Predicting Properties of Quantum Systems by Regression on a Quantum Computer

Andrey Kardashin[1,*], Yerassyl Balkybek[1], Vladimir V. Palyulin[1], Konstantin Antipin[1,2]

*andrey.kardashin@skoltech.ru

[1]Skolkovo Institute of Science and Technology, Moscow, Russia

[2]Lomonosov Moscow State University, Moscow, Russia

Problem statement

Suppose we are given the following training set:

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

where ρ_{α_j} are labeled quantum states and $\alpha_j \in \mathbb{R}$ are their corresponding labels. Hereinafter, we assume that ρ_α describes a state of n qubits.

Our goal is therefore solving a regression problem, i.e., using the given training set $\mathcal{T}$ for learning how to estimate the parameter α for an unseen datum ρ_α.

There could be various connections between the data points ρ_α and their labels α, e.g.:

- α quantifies the entanglement of ρ_α;
- ρ_α is an output state of a parametrized channel $\Phi_\alpha[\rho]$ acting on some fixed input ρ;
- $\rho_\alpha = |\psi_\alpha\rangle\langle\psi_\alpha|$ is the ground state of a parametrized Hamiltonian H_α.

Estimation

Given a labeled state ρ_α, one can obtain the estimation $\hat{\alpha}$ of the label α from the expected value of an observable H in the state ρ_α. Generally, such expectation would give a function $f(\alpha)$, which can be written as the label α itself adjusted by a bias $b(\alpha)$:

$$f(\alpha) \equiv \operatorname{Tr} H\rho_\alpha = \alpha + b(\alpha). \tag{2}$$

We parametrize the Hermitian operator H by $\boldsymbol{x}, \boldsymbol{\theta} \subset \mathbb{R}$ and represent this observable as a spectral decomposition

$$H(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_i x_i \Pi_i(\boldsymbol{\theta}), \tag{3}$$

where $\boldsymbol{x} = \{x_i\}_i$ are the eigenvalues, and the eigenprojectors $\Pi_i(\boldsymbol{\theta}) = U^\dagger(\boldsymbol{\theta})\,|i\rangle\langle i|\,U(\boldsymbol{\theta})$ are the projectors onto the ith state of the computational basis transformed by a variational circuit $U(\boldsymbol{\theta})$.

Schematically, the label prediction can be depicted as follows:

$$\rho_\alpha \;\overset{/n}{—}\; \boxed{U(\boldsymbol{\theta})} \;—\; \text{[measurement]} \xrightarrow{p_i} i \mapsto x_i$$

That is, we transform an n-qubit labeled state ρ_α by a parametrized unitary $U(\boldsymbol{\theta})$, measure the resultant state $\rho_\alpha(\boldsymbol{\theta}) \equiv U(\boldsymbol{\theta})\rho_\alpha U^\dagger(\boldsymbol{\theta})$ in the computational basis, and with probability $p_i = \langle i|\rho_\alpha(\boldsymbol{\theta})|i\rangle$ get the outcome i associated with x_i, which gives $f(\alpha) = \sum_i x_i p_i$.

Optimization

To find optimal parameters $\boldsymbol{x}^*$ and $\boldsymbol{\theta}^*$, we solve the following minimization problem:

$$\boldsymbol{x}^*, \boldsymbol{\theta}^* = \arg\min_{\boldsymbol{x},\boldsymbol{\theta}} \Big(w_{\text{ls}} F_{\text{ls}}(\boldsymbol{x}, \boldsymbol{\theta}) + w_{\text{var}} F_{\text{var}}(\boldsymbol{x}, \boldsymbol{\theta})\Big), \tag{4}$$

where

$$F_{\text{ls}}(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{j=1}^{T} \Big(\alpha_j - \hat{f}(\rho_{\alpha_j}, \boldsymbol{x}, \boldsymbol{\theta})\Big)^2, \quad F_{\text{var}}(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{j=1}^{T} \Delta^2_{\rho_{\alpha_j}} H(\boldsymbol{x}, \boldsymbol{\theta}),$$

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

Here, F_{ls} is the sum of the squared differences between the given labels α and estimations $\hat{f}$ of $f(\alpha) = \langle H\rangle_{\rho_\alpha} \equiv \operatorname{Tr} H\rho_\alpha$, while F_{var} is the sum of variances $\Delta^2_{\rho_\alpha} H \equiv \langle H^2\rangle_{\rho_\alpha} - \langle H\rangle^2_{\rho_\alpha}$.

