

VQBR: Variational Quantum Bayesian Regression via Measurement-Based MAP Direction Recovery

Van Tien Nguyen¹ Panagiotis Markopoulos^{2,3}

¹Department of Electrical Engineering, The University of Texas at San Antonio, San Antonio, TX, USA

²Department of Computer Engineering, The University of Texas at San Antonio, San Antonio, TX, USA

³Department of Computer Science, The University of Texas at San Antonio, San Antonio, TX, USA

tien.nguyen@utsa.edu, panagiotis.markopoulos@utsa.edu

© 2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Accepted for publication in the 2026 IEEE International Conference on Quantum Computing and Engineering (QCE).

Abstract

Bayesian linear regression (BLR) with Gaussian priors admits a closed-form *maximum a posteriori* (MAP) estimator, but a dense direct implementation costs $\mathcal{O}(ND^2 + D^3)$ for N samples and D features. We propose *Variational Quantum Bayesian Regression* (VQBR), a hybrid NISQ-compatible formulation that represents a candidate MAP-weight direction in the amplitudes of a parameterized circuit on $d = \lceil \log_2 D \rceil$ qubits and eliminates its scalar magnitude analytically. For a fixed ansatz family, this scale elimination is exact: optimizing the resulting direction-only objective is equivalent to jointly optimizing scale and direction, although recovery of the unrestricted MAP solution still depends on ansatz expressivity and classical optimization. VQBR estimates the objective using row-wise overlap circuits and Pauli expectation measurements without forming the Gram matrix or block-encoding the MAP system. On synthetic data, VQBR attains cosine similarity 0.998 ± 0.005 with the direct MAP solution at $D=8$, while on the Energy Efficiency dataset it attains 0.927 ± 0.039 and predictive metrics close to an SVD-based direct solve. At $D=16$, increasing the `EfficientSU2` repetitions from $p=1$ to $p=5$ raises cosine similarity from 0.864 ± 0.058 to 0.9999 ± 0.0001 , demonstrating the effect of ansatz expressivity in this setting. A single proof-of-concept run on `ibm_kingston` improves from 0.8871 without mitigation to 0.9478 with zero-noise extrapolation, compared with 0.9956 in statevector simulation. VQBR does not claim a runtime advantage over matrix-free classical solvers; establishing such a regime is left to an extended study.

Keywords: Variational quantum algorithms, Bayesian linear regression, quantum machine learning, NISQ, Hadamard test, zero-noise extrapolation

1 Introduction

Bayesian linear regression (BLR) with Gaussian priors provides a principled, uncertainty-aware prediction framework central to signal processing and machine learning [1, 2]. Given data $(\mathbf{X}, \mathbf{y}) \in$

$\mathbb{R}^{N \times D} \times \mathbb{R}^N$, prior $(\mathbf{m}_0, \Sigma_0)$, and noise variance V , BLR yields the *maximum a posteriori* (MAP) estimation problem

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + (\mathbf{w} - \mathbf{m}_0)^\top V \Sigma_0^{-1} (\mathbf{w} - \mathbf{m}_0), \quad (1)$$

whose closed-form solution is

$$\mathbf{w}^* = \left(\mathbf{X}^\top \mathbf{X} + V \Sigma_0^{-1} \right)^{-1} \left(\mathbf{X}^\top \mathbf{y} + V \Sigma_0^{-1} \mathbf{m}_0 \right). \quad (2)$$

A dense direct evaluation of (2), including construction of $\mathbf{X}^\top \mathbf{X}$ and solution of the resulting dense system, costs $\mathcal{O}(ND^2 + D^3)$. This is not a universal lower bound: sparse factorizations and matrix-free Krylov methods can solve the system without explicitly constructing the Gram matrix.

Equation (1) is a generalized Tikhonov-regularized least-squares problem [3]. In the special case of $V = 0$, it reduces to Ordinary Least Squares (OLS). In the special case of $V \Sigma_0^{-1} = \lambda \mathbf{I}$ and $\mathbf{m}_0 = \mathbf{0}$, it reduces to ridge regression [4]. Matrix-free conjugate gradient (CG) and LSQR are therefore important classical alternatives [5, 6].

In this work, we present *Variational Quantum Bayesian Regression* (VQBR), a hybrid NISQ-compatible framework for MAP estimation in Bayesian linear regression. VQBR represents the regression-weight direction using the amplitudes of a parameterized quantum state, eliminates the corresponding scalar magnitude analytically, and evaluates the resulting direction-only objective through row-wise overlap measurements and Pauli expectation values. This formulation avoids explicitly constructing the Gram matrix or block-encoding the MAP linear system.

VQBR is not intended to replace matrix-free classical methods through a proven asymptotic speedup. Rather, it provides a measurement-based formulation that is compatible with shallow parameterized quantum circuits and suitable for investigation on near-term quantum hardware.

Amplitude encoding represents a normalized D -component vector using $d = \lceil \log_2 D \rceil$ qubits. A fixed-depth parameterized circuit, however, spans only a restricted variational manifold and therefore does not guarantee that every direction in $\mathbb{R}^D$ is reachable. In our implementation, the number of trainable angles is $\mathcal{O}(d)$ for fixed circuit depth, providing a compact but restricted parameterization. Hybrid variational algorithms offer a natural near-term framework for exploring such representations [7–9].

