

Evaluation of Parameterized Quantum Circuits: on the design, and the relation between classification accuracy, expressibility and entangling capability

Thomas Hubregtsen^{1,2} · Josef Pichlmeier¹ · Koen Bertels¹

arXiv:2003.09887v1 [quant-ph] 22 Mar 2020

Received: March 2020 / Accepted: [under review]

Abstract Quantum computers promise improvements in terms of both computational speedup and increased accuracy. Relevant areas are optimization, chemistry and machine learning, of which we will focus on the latter. Much of the prior art focuses on determining computational speedup, but how do we know if a particular quantum circuit shows promise for achieving high classification accuracy? Previous work by Sim et al. (2019) proposed descriptors to characterize and compare Parameterized Quantum Circuits. In this work, we will investigate any potential relation between the classification accuracy and two of these descriptors, being expressibility and entangling capability. We will first investigate different types of gates in quantum circuits and the changes they incur on the decision boundary. From this, we will propose design criteria for constructing circuits. We will also numerically compare the classifications performance of various quantum circuits and their quantified measure of expressibility and entangling capability, as derived in previous work. From this, we conclude that the common approach to layer combinations of rotational gates and conditional rotational gates provides the best accuracy. We also show that, for our experiments on a limited number of circuits, a coarse-grained relationship exists between entangling capability and classification accuracy, as well as a more fine-grained correlation between expressibility and classification accuracy. Future research will need to be performed to quantify this relation.

Corresponding author: Thomas Hubregtsen
E-mail: t.s.hubregtsen@tudelft.nl

¹Delft University of Technology, Delft, Netherlands

²BMW Group Research, New Technologies, Innovations, Garching bei München, Germany

Keywords Parameterized Quantum Circuits · Expressibility · Entangling capability · Quantum Machine Learning · Quantum Computing

1 Introduction

Quantum Computing is said to bring advances in computational speedup. Furthermore, the ability to represent states in larger state spaces could boost expressive power. One of the key areas in which expressivity holds promise is Quantum accelerated Machine Learning. Various quantum algorithms for Machine Learning have been proposed, such as quantum Support Vector Machines(Rebentrost et al., 2014), quantum Neural Networks(Purushothaman and Karayiannis, 1997) and quantum (Restricted) Boltzmann Machines(Denil and Freitas, 2011). Quantifying the absolute or relative expressive power of quantum Machine Learning models is the focus of current academic work. In particular, Sim et al. (2019) made significant progress. In their work, they defined various descriptors as measures of evaluation of Parameterized Quantum Circuits. Parameterized circuits are the key component in variational quantum algorithms, which have been linked in functioning to neural networks(Schuld et al., 2018). Key descriptors that Sim et al. introduced are expressibility and entangling capability. Expressibility quantifies the ability of a circuit to map data from a ground state to any region of the higher-dimensional Hilbert Space, and entangling capability captures non-trivial correlation in the quantum data(Kandala et al., 2017). Although classical systems do expect that mapping input data through a high-dimensional feature space provides a more accurate space for finding a separating hyperplane, we are not aware of any study exploring the cor-

relation in quantum circuits between classification accuracy in relation to expressibility and entangling capability. In this work, we set out to find design criteria for quantum circuits by investigating the effects of gates in various circuits on classification performance in terms of the ability to change the decision boundary. We also create multiple sets of data with different challenges embedded in it, which we will use to evaluate classification performance of Parameterized Quantum Circuits. We will use these findings to investigate any correlation between the classification performance of a circuit with the expressibility and entangling capability of this circuit. In particular, we pose the following research questions:

- RQ1** What are design criteria for constructing a Parameterized Quantum Circuit?
- RQ2** What are the effects of individual types of gates on the decision boundary?
- RQ3** Do changes in expressibility and entangling capability correlate with changes in classification accuracy for variational quantum circuits?

This paper provides the following scientific contributions, as well as the following contributions to the current state of practice:

- (i) Visual insights into the effects of individual types of gates on the decision boundary
- (ii) Performance evaluation of variational quantum circuits
- (iii) Correlation evaluation between expressibility, entangling capability and classification accuracy

2 Related work

The effectiveness of Parameterized Quantum Circuits has been investigated in several studies. Benedetti et al. (2019) provided a detailed overview of current algorithms and their applications in the field of quantum machine learning. In their review, several state of the art data embeddings using quantum feature maps are listed. The task of variational circuits is explained and different design structures are shown. Schuld et al. (2018) proposed a low-depth quantum circuit for supervised learning tasks. In their work, the data gets embedded through amplitude encoding and a variational circuit is used for classification. The algorithm is benchmarked with several datasets and the performance is compared with that of classical Neural Networks and Support Vector Machines. Additionally, the decision boundaries of the classical SVM and the quantum classifier are compared. Killoran et al. (2019) proposed to construct continuous Neural Networks in a layered structure.

The expressive power of Parameterized Quantum Circuits has been investigated by Du et al. (2018). The group proved that highly entangled Parameterized Quantum Circuits have more expressive power than classical neural networks, with the restriction that only a polynomial number of parameters are used. The expressibility of different Parameterized Quantum Circuits has been investigated by Sim et al. (2019). The study evaluates the expressibility and entangling capability of various parameterized circuits with increasing number of layers. It shows that the expressiveness saturates with increasing circuit depth. The notion of expressive power has also been used in the field of tensor-networks for instances by Glasser et al. (2019). The work evaluates the expressive power of several tensor-network models and provides a basis for the analysis of Parameterized Quantum Circuit.

3 Approach

We will evaluate the effect of individual gates in Parameterized Quantum Circuits by freezing all but the relevant parameters and swiping over the parameters of the particular gates. We will then plot the decision boundary in the output space, and observe characteristics when comparing the different plots.