Cramer-Rao bound

The accuracy of the estimation $\hat{\alpha} \equiv f^{-1}(\hat{f})$ can be characterized by the mean-squared error (MSE) $\Delta^2\hat{\alpha} \equiv \langle(\alpha - \hat{\alpha})^2\rangle$, for which one can write

$$\Delta^2\hat{\alpha} = \frac{\Delta^2_{\rho_\alpha} H}{\mu\left|\partial_\alpha \langle H\rangle_{\rho_\alpha}\right|^2} = \frac{\Delta^2 \hat{f}}{\left|\partial_\alpha \langle H\rangle_{\rho_\alpha}\right|^2} \geqslant \frac{1}{\mu I_c(\Pi, \rho_\alpha)} \geqslant \frac{1}{\mu I_q(\rho_\alpha)}. \tag{5}$$

where μ is the number of measurements.

The first equality is known as the error propagation formula. The first and the second inequalities are, respectively, the classical and the quantum Cramer-Rao bounds (CRB), where $I_c(\Pi, \rho_\alpha)$ is the classical Fisher information (FI) and $I_q(\rho_\alpha)$ is the quantum FI.

Skoltech

preprint

Predicting the transverse field of the Ising model

First, we demonstrate the performance of our method in predicting the label h of a labeled state $\rho_h = |\psi_h\rangle\langle\psi_h|$ being the ground state of the 8-qubit transverse field Ising Hamiltonian

$$H_h = -\sum_{i=1}^{8} \left(\sigma_z^i \sigma_z^{i+1} + h\sigma_x^i\right). \tag{6}$$

We trained the observable H on a set $\mathcal{T} = \left\{|\psi_j\rangle, h_j\right\}_{j=1}^{20}$ with random h_j.

In Fig. 1, we show the performance for the observables trained with different weights w_{var} in (4) and setting $w_{\text{ls}} = 1$. As expected, the greater is the weight w_{var}, the less accurate predictions we get, but also lower is the variance.