Detailed contributions. We propose *VQBR* (Variational Quantum Bayesian Regression), a hybrid algorithm with the following contributions.

(i) *Exact scale elimination within the ansatz family.* VQBR parameterizes $\mathbf{w} = t \phi(\boldsymbol{\theta})$ and eliminates t through a closed-form one-dimensional minimization. The resulting direction-only objective is exactly equivalent to jointly optimizing scale and direction over the chosen ansatz family. It does not, by itself, guarantee that the unrestricted MAP direction is representable.

(ii) *Gram-matrix-free estimator.* The direction-only objective is estimated from $N+L+1$ row-wise Hadamard-test and Pauli-measurement circuits on $d+1$ qubits, with no block-encoding or explicit Gram-matrix construction.

(iii) *Simulation and hardware validation.* Statevector simulations characterize direction recovery, ansatz depth, and finite-shot effects. A single run on `ibm_kingston` shows that the measurement primitives are executable and compatible with zero-noise extrapolation (ZNE).

2 Related Work

Quantum regression. Fault-tolerant quantum algorithms for least-squares regression include HHL-style methods [10–12] and regularized/QSVT-based formulations [13]. Dequantization results

show that classical algorithms with comparable access assumptions can match or approach some proposed speedups [14], while more recent work establishes quadratic quantum speedups under different query models [15]. VQBR targets the near-term regime and makes no asymptotic-advantage claim.

Variational linear-algebra solvers. The variational quantum linear solver (VQLS) [16] trains $U(\boldsymbol{\theta})|0\rangle$ so that $A|x(\boldsymbol{\theta})\rangle$ aligns with $|b\rangle$. VQBR similarly uses overlap estimation but evaluates the BLR MAP objective directly from row-state preparations, avoiding a block-encoding of $\mathbf{X}^\top \mathbf{X} + V \Sigma_0^{-1}$.

Quantum Bayesian learning. Quantum circuits have been studied as variational posterior families [17], and quantum-kernel Gaussian processes retain predictive variance [18]. VQBR instead targets MAP point estimation under an explicit Gaussian prior; full posterior uncertainty is left to future work.

3 Proposed Method: Variational Quantum Bayesian Regression

3.1 Scale Elimination and Direction-Only Objective

We seek a circuit $U(\boldsymbol{\theta})$ on $d=\lceil \log_2 D \rceil$ qubits whose output $|\phi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle^{\otimes d}$ encodes a candidate MAP direction. Feature vectors are zero-padded to length 2^d .

Scale elimination We parameterize a candidate regression-weight vector using the two independent optimization variables $t \in \mathbb{R}$ and $\boldsymbol{\theta} \in \Theta$ as

$$\mathbf{w}(t, \boldsymbol{\theta}) = t \phi(\boldsymbol{\theta}), \quad (3)$$

where $\phi(\boldsymbol{\theta}) \in \mathbb{R}^{2^d}$ is the real amplitude vector generated by the parameterized quantum circuit. Substituting this parameterization into (1) gives

$$\mathcal{L}(t, \boldsymbol{\theta}) = \alpha(\boldsymbol{\theta})t^2 - 2\beta(\boldsymbol{\theta})t + \text{const}, \quad (4)$$

where

$$\begin{aligned} \alpha(\boldsymbol{\theta}) &= \phi(\boldsymbol{\theta})^\top \mathbf{X}^\top \mathbf{X} \phi(\boldsymbol{\theta}) + \phi(\boldsymbol{\theta})^\top V \Sigma_0^{-1} \phi(\boldsymbol{\theta}), \\ \beta(\boldsymbol{\theta}) &= \phi(\boldsymbol{\theta})^\top \mathbf{X}^\top \mathbf{y} + \phi(\boldsymbol{\theta})^\top V \Sigma_0^{-1} \mathbf{m}_0. \end{aligned} \quad (5)$$

For any fixed $\boldsymbol{\theta}$, $\mathcal{L}(t, \boldsymbol{\theta})$ is a one-dimensional quadratic function of t . Since $V \Sigma_0^{-1} \succ 0$, we have $\alpha(\boldsymbol{\theta}) > 0$, and the minimizing scale is therefore a function of the selected direction:

$$t^*(\boldsymbol{\theta}) = \underset{t}{\operatorname{argmin}} \mathcal{L}(t, \boldsymbol{\theta}) = \frac{\beta(\boldsymbol{\theta})}{\alpha(\boldsymbol{\theta})}. \quad (6)$$

Substituting $t^*(\boldsymbol{\theta})$ into the original objective gives the direction-only objective

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}) := \mathcal{L}(t^*(\boldsymbol{\theta}), \boldsymbol{\theta}) = \min_t \mathcal{L}(t, \boldsymbol{\theta}) = -\frac{\beta(\boldsymbol{\theta})^2}{\alpha(\boldsymbol{\theta})} + \text{const}. \quad (7)$$

Consequently,

$$\min_{\boldsymbol{\theta}, t} \mathcal{L}(t, \boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \tilde{\mathcal{L}}(\boldsymbol{\theta}) \quad (8)$$

is exact over the ansatz family

$$\mathcal{A} := \{t\phi(\boldsymbol{\theta}) : t \in \mathbb{R}, \boldsymbol{\theta} \in \Theta\}. \quad (9)$$