In order to evaluate the classification performance of various circuits, we define a set of classes of datasets. Every class represents a major change in dataset characteristics, every subclass represents a minor change in dataset characteristics. In these datasets, we embed data that is linearly separable, separable by a simple curved line, and data that requires more complex decision boundaries, such as enclosed clusters of data.

We will compare the relative classification performance of the circuits with the level of expressibility and entanglement capability found in prior art.

4 Evaluation

Machine learning algorithms can be broadly separated into regression algorithms, which predict a continuous or discretized continuous value, and classification algorithms, which predict a binary or multi-class label. In this work, we will focus on classification, in particular binary classification. We believe our work does not hold any fundamental limitations to be extended towards multi-class classification. In this section, we will first introduce the problem of binary classification and the datasets we will use. Subsequently, the classical and quantum algorithms which are used for deriving the classification accuracy will be presented. We will then

summarize the way prior art quantified expressibility and entangling capability and correlate that with the classification performance.

4.1 Binary classification

The goal of any binary classification algorithm is to take a normalized input set $\{x_0, \dots, x_n\} \in \mathbb{Z}, \mathbb{R}$ or \mathbb{B} and perform a transformation $\hat{y} = f(x)$ such that the predicted outcome \hat{y} matches the true outcome y . The transformation $f(\cdot)$ maps the input data from the input space into a, typically higher dimensional, feature space. This mapping can be performed multiple times between different features spaces before a hyperplane separates the data into a zero-class and a one-class which can be mapped to a label (e.g. true/false, positive/negative).

4.2 Dataset classes

In order to evaluate the classification performance, we have constructed an artificial dataset with varying levels of difficulty. Each dataset consists of around 1000 to 1200 samples, divided into 70% training data and 30% validation data. No data is split into test data, as we do not optimize hyperparameters. Certain datasets are slightly biased.

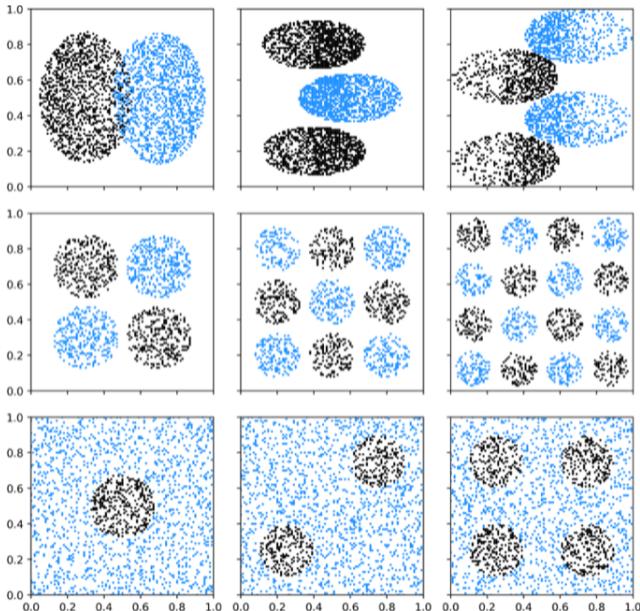


Fig. 1 Top to bottom: dataset classes 1 to 3. Left to right: dataset subclasses a to c. Black (dark) datapoints are labelled 1, blue (light) datapoints are labelled 0

All datasets are shown in Figure 1. Datasets of class 1 contain examples that can be classified with a single decision boundary. In the subsets, the difficulty is increased from requiring a straight decision boundary to a curved to a multi-curved boundary. Datasets of class 2 contain examples that require either a double decision boundary or more, favoring algorithms that produce a decision boundary that is also at least slightly curved. Datasets of class 3 require algorithms that can deal with one or more encapsulated clusters of data.

4.3 Algorithms

We will address the binary classification problem using Variational Quantum Circuits in a Hybrid Quantum/Classical setup. We will also implement a classical Neural Network as a sanity check and in order to provide grounding to the analysis of the decision boundary.

4.3.1 Classical Neural Network

Classical Neural Networks, including the breakthrough AlexNet network proposed by Krizhevsky and Hinton (2012) consists of the following components:

- An input layer. The input layer has the same dimension as the input. Sometimes, a cast of datatype or resolution is performed, but no real mapping takes place in this layer
- A hidden layer. The hidden layer takes the input and maps it from the input space into the feature space using a linear multiplication and addition $y = wx + b$. Different types of connectivity between the layers can exist, such as fully connected layers, which connect every neuron in the current layer to the previous layer, as well as convolutional layers, which apply a kernel
- An activation function. The activation function creates non-linearity in the decision boundary.
- A readout layer. This layer typically is a fully connected layer and maps the data from the feature space into the output space
- A loss function to quantify the difference between the prediction \hat{y} and the true value y
- An optimizer to optimize the weights in order to reduce the loss in future predictions

In particular, we will implement the following Neural Networks:

- One Neural Network consisting of a single hidden fully connected layer with 16 neurons and a Rectified Linear Unit (RELU) for the activation function

- One Neural Network consisting of two hidden fully connected layers, each with 16 neurons and a RELU for the activation function

Both are optimized using the adam optimizer(Kingma and Ba, 2014). All is implemented in Tensorflow(Abadi, 2015).

