Information Theoretic Error Bounds on NISQ Learning Systems

B.Tech Project I

Sankalp Gambhir

Advisor: Prof. Sai Vinjanampathy

December 2, 2021

Noisy Intermediate-Scale Quantum Devices

- Near term devices
- Not large enough to do error checking and correction — 'noisy'
- Few hundred qubits expected — 'intermediate-scale'
- Low coherence times

These properties leave devices manufacturable in the forseeable future unable to live up to the 'quantum hype', atleast in a purist's mind.

NISQ Devices — Utility

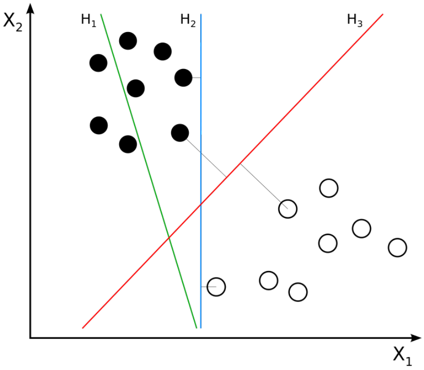
What are NISQ devices good for then?

- Computations resilient to noise
- Computations with low processing times

This suggests, rather than using these devices for independent computation, perhaps it's better to use them to run subroutines. Any actual examples to speak of?

(Supervised) Classification Problem

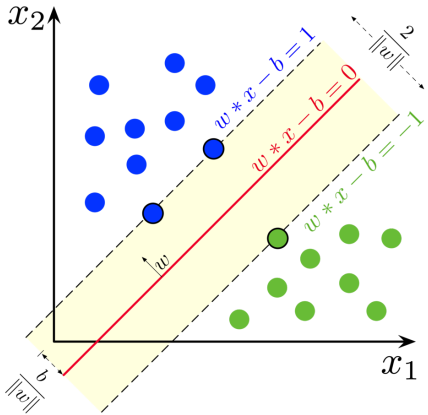
Given a set of input vectors with labels $\vec{x_i}, y_i$, learn a model, and attempt to predict the labels for arbitrary inputs. [1]



[1] Image: `User:ZackWeinberg` on Wikimedia Commons,

Support Vector Machines (SVM)

Support Vector Machine or Maximum Margin Classifier attempts to find a separating plane between labels and optimizes the margin, i.e., distance from inputs on either side to it. [2]



Hyperplane characterized by a normal vector and a bias $(\vec{w}, b) \in \mathbb{R}^{n+1}$.

$$\langle \vec{w}, \vec{x_i} \rangle + b \geq 1 \text{ if } y_i = 1 \text{ , and}$$
$$\langle \vec{w}, \vec{x_i} \rangle + b \leq -1 \text{ if } y_i = -1 \text{ .}$$

We minimize

$$\mathcal{L}(\vec{w}, \vec{\alpha}) = \frac{1}{2} \langle \vec{w}, \vec{w} \rangle + \sum_i \alpha_i \left[y_i \cdot \langle \vec{w}, \vec{x_i} \rangle \right] \text{ .}$$

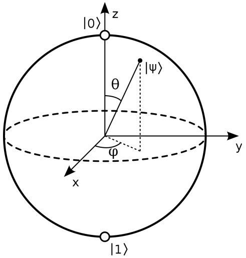
[2] Image: `User:Larhmam` on Wikimedia Commons, CC BY-SA 3.0

Optimization is hard

Common techniques — gradient descent, Hessian-based descent, etc. Lots of linear algebraic computation!
Matrix multiplication: $\mathcal{O}\left(n^{2.37}\right)$
How many dimensions do we have? Consider a simple case of classifying 200×200 sized images, 40,000 dimensional linear algebra!

Quantum Relief?

Qubits scale exponentially in the amount of information they can contain. An n-qubit state can carry information equivalent to $\sim 2^n$ complex numbers!