Equality with the unrestricted solution $\mathbf{w}^*$ additionally requires its direction to be representable by the ansatz and the classical optimizer to reach the corresponding optimum. After obtaining the optimized direction $\boldsymbol{\theta}^*$, the corresponding weight estimate is reconstructed as

$$\hat{\mathbf{w}} = t^*(\boldsymbol{\theta}^*)\phi(\boldsymbol{\theta}^*). \quad (10)$$

Measurement-based decomposition Let

$$\mathbf{M} := V\boldsymbol{\Sigma}_0^{-1}, \quad \mathbf{c} := \mathbf{M}\mathbf{m}_0. \quad (11)$$

Define

$$\begin{aligned} a(\boldsymbol{\theta}) &:= \phi(\boldsymbol{\theta})^\top \mathbf{X}^\top \mathbf{X} \phi(\boldsymbol{\theta}), \\ c(\boldsymbol{\theta}) &:= \phi(\boldsymbol{\theta})^\top \mathbf{X}^\top \mathbf{y}, \\ d(\boldsymbol{\theta}) &:= \phi(\boldsymbol{\theta})^\top \mathbf{M} \phi(\boldsymbol{\theta}), \\ e(\boldsymbol{\theta}) &:= \phi(\boldsymbol{\theta})^\top \mathbf{c}. \end{aligned} \quad (12)$$

With these definitions,

$$\alpha(\boldsymbol{\theta}) = a(\boldsymbol{\theta}) + d(\boldsymbol{\theta}), \quad \beta(\boldsymbol{\theta}) = c(\boldsymbol{\theta}) + e(\boldsymbol{\theta}). \quad (13)$$

Therefore,

$$t^*(\boldsymbol{\theta}) = \frac{c(\boldsymbol{\theta}) + e(\boldsymbol{\theta})}{a(\boldsymbol{\theta}) + d(\boldsymbol{\theta})}, \quad (14)$$

and the direction-only objective becomes

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}) = -\frac{(c(\boldsymbol{\theta}) + e(\boldsymbol{\theta}))^2}{a(\boldsymbol{\theta}) + d(\boldsymbol{\theta})} + \text{const.} \quad (15)$$

These quantities can be estimated without explicitly forming $\mathbf{X}^\top \mathbf{X}$.

3.2 Quantum Measurement Routines

State preparation For each row $\mathbf{x}_i$, prepare $U_{\mathbf{x}_i}|0\rangle^{\otimes d} = |\mathbf{x}_i/\|\mathbf{x}_i\|_2\rangle$; similarly, $U_{\mathbf{c}}|0\rangle^{\otimes d} = |\mathbf{c}/\|\mathbf{c}\|_2\rangle$. If $\mathbf{c} = \mathbf{0}$, then $e(\boldsymbol{\theta}) = 0$ exactly, and the $U_{\mathbf{c}}$ overlap circuit is omitted.

Hadamard-test estimation of a, c, e For $U_{\mathbf{v}} \in \{U_{\mathbf{x}_i}, U_{\mathbf{c}}\}$, the Hadamard-test circuit in Fig. 1 yields $\langle \mathbf{v} | \phi(\boldsymbol{\theta}) \rangle = 2\Pr(z=0) - 1$. For notational simplicity, we suppress the explicit dependence of $\hat{s}_i$, $\hat{a}$, $\hat{c}$, $\hat{d}$, and $\hat{e}$ on $\boldsymbol{\theta}$ throughout the measurement and training expressions. With S shots,

$$\hat{s}_i = 2\frac{n_0^{(i)}}{S} - 1 \approx \left\langle \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}, \phi(\boldsymbol{\theta}) \right\rangle, \quad (16)$$

and

$$\begin{aligned} \hat{a} &= \sum_{i=1}^N (\|\mathbf{x}_i\|_2 \hat{s}_i)^2, \\ \hat{c} &= \sum_{i=1}^N y_i \|\mathbf{x}_i\|_2 \hat{s}_i, \quad \hat{e} = \|\mathbf{c}\|_2 (2n_0^{(c)}/S - 1). \end{aligned} \quad (17)$$

Although $\mathbb{E}[\hat{s}_i] = s_i$, squaring gives $\mathbb{E}[\hat{s}_i^2] = s_i^2 + (1 - s_i^2)/S$; hence the plug-in $\hat{a}$ has an $\mathcal{O}(S^{-1})$ upward bias. An unbiased componentwise alternative is $\tilde{s}_i^2 = (S\hat{s}_i^2 - 1)/(S - 1)$ for $S > 1$. We retain the direct plug-in estimator in the reported experiments and characterize its aggregate effect empirically through the shot-budget ablation.

Pauli estimation of d Decompose $\mathbf{M} = \sum_{\ell=1}^L \gamma_{\ell} P_{\ell}$. For a diagonal prior, at most $L \leq 2^d$ diagonal Pauli strings suffice. For the power-of-two feature dimensions used in our experiments, $2^d = D$. Estimate

$$\hat{d} = \sum_{\ell=1}^L \gamma_{\ell} \hat{\mu}_{\ell}, \quad \hat{\mu}_{\ell} \approx \langle \phi(\boldsymbol{\theta}) | P_{\ell} | \phi(\boldsymbol{\theta}) \rangle, \quad (18)$$

using S shots per Pauli string.