4.4 Variational Quantum Circuits in a Hybrid Quantum/Classical setup

A Hybrid Quantum-Classical algorithm(Xiang et al., 2013) consists of the following components:

- A data embedding circuit $\psi = U(x)$ to embed the classical data into the quantum system
- A Parameterized Quantum Circuit $\phi = V(\psi, \theta)$ that follows the embedding circuit and performs a mapping from the input space to the feature space
- Measurement of an observable \hat{B} that collapses the quantum state into a classical observable prediction point
- A loss function to quantify the difference between the predicted value \hat{y} and the true value y
- An optimizer to optimize free parameters θ in order to reduce the loss

We will evaluate the ability of Parameterized Quantum Circuits using qubits in finite dimensional Hilbert space to perform predictions. We implement and simulate this using Qiskit(Abraham, 2019).

4.4.1 Embedding circuit

Various embeddings have been proposed:

- Basis state encoding:
 $x \in \{0, 1\}^n \rightarrow |j\rangle$ (Farhi and Neven, 2018; Schuld and Killoran, 2019)
- Amplitude encoding:
 $x \in \mathbb{R}^N \rightarrow |\psi_x\rangle = \sum_{n=0}^{N-1} x_j |j\rangle$ (Schuld and Killoran, 2019)
- Product encoding:
 $x \in \mathbb{R}^N \rightarrow |\psi_x\rangle = \cos(x_j) |0\rangle + \sin(x_j) |1\rangle$ (Stoudenmire and Schwab, 2016)
- Squeezed vacuum state encoding:
 $x \in \mathbb{R}^N \rightarrow |\psi_{(c, x_j)}\rangle = \frac{1}{\sqrt{\cosh(c)}} \sum_{n=0}^{\infty} \frac{\sqrt{(2n)!}}{2^n n!} (-e^{ix_j} \tanh(c))^n |2n\rangle$ (Schuld and Killoran, 2019)
- Fourier-Walsh encoding
 $x \in \mathbb{R}^N \rightarrow |\Phi(x)\rangle = U_{\Phi(x)} H^{\otimes n} U_{\Phi(x)} H^{\otimes n} |0\rangle^{\otimes n}$

with $\Phi(\mathbf{x}) = e^{i \sum_{S \subseteq [n]} (\phi(\mathbf{x})) \prod_{i \in S} Z_i}$ (Havlicek et al., 2018)

These embeddings provide different ways to map the classical data into the Hilbert space. Some of the aforementioned methods directly produce an output distribution that is hard to simulate with classical computers (Havlicek et al., 2018; Schuld and Killoran, 2019). However, the purpose of this study is to evaluate the power of the Parameterized Quantum Circuit, not the power of the different embeddings. For this reason, we use a circuit that a) embeds the data evenly in every computational basis state and b) embeds the data in a minimally expressive state. The applied circuit can be seen in 2. All other circuits are preceded with the embedding circuit.

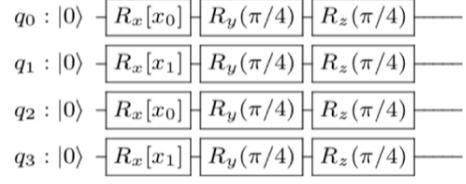


Fig. 2 Embedding circuit

It is similar to the circuit proposed by Stoudenmire and Schwab (2016) and can therefore be categorized as a product state encoding. Each component of the feature-vector \mathbf{x} is embedded as a single qubit rotation $R_x(\phi(x_i))$ with $\phi : \mathbf{x} \rightarrow [0, 2\pi]$. As we work with two-dimensional data, we duplicate x_0 and x_1 on x_2 and x_3 . After embedding the data in the parameters of the R_x gates, we apply a set of constant R_y and R_z rotations. That ensures that if projected down onto the computational basis states, the output states span approximately the same range. This is shown in 3, 4, and 5.

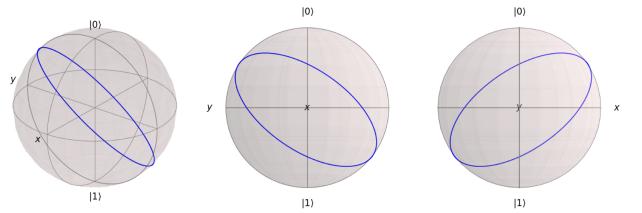


Fig. 3 XYZ View **Fig. 4** YZ View **Fig. 5** XZ View

4.4.2 Parameterized Quantum Circuit

Parameterized Quantum Circuits contain one or two qubit gates with free parameters that can be optimized to fit a probability distribution. In this work, we will evaluate a subset of the circuits addressed in the work of Sim et al. (2019).

Circuit number 1 consists of a set of R_x and R_z gates as shown in Figure 6. It is not using any two qubit operations and therefore does not project the input state into a higher dimensional space. This circuit is characterized as having lower expressive power. Additionally, the expressiveness of this circuit is not significantly increased when using a higher number of layers.

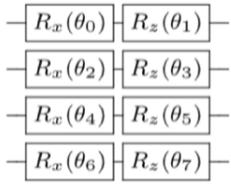


Fig. 6 Circuit number 1 (Sim et al., 2019)

Circuit 16 is an extension of circuit 1, constructed to show an increase of expressibility caused by additional controlled gates (Sim et al., 2019). Hence, it consists of the same layers of single R_x and R_z rotation gates which are then followed by controlled R_z gates. In theory, these circuits are capable of mapping the input state $|\phi(x_i)\rangle$ into a higher dimensional feature space in which the data points \mathbf{x} are more likely to be linearly separable(Schuld et al., 2018).