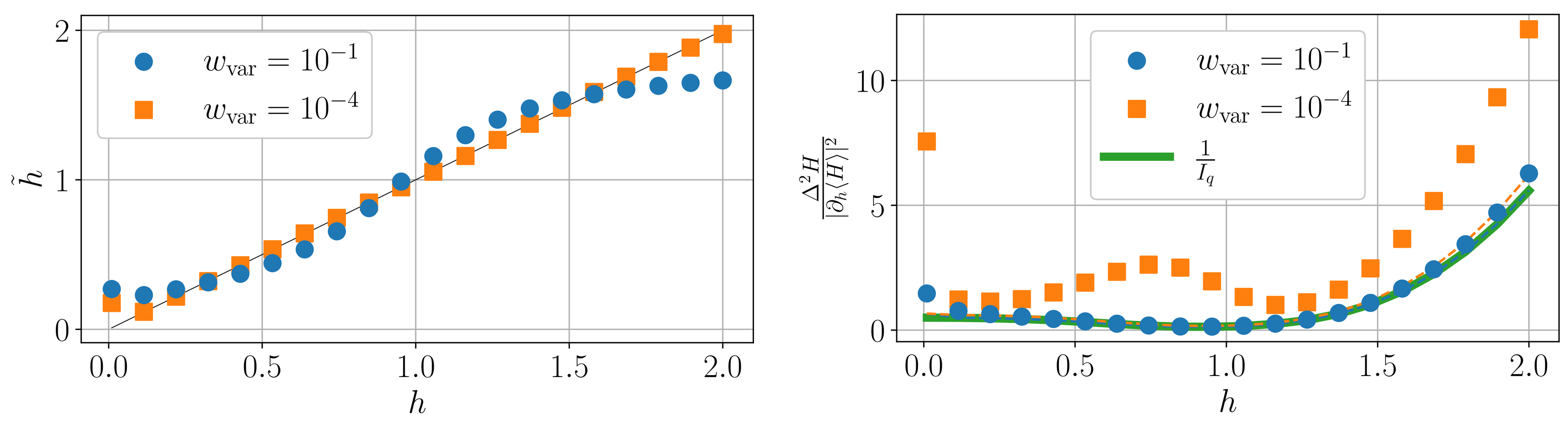

Fig. 1: Left: Predicted $\hat{h} = \operatorname{Tr} H(\boldsymbol{x}^*, \boldsymbol{\theta}^*)\rho_h$ vs. true h transverse field of the 8-qubit Ising Hamiltonian (6). Right: Error propagation and CRB (5) vs. α; the dashed lines indicate classical CRB.

Connection to the Bayesian approach

For a large training set size T and $w_{\text{ls}} = w_{\text{var}} = 1$, the problem (4) can be reduced to $\min_H \int_a^b \operatorname{Tr} \rho_\alpha (H - \alpha\mathbb{1})^2\, d\alpha$, which is equivalent to minimizing the Bayesian MSE

$$\Delta^2_B \hat{\alpha} = \int_a^b \Pr(\alpha) \operatorname{Tr} \rho_\alpha (H - \alpha\mathbb{1})^2\, d\alpha \tag{7}$$

with the flat prior $\Pr(\alpha) = 1/(b - a)$.

Consider the amplitude-damping (AD) channel $\Phi_\alpha[\rho] = \sum_{k=1}^2 V_k(\alpha)\rho V_k^\dagger(\alpha)$, where $V_1(\alpha) = \sqrt{\alpha}\,|0\rangle\langle 1|$, $V_2(\alpha) = |0\rangle\langle 0| + \sqrt{1-\alpha}\,|1\rangle\langle 1|$ and the input state $\rho = |+\rangle\langle +|$.

In Fig. 2, we compare the predictions of α of the AD channel via (4) and via the Bayesian approach with the uniform prior. With $w_{\text{ls}} = w_{\text{var}} = 1$, our procedure indeed coincides with the Bayesian one with the flat prior.

