Predicting properties of quantum systems by regression on a quantum computer

Konstantin Antipin (joint work with Andrey Kardashin, Yerassyl Balkybek, Vladimir Palyulin)

October 4, 2025

Regression problem

Task: given a training set

$$\mathcal{T} = \left\{(\rho_{\alpha_j}, \alpha_j)\right\}_{j=1}^{T}, \tag{1}$$

learn to predict the label $\alpha \in \mathbb{R}$ of a given ρ_α.

Let ρ_α be a labeled *quantum state* with the label α.

The connection between ρ_α and α could be such that:

- ρ_α is the output state of a parametrized channel Φ_α acting on some input ρ, e.g., the single-qubit depolarizing channel

$$\rho_\alpha \equiv \Phi_\alpha[\rho] = (1-\alpha)\rho + \frac{\alpha}{2}\mathbb{1}$$

- $\rho_\alpha = |\psi_\alpha\rangle\langle\psi_\alpha|$ is such that $|\psi_\alpha\rangle$ is the ground state of a parametrized Hamiltonian H_α
- α is an entanglement measure of ρ_α

Methods

Expectation as the prediction

For a given ρ_α, our prediction a for α is the expectation

$$\mathsf{a}(\rho_\alpha, H) = \mathsf{Tr}\, H\rho_\alpha. \tag{2}$$

We want to find an H giving a prediction a as close to α as possible.

The observable H can be represented by a spectral decomposition

$$H(\mathsf{x}, \Pi) = \sum_i x_i \Pi_i, \tag{3}$$

where x_i are the eigenvalues with the eigenprojectors Π_i.

Preparing eigenprojectors

For a given ρ_α, our prediction a for α is the expectation

$$\mathsf{a}(\rho_\alpha, H) = \mathsf{Tr}\, H\rho_\alpha, \tag{4}$$

$$H(\mathsf{x}, \Pi) = \sum_i \mathsf{x}_i \Pi_i. \tag{5}$$

The eigenprojectors Π_i can be constructed as

$$\Pi_i(\boldsymbol{\theta}) = U_{\boldsymbol{\theta}}^\dagger \left|i\right\rangle\!\left\langle i\right| U_{\boldsymbol{\theta}}. \tag{6}$$

where $\{|i\rangle\}_i$ is the computational basis, and $U_{\boldsymbol{\theta}}$ is a unitary operator parametrized by $\boldsymbol{\theta} \subset \mathbb{R}$.

Prediction

For a given ρ_α, our prediction a for α is the expectation

$$\mathsf{a}(\rho_\alpha, H) = \mathsf{Tr}\, H\rho_\alpha, \tag{7}$$

$$H(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_i x_i \Pi_i(\boldsymbol{\theta}), \tag{8}$$

$$\Pi_i(\boldsymbol{\theta}) = U_{\boldsymbol{\theta}}^\dagger \left|i\right\rangle\!\left\langle i\right| U_{\boldsymbol{\theta}}. \tag{9}$$

Essentially, we take ρ_α, transform it with $U_{\boldsymbol{\theta}}$, and measure it in the computational basis:

The prediction therefore becomes

$$\mathsf{a}(\rho_\alpha, \boldsymbol{x}, \boldsymbol{\theta}) = \sum_i x_i \left\langle i\right| U_{\boldsymbol{\theta}} \rho_\alpha U_{\boldsymbol{\theta}}^\dagger \left|i\right\rangle \tag{10}$$

$$= \sum_i x_i p_i \tag{11}$$

Hardware-efficient ansatz

As an instance, if ρ_α is a three-qubit state, the unitary $U_{\boldsymbol{\theta}}$ can be represented in the form of the hardware-efficient ansatz[1].

The more layers the ansatz has, the more expressive it is.