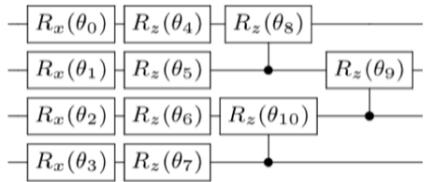


Fig. 7 Circuit number 16 (Sim et al., 2019)

Circuit number 19 was constructed for the task of data classification. It consists of the same set of single rotation gates as circuit number 1 and is extended by a set of controlled R_x gates. The design criteria of the circuit is to generate a highly entangled state which allows an efficient projection of the data into a space where it can be separated (Schuld et al., 2018). Additionally, it has been shown that this circuit group can identify correlations in the input data (Schuld et al.,

2018; Sim et al., 2019). The circuit is characterized as having higher expressive power (Sim et al., 2019).

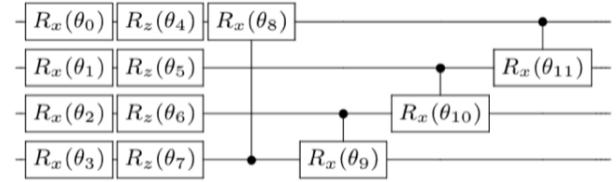


Fig. 8 Circuit number 19 (Sim et al., 2019)

4.4.3 Measurement

By measuring the state of the n qubits, a bitstring $j \in \{0, 1\}^n$ can be derived (Havlicek et al., 2018). This bit-string is forwarded to a boolean function such that an label y is associated with it $y : j \rightarrow \{0, 1\}$. This process is repeated r times in order to derive a probability distribution $\hat{p}_y(x_j)$ for one specific set of θ values. The label with the highest probability is chosen (Havlicek et al., 2018).

4.4.4 Optimization algorithm

The optimizer in the variational quantum algorithm is searching for an optimal set of parameters θ in order to minimize the loss. For the optimization of our algorithm we use the Simultaneous Perturbation Stochastic Approximation algorithm (SPSA)(Li et al., 2013). We vary the learning rate by tweaking c_0 , for which we start high and lower the rate every 30 epochs. We cycle through the following values: 4, 3, 2, 1.5, 1, 0.5. We have chosen the starting value 4 from Havlicek et al. (2018) and ended the optimization at the recommended learning rate according to the VQC algorithm in the Qiskit software(Abraham, 2019).

4.5 Expressibility and entangling capability of quantum circuits

In our work, we will rely on the definition and quantification of expressibility and entangling capability as proposed by Sim et al. (2019). In their work, they define the operational meaning of expressibility to be "the amount of information that is lost if we were to approximate the distribution of state fidelities generated by a Parameterized Quantum Circuit using that of Haar random states". Sim et al. (2019) quantify this by calculating the Kullback-Leibler divergence(Kullback and

Leibler, 1951) between "the distribution of states obtained from sampling the parameters of a Parameterized Quantum Circuit and the (expressive) uniform distribution of state, i.e. the ensemble of Haar-random state" (Sim et al., 2019). For the entangling capability, they propose to use the Meyer-Wallach entanglement measure (Meyer and Wallach, 2002). The quantification of both expressibility and entangling capability for the circuits, as found by Sim et al. (2019), is shown in Table 1.

One would expect that a circuit that is able to map input data towards any part of the Hilbert space, is also able to map states in such a way that a separating hyperplane can be found. However, it is unclear to what extent this relation holds. For instance, one could argue that merely mapping all data labeled zero to one "corner" and all data labeled one to another "corner" is sufficient, and the ability to map data to other parts of the space is overkill. Still, classical Neural Networks are said to work by the ability of mapping data first in higher dimensions before separating. With this reason in mind, we will evaluate the decision boundary characteristics and classification performance of various circuits, and correlate the findings with the circuit descriptors expressibility and entangling capability as previously described.

5 Results and discussion

In this section, we present and discuss the results of our experiments. We will first address the feasibility of our datasets by fitting them with a classical Neural Network. We will then evaluate the influence on the decision boundary by various types of gates in the different circuits. This will be followed by an evaluation of the classification performance of our circuits on our datasets. We will end the section by linking the expressibility and entangling capability of the circuits, as found in prior art, to the average classification accuracy of the circuits on our datasets.

5.1 Classical ML algorithms

5.1.1 Neural Network with 1 layer

The first Neural Network, consisting of only a single hidden layer with 16 neurons, achieves an average accuracy of 92%. We see the convergence of the algorithm in Figure 9. When examining the decision boundary, shown in Figure 10, we clearly see two components at play. The straightened lines at different angle are due to the various neurons calculating $y = w * x + b$. The

non-linear breaks in the straight lines originate from the activation function.

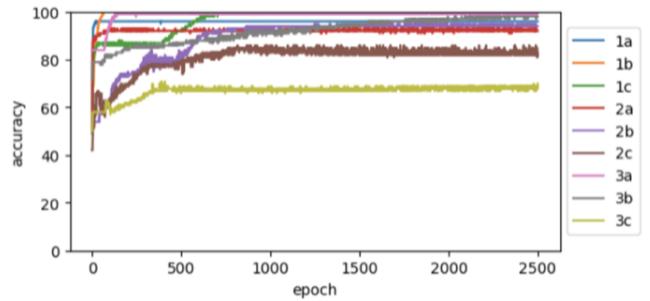


Fig. 9 Convergence for the Neural Network (1 layer)

