2022-04

Information Theoretic Error Bounds on NISQ Learning Systems

B.Tech Project

Sankalp Gambhir

Advisor **Prof. Sai Vinjanampathy**

Email sgambhir@iitb.ac.in

Abstract In this report, we review the classical classification problem, common techniques to solve it, then identify computational bottle necks. We introduce and review Variational Quantum Algorithms (VQA), which form a model of computation on Noisy Intermediate-Scale Quantum (NISQ) computers, to perform relevant computation on quantum computers instead. We present a mathematical review of the structure of involved spaces, and discuss the constraints on computational precision imposed by the architecture. Finally, we discuss these information-theoretic bounds from contexts in optimal control to generalize them to VQAs, and present a resulting bound on Quantum Support Vector Machines (QSVMs).

Todo list

Contents

1 Introduction

There has been long-standing interest in constructing systems capable of learning from experience since even before computers in their modern form have existed. In the last few decades, with computing power skyrocketing exponentially coupled with leaping advances in theory of learning systems and statistical inference, these problems became tractable and eventually came into use ubiquitously. With applications ranging from image recognition systems for surveillance to identifying cosmic objects for astrophysical applications, they have found widespread adoption in industry and academia. These systems excel in problems where producing a precise mathematical model for the problem at hand is intractable, exploiting general techniques to instead infer a model from available data. With their advent, however, has come an ever rising need for computing power to facilitate their operation. This has found data centers of unprecedented scales consuming enormous amounts of power to provide the instant predictions we've come to rely on.

With snowballing energy and space requirements of classical computers in the form of GPU clusters and Application Specific Integrated-Circuits (ASICs), there has been a spark of interest in offloading this computation onto quantum computers, which, till recently, have largely remained a rare species spotted only in labs surrounded by helium-cooled superconductors and white-coated predators. Current scales of available quantum computers (100 physical qubits), however, still lack the power required to fully tackle these challenges while maintaining reliable error-levels or adding their own error checking and correction. This has motivated using quantum computers to run bottle necked computational subroutines with classical control systems. As such, these systems generally lack error correction, and thus earn themselves the title of 'noisy'. These form the basis of computation considered in this thesis, Noisy Intermediate-Scale Quantum (NISQ) computers.

However, making these computers fault-tolerant in practice has been a tall order, and seems to be at least a few years away. The strategy today is thus to explore how we can use NISQ systems to achieve a quantum advantage. However, working with NISQ, one must account for the limited number and connectivity of qubits, as well as the incoherence issues that limit circuit depth [1]. To tackle these issues and approach a quantum advantage, variational quantum algorithms (VQAs) have emerged as the leading candidate. As an analogue to classical machine learning techniques, they leverage the toolbox of optimization

techniques, outsourcing the memory requirement and parameter control to a classical counterpart.

VQAs are characterized by parametrized quantum circuits, wherein the parameters are controlled by a classical computer running an optimization routine, updating them based on the measurement outcomes of the circuit. The technique has shown great promise and bypasses several of the issues arising from the lack of capability for error checking and correction, in turn arising from the small number of available qubits, in quantum computers expected to be manufactured within the next few years. The bulk of the computation requiring memory, such as parameter updation and gradient computation for optimization is generally performed on the classical controller. Effectively, VQAs trade the coherence required for independent quantum computation against access to classical memory.

Recently, these computers have been developed and used for experiments on supervised learning tasks, either by performing the learning task on a quantum computer, or by using on to speed up subtasks such as kernel-estimation [2, 3]. Recently, there has been significant progress towards solving currently looming problems such as barren plateus [4], and these models have also been show to be reasonably robust and error-tolerant in simulations [5].

Despite the promising advances in VQAs driven by access to a classical puppeteer, the benefits cannot be taken for granted. Communication with a classical system comes at its own cost, one paid in information entropy. These costs have been long studied in purely classical information channels, bounding the 'amount of data' that can be transferred over a channel between two parts of a system communicating with each other. This cost fundamentally limits the size of problems that can be computed by the system. Communicating over a classical channel, there is no basis to expect VQAs to be free from these chains either.

1.1 Outline of New Results

In this thesis, we study the architecture of a VQA, the processes, and mathematical structures involved in its functioning. Finally, we establish an information theoretic uncertainty theorem that bounds the expressiveness of a VQA ansatz choice in practice, limiting the problems computable within VQAs, by establishing a tradeoff with its trainability.