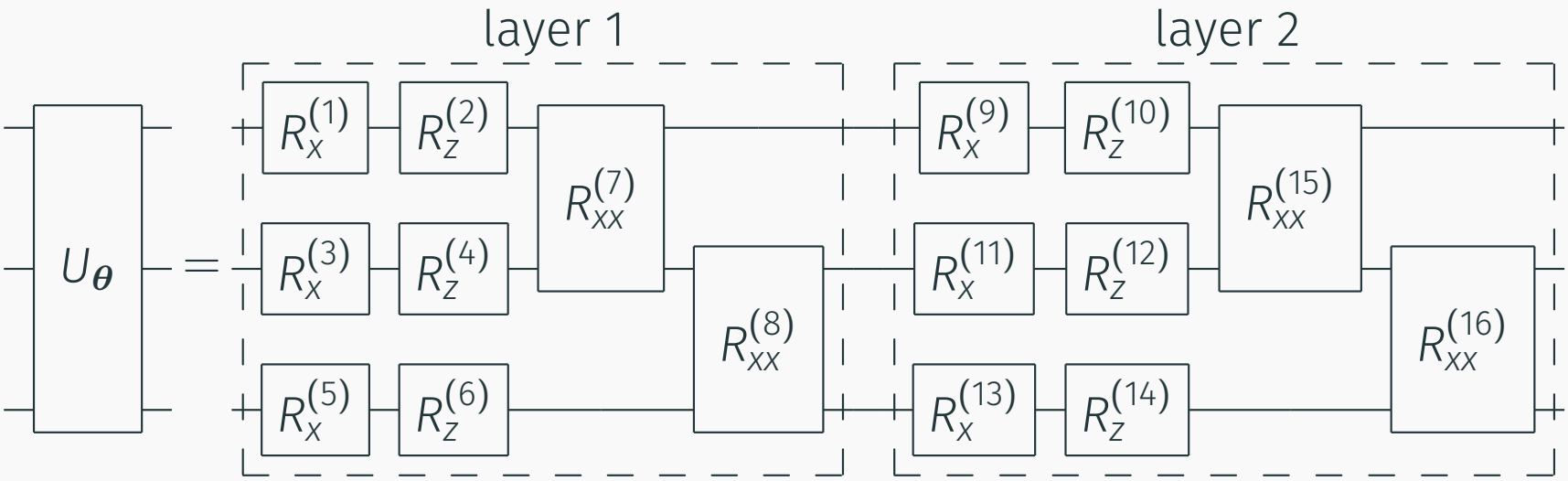

Figure 1: Example of a two-layered hardware-efficient ansatz for three qubits. Here, $R_i^{(k)} = e^{-i\theta_k \sigma_i}$ with σ_i being a Pauli operator, and the rotation angles $\theta_k \in \boldsymbol{\theta}$.

[1] A. Kandala, et al. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets." Nature 549.7671 (2017): 242-246.

Least squares cost

For a given ρ_α, our prediction a for α is the expectation

$$\mathsf{a}(\rho_\alpha, x, \boldsymbol{\theta}) = \sum_i x_i \langle i | U_{\boldsymbol{\theta}} \rho_\alpha U_{\boldsymbol{\theta}}^\dagger | i \rangle. \tag{12}$$

Given a training set $\mathcal{T} = \left\{(\rho_{\alpha_j}, \alpha_j)\right\}_{j=1}^T$, the optimal $\boldsymbol{x}$ and $\boldsymbol{\theta}$ can be found as

$$(\boldsymbol{x}^*, \boldsymbol{\theta}^*) = \arg\min_{\boldsymbol{x}, \boldsymbol{\theta}} \sum_{j=1}^T \Big(\alpha_j - \mathsf{a}_j\big(\boldsymbol{x}, \boldsymbol{\theta}\big)\Big)^2, \tag{13}$$

where $\mathsf{a}_j\big(\boldsymbol{x}, \boldsymbol{\theta}\big) \equiv \mathsf{a}\big(\rho_{\alpha_j}, \boldsymbol{x}, \boldsymbol{\theta}\big)$.

Prediction variance

We are interested not only in high accuracy of the prediction, but also in low variance of it.