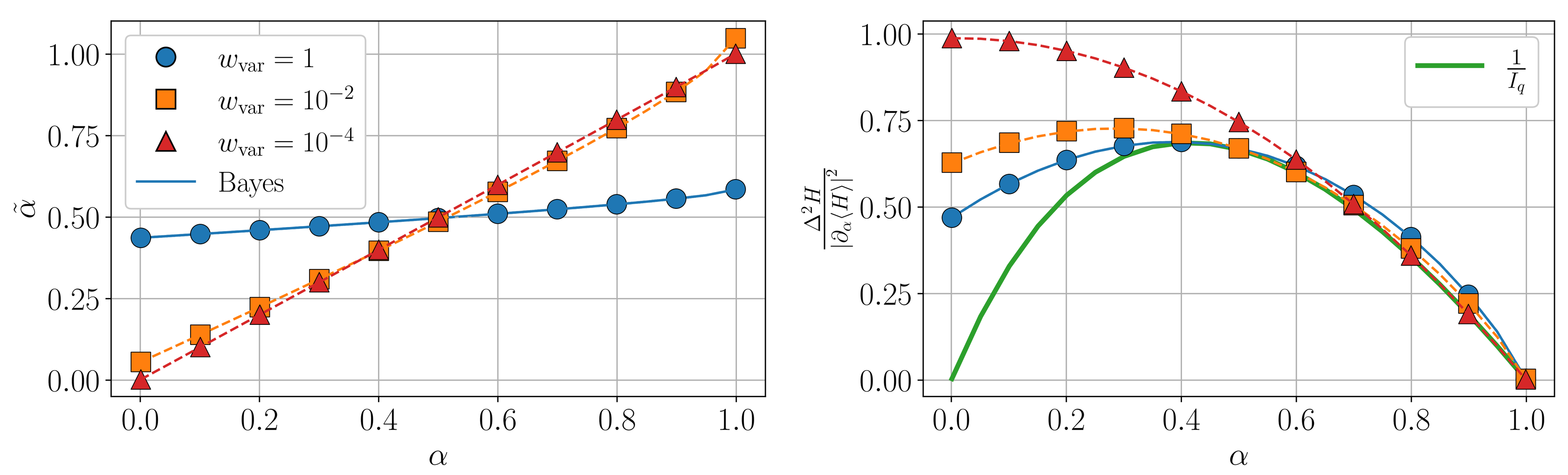

Fig. 2: Left: Predicted $\hat{\alpha} = \operatorname{Tr} H(\boldsymbol{x}^*, \boldsymbol{\theta}^*)\rho_\alpha$ vs. true α amplitude damping parameter for different weights w_{var}. Right: Error propagation and CRB (5) vs. α. The models (4) are trained on a set $\mathcal{T} = \{\rho_{\alpha_j}, \alpha_j\}_{j=1}^{500}$ with equidistant α_j.

Predicting the entanglement of two-qubit states

Finally, we apply our method for entanglement learning for two-qubit random mixed states. As a measure of entanglement, we chose the negativity $N(\rho_{AB}) = \left\|\rho_{AB}^{T_B}\right\|_1 - 1$.

We allow our model to process $c = 4$ copies of the labeled states, so that we train it on a set $\mathcal{T} = \left\{\rho_j^{\otimes 4}, N_j\right\}_{j=1}^{1000}$ with random mixed two-qubit states ρ_j and their negativities N_j.

As can be seen in Fig. 3, with our method, one is able to predict the entanglement of two-qubit states with a good accuracy.

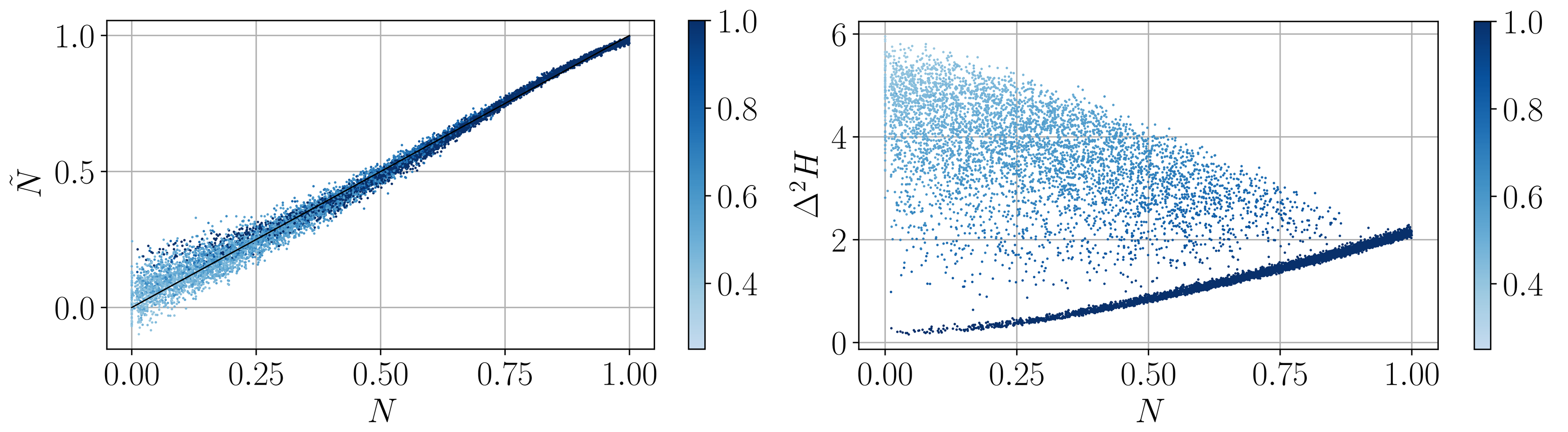

Fig. 3: Left: Predicted negativity $\tilde{N} = \operatorname{Tr} H(\boldsymbol{x}^*, \boldsymbol{\theta}^*)\rho_N$ of 10^4 random mixed states. Right: Variance of the trained observable H vs. the true negativity N. The color of points indicates the purity $P(\rho) = \operatorname{Tr}\rho^2$ of the corresponding states ρ_N.