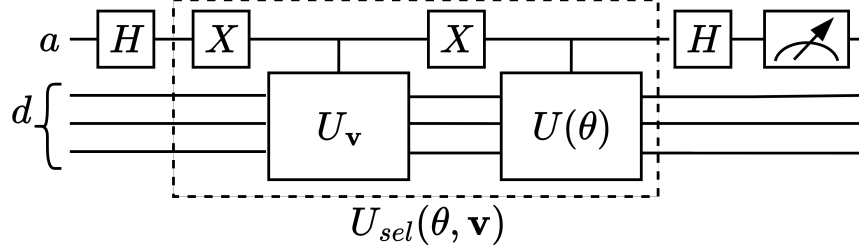


Figure 1: Hadamard-test circuit for estimating $\langle \mathbf{v} | \phi(\boldsymbol{\theta}) \rangle$. The ancilla probability is $\Pr(z=0) = \frac{1+\langle \mathbf{v} | \phi(\boldsymbol{\theta}) \rangle}{2}$. The circuit uses $d+1$ qubits.

3.3 Training Algorithm

We minimize the stabilized log-ratio surrogate

$$\hat{\mathcal{J}}_{\log}(\boldsymbol{\theta}) = \log(\hat{a} + \hat{d} + \delta) - \log((\hat{c} + \hat{e})^2 + \delta), \quad (19)$$

where $\delta > 0$ prevents evaluation of $\log 0$. At $\delta = 0$, (19) is a monotone transformation of the corresponding plug-in form of (15), wherever the estimated numerator and denominator are nonzero.

Algorithm 1 VQBR

Require: $(\mathbf{X}, \mathbf{y})$; prior $(\mathbf{m}_0, \Sigma_0, V)$; ansatz $U(\boldsymbol{\theta})$; shots S ; $K_{\max}$; tol ; $\delta > 0$.

Ensure: Reconstructed estimate $\hat{\mathbf{w}}$

- 1: Set $d = \lceil \log_2 D \rceil$ and compute $\mathbf{M} \leftarrow V \Sigma_0^{-1}$, $\mathbf{c} \leftarrow \mathbf{M} \mathbf{m}_0$.
 - 2: Build $\{U_{\mathbf{x}_i}\}_{i=1}^N$ and $U_{\mathbf{c}}$; decompose $\mathbf{M} = \sum_{\ell} \gamma_{\ell} P_{\ell}$.
 - 3: Initialize $\boldsymbol{\theta}^{(0)}$ and $\mathcal{J}_{\text{prev}} \leftarrow +\infty$.
 - 4: **for** $k = 0, \dots, K_{\max} - 1$ **do**
 - 5: Estimate $\hat{a}, \hat{c}, \hat{e}, \hat{d}$ using (17) and (18).
 - 6: Compute $\hat{\mathcal{J}}_{\log}(\boldsymbol{\theta}^{(k)})$.
 - 7: **if** $k > 0$ **and** $|\hat{\mathcal{J}}_{\log}^{(k)} - \mathcal{J}_{\text{prev}}| \leq tol$ **then**
 - 8: **break**
 - 9: **end if**
 - 10: $\boldsymbol{\theta}^{(k+1)} \leftarrow \text{Opt}(\hat{\mathcal{J}}_{\log}, \boldsymbol{\theta}^{(k)})$; $\mathcal{J}_{\text{prev}} \leftarrow \hat{\mathcal{J}}_{\log}^{(k)}$.
 - 11: **end for**
 - 12: Set $\boldsymbol{\theta}^*$ to the final iterate and re-estimate $\hat{a}, \hat{c}, \hat{d}, \hat{e}$.
 - 13: $\hat{t}^* \leftarrow (\hat{c} + \hat{e}) / (\hat{a} + \hat{d})$.
 - 14: **return** $\hat{\mathbf{w}} = \hat{t}^* \phi(\boldsymbol{\theta}^*)$
-

The complete hybrid optimization and reconstruction procedure is summarized in Algorithm 1. Starting from an initial parameter vector $\boldsymbol{\theta}^{(0)}$, each objective evaluation prepares the variational

state $|\phi(\boldsymbol{\theta})\rangle$ and estimates $\hat{a}$, $\hat{c}$, $\hat{d}$, and $\hat{e}$ using the overlap and Pauli-measurement routines described above. These estimates are then combined to compute $\hat{\mathcal{J}}_{\log}(\boldsymbol{\theta})$, which is supplied to a classical optimizer. The optimizer proposes updated circuit parameters, and the quantum-measurement and classical-update steps are repeated until the change in the objective falls below tol or the maximum number of objective evaluations $K_{\max}$ is reached. In the reported experiments, we use the derivative-free COBYLA optimizer, so the procedure requires only objective-function evaluations and does not require analytic gradients or parameter-shift measurements. After optimization, the objective components are re-estimated at $\boldsymbol{\theta}^*$, and the scale and regression-weight estimate are reconstructed as specified in Algorithm 1.