As an example, consider the prediction of α for

$$\rho_\alpha = (1-\alpha)\,|+\rangle\langle+| \;+\; \frac{\alpha}{2}\mathbb{1}, \qquad |+\rangle = \frac{1}{\sqrt{2}}\big(\,|0\rangle + |1\rangle\,\big) \tag{14}$$

The observable

$$H = \mathbb{1} - \sigma_x + h_y\sigma_y + h_z\sigma_z, \qquad h_y, h_z \in \mathbb{R}, \tag{15}$$

gives the desired expectation $\langle H\rangle_{\rho_\alpha} \equiv \mathsf{Tr}\, H\rho_\alpha = \alpha$ for any h_y and h_z.

At the same time, it has the variance

$$\Delta^2_{\rho_\alpha} H \equiv \langle H^2\rangle_{\rho_\alpha} - \langle H\rangle^2_{\rho_\alpha} \tag{16}$$

$$= -\alpha^2 + 2\alpha + h_y^2 + h_z^2, \tag{17}$$

which is dependent on h_y and h_z.

Accounting for the variance

For a given ρ_α, our prediction a for α is the expectation

$$\mathsf{a}(\rho_\alpha, x, \theta) = \operatorname{Tr} H(x, \theta)\, \rho_\alpha. \tag{18}$$

Given a training set $\mathcal{T} = \left\{(\rho_{\alpha_j}, \alpha_j)\right\}_{j=1}^{T}$, we seek to minimize both the least squares and the variance,

$$(x^*, \theta^*) = \arg\min_{x,\theta} \Big(w_{\mathrm{ls}} f_{\mathrm{ls}}(x, \theta) + w_{\mathrm{var}} f_{\mathrm{var}}(x, \theta) \Big), \tag{19}$$

where

$$f_{\mathrm{ls}}(x, \theta) = \sum_{j=1}^{T} \Big(\alpha_j - \mathsf{a}_j(x, \theta) \Big)^2, \qquad f_{\mathrm{var}}(x, \theta) = \sum_{j=1}^{T} \Delta^2_{\rho_{\alpha_j}} H(x, \theta), \tag{20}$$

with $w_{\mathrm{ls}}, w_{\mathrm{var}} > 0$ being weights.

Quantum Cramer-Rao bound

Using certain relations known in quantum estimation theory, we can write

$$\frac{\Delta^2_{\rho_\alpha} H}{\left|\partial_\alpha \langle H \rangle_{\rho_\alpha}\right|^2} \geqslant \frac{1}{I_q(\rho_\alpha)}, \tag{21}$$

which is essentially the quantum Cramer-Rao bound[2].

The quantum Fisher information $I_q(\rho_\alpha)$ can be defined as

$$I_q(\rho_\alpha) = 8\frac{1 - F(\rho_\alpha, \rho_{\alpha+\mathrm{d}\alpha})}{\mathrm{d}\alpha^2}, \tag{22}$$

where $F(\rho, \tau) = \mathsf{Tr}\,\sqrt{\rho\sqrt{\tau}\rho}$.

Therefore, we can use $I_q(\rho_\alpha)$ for characterizing an observable H in the task of predicting the label α of a labeled state ρ_α.

[2] S. Sidhu, P. Kok. "Geometric perspective on quantum parameter estimation." AVS Quantum Science 2.1 (2020).

Analytical results

Equation for optimal observable

Our method is described by the following minimization problem:

$$H^* = \arg\min_H \left[w_{\text{ls}} \sum_{j=1}^{T} \left(\operatorname{Tr} H\rho_{\alpha_j} - \alpha_j \right)^2 + w_{\text{var}} \sum_{j=1}^{T} \left(\operatorname{Tr} H^2\rho_{\alpha_j} - \operatorname{Tr}^2 H\rho_{\alpha_j} \right) \right]. \quad (23)$$

Optimal observable H^* satisfies the following equation:

$$\frac{1}{2}\left(\tilde{\rho}_\alpha H^* + H^*\tilde{\rho}_\alpha\right) - \frac{k}{T}\sum_{j=1}^{T}\alpha_j\rho_{\alpha_j} + \frac{(k-1)}{T}\sum_{j=1}^{T}\operatorname{Tr}\left(H^*\rho_{\alpha_j}\right)\rho_{\alpha_j} = 0, \quad (24)$$

where

$$\tilde{\rho}_\alpha = \frac{1}{T}\sum_{j=1}^{T}\rho_{\alpha_j}, \qquad L = b - a, \qquad k = \frac{w_{\text{ls}}}{w_{\text{var}}}. \quad (25)$$

