuristic criteria to select ϵ. In this work we adopt the L-curve heuristic which we compute using the algorithm in Ref. [12].

The L-curve method is a graphical tool for selecting the regularization parameter ϵ in ill-posed or ill-conditioned inverse problems, most notably in Tikhonov regularization. Choosing ϵ too small can lead to under-regularization, resulting in solutions that overfit noise, exhibit large norms, or become highly oscillatory, often causing exploding gradient problems manifesting as exponentially large values of the solution norm $n(\epsilon) = \|\delta\theta_\epsilon\|^2$. Conversely, overly large values of ϵ over-smooth the solution, weakening the effectiveness of natural gradient strategies and increasing the residual error, defined as $R(\epsilon) = \|\mathbf{G}\delta\theta_\epsilon - \boldsymbol{F}\|^2$ [4].

The L-curve is a parametric plot of $(\log R(\epsilon), \log n(\epsilon))$. Its characteristic "L" shape captures the tradeoff between fitting the data and applying sufficient regularization. The optimal regularization parameter ϵ_{opt} is identified as the point of maximum positive curvature, i.e. the "corner" of the curve, balancing these competing effects.



Supplementary Figure 5. Spectra of the QGT for VMC on the spin-1 BLBQ chain. The distribution of the eigenvalues of the QGT is shown after convergence of the VMC optimization on the spin-1 BLBQ chain (with open boundary conditions) with length $L = 10$ for the (a) AFH, (b) AKLT, and (c,d) the two critical phases in the BLBQ phase diagram (corresponding to the four marked points in Fig. 1 in the main text). The eigenvalues are shown in $\log_{10}$ scale. Note that while the VMC procedure involves MC sampling for computing the energies, gradients, and the QGT to implement the stochastic reconfiguration method, we compute the infidelity and the QGT at the end of the optimization exactly by full summation over the total $S_z = 0$ subspace of the Hilbert space. The insets show the cumulative distribution of the eigenvalues (number of eigenvalues exceeding the value on the x-axis) for each case. The dashed line marks the cutoff 10^{-5}, used to compute the rank of the QGT d_r.

To efficiently locate this corner, we adopt the algorithm from Ref. [12]. The method involves computing solutions for a range of ϵ values, typically logarithmically spaced between suitable lower and upper bounds. Crucially, the necessary points $P(\epsilon)$ can be obtained at no additional computational cost beyond a single NGD iteration by leveraging the singular value decomposition used in computing the natural gradient. The code used for our simulations is provided in the GitHub repository [13].

Our later investigations have shown that the L-curve is a suboptimal method for selecting ϵ within NGD. Alternative approaches, specifically designed for NGD strategies and outperforming the present approach, have been recently explored in Ref. [14].

SUPPLEMENTARY NOTE 4: HESSIAN OF THE INFIDELITY LOSS FUNCTION AND ITS RELATION TO QGT

The loss function (Eq. 12 in the main text) is a scalar real valued function from $\mathbb{C}^N$ to $\mathbb{R}$, $\mathscr{L} : \mathbb{C}^N \to \mathbb{R}$. It is straightforward to see that the loss function depends on both θ and θ^*, and hence is non-holomorphic. For convenience while working with complex derivatives of non-holomorphic functions, we use the following notations [15]:

$$\boldsymbol{\theta}_c = \begin{bmatrix} \theta \\ \theta^* \end{bmatrix} \ , \ \frac{\partial}{\partial \boldsymbol{\theta}_c} = \begin{bmatrix} \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \theta^*} \end{bmatrix} \ ; \ \theta, \theta^* \in \mathbb{C}^N. \tag{16}$$

Then, we can write the complex Hessian of the loss function, following ref. [15], as

$$\mathbb{H} = \left(\frac{\partial}{\partial \boldsymbol{\theta}_c} \right)^\dagger \frac{\partial \mathscr{L}}{\partial \boldsymbol{\theta}_c} \tag{17}$$

$$= \begin{bmatrix} \frac{\partial^2 \mathscr{L}}{\partial \theta^* \partial \theta} & \frac{\partial^2 \mathscr{L}}{\partial \theta^* \partial \theta^*} \\ \frac{\partial^2 \mathscr{L}}{\partial \theta \partial \theta} & \frac{\partial^2 \mathscr{L}}{\partial \theta \partial \theta^*} \end{bmatrix}. \tag{18}$$