3.4 Resource Comparison

Table 1 distinguishes classical arithmetic, circuit shots, and state preparation. Matrix-free CG applies $\mathbf{X}^\top(\mathbf{X}\mathbf{v}) + \mathbf{M}\mathbf{v}$ without constructing $\mathbf{X}^\top\mathbf{X}$, costing $\mathcal{O}(ND + D)$ per iteration for diagonal $\mathbf{M}$. Mini-batch SGD similarly avoids the Gram matrix and costs $\mathcal{O}(bD)$ per step. These are strong classical baselines, and VQBR is not claimed to dominate them in runtime.

For the diagonal prior used in all experiments, $\mathbf{M} = V\boldsymbol{\Sigma}_0^{-1}$ and $\mathbf{c} = \mathbf{M}\mathbf{m}_0$ are formed in $\mathcal{O}(D)$ time and require at most $L = 2^d = D$ diagonal Pauli strings for the power-of-two feature dimensions considered here. Each VQBR objective evaluation uses $N+L+1$ circuit families with S shots each. Generic exact amplitude-state preparation can require $\mathcal{O}(D)$ gates [19] and may dominate the shallow trainable ansatz. Dense, unstructured priors would require larger preprocessing and Pauli-decomposition costs and are outside the present validation.

Table 1: Resource comparison for the diagonal-prior setting used in this work. K_{CG} , K_{SGD} , and K_V denote iteration or objective-evaluation counts; b is mini-batch size.

Method	Dominant resource count
Dense direct MAP	$\mathcal{O}(ND^2 + D^3)$ classical operations
Matrix-free CG	$\mathcal{O}(K_{\text{CG}}(ND + D))$ classical operations
Mini-batch SGD	$\mathcal{O}(K_{\text{SGD}}bD)$ classical operations
VQBR	$\mathcal{O}(D)$ prior setup and $K_V(N + L + 1)S$ circuit shots; generic exact state preparation may require $\mathcal{O}(D)$ gates per circuit

4 Numerical Studies

4.1 Experimental Setup

Datasets **Synthetic:** $\mathbf{X} \in \mathbb{R}^{N \times D}$ has i.i.d. $\mathcal{N}(0, 1)$ entries, $\mathbf{y} = \mathbf{X}\mathbf{w}_{\text{true}} + \boldsymbol{\eta}$, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, 0.01\mathbf{I})$, and $\mathbf{w}_{\text{true}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, with $N=200$ and $D \in \{8, 16, 32\}$. **Energy Efficiency** [20]: we predict heating load (target Y1) from eight building-design variables, giving $(N, D)=(768, 8)$. Each dataset is split into prior/train/test subsets at 0.2:0.6:0.2. The prior subset estimates $(\mathbf{m}_0, \boldsymbol{\Sigma}_0, V)$ through bootstrap ridge regression ($B=200$, $\lambda=10^{-3}$), avoiding test leakage. A diagonal $\boldsymbol{\Sigma}_0$ limits the Pauli count to $L \leq 2^d$; because all tested feature dimensions are powers of two, $2^d = D$ in our experiments.

Implementation VQBR is implemented in Qiskit using statevector simulation and IBM Quantum hardware. Statevector experiments use exact expectation values and the derivative-free COBYLA

optimizer [21,22], with $K_{\max}=1000$ and $tol=2\times 10^{-3}$. The ansatz is **EfficientSU2** with R_y rotations, linear CNOT entanglement, and p entangling repetitions (Fig. 2). With the final rotation layer retained, it has $(p+1)d$ trainable angles; default experiments use $p=2$. The reference denoted CF is the MAP solution in (2) computed through an SVD-based linear solve rather than an explicit inverse. All simulation results are averaged over 20 random seeds.

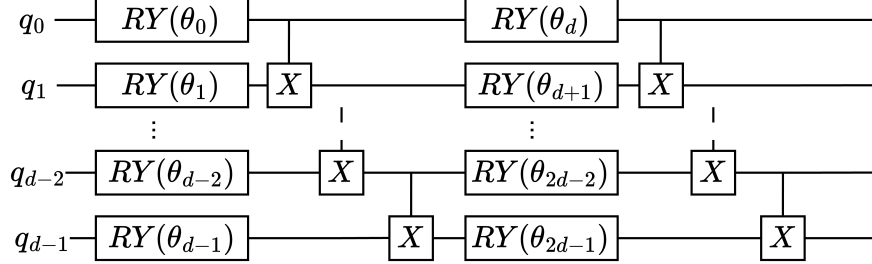


Figure 2: Ansatz $U(\theta)$: **EfficientSU2** with $p=2$ entangling repetitions, R_y rotations, and linear CNOT entanglement on $d=\lceil \log_2 D \rceil$ qubits.

Metrics We report (i) cosine similarity between $\phi(\theta^*)$ and $\mathbf{w}^*/\|\mathbf{w}^*\|_2$, and (ii) RMSE and R^2 from $\hat{\mathbf{w}} = \hat{t}^*\phi(\theta^*)$, where $\hat{t}^* = (\hat{c} + \hat{e})/(\hat{a} + \hat{d})$.

Scope of validation The experiments ask: (Q1) Does VQBR recover the direct MAP direction at the tested dimensions? (Q2) How does ansatz depth affect recovery as D increases? (Q3) Can the required circuits be executed on current hardware, and does ZNE improve this proof-of-concept run? Comparisons against CG, LSQR, SGD, and random-feature models, together with larger real datasets, are reserved for the extended version.