Reduction to the Bayesian approach

$$H^* = \arg\min_H \left[w_{\mathrm{ls}} \sum_{j=1}^{T} \left(\mathsf{Tr}\, H\rho_{\alpha_j} - \alpha_j \right)^2 + w_{\mathrm{var}} \sum_{j=1}^{T} \left(\mathsf{Tr}\, H^2\rho_{\alpha_j} - \mathsf{Tr}^2\, H\rho_{\alpha_j} \right) \right]. \quad (26)$$

When $w_{\mathrm{ls}} = w_{\mathrm{var}}$, we have

$$H^* = \arg\min_H \sum_{j=1}^{T} \mathsf{Tr}\, \rho_{\alpha_j} \left(H - \alpha_j \mathbb{1} \right)^2 \quad (27)$$

This is similar to the minimization of the Bayesian mean squared error (BMSE)

$$\Delta_B^2 \alpha \equiv \int_a^b \Pr(\alpha)\, \mathsf{Tr}\, \rho_\alpha \left(H - \alpha \mathbb{1} \right)^2 \, d\alpha, \quad (28)$$

when the prior distribution $\Pr(\alpha)$ of the labels α is uniform, i.e., $\Pr(\alpha) = 1/(b-a)$.

Numerical results

Entanglement of Bell-type states

Consider the two-qubit state

$$\left|\Phi_p^+\right\rangle = \sqrt{p}\,|00\rangle + \sqrt{1-p}\,|11\rangle\,, \tag{29}$$

with $|ab\rangle \equiv |a\rangle \otimes |b\rangle$.

We want to learn to predict the *negativity* of $\rho_p = \left|\Phi_p^+\right\rangle\!\left\langle\Phi_p^+\right|$, an entanglement measure defined as

$$N(\rho) = \left\|\rho^{T_2}\right\|_1 - 1, \tag{30}$$

where $\|A\|_1 = \mathbf{Tr}\,\sqrt{A^\dagger A}$ and $\rho^{T_2} \equiv (\mathbb{1} \otimes T)[\rho]$ is the partial transpose of ρ with respect to the second qubit.

We train our model on a set $\mathcal{T} = \left\{|\Phi_{p_j}^+\rangle, N_j\right\}_{j=1}^{5}$, where the coefficients p_j were picked randomly and uniformly.

Entanglement of Bell-type states

$$\left|\Phi_p^+\right\rangle = \sqrt{p}\,|00\rangle + \sqrt{1-p}\,|11\rangle\,,$$
$$N(\rho) = \left\|\rho^{T_2}\right\|_1 - 1,$$

With no assumptions on the data, our model finds H that accurately predicts N for $\left|\Phi_p^+\right\rangle$ and saturates the Cramer-Rao bound.