The generic elements of the blocks (1,1) and (1,2) in the above equation, when the loss function is given by the infidelity (Eq. 12 in the main text), are given by

$$\frac{\partial^2 \mathscr{L}}{\partial \theta_i^* \partial \theta_j} = -\frac{\left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \Omega \right\rangle \left\langle \Omega \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle} + \frac{\langle \psi_\theta | \Omega \rangle \left\langle \Omega \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \psi_\theta \right\rangle + \frac{\left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \Omega \right\rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle$$
$$+ \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle - 2 \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^3 \langle \Omega | \Omega \rangle} \left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \psi_\theta \right\rangle, \tag{19}$$

$$\frac{\partial^2 \mathscr{L}}{\partial \theta_i^* \partial \theta_j^*} = -\frac{\left\langle \frac{\partial^2 \psi_\theta}{\partial \theta_i \partial \theta_j} \middle| \Omega \right\rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle} + \frac{\left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \Omega \right\rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \psi_\theta \right\rangle + \frac{\left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \Omega \right\rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle} \left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \psi_\theta \right\rangle$$
$$+ \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \left\langle \frac{\partial^2 \psi_\theta}{\partial \theta_i \partial \theta_j} \middle| \psi_\theta \right\rangle - 2 \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^3 \langle \Omega | \Omega \rangle} \left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \psi_\theta \right\rangle \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \psi_\theta \right\rangle \tag{20}$$

respectively. The blocks (1,1) and (2,2) are complex conjugates of each other, and are both Hermitian. The blocks (1,2) and (2,1) are Hermitian conjugates (as well as complex conjugates) of each other. As a result, the complex Hessian matrix Eq. (18) is Hermitian.

It is interesting to note that when we are at the minimum of the infidelity landscape, i.e. when $|\Omega\rangle = |\psi_\theta\rangle$,

$$\frac{\partial^2 \mathscr{L}}{\partial \theta_i^* \partial \theta_j} = \frac{\left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle}{\langle \psi_\theta | \psi_\theta \rangle} - \frac{\left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \psi_\theta \right\rangle \left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle}{\langle \psi_\theta | \psi_\theta \rangle^2} = \mathrm{G}_{ij},$$

which is the quantum geometric tensor (QGT), and

$$\frac{\partial^2 \mathscr{L}}{\partial \theta_i^* \partial \theta_j^*} = \frac{\partial^2 \mathscr{L}}{\partial \theta_i \partial \theta_j} = 0.$$

In this case, the Hessian becomes

$$\mathbb{H}|_{|\Omega\rangle = |\psi_\theta\rangle} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^* \end{bmatrix}, \tag{21}$$

Hence, at the minimum of the infidelity landscape, the eigenvalues of the Hessian (of the infidelity loss function) are the same as that of the quantum geometric tensor (QGT), but with a degeneracy 2.

SUPPLEMENTARY NOTE 5: SPECTRA OF THE HESSIAN FOR INFIDELITY OPTIMIZATION ON THE SPIN-1 BLBQ CHAIN WITH $L = 8$

We plot the spectra of the Hessian (Eq. (18)) as histograms in Supplementary Fig. 4, at the end of the infidelity minimization for the four phases of the spin-1 BLBQ chain (Fig. 1 in the main text) with length $L = 8$. Eigenvalues are normalized w.r.t. the maximum eigenvalue. The positive and negative eigenvalues are shown in the right and left panels respectively in each subplot of Supplementary Fig. 4, for clarity on the nature of the landscape around the solution.