$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,2^n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,2^n} \\ \vdots & \ddots & \ddots & \vdots \\ c_{2^n,1} & c_{2^n,2} & \cdots & c_{2^n,2^n} \end{bmatrix}$$

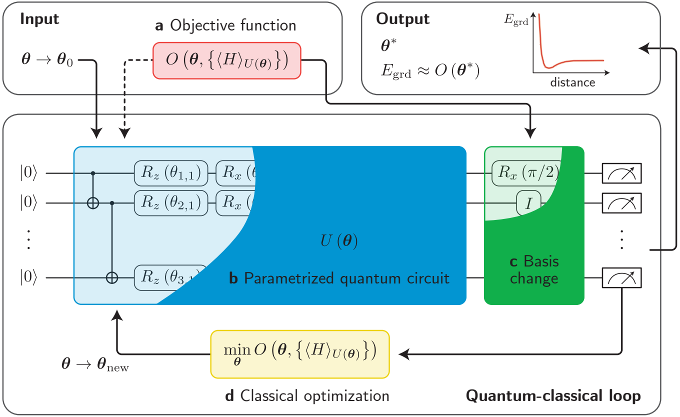
The scale of computation is suddenly reduced. Instead of 40,000 dimension classical computation, we may only need $\lceil \log_2(40000) \rceil = 16$ qubit sized computation systems.

Quantum, sure, but NISQ?

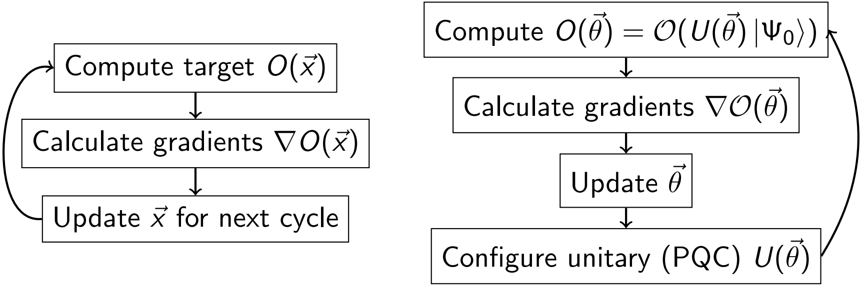
In the near term, it seems the required number of qubits to outpace classical computers may not be out of reach. But can we reliably perform those computations on our quantum computers? Short coherence times make this impossible to do directly. Perhaps the constrained calculations can be processed as a subroutine? Yes, with *Variational Quantum Algorithms*!

Variational Quantum Algorithms

Variational Quantum Algorithms (VQA) form the idea of using a proposed architecture for generalised learning problems on NISQ systems [1].



Transitioning



$$\vec{\theta} \in \mathbb{R}^M, \vec{x} \in \mathbb{R}^N, M \sim \lceil \log_2 N \rceil$$

Parametrized Quantum Circuits (PQCs)

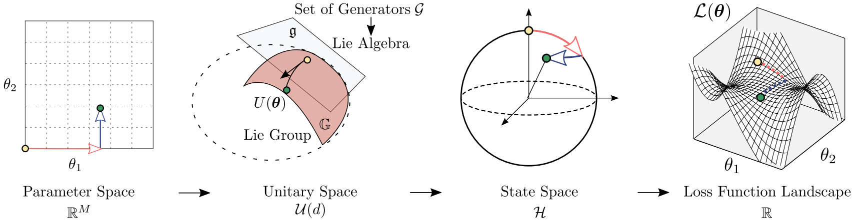
The actual unitary computation happens in an L-layered structure with the mathematical form

$$U(\vec{\theta}) = \prod_{l=1}^{L} U_l(\vec{\theta}_l), \quad U_l(\vec{\theta}) = \prod_{k=1}^{K} e^{-\iota \theta_{lk} H_k} \quad . \tag{1}$$

for a set of generators $\{H_k\}$ with $\vec{\theta} = (\vec{\theta}_1, \vec{\theta}_2, \ldots, \vec{\theta}_k)$.
Formally, the generators create a Dynamical Lie Algebra, which determines the set of reachable unitaries. The choice of generators can strongly affect the optimization process [2]. See the report for a detailed study.