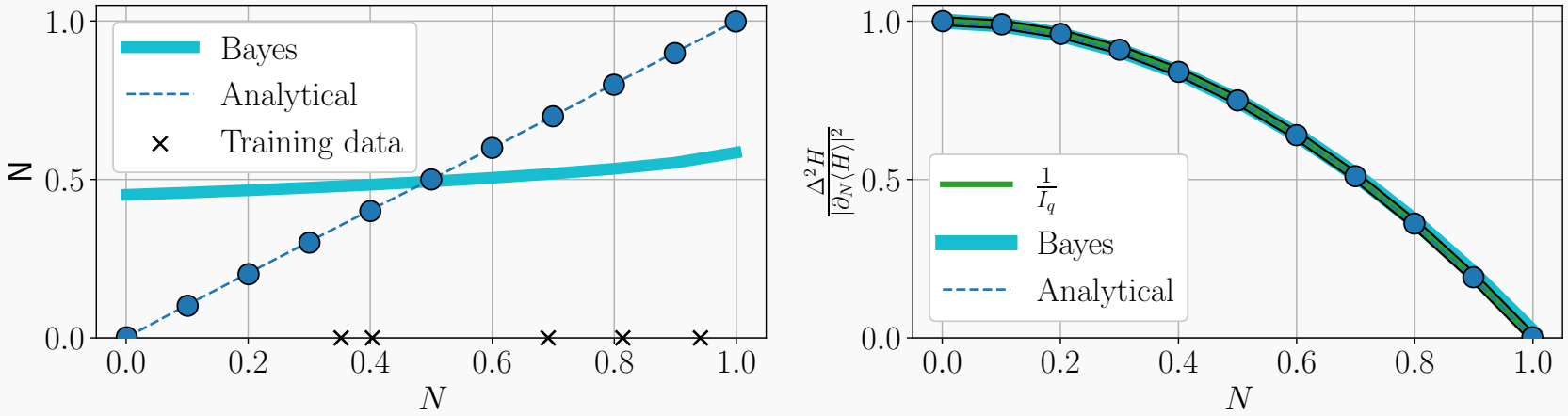

Figure 2: Left: predicted **N** vs. true N negativity of the Bell state. Right: Variance and quantum Cramer-Rao bound vs. N. The model was trained with $w_{\mathbf{var}} = 10^{-4}$ and $w_{\mathbf{ls}} = 1$.

Results: negativity of random mixed states

Consider $\mathcal{T} = \left\{\rho_j^{\otimes 4}, N_j\right\}_{j=1}^{1000}$ with ρ_j being *random mixed* states with the negativity N_j.

The more state is mixed, the less accurate is the prediction, and the greater is the variance.

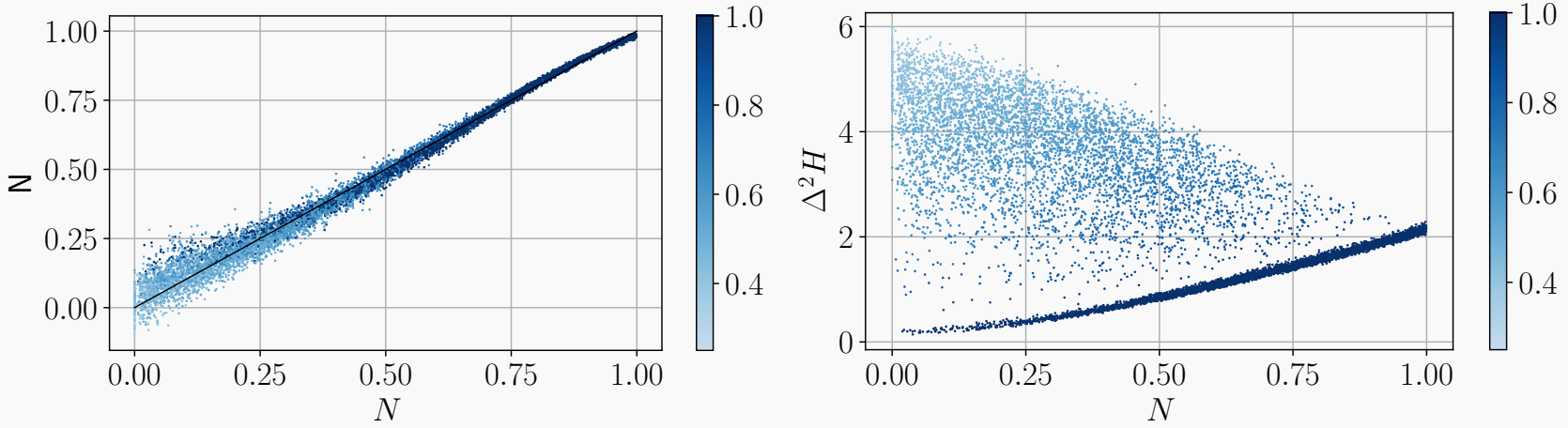

Figure 3: Predicted negativity N of 10^4 random mixed states (left) and variance of the optimized observable H (right) vs. the true negativity N for $c = 4$ copies. The color of points indicates the purity $P(\rho_N) = \mathrm{Tr}\,\rho_N^2$ of the corresponding states ρ_N. The model was trained with $w_{\mathrm{var}} = 10^{-4}$ and $w_{\mathrm{ls}} = 1$.