1.2 Structure

In section 2, definitions and relevant results in classical computing, physics, and quantum information are presented. section 3 reviews Variational Quantum Algorithms and their architecture, while section 4 discusses bounds on their expressiveness imposed by the architecture.

2 Preliminaries

2.1 Classical Computing

2.1.1 Optimisation Techniques

The discussion of optimization techniques in classical computing is a long and arduous one. We refer the reader to a common text on the matter for a detailed discussion [6, 7], while reviewing the general ideas briefly here.

The goal of an optimization problem, given (generally) a function $\mathcal{L} : X \to \mathbb{R}$ is to find the minimum value in its range, and sometimes an inverse in the domain, i.e., output a point $x_* \in X$, such that

$$x_* = \arg\min \mathcal{L}(x) ,$$

or the argument to $\mathcal{L}$ which minimizes it. In some problems, a local minima may suffice, and in others, a global minimum may be the requirement. Depending on the nature of the domain X, different techniques may be employed to find x_*. These include gradient descent and its reductions (finite element methods, etc.), Hessian-aided descent, stochastic gradient descent, among others. The rest of the discussion is agnostic of the optimization routine, beyond, of course, the existence of one.

In a NISQ system, there is generally little to no attempt at error correction, and the general goal is to capitalize on what is possible with the short available coherence times, without devoting a majority of the system's resources to error checking and correction. As such, these systems are unable to support high-depth circuits with computationally involved analytical gradient based approaches. An effective optimizer in control of such a temporally-bound circuit

should try to utilize techniques minimizing the number of measurements or function evaluations, as the relevant modules generally form the bottleneck of the computation [see 8, chapter II.D].

2.2 Quantum Regime

2.2.1 Hilbert Space

Definition 1. *A Hilbert space is a vector space $\mathcal{H}$ equipped with an inner product $\langle f, g\rangle \; \forall f, g \in \mathcal{H}$ such that the norm defined by*

$$\|f\| = \sqrt{\langle f, f\rangle}$$

turns $\mathcal{H}$ into a complete metric space [9].

Physical quantities — such as energy, momentum, and position — are represented as operators over a Hilbert space $\mathcal{H}$ to which the wavefunctions belong [10]. For our purposes, we will assume $\mathcal{H}$ to always have as its base field the field of complex numbers, $\mathbb{C}$.

Herein, the complex inner product is assumed to be linear in the second factor, i.e.,

$$\langle f, \lambda g\rangle = \lambda\langle f, g\rangle; \quad \langle \lambda f, g\rangle = \bar{\lambda}\langle f, g\rangle$$

$\forall f, g \in \mathcal{H}$ and $\lambda \in \mathbb{C}$.

For every bounded operator A acting on a Hilbert Space, there is a unique bounded operator A^* called its *adjoint* such that

$$\langle f, Ag\rangle = \langle A^* f, g\rangle \,.$$

We will assume relevant quantities to be linear operators $\cdot : \mathcal{H} \to \mathcal{H}$ with adjoints where necessary, [see 10, Appendix A] for details.

2.2.2 Quantum Computation

In a quantum system, 'computation' in its essence is jugglery of probability amplitudes of states using unitary actions. After action of unitaries as necessary, the coefficients are estimated using a measurement schema and post-processed to recover the computational result. See [11] for a detailed study.

2.2.3 State Construction and Embedding

To perform computation in the quantum regime, data first needs to be converted to a format in which it can be acted upon by quantum operators. This is again an embedding function, specifically from the input domain to the Hilbert space of quantum states for the system ansatz. See [12] for details.