4.2 Main Results

Table 2: VQBR versus the SVD-based direct MAP solve (CF). Simulation entries are mean $\pm$ standard deviation over 20 seeds, not confidence intervals; thus mean $\pm$ standard deviation may cross a metric’s natural bound even though each individual $R^2 \leq 1$.

Dataset	(N, D)	Method	Cosine $\uparrow$	Train RMSE $\downarrow$	Test RMSE $\downarrow$	Train $R^2 \uparrow$	Test $R^2 \uparrow$
Synthetic	(200, 8)	CF	—	0.097 ± 0.008	0.100 ± 0.011	0.998 ± 0.001	0.998 ± 0.002
		VQBR	0.998 ± 0.005	0.135 ± 0.105	0.139 ± 0.108	0.995 ± 0.010	0.995 ± 0.009
	(200, 16)	CF	—	0.093 ± 0.006	0.104 ± 0.010	0.999 ± 0.000	0.999 ± 0.000
		VQBR	0.952 ± 0.029	0.946 ± 0.320	1.104 ± 0.391	0.928 ± 0.040	0.901 ± 0.068
	(200, 32)	CF	—	0.100 ± 0.009	0.132 ± 0.017	1.000 ± 0.000	0.999 ± 0.000
		VQBR	0.820 ± 0.040	2.644 ± 0.391	3.119 ± 0.558	0.772 ± 0.052	0.652 ± 0.129
Energy	(768, 8)	CF	—	2.925 ± 0.063	2.943 ± 0.162	0.916 ± 0.005	0.914 ± 0.011
		VQBR	0.927 ± 0.039	2.958 ± 0.064	2.980 ± 0.173	0.914 ± 0.005	0.912 ± 0.012

Table 2 reports the main results and Fig. 3 shows convergence.

Q1: MAP-direction recovery at $D=8$. VQBR obtains cosine similarity 0.998 ± 0.005 on the synthetic benchmark and 0.927 ± 0.039 on Energy Efficiency. On Energy, its test RMSE and

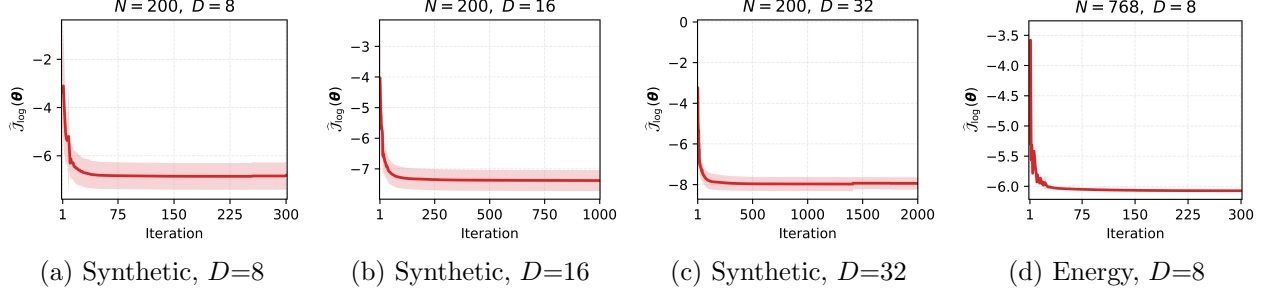


Figure 3: Convergence of $\hat{\mathcal{J}}_{\log}(\theta)$. Shading indicates ± 1 standard deviation over 20 seeds.

R^2 are 2.980 ± 0.173 and 0.912 ± 0.012 , compared with 2.943 ± 0.162 and 0.914 ± 0.011 for CF. These results support accurate recovery at the tested low-dimensional settings, but do not imply that every unrestricted MAP direction is reachable by every ansatz.

Q2: Expressivity–dimension trade-off. For the fixed $p=2$ ansatz, cosine similarity is 0.998, 0.952, and 0.820 at $D=8, 16, 32$, respectively. The $D=16$ depth ablation below indicates that much of the loss at that dimension is due to ansatz expressivity. We have not demonstrated the same recovery at $D=32$, so the sweep is an expressivity study rather than a general scalability result.

4.3 Ablation Studies

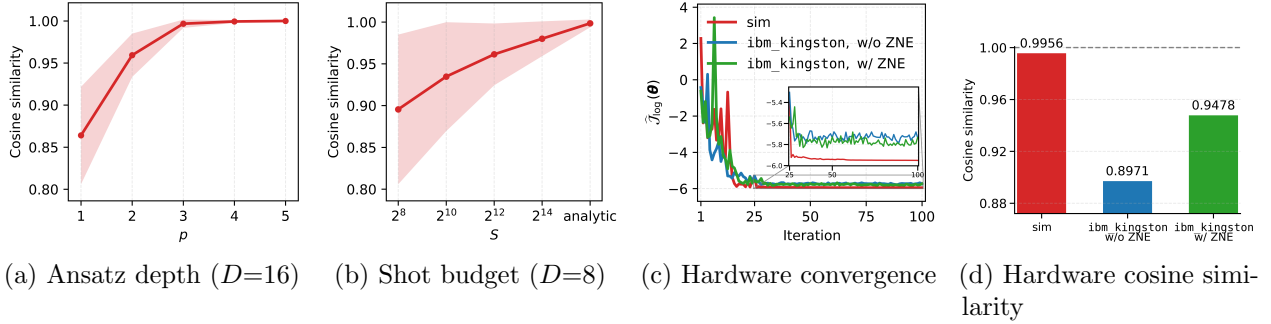


Figure 4: Ablations. (a) Increasing p improves recovery at $D=16$. (b) Accuracy increases with shot budget. (c–d) ZNE reduces the hardware–simulation cosine gap in one hardware run.