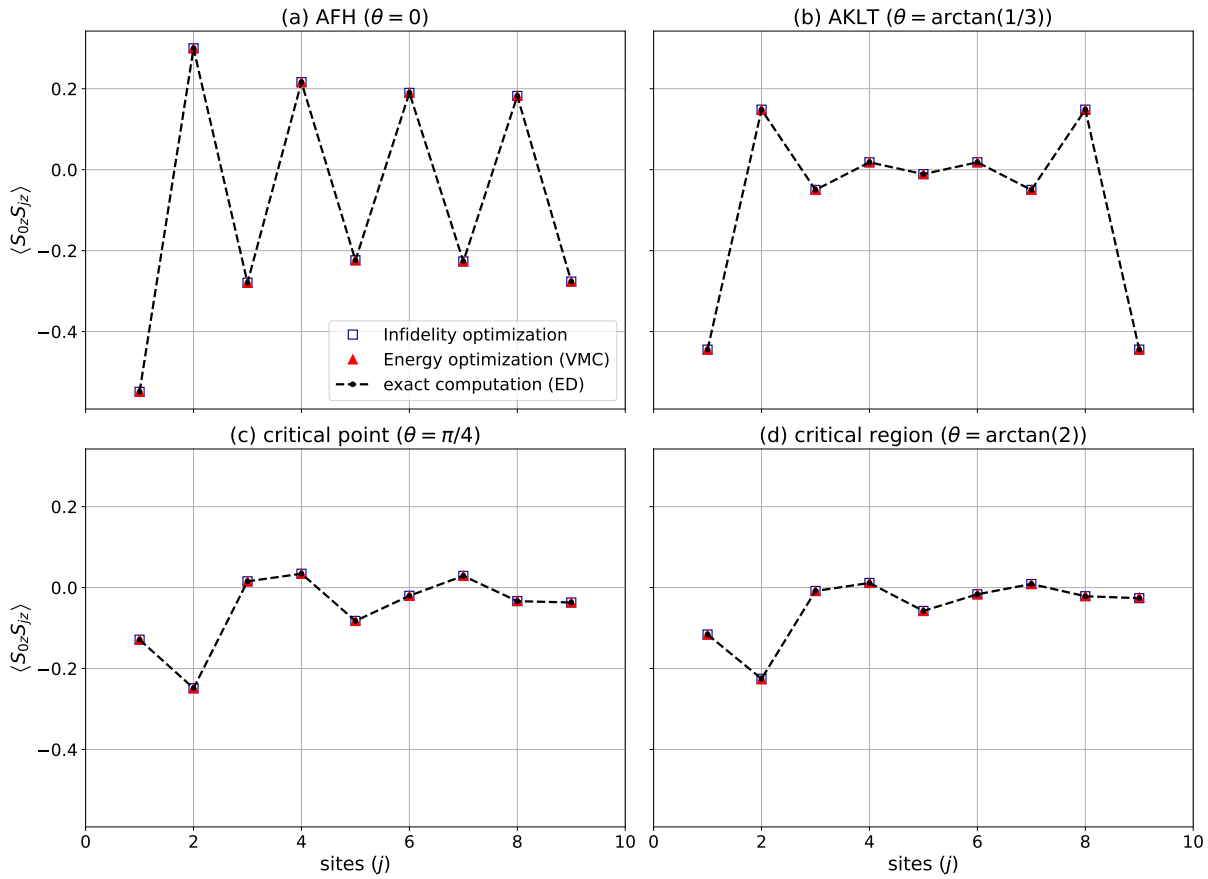
We observe that the positive eigenvalues dominate around the converged NQSs, for all the phases that we studied. The magnitudes of the negative eigenvalues are smaller than the largest positive eigenvalue at most by a factor $\sim 10^{-6}$. This suggests that we have converged reasonably in a valley with steep positive curvatures along most directions, and only a few almost flat directions with very small negative curvatures. This holds for all values of α that we studied, $\alpha = 2$, 4, 6, 8, and 10.