2.2.4 Information Matrices and Distance Measures

Consider a parametrized distance function on quantum states

$$d(\boldsymbol{\theta}, \boldsymbol{\theta}') = d(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta}')) \,. \tag{1}$$

For close enough $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ we can write a Taylor expansion of the distance function. Clearly, the zeroth and first terms vanish, as the distance of a state from itself is zero, and it forms a minimum for the distance function. We have then the second order term

$$d(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta} + \boldsymbol{\delta})) = \frac{1}{2} \boldsymbol{\delta}^\top F(\boldsymbol{\theta}) \boldsymbol{\delta} \,, \tag{2}$$

with F representing the metric tensor

$$F(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \delta_i \partial \delta_j} d(\boldsymbol{\theta}, \boldsymbol{\theta} + \boldsymbol{\delta}) \Big|_{\delta = 0} \,. \tag{3}$$

Therefore, F captures all the information necessary to define a distance metric. F is called the Quantum Fisher Information Matrix [13], and it is defined as

$$\Big[F_{ij}\Big] = 4\mathrm{Re}\Big[\langle \partial_i \psi(\boldsymbol{\theta}) | \partial_j \psi(\boldsymbol{\theta}) \rangle - \langle \partial_i \psi(\boldsymbol{\theta}) | \psi(\boldsymbol{\theta}) \rangle \langle \psi(\boldsymbol{\theta}) | \partial_j \psi(\boldsymbol{\theta}) \rangle\Big] \,. \tag{4}$$

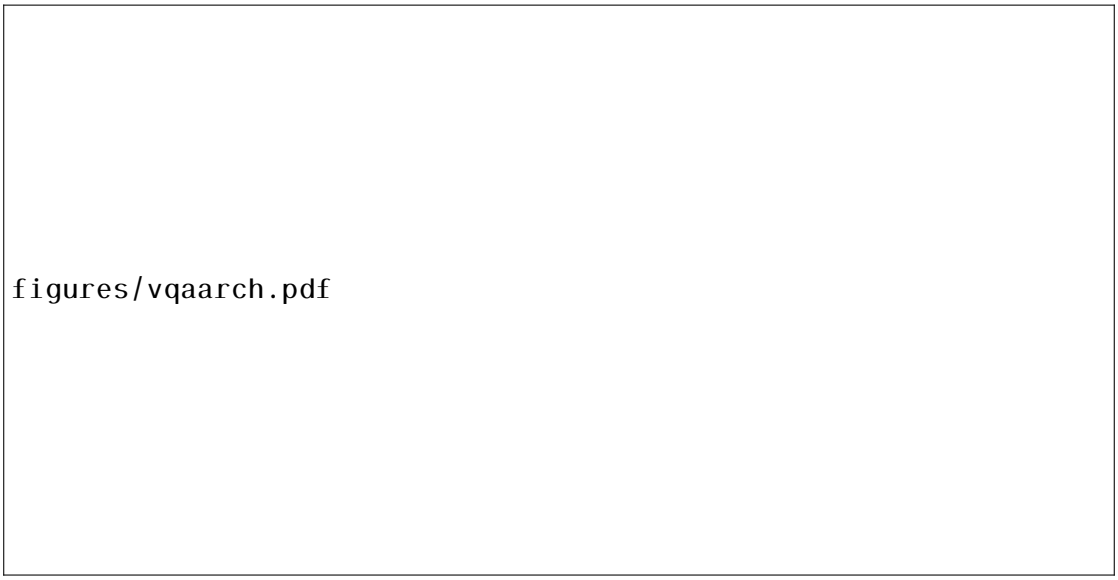

Figure 1: *Diagrammatic representation of a Variational Quantum Algorithm (VQA) [taken from 8, Figure 2].*

3 Variational Quantum Algorithms

bls better

With the goal of optimizing learning computations using quantum computers in mind, we need an abstract idea of how to implement this connection. A *Variational Quantum Algorithm* (VQA) is any such system based on a proposed architecture for a classically controlled quantum computer [8]. Figure 1 presents the proposed architecture. The following subsection presents an expanded view of the computation.

3.1 Building Blocks

A VQA computation has 4 major components, as shown in Figure 1:

- objective function — the encoding of the problem at hand as an optimization,
- parametrized quantum circuit (PQC) — circuit encoding a unitary operator parametrized by classically controlled parameters θ,

- measurement scheme — the system performing basis changes and transferring outputs to the control system, and
- classical optimizer — a classical objective minimizer which controls the PQC.

These components form a modular computation model where each of the components can be swapped and improved individually to relieve bottle necks and adapt to the problem at hand, to control the expressiveness of the system or avoid treacherous optimization landscapes [14].