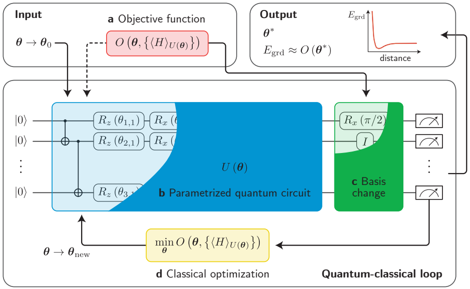
Quantum Landscape Theory

Spaces and maps involved in VQA calculation [3]:



This structure forms the playground for our analysis. The study of these spaces forms an important part of establishing relative bounds at stages of the calculation (which are the spaces). We discuss the particular case of $\mathcal{U}(d) \to \mathcal{H}$ here. See the report for details on the other maps.

Information Theoretic Limits



The classical system transferring information about the parameters to the PQC has a limit on how much information it can transfer in a given time T given by

$$b = T \Delta\Omega \kappa_s$$

where $\Delta\Omega$ is the channel bandwidth and κ_s is the bit depth of the control signal (amount of information *per* trasnfer).

Example — Bounds on Quantum Optical Control

For the specific example of a control pulse for a quantum system

$$\dot{\rho} = \mathcal{L}(\rho, \gamma(t)) \ ,$$

where ρ is the state density matrix, $\mathcal{L}$ is the Liouvillian operator, and $\gamma(t)$ is the control pulse. We have the Hamiltonian

$$\hat{H} = \hat{H}_D + \gamma(t) \cdot \hat{H}_C \ .$$

Example — Bounds on Quantum Optical Control

We have

$$\kappa_s = \log\left(1 + \frac{\Delta\gamma}{\delta\gamma}\right) ,$$

where $\Delta\gamma$ and $\delta\gamma$ are the maximum and minimum variation in the control field.
We have the error bound [4] $\|\rho - \rho_*\| > \epsilon$, with

$$\epsilon \geq 2^{-\frac{T\Delta\Omega\kappa_s}{\mathcal{D}_{\mathcal{W}^+}}} ,$$

where $\mathcal{D}_{\mathcal{W}^+}$ is the dimension of the relevant (polynomially reachable) space of density matrices.
We intend to employ a similar technique to establish bounds on PQC errors.

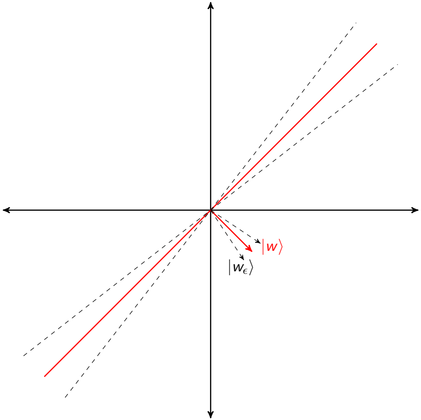
Quantum SVM

Given training data embedded as n-qubit quantum states $\{\,|x_i\rangle\}$ with corresponding labels $y_i = \pm 1$, a QSVM implemented as a VQA attempts to learn a unitary $U(\vec{\theta})$ such that

$$\operatorname{sgn}\ \langle 0|^{\otimes n}\, U(\vec{\theta})^* \,|x_i\rangle = y_i \forall i\ .$$

Setting $|w\rangle = U(\vec{\theta})\,|0\rangle^{\otimes n}$ recovers the familiar classical SVM.

Quantum SVM — Bounds



Picking a point randomly in the space outside the training data and attempting to classify it we find [5]

$$
\begin{aligned}
p_{\text{error}} &= \lim_{r\to\infty} \frac{2 \cdot V_{\text{sector}}(r)}{V_{\text{sphere}}(r)} \\
&= \lim_{r\to\infty} \frac{2 \cdot V_{\text{sphere}}(r) \cdot 0.5 \cdot I_{\sin^2\phi}(\frac{n-1}{2}, \frac{1}{2})}{V_{\text{sphere}}(r)} \\
&= I_{\sin^2\phi}(\frac{n-1}{2}, \frac{1}{2}) \ . \qquad (2)
\end{aligned}
$$