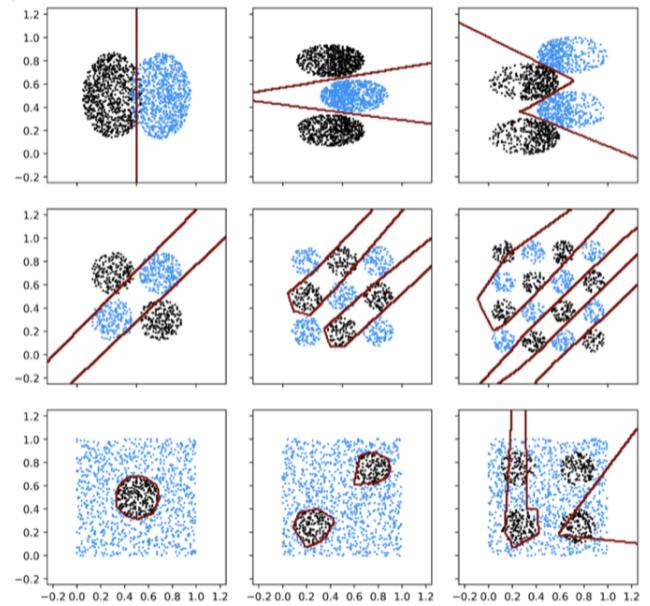


Fig. 10 Decision boundary for the Neural Network (1 layer)

5.1.2 Neural Network with 2 layers

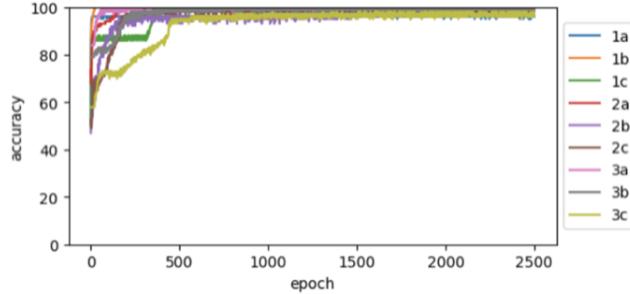
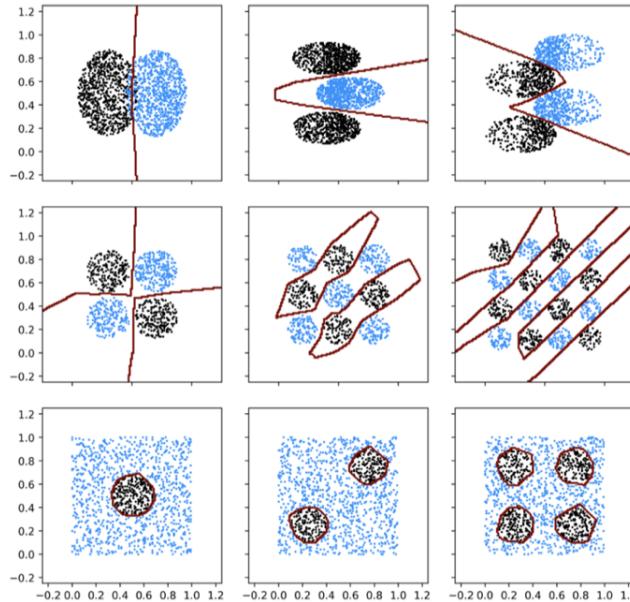
The second Neural Network, with two hidden layers of 16 neurons each, achieves an average accuracy of 99% in less than 500 epochs as shown in Figure 11. Even though the decision boundary still consists of many straight parts, as shown in Figure 12, it is capable of finding a near-perfect fit.

5.2 Variational Quantum Circuit - understanding the effect of gates

In this subsection, we will examine the impact of various gates in various circuits on the decision boundary. We

Table 1 Expressibility and entangling capability of the circuits, as quantified by Sim et al. (2019)

Circuit	1 (1 layer)	1 (2 layers)	16 (1 layer)	16 (2 layers)	19 (1 layer)	19 (2 layers)
Expressibility	0.3	0.2	0.25	0.09	0.08	0.009
Entangling capability	0	0	0.35	0.5	0.6	0.75

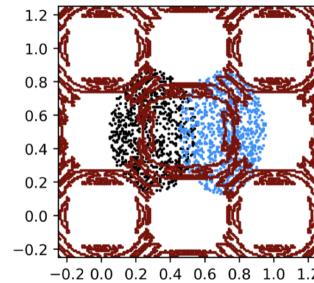
**Fig. 11** Convergence for the Neural Network (2 layers)**Fig. 12** Decision boundary for the Neural Network (2 layers)

do this by providing the circuit under examination an input on all qubits in the range of $[0, 1]$, and observed the output for all sets of input values. The decision boundary, plotted in red, denotes the change of output label from either 0 to 1, or from 1 to 0.

5.2.1 Embedding

Before we get started examining the gates, we want to understand the state of the system and the associated decision boundary right after the embedding of the data in the quantum system. This embedding circuit contains no parameters and is not followed by a Parameterized Quantum Circuit. This is shown in Figure 13. We can see that it creates a decision boundary

with a high level of symmetry, but also large regions of no change. A total of 9 clusters are shown. This is very different from a typical decision boundary, which generally would consist of a straight or curved line. This already tells us that the optimizer is dealing with a difficult solution landscape.

**Fig. 13** Decision boundary using a plain embedding without any Parameterized Quantum Circuit

5.2.2 Circuit 1 with 1 layer

In order to investigate the effect of the R_x gates, we plotted the decision boundary when sweeping over the values $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ for all parameters of the R_x gates, while keeping all parameters of the R_z gates to 0.25. This is shown, left to right and top to bottom, in Figure 14. We see that with negative values (top row) the decision boundary starts to curve. Four out of our nine pattern blocks get squeezed, and there is an upward shift. The decision boundary clears up, and becomes circular. Positive values (bottom row) make the decision boundary straighten out, and shift in the opposite direction.

We have also swept over the R_z gates in similar fashion while keeping the parameters of the R_x gates at 0.25. This is shown in Figure 15. In this, we only observe the effects of the embedding circuit and the R_x gates. This is because the R_z gates turn the vector around the z -axis, which will not have any effect when performing a measurement in this same computational basis. This implies that only a small subset of the Hilbert space can be reached. Only when followed by a rotation in a different axis, can the effects on the decision boundary be observed.