3.1.1 Objective Function

The *objective* or *loss function* [14] forms the target of the optimization problem at hand. This can be any function that can be encoded in an operational form, i.e., written as or decomposed into quantum operators. In most cases, this can be expected to be something akin to the Hamiltonian of a system [8], thus making the minimal, ground state energy, the optimization target. This may also be called the *parametrized cost* of the computation. Subject to the optimization constraints, the target of the system is then to find the optimal parameter input

$$\boldsymbol{\theta}_* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, p_0(\boldsymbol{\theta}))\,,$$

where $p_0(\boldsymbol{\theta})$ represents the parametrized probability to measure the output in the state $|0\rangle$.

3.1.2 Parametrized Quantum Circuits (PQCs)

The module central to the design of a VQA is the parametrized quantum circuit, denoted by $U(\boldsymbol{\theta})$. It is the component of the circuit which performs the actual 'computation' and outputs the state that best meets the objective. It does so by acting on the input state a series of unitary transformations parametrized by controllable inputs. We assume the circuit to have an L-layered structure as

$$U(\boldsymbol{\theta}) = \prod_{l=1}^{L} U_l(\boldsymbol{\theta}_l), \quad U_l(\boldsymbol{\theta}) = \prod_{k=1}^{K} e^{-i\theta_{lk}H_k}\,, \tag{5}$$

where the index l indicates the layer, and the index k spans the traceless Hermitian operators $\{H_k\}$ that generate the space of unitaries for the chosen ansatz. Here, $\boldsymbol{\theta}$ decomposes as a set of vectors of parameters $\boldsymbol{\theta}_l$ for each of the indexed layers, which in turn map to individual parametrized unitary actions indexed by k. Finally, $M = K \cdot L$ gives the number of trainable parameters of the system [see 14, section II.A]. The experimental apparatus to tune these parameters depends heavily on the hardware design chosen for the PQC, and may be mechanical, electronic, or optoelectronic in practice . Cite

This general description of a PQC subsumes most ansatzes studied in literature [15]. These include the hardware-efficient ansatz [16], quantum alternating operator ansatz (QAOA) [17], Hamiltonian variational ansatz [18], quantum optimal control ansatz [19], among others [20, 21, 22]. These correspond to specific configurations of layer sizes and choices of the generators. This generic hardware structure allows us to use the well-formed foundations of landscape theory (subsection 3.2) to discuss advantages and limitations of VQAs independent of the specific problem being tackled with them.

The choice of generators is intimately tied to the reachable states of the system, and the landscape needed to be traversed to get there. This determines many things about the VQA, including, but not limited to, the problems solvable within the framework, and the time and hardware constraints required to train the system. For further detail, see discussion in subsection 3.2.

Assuming for now that the space spanned by the generators contains our target unitary, we proceed with the discussion of the computation. After the application of the PQC, the initial state $|\Psi_0\rangle$ is transformed as

$$|\Psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\Psi_0\rangle \ . \tag{6}$$

Typically, the input state is chosen to be a zero-valued product state in the computational basis representation, i.e., $|\Psi_0\rangle = |00\ldots00\rangle = |0\rangle^{\otimes n}$. Other choices of the initial state may be made based on the problem requirements, possibly even to depend on some variational parameters itself as $|\Psi_0\rangle = P(\phi)|0\rangle^{\otimes n}$, with $P(\phi)$ a parametrized unitary, and ϕ the set of variational parameters. We discuss these as subjects of study in the future in section 5.

3.1.3 Measurement Scheme

The measurement scheme is chosen to evaluate probabilities and determine the relevant coefficients using basis changes as necessary.

new material bls

3.1.4 Parameter Optimization and Classical Control

After obtaining the loss function from the measurement and post-processing, a classical control system may treat it as output from a black box, at which point the optimizer may be oblivious of the fact the computation is sourced from a quantum computer, and apply any optimization technique of choice. Based on the chosen scheme, the controller readjusts the parameters.

The optimization technique may be a classical one — gradient based, Hessian based, etc — or quantum-aware, taking advantage of the hardware structure, or to combat specific issues such as quantum noise [23, 24, 25, 26, 27].

3.2 Quantum Landscape Theory

Better intro here

In subsubsection 3.1.2, we suggested the problem of the target unitary not existing in the space reachable in our circuit configuration. In this section, we elaborate on this issue, and discuss the related problems of studying the loss landscape, how it emerges, and how it affects the optimization process. We begin with a review of Quantum Landscape Theory [see 14, chapter II.B].