SUPPLEMENTARY NOTE 6: SPIN-SPIN CORRELATION FUNCTION

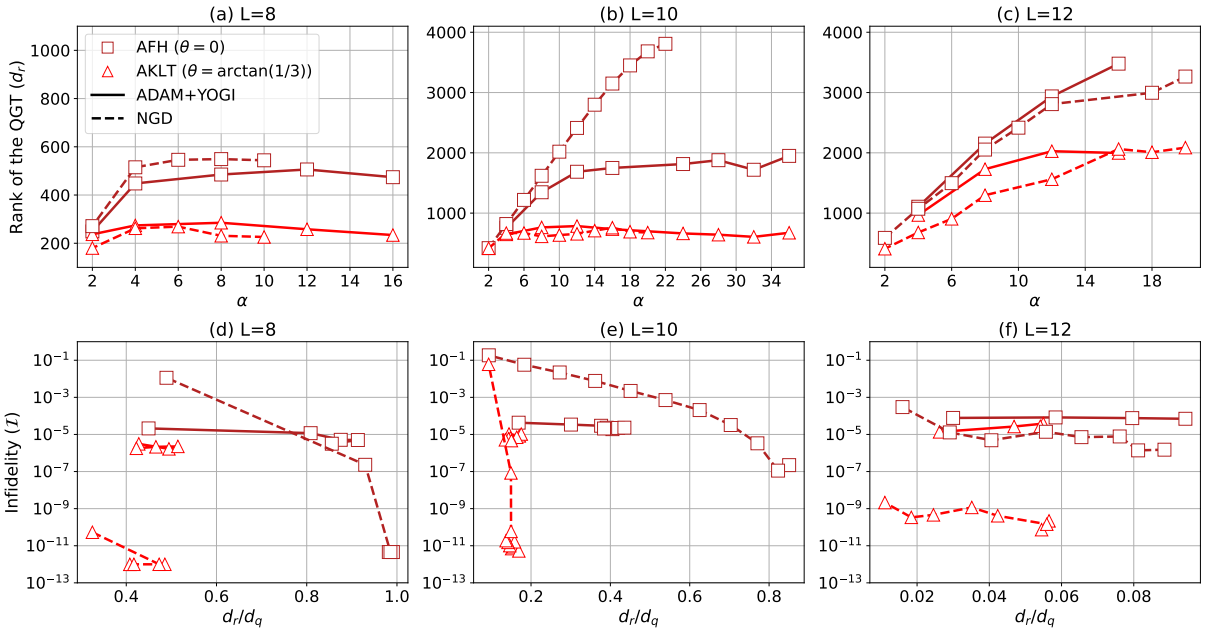
We plot the spin-spin correlation function in Supplementary Fig. 6, computed from the optimized NQSs for the different phases of the spin-1 BLBQ chain with length $L = 10$. The correlation functions for the AFH and AKLT phases show a modulation with wave-vector $k = \pi$. This is different from that in the critical phases, where the modulation wave-vector $k \sim 2\pi/3$. Note the enhancement of the spin-spin correlation $\langle S_{0z} S_{jz} \rangle$ at the end of the chain in the AKLT state, despite the open boundary condition. This is due to the topological string order in the Haldane phase [16], which is strongest for the AKLT state.

SUPPLEMENTARY NOTE 7: COMPARISON OF TWO DIFFERENT OPTIMIZATION SCHEMES

We also performed the infidelity minimization with a different algorithm for the AFH and the AKLT phases combining ADAM [17] and YOGI [18] (see caption of Supplementary Fig. 7 for details), but the NQS approximations were generally much worse for large network widths compared to that with the natural gradient descent (NGD) optimization used in the main text. We show the rank of the QGT as a function of α after convergence of the infidelity minimization in Supplementary Fig. 7(a)-(c) for both optimization schemes. In addition, we show the converged infidelities as a function of the ra-



Supplementary Figure 6. The real space spin-spin correlation from infidelity minimization, VMC, and exact diagonalization on the spin-1 BLBQ chain. The spin-spin correlation function $\langle S_{0z} S_{jz} \rangle$ is shown as a function of the site j, for the (a) AFH, (b) AKLT, and (c, d) the two critical phases of the spin-1 BLBQ chain with length $L = 10$ (with open boundary conditions). The correlation functions are computed from the optimized NQSs with hidden layer density $\alpha = 16$ for the case of infidelity minimization, and VMC, and are compared with that from exact diagonalization (ED). Note that, while, the VMC procedure involves an MC sampling, we compute the correlation functions at convergence by a full summation over the total $S_z = 0$ subspace of the Hilbert space.