To further investigate the rotational gates, we perform a similar experiment, but with alternating the sign for the parameters of the R_x gates acting on x_0 and x_1 .

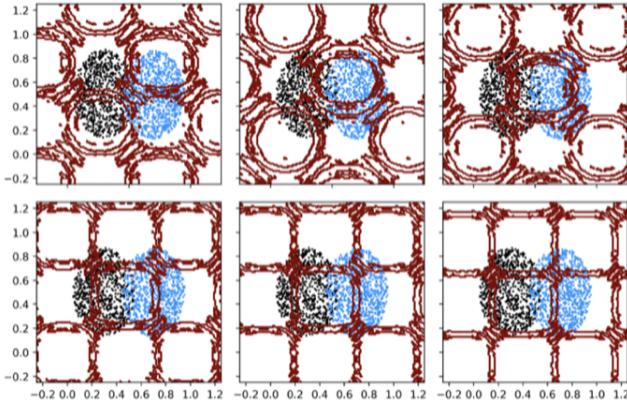


Fig. 14 Decision boundary for circuit 1 (1 layer) when sweeping with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ over R_x while keeping all R_z fixed to 0.25

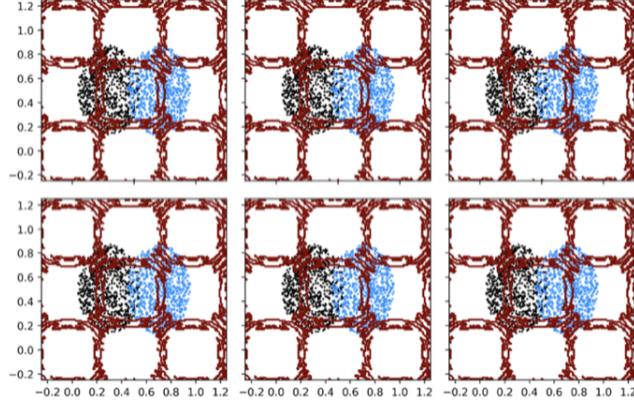


Fig. 15 Decision boundary for circuit 1 (1 layer) when sweeping $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ for R_z while keeping R_x at 0.25

We do this by sweeping over $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ for p_0 and p_2 , and sweeping over $[1, 0.5, 0.25, -0.25, -0.5, -1]$ for p_1 and p_3 of the R_x gates. We keep R_z constant at 0.25. This is shown in Figure 16. We observe a similar pattern for subfigures a and f, b and e as well as c and d. In this, the decision boundaries separate and straighten out.

This investigation is also representative for the R_x and R_z gates in circuit 16 and 19 when applying a single layer.

5.2.3 Circuit 1 with 2 layers

In order to examine the effects in the same circuits with 2 layers, we swept over both sets of R_x gates with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ for the parameters of R_z while keeping all parameters for R_x at 0.25. The observed effect is an amplification of the effects we saw in circuit 1 with 1 layer and is shown in Figure 17. What we do notice, is that a crossover starts to happen when

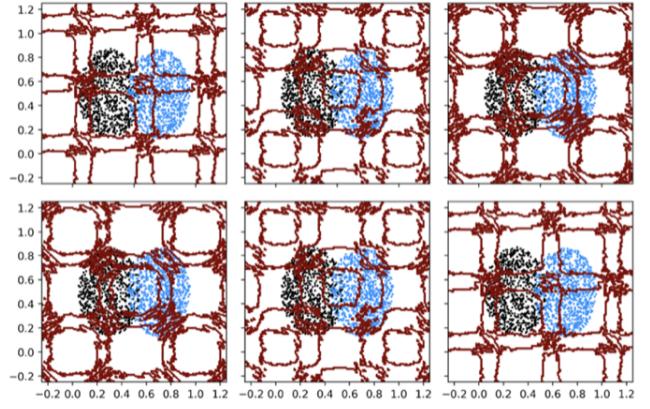


Fig. 16 Decision boundary for circuit 1 (1 layer) when sweeping p_0 and p_2 of R_x over $[-1, -0.5, -0.25, 0.25, 0.5, 1]$, p_1 and p_3 over the inverse, while keeping all R_z fixed to 0.25

going further in the negative or positive. This is because of the period of the gates that eventually meet.

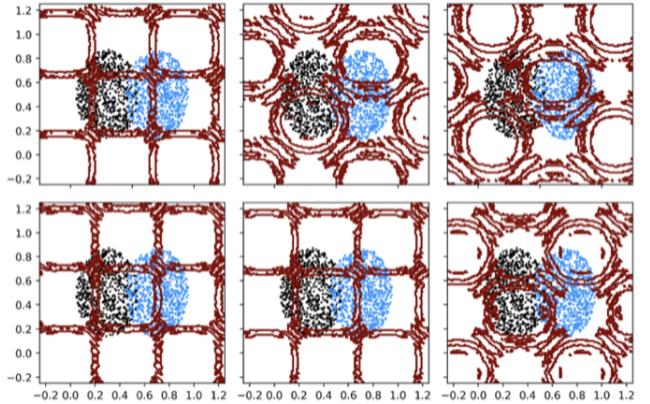


Fig. 17 Decision boundary for circuit 1 (2 layers) when sweeping over both sets of R_x gates with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ while keeping all R_z at 0.25

Sweeping over the R_z gates with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ while keeping the R_x gates at 1 is shown in Figure 18. Here, the latter R_z gates again have no effect due to the rotation happening in the same computational basis as the readout, so no crossover takes place yet. What we do see, is that negative values straighten out the boundary, and positive values increase the curvature.