Quantum SVM — Bounds

$$p_{\text{error}} = I_{\sin^2\phi}(\frac{n-1}{2}, \frac{1}{2}) \ , \tag{3}$$

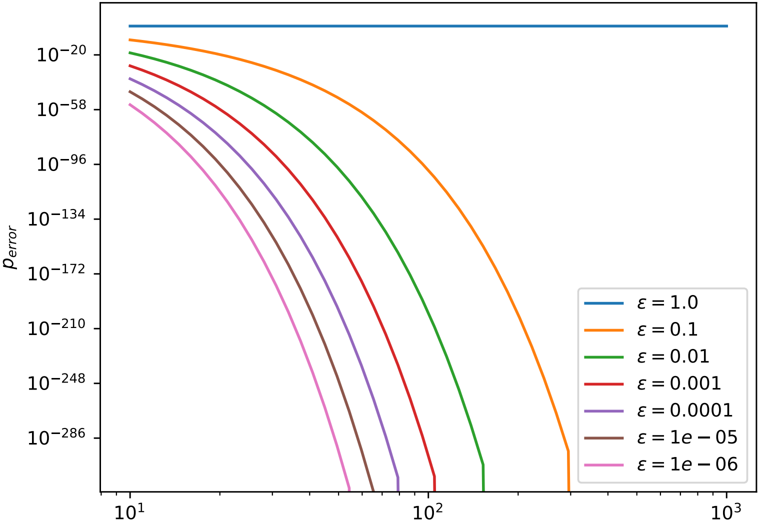
where I is the incomplete Beta function,

$$I_x(a,b) = \frac{B(x;a,b)}{B(a,b)} = \frac{\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t}{\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t} \ , \tag{4}$$

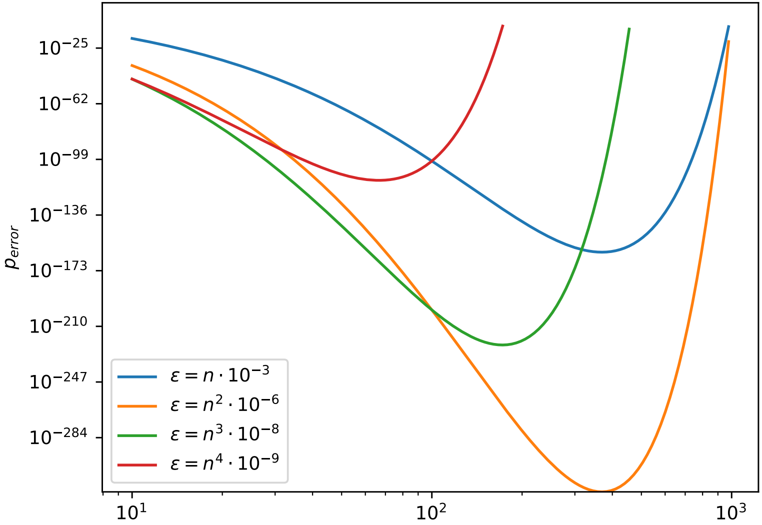
and ϕ is the angular distortion, and it is seen from $|w\rangle = U(\vec{\theta})\,|0\rangle$ that $\sin\phi \sim \epsilon$. Finally,

$$p_{\text{error}} = I_{\epsilon^2}(\frac{n-1}{2}, \frac{1}{2}) \ . \tag{5}$$

Quantum SVM — Bound plots



Quantum SVM — Bound plots



References

[1] Kishor Bharti et al. "Noisy intermediate-scale quantum (NISQ) algorithms". In: *arXiv preprint arXiv:2101.08448* (2021).

[2] Martin Larocca et al. *Diagnosing barren plateaus with tools from quantum optimal control*. 2021. arXiv: `2105.14377 [quant-ph]`.

[3] Martin Larocca et al. "Theory of overparametrization in quantum neural networks". In: *arXiv preprint arXiv:2109.11676* (2021).

[4] S Lloyd and Simone Montangero. "Information theoretical analysis of quantum optimal control". In: *Physical review letters* 113.1 (2014), p. 010502.

[5] Shengqiao Li. "Concise formulas for the area and volume of a hyperspherical cap". In: *Asian Journal of Mathematics and Statistics* 4.1 (2011), pp. 66–70.