Ansatz expressivity For $N=200, D=16$, cosine similarity improves from 0.864 ± 0.058 at $p=1$ to 0.959 ± 0.026 at $p=2$, 0.997 ± 0.005 at $p=3$, 0.999 ± 0.002 at $p=4$, and 0.9999 ± 0.0001 at $p=5$. Test RMSE falls from 1.814 to 0.107, matching CF. Thus, at $D=16$, greater depth recovers nearly perfect direction alignment; this does not establish that arbitrary D can be handled without additional trainability or hardware costs.

Shot noise For $N=200, D=8$, increasing $S \in \{256, 1024, 4096, 16384\}$ raises cosine similarity from 0.895 ± 0.090 to 0.980 ± 0.021 , approaching the analytic baseline 0.998 ± 0.005 . The trend is consistent with the $\mathcal{O}(S^{-1/2})$ sampling uncertainty of overlap estimators [23]; $S=4096$ already exceeds 0.96 cosine similarity.

Proof-of-concept on IBM hardware We execute one run on `ibm_kingston` with $(N, D)=(50, 8)$ and $S=1024$. The circuits use $d+1=4$ qubits. The trainable `EfficientSU2` component contributes $\mathcal{O}(pd)$ entangling gates, while controlled data-state preparation adds data-dependent gates and can require $\mathcal{O}(D)$ gates for generic exact loading. The mitigation configuration uses IBM Runtime resilience level 2, probabilistic error amplification with noise factors $\{1, 3, 5\}$, linear extrapolation, measurement mitigation, and gate/measurement twirling with 32 randomizations.

Statevector, unmitigated hardware, and ZNE produce cosine similarities 0.9956, 0.8871, and 0.9478, respectively. ZNE reduces the hardware–simulation gap from 0.1085 to 0.0478, a 56% reduction on this single run. This establishes circuit executability and mitigation compatibility, not statistically robust hardware performance or scalability beyond $D=8$.

5 Discussion and Limitations

Position relative to classical solvers. VQBR presently has no demonstrated runtime advantage over matrix-free CG, LSQR, or mini-batch SGD. Its structural distinction is that a D -component candidate direction is stored in the amplitudes of $\lceil \log_2 D \rceil$ qubits and its objective is evaluated through measurement circuits without block-encoding the MAP system. A shallow ansatz supplies a compact but restricted parameterization; logarithmic qubit count alone implies neither logarithmic runtime nor universal representability. A matched benchmark must include wall-clock time, state preparation, preconditioning, stopping tolerance, and solution quality.

Random-feature comparison. Random Fourier features (RFF) approximate shift-invariant kernels with a randomized finite-dimensional feature map [24]. RFF therefore changes the model representation, whereas VQBR acts on a given BLR design matrix and restricts the weight direction through a quantum ansatz. A matched RFF-based Bayesian-regression comparison is left to the extended version.

Barren plateaus. Deep variational circuits may exhibit exponentially vanishing gradients [7]. We observe no plateau at $p=5$, $D=16$, but trainability may worsen with dimension. Problem-informed ansatzes and local-cost strategies are natural mitigations.

Limitations. The present framework (i) estimates a MAP point and not posterior predictive variance; (ii) is validated only to $D=32$ in simulation and $D=8$ in one hardware run; (iii) depends on ansatz representability and optimizer performance; (iv) incurs state-preparation and finite-shot costs; and (v) lacks matched empirical comparisons with matrix-free classical solvers and random-feature models.

6 Conclusion

We introduced Variational Quantum Bayesian Regression (VQBR), a hybrid NISQ formulation for BLR. Its core is an analytic scale elimination that introduces no additional scale-optimization error within the chosen ansatz family, followed by a Gram-matrix-free estimator using row-wise Hadamard tests and Pauli measurements on $\lceil \log_2 D \rceil + 1$ qubits. VQBR reaches 0.998 ± 0.005 cosine similarity at $D=8$; at $D=16$, increasing ansatz depth yields 0.9999 ± 0.0001 . Finite-shot experiments show graceful degradation, and one `ibm_kingston` run shows that ZNE reduces the hardware–simulation gap.

Future work will conduct matched comparisons with CG, LSQR, SGD, and random-feature regression; test larger-feature real datasets and repeated hardware runs; derive nonasymptotic finite-shot bounds; develop problem-informed ansatzes; and extend VQBR from MAP estimation to posterior or predictive uncertainty.