To study the landscapes, one must first be aware of the spaces each of the objects relevant to the computation belong to. This is diagrammatically illustrated in Figure 2. First, the input parameter set $\boldsymbol{\theta}$ is seen as a vector in $\mathbb{R}^M$. The PQC then represents an embedding of $\mathbb{R}^M$ into the unitary space of appropriate size $d \in \mathbb{N}$, $\mathcal{U}(d)$. Its action on the input state is the map $U(\boldsymbol{\theta}) : \mathcal{H} \to \mathcal{H}$. Finally, the measurement scheme maps the output state to a real-valued loss, which is used by the optimizer to recompute the parameters. Succinctly, the action of the model arises from the following transformations

$$\mathbb{R}^M \to \mathcal{U}(d) \to \mathcal{H} \to \mathbb{R}\,. \tag{7}$$

Figure 2: *Relevant mathematical spaces for VQA [taken from 14, Figure 2].*

3.2.1 Parameter Space to Unitary Group ($\mathbb{R}^M \to \mathcal{U}(d)$)

The first map, i.e., the map from the space of parameters to the unitary group, will be the focal point of the rest of this section. The unitaries generated by this map, and thus the chosen ansatz, are characterized by an object called the Dynamical Lie Algebra (DLA) of the system [see 28, chapter 3]. This represents the space formed by (Lie) closure of the individual operators in the architecture, under repeated application and commutation. These operators form the *generators* of the DLA.

Definition 2 (Set of Generators)**.** *Consider a PQC of the form Equation 5. The set of generators $\mathcal{G} = \{H_k\}_{k=0}^{K}$ is defined as the set (of size K) of the Hermitian operators that generate the unitaries in a single layer of $U(\boldsymbol{\theta})$.*

Definition 3 (Dynamical Lie Algebra (DLA))**.** *Given a set of generators $\mathcal{G}$, its DLA $\mathfrak{g}$ is defined as the span of its Lie closure, or the space generated by $\mathcal{G}$ after closure with repeated nested commutation. Mathematically,*

$$\mathfrak{g} = \text{span}\,\langle iH_1, iH_2, \ldots, iH_k \rangle_{Lie}\,,$$

where $\langle S \rangle_{Lie}$ denotes the Lie or the nested-commutator closure of S.

The set of reachable unitaries is then a subset of the Lie group $\mathbb{G}$ generated by $\mathfrak{g}$,

$$\{U(\boldsymbol{\theta})\}_{\boldsymbol{\theta}} \subseteq \mathbb{G} \subseteq \mathcal{SU}(d)\,. \tag{8}$$

$\mathbb{G}$ can also be generated completely from the underlying Lie algebra as $e^{\mathfrak{g}}$.

It would seem at first glance that a configuration of generators should be chosen to be as expressive as possible, which is to have $\mathbb{G}$ be as close to $\mathcal{SU}(d)$ as possible,

however, this often leads to trainability issues such as barren plateaus due to randomly chosen initial parameters [15, 29, 30]. As such, the ansatz is generally either chosen to make the problem convenient, i.e. problem-inspired ansatz [19], or to make the implementation convenient, i.e. hardware-efficient ansatz [31].

3.2.2 Unitary Group to State Space ($\mathcal{U}(d) \to \mathcal{H}$)

Recall that for the map from the unitary group to the state space, the unitary output from the first map acts on states in the input set. Specifically, choosing the input set to be a training set $\mathcal{S} = |\psi_\mu\rangle$. Then, the second map (now parametrized by μ) is defined as

$$U(\boldsymbol{\theta}) \mapsto U(\boldsymbol{\theta})|\psi_\mu\rangle \ . \tag{9}$$

The reachable set from each starting state is called its *orbit*. In many cases, when the states in $\mathcal{S}$ have certain symmetries, the DLA in turn decomposes as the direct sum of the subspaces invariant under the symmetries

$$\mathfrak{g} = \bigoplus_\nu \mathfrak{g}_\nu \ . \tag{10}$$

There is no restriction on whether the states in the training set share or respect any symmetries of the PQC itself. In this way, the DLA serves as a focal point to determine the expressiveness in terms of unitaries as well as the set of reachable states in the Hilbert space.