5.2.4 Circuit 16 with 1 layer

Sweeping over the parameters of the conditional R_z gates in circuit 16 with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ and the parameters of all other gates fixed to 0.25 incurs the same effect of sweeping over the regular R_z gates of circuit 1 as shown in Figure 15. As none are followed by

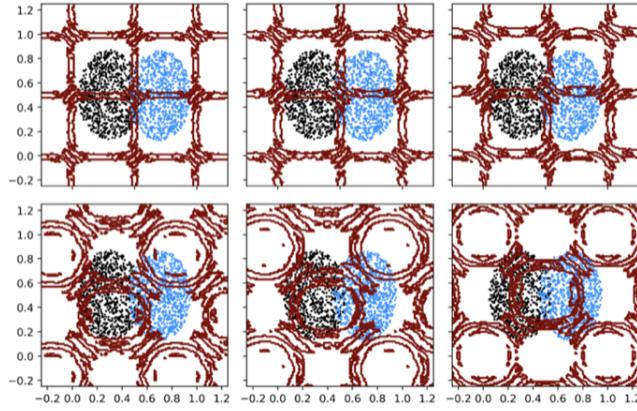


Fig. 18 Decision boundary for circuit 1 (2 layers) when sweeping over both sets of R_z gates with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ while keeping the R_x gates at 1

an R_x or R_y gate, the rotation along the z axis is not observed in the measurement.

5.2.5 Circuit 16 with 2 layers

Sweeping over the parameters of the conditional R_z gates in circuit 16 with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$, with the parameters of all other gates fixed to 0.25, only results in an observable effect from the first set of conditional R_z gates due to reasons previously discussed. These effects are shown in Figure 19. Here, we see a clear divergence from previous patterns. In particular, we see an alteration that does not follow the strict pattern of similar sections forming in relatively identical clusters, but a grouping in blocks of 4. In these blocks of 4, there is also less of strict symmetry, and the decision boundary breaks away. This is a pattern that can be linked more towards a non-linear activation function in a Neural Network.

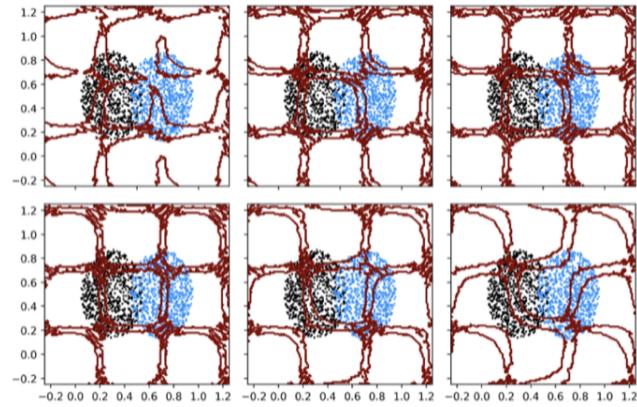


Fig. 19 Decision boundary when sweeping over all conditional R_z gates in circuit 16 (2 layers) with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ and while keeping all other gates to 0.25

5.2.6 Circuit 19 with 1 layer

Sweeping over the parameters of the conditional R_x gates with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ while keeping the parameters of all other gates fixed to 0.25 is shown in Figure 20. In it, we see a similar break from the regular pattern as we saw with the conditional R_z gates for circuit 16. We again see a less firm structure of clusters of blocks, creating larger fields for the positive values and smaller circles for the negative values.

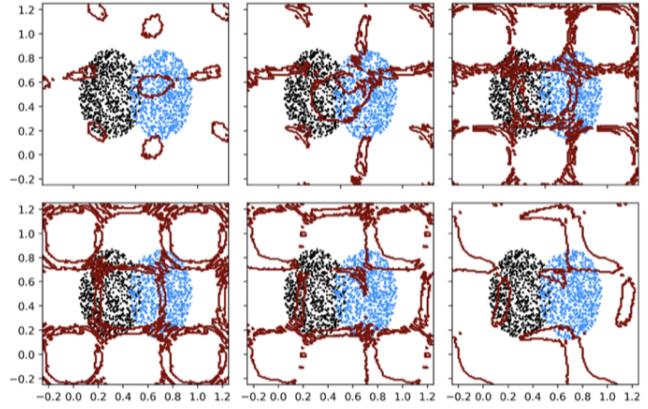


Fig. 20 Decision boundary when sweeping over all conditional R_x gates in circuit 19 (1 layer) with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ while keeping all other gates to 0.25

5.2.7 Circuit 19 with 2 layers

Sweeping over both layers of conditional R_x gates with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ for the parameters, while keeping the parameters for all other gates fixed to 0.25, is shown in Figure 21. Concatenating layers again amplifies the effect, this time creating clusters spanning up to 9 blocks. For negative, the decision boundary starts to fade away, consisting of small circles. For large values, we see larger trapezoid structures forming.