Results: transverse field in the Ising Hamiltonian

Let $\mathcal{T} = \left\{ |\psi_{h_j}\rangle,\, h_j \right\}_{j=1}^{10}$, where $|\psi_h\rangle$ is the ground state of

$$H_h = \sum_{i=1}^{8} \left(\sigma_z^i \sigma_z^{i+1} + h\sigma_x^i \right) \tag{31}$$

For fixed $w_{\mathrm{ls}} = 1$, smaller w_{var} give more precise predictions of h, but also greater variances.

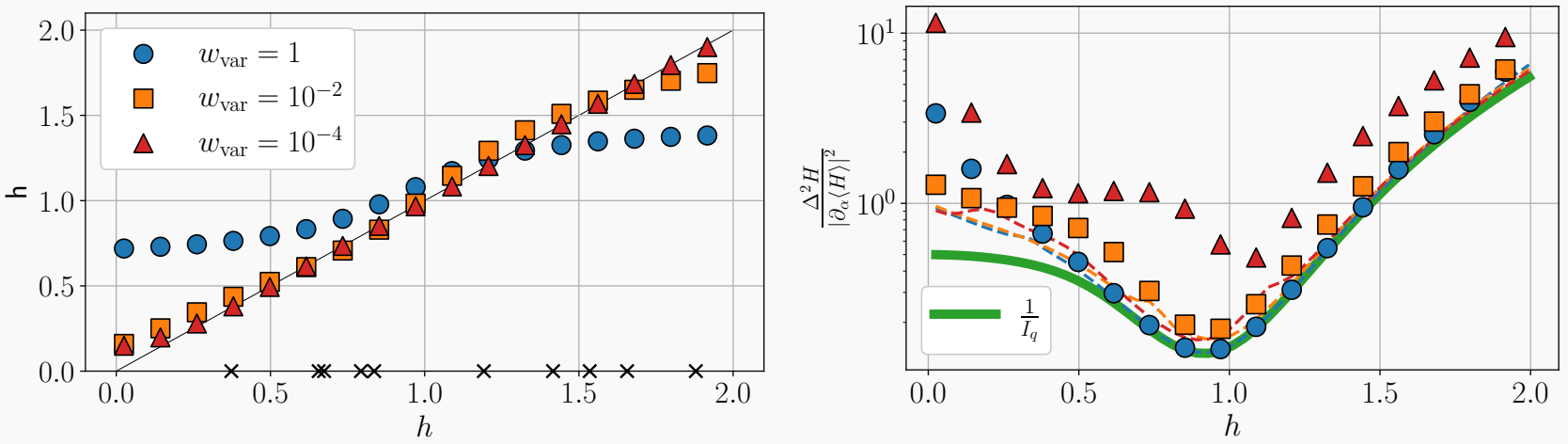

Figure 4: Left: predicted **h** vs. true h transverse field of the 8-qubit Ising Hamiltonian. Right: Variance of the trained observable, classical (dashed lines) and quantum (solid line) Cramer-Rao bounds vs. h.

Conclusion

Conclusion

Given a training set $\mathcal{T} = \left\{(\rho_{\alpha_j}, \alpha_j)\right\}_{j=1}^{T}$, we can train an observable H which may give good predictions for α and saturate the Cramer-Rao bound.

The proposed method assumes no knowledge about the connection between ρ_α and α.

Such connection could be:

- ρ_α is a random state with entanglement α
- $\rho_\alpha = \Phi_\alpha[\rho]$, where Φ_α is a parametrized channel and ρ is some fixed input state
- $\rho_\alpha = |\psi_\alpha\rangle\langle\psi_\alpha|$, where $|\psi_\alpha\rangle$ is the ground state of a parametrized Hamiltonian H_α

Under certain assumptions and conditions, our method reduces to the minimization of the Bayesian MSE.

The Bayesian approach produces observables with a large bias, as it prefers to minimize the variance.

Thank you!

Predicting properties of quantum systems by regression on a quantum computer

Physical Review Research 7 (1), 013201 (2025)

doi.org/10.1103/PhysRevResearch.7.013201