References

- [1] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006. [Online]. Available: <https://link.springer.com/book/9780387310732>
- [2] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: The MIT Press, 2012. [Online]. Available: <https://dl.acm.org/doi/10.5555/2380985>
- [3] A. N. Tikhonov and V. Y. Arsenin, “Solutions of ill-posed problems,” *SIAM Review*, vol. 21, no. 2, pp. 266–267, 1979. [Online]. Available: <https://doi.org/10.1137/1021044>
- [4] A. E. Hoerl and R. W. Kennard, “Ridge regression: biased estimation for nonorthogonal problems,” *Technometrics*, vol. 42, no. 1, p. 80–86, Feb. 2000. [Online]. Available: <https://doi.org/10.2307/1271436>
- [5] M. R. Hestenes and E. Stiefel, “Methods of conjugate gradients for solving linear systems,” *Journal of Research of the National Bureau of Standards*, vol. 49, no. 6, pp. 409–436, 1952. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/jres/049/jresv49n6p409_a1b.pdf
- [6] C. C. Paige and M. A. Saunders, “Lsqr: An algorithm for sparse linear equations and sparse least squares,” *ACM Trans. Math. Softw.*, vol. 8, no. 1, p. 43–71, Mar. 1982. [Online]. Available: <https://doi.org/10.1145/355984.355989>
- [7] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, “Variational quantum algorithms,” *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, Sep 2021. [Online]. Available: <https://doi.org/10.1038/s42254-021-00348-9>
- [8] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, “Noisy intermediate-scale quantum algorithms,” *Rev. Mod. Phys.*, vol. 94, p. 015004, Feb 2022. [Online]. Available: <https://link.aps.org/doi/10.1103/RevModPhys.94.015004>
- [9] J. Preskill, “Quantum Computing in the NISQ era and beyond,” *Quantum*, vol. 2, p. 79, Aug. 2018. [Online]. Available: <https://doi.org/10.22331/q-2018-08-06-79>
- [10] N. Wiebe, D. Braun, and S. Lloyd, “Quantum algorithm for data fitting,” *Phys. Rev. Lett.*, vol. 109, p. 050505, Aug 2012. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.109.050505>
- [11] M. Schuld, I. Sinayskiy, and F. Petruccione, “Prediction by linear regression on a quantum computer,” *Phys. Rev. A*, vol. 94, p. 022342, Aug 2016. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.94.022342>
- [12] G. Wang, “Quantum algorithm for linear regression,” *Phys. Rev. A*, vol. 96, p. 012335, Jul 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.96.012335>
- [13] S. Chakraborty, A. Morolia, and A. Peduri, “Quantum Regularized Least Squares,” *Quantum*, vol. 7, p. 988, Apr. 2023. [Online]. Available: <https://doi.org/10.22331/q-2023-04-27-988>
- [14] A. Gilyén, Z. Song, and E. Tang, “An improved quantum-inspired algorithm for linear regression,” *Quantum*, vol. 6, p. 754, Jun. 2022. [Online]. Available: <https://doi.org/10.22331/q-2022-06-30-754>

- [15] Z. Song, J. Yin, and R. Zhang, “Revisiting quantum algorithms for linear regressions: Quadratic speedups without data-dependent parameters,” 2023, arXiv:2311.14823. [Online]. Available: <https://arxiv.org/abs/2311.14823>
- [16] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, “Variational Quantum Linear Solver,” *Quantum*, vol. 7, p. 1188, Nov. 2023. [Online]. Available: <https://doi.org/10.22331/q-2023-11-22-1188>
- [17] M. Benedetti, B. Coyle, M. Fiorentini, M. Lubasch, and M. Rosenkranz, “Variational inference with a quantum computer,” *Phys. Rev. Appl.*, vol. 16, p. 044057, Oct 2021. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevApplied.16.044057>
- [18] F. Rapp and M. Roth, “Quantum gaussian process regression for bayesian optimization,” *Quantum Machine Intelligence*, vol. 6, no. 1, p. 5, Jan 2024. [Online]. Available: <https://doi.org/10.1007/s42484-023-00138-9>
- [19] M. Möttönen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, “Transformation of quantum states using uniformly controlled rotations,” *Quantum Info. Comput.*, vol. 5, no. 6, p. 467–473, Sep. 2005. [Online]. Available: <https://dl.acm.org/doi/abs/10.5555/2011670.2011675>
- [20] A. Tsanas and A. Xifara, “Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools,” *Energy and Buildings*, vol. 49, pp. 560–567, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S037877881200151X>
- [21] X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Bäck, V. Dunjko, and T. E. O’Brien, “Performance comparison of optimization methods on variational quantum algorithms,” *Phys. Rev. A*, vol. 107, p. 032407, Mar 2023. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.107.032407>
- [22] H. Singh, S. Majumder, and S. Mishra, “Benchmarking of different optimizers in the variational quantum algorithms for applications in quantum chemistry,” *The Journal of Chemical Physics*, vol. 159, no. 4, p. 044117, 07 2023. [Online]. Available: <https://doi.org/10.1063/5.0161057>
- [23] K. Plekhanov, M. Rosenkranz, M. Fiorentini, and M. Lubasch, “Variational quantum amplitude estimation,” *Quantum*, vol. 6, p. 670, Mar. 2022. [Online]. Available: <https://doi.org/10.22331/q-2022-03-17-670>
- [24] A. Rahimi and B. Recht, “Random features for large-scale kernel machines,” in *Advances in Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds., vol. 20. Curran Associates, Inc., 2007. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2007/file/013a006f03dbc5392effeb8f18fda755-Paper.pdf