Supplementary Figure 7. Performance comparison of two optimization schemes: ADAM+YOGI and natural gradient descent (NGD). (a)-(c) show the rank of the QGT as a function of the hidden layer density α after convergence of the infidelity minimization procedure with two different optimization schemes i) ADAM+YOGI (solid lines), and ii) NGD (dashed lines), for the AFH and the AKLT phase. (d)-(f) show the converged infidelity as a function of the ratio d_r/d_q for the two optimization schemes. The rank d_r is defined by the number of QGT eigenvalues greater than the cutoff 10^{-5} for both cases. i) For the infidelity minimization with ADAM+YOGI, first 3000 steps were performed with the ADAM algorithm [17] with a learning rate 5×10^{-4}, and then 117000 minimization steps were performed with the YOGI algorithm [18] with a learning rate 3×10^{-4}. The optimization was performed by an exact summation over the total $S_z = 0$ subspace, and the NQS ansatz, given by the spin-1 RBM (Eq. 11 in the main text), has the same symmetry (Eq. 11) as used in the main text. ii) The details of the NGD optimization are given in Supplementary Note 3.

tio d_r/d_q in Supplementary Fig. 7(d)-(f) for both optimization schemes.

We observe that the rank of the QGT saturates with α for the optimizations with ADAM+YOGI, as also seen for optimizations with NGD. However, in contrast to the NGD optimizations, the converged infidelity with the ADAM+YOGI scheme decreases very slowly with the rank of the QGT (or d_r/d_q), even increasing slowly in some cases, for example for the AKLT phase with $L = 12$. The superior performance of the NGD optimization arises from preconditioning the gradient with the inverse of the QGT (Eq. (15)). This makes the local updates curvature-aware, which makes it less likely to get stuck in local minimas or saddle points. Consequently, the only limitation of the NGD comes from the rank deficiency of the QGT matrix. Therefore, we use the NGD optimization in the main text.

SUPPLEMENTARY REFERENCES

[1] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Valence bond ground states in isotropic quantum antiferromagnets, Communications in Mathematical Physics **115**, 477 (1988).

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

[3] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, Tikhonov regularization and the l-curve for large discrete ill-posed problems, Journal of Computational and Applied Mathematics **123**, 423–446 (2000).

[4] P. Hansen and D. T. U. I. for Matematisk Modellering, *The L-curve and Its Use in the Numerical Treatment of Inverse Problems*, IMM-REP (IMM, Department of Mathematical Modelling, Technical Universityof Denmark, 1999).

[5] P. C. Hansen and D. P. O'Leary, The use of the l-curve in the regularization of discrete ill-posed problems, SIAM Journal on Scientific Computing **14**, 1487–1503 (1993).

[6] P. C. Hansen, Analysis of discrete ill-posed problems by means of the l-curve, SIAM Review **34**, 561–580 (1992).

[7] A. Dawid, J. Arnold, B. Requena, A. Gresch, M. Płodzień, K. Donatella, K. A. Nicoli, P. Stornati, R. Koch, M. Büttner, *et al.*, Modern applications of machine learning in quantum sciences, arXiv preprint arXiv:2204.04198 (2022).
[8] Y. Nomura, Helping restricted boltzmann machines with quantum-state representation by restoring symmetry, Journal of Physics: Condensed Matter **33**, 174003 (2021).
[9] G. H. Golub, M. Heath, and G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics **21**, 215 (1979).
[10] D. Gerth, A new interpretation of (tikhonov) regularization, Inverse Problems **37**, 064002 (2021).
[11] C. Bi, M. Y. Ou, M. Bouhrara, and R. G. Spencer, Span of regularization for solution of inverse problems with application to magnetic resonance relaxometry of the brain, Scientific Reports **12**, 10.1038/s41598-022-22739-3 (2022).
[12] A. Cultrera and L. Callegaro, A simple algorithm to find the l-curve corner in the regularisation of ill-posed inverse problems, IOP SciNotes **1**, 025004 (2020).
[13] S. Dash and L. Gravina, Efficiency of neural quantum states in light of the quantum geometric tensor (2024).
[14] L. Gravina, V. Savona, and F. Vicentini, Neural projected quantum dynamics: a systematic study, arXiv preprint arXiv:2410.10720 (2024).
[15] K. Kreutz-Delgado, The complex gradient operator and the cr-calculus, arXiv preprint arXiv:0906.4835; see page 14, 35 (2009).
[16] F. Pollmann, E. Berg, A. M. Turner, and M. Oshikawa, Symmetry protection of topological phases in one-dimensional quantum spin systems, Physical review b **85**, 075125 (2012).
[17] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[18] M. Zaheer, S. Reddi, D. Sachan, S. Kale, and S. Kumar, Adaptive methods for nonconvex optimization, Advances in neural information processing systems **31** (2018).