Next, we wish to see how the output state changes with varying parameters $\boldsymbol{\theta}$. So, consider an infinitesimal perturbation to the parameters $\boldsymbol{\delta} \in \mathbb{R}^M$, and we can then quantify the distance between the initial and perturbed state. Define $|\psi_\mu(\boldsymbol{t})\rangle = U(\boldsymbol{t})|\psi_\mu\rangle \ \forall t \in \mathbb{R}^m$. Writing the distance function (second order) as discussed in subsubsection 2.2.4

$$d(|\psi_\mu(\boldsymbol{\theta})\rangle, |\psi_\mu(\boldsymbol{\theta}+\boldsymbol{\delta})\rangle) = \frac{1}{2}\boldsymbol{\delta}^\top F_\mu(\boldsymbol{\theta})\boldsymbol{\delta} \ , \tag{11}$$

where $F_\mu(\boldsymbol{\theta})$ is the Quantum Fisher Information Matrix (QFIM) for $|\psi_\mu(\boldsymbol{\theta})\rangle$. The QFIM plays a crucial role in quantum-aware optimizers such as the quantum natural gradient descent [24, 25, 26, 27]. Further, the rank of the QFIM quantifies

the number of independent parameters in the state space that changing the parameters can allow us to explore. The rank of the QFIM, observed over the whole parameter space, thus gives us a lower bound on the number of parameters that must be communicated to the PQC to explore the entire unitary space available to it.

3.2.3 State Space to Loss Landscape ($\mathcal{H} \to \mathbb{R}$)

Finally, the loss landscape generated by the composition map is characterized by the classically computed landscape of the real-valued loss function, i.e., using the $M \times M$ Hessian matrix with indexed entries

$$\left[\nabla^2 \mathcal{L}(\boldsymbol{\theta})\right]_{ij} = \partial_i \partial_j \mathcal{L}(\boldsymbol{\theta}) \,. \tag{12}$$

Computing the gradient and Hessian matrix allows us to form a quadratic model of the loss function, with the Hessian's eigenvectors characterizing curvature at each point. The rank of the Hessian once again gives us the number of independent directions explorable by change in parameters, emphasizing similarly how the QFIM functions as a measure of curvature in the state space.

4 Information Theoretic Limits

5 Conclusion and Future Work

In this report we reviewed Variational Quantum Algorithms and optimization theory, finally establishing bounds on the expressiveness achievable by choice of ansatz in a VQA. These bounds fundamentally limit the set of problems finding whose solutions within the VQA framework may be tractable in practice. We are continuing work on studying specific ansatzes and the limits imposed on them to develop a precise idea of the intractable problems. The effects of this bound on the possibility of achieving a quantum advantage on NISQ systems remains unclear, and is a subject for further study. A more mathematically precise discussion of the landscapes generated within the framework, and a study of their transformations is expected to allow a more through understanding of their limits at each stage.

Acknowledgements

I am very grateful to Prof. Sai Vinjanampathy for his unrelenting support in matters academic and otherwise. This project would not have been possible without his guidance. I thank Karthik Dasigi and Drishti Baruah for their constant involvement in discussions regarding the project and report.

References

[1] M. Cerezo et al. 'Variational quantum algorithms'. In: *Nature Reviews Physics* 3.9 (Sept. 2021), pp. 625–644. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00348-9. URL: https://doi.org/10.1038/s42254-021-00348-9.

[2] Vojtěch Havlíček et al. 'Supervised learning with quantum-enhanced feature spaces'. In: *Nature* 567.7747 (Mar. 2019), pp. 209–212. ISSN: 1476-4687. DOI: 10.1038/s41586-019-0980-2. URL: https://doi.org/10.1038/s41586-019-0980-2.

[3] Edward Farhi and Hartmut Neven. *Classification with Quantum Neural Networks on Near Term Processors*. 2018. DOI: 10.48550/ARXIV.1802.06002. URL: https://arxiv.org/abs/1802.06002.

[4] Kaining Zhang et al. *Toward Trainability of Quantum Neural Networks*. 2020. DOI: 10.48550/ARXIV.2011.06258. URL: https://arxiv.org/abs/2011.06258.

[5] Maria Schuld et al. 'Circuit-centric quantum classifiers'. In: *Phys. Rev. A* 101 (3 Mar. 2020), p. 032308. DOI: 10.1103/PhysRevA.101.032308. URL: https://link.aps.org/doi/10.1103/PhysRevA.101.032308.

[6] Stephen Boyd, Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[7] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

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

[9] Giovanni Sansone. *Orthogonal functions*. Vol. 9. Courier Corporation, 1959.

[10] Brian C Hall. *Quantum theory for mathematicians*. Vol. 267. Springer, 2013.

[11] Michael A Nielsen and Isaac Chuang. *Quantum computation and quantum information*. 2002.

[12] Seth Lloyd et al. *Quantum embeddings for machine learning*. 2020. arXiv: 2001.03622 [quant-ph].

[13] Johannes Jakob Meyer. 'Fisher Information in Noisy Intermediate-Scale Quantum Applications'. In: *Quantum* 5 (Sept. 2021), p. 539. ISSN: 2521-327X. DOI: 10.22331/q-2021-09-09-539. URL: http://dx.doi.org/10.22331/q-2021-09-09-539.
[14] Martin Larocca et al. 'Theory of overparametrization in quantum neural networks'. In: *arXiv preprint arXiv:2109.11676* (2021).
[15] Martin Larocca et al. *Diagnosing barren plateaus with tools from quantum optimal control*. 2021. arXiv: 2105.14377 [quant-ph].
[16] Abhinav Kandala et al. 'Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets'. In: *Nature* 549.7671 (2017), pp. 242–246.
[17] Edward Farhi, Jeffrey Goldstone and Sam Gutmann. 'A quantum approximate optimization algorithm'. In: *arXiv preprint arXiv:1411.4028* (2014).
[18] Dave Wecker, Matthew B Hastings and Matthias Troyer. 'Progress towards practical quantum variational algorithms'. In: *Physical Review A* 92.4 (2015), p. 042303.
[19] Alexandre Choquette et al. 'Quantum-optimal-control-inspired ansatz for variational quantum algorithms'. In: *Physical Review Research* 3.2 (2021), p. 023092.
[20] Stuart Hadfield et al. 'From the quantum approximate optimization algorithm to a quantum alternating operator ansatz'. In: *Algorithms* 12.2 (2019), p. 34.
[21] Linghua Zhu et al. 'An adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer'. In: *arXiv preprint arXiv:2005.10258* (2020).
[22] Juneseo Lee et al. 'Progress toward favorable landscapes in quantum combinatorial optimization'. In: *Physical Review A* 104.3 (2021), p. 032401.
[23] Wim Lavrijsen et al. 'Classical Optimizers for Noisy Intermediate-Scale Quantum Devices'. In: (Apr. 2020). URL: https://www.osti.gov/biblio/1615327.
[24] James Stokes et al. 'Quantum natural gradient'. In: *Quantum* 4 (2020), p. 269.
[25] Bálint Koczor and Simon C Benjamin. 'Quantum natural gradient generalised to non-unitary circuits'. In: *arXiv preprint arXiv:1912.08660* (2019).
[26] Julien Gacon et al. 'Simultaneous perturbation stochastic approximation of the quantum fisher information'. In: *arXiv preprint arXiv:2103.09232* (2021).
[27] Tobias Haug and MS Kim. 'Natural parameterized quantum circuit'. In: *arXiv preprint arXiv:2107.14063* (2021).
[28] Domenico d'Alessandro. *Introduction to quantum control and dynamics*. Chapman and Hall/CRC, 2021.

[29] Zoë Holmes et al. 'Connecting ansatz expressibility to gradient magnitudes and barren plateaus'. In: *arXiv preprint arXiv:2101.02138* (2021).
[30] Jarrod R McClean et al. 'Barren plateaus in quantum neural network training landscapes'. In: *Nature communications* 9.1 (2018), pp. 1–6.
[31] Marcello Benedetti, Mattia Fiorentini and Michael Lubasch. 'Hardware-efficient variational quantum algorithms for time evolution'. In: *Physical Review Research* 3.3 (July 2021). ISSN: 2643-1564. DOI: 10.1103/physrevresearch.3.033083. URL: http://dx.doi.org/10.1103/PhysRevResearch.3.033083.