5.2.8 Summary on the effect of individual gates in Variational Quantum Circuits

In the previous subsections, we have seen the following:

- Rotational gates can straighten out or curve the decision boundary, but do not alter the overall structure
- Concatenating layers amplify the effects observed in single layers
- Rotations along the z axis are not observed when measuring in the z basis

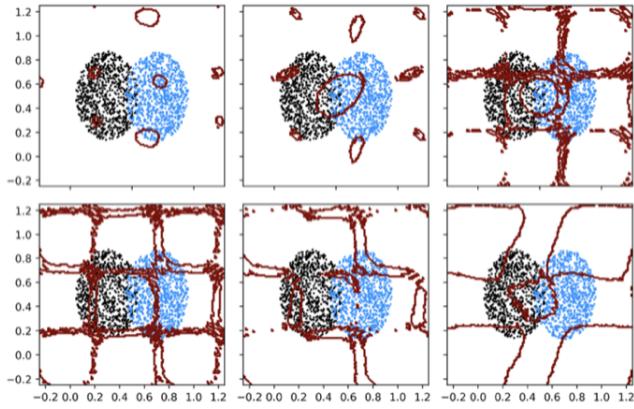


Fig. 21 Decision boundary when sweeping over all conditional R_x gates in circuit 19 (2 layers) with $[-1, -0.5, -0.25, 0.25, 0.5, 1]$ while keeping all other gates to 0.25

- Conditional gates provide effects similar to non-linear activation functions in a neural network. It changes the overall structure of the decision boundary, and breaks symmetries

5.3 Variational Quantum Circuit - classification performance

In this subsection, we will present and discuss the ability of the various circuits to fit to our datasets. All results are summarized in Table 2, located at the end of the section. The results for the Neural Network have been discussed previously in 5.1. All other results are discussed in the following sections, with use of convergence graphs and decision boundary plots.

5.3.1 Circuit 1 with 1 layer

The convergence of circuit 1 with 1 layer is shown in Figure 22. Here we see that convergence is reached early, as the epochs with the smaller learning rates do not improve the overall accuracy. We also see the start of every major cycle by a peak in accuracy. This is because every major cycle picks the set of parameters that yielded the best performance in the previous cycle. All accuracies roughly lay between 55% and 85%, and the total average accuracy is 69%. This is relatively low, especially because random guessing would yield an accuracy of over 50% due to the unbalanced dataset. This can likely be attributed to the low number of gates that can affect the circuit and the capabilities of these gates. In particular, the only effective gates are the R_x gates, as the R_z gates do not provide impact when measurement immediately in this same z basis. As we have previously seen, this circuit is only capable of straightening

out or curving the boundary, but not able to provide a breakage of the pattern, nor offset. Datasets that contain alternating symmetry, such as dataset 2a, are beyond reach for this circuit. We can also observe this in the decision boundary plot, as shown in Figure 13. Here we see that the optimizer was not able to gain any traction, resulting in a decision boundary equal to that of the embedding circuit. The optimizer did change for the other datasets, but none to an extend that show a remote proper fitting of the data structure.

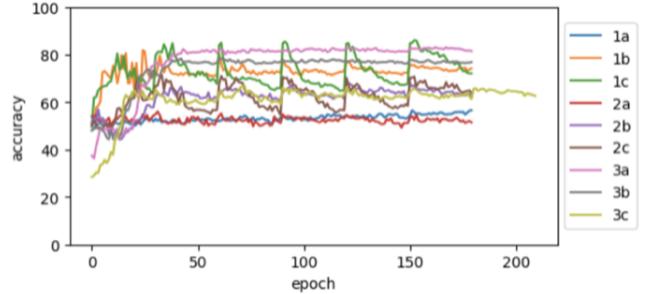


Fig. 22 Convergence for circuit 1 (1 layer)

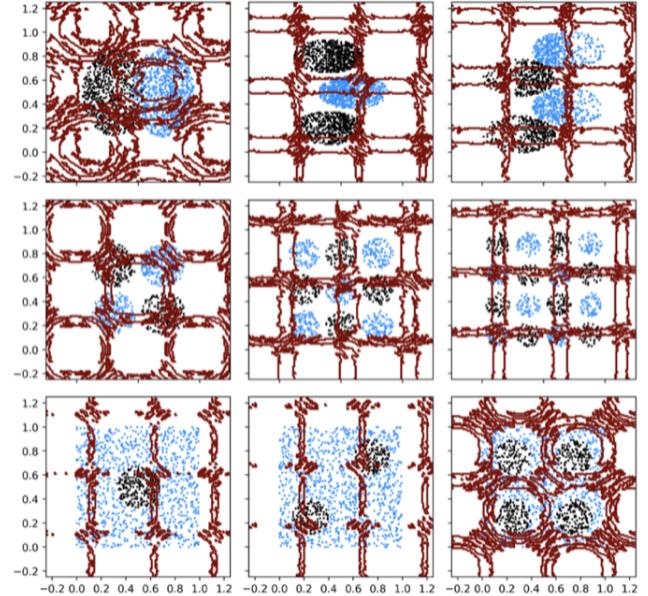


Fig. 23 Decision boundary for circuit 1 (1 layer) optimized on our dataset

5.3.2 Circuit 1 with 2 layers

Adding an extra layer "activates" the effects of the R_z gates in the first layer, as well as adding another set of R_x gates. Although many datasets are still challenging

for the optimizer to find a fit, as shown in Figure 25, we see a significant performance increase in dataset 1b and 2c. Here, the circuit is able to score 80% and 100% accuracy respectively. The fact that these circuits line up horizontally and vertically is helpful for the circuit, that so far has not shown any straight lines in a diagonal orientation. The overall average classification accuracy is raised to 74%. The optimizer does require more time to find these fits, as shown in Figure 24.

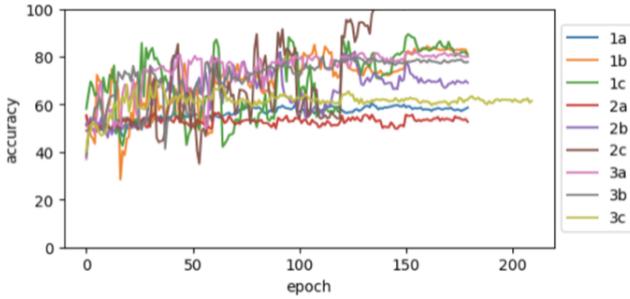


Fig. 24 Convergence for circuit 1 (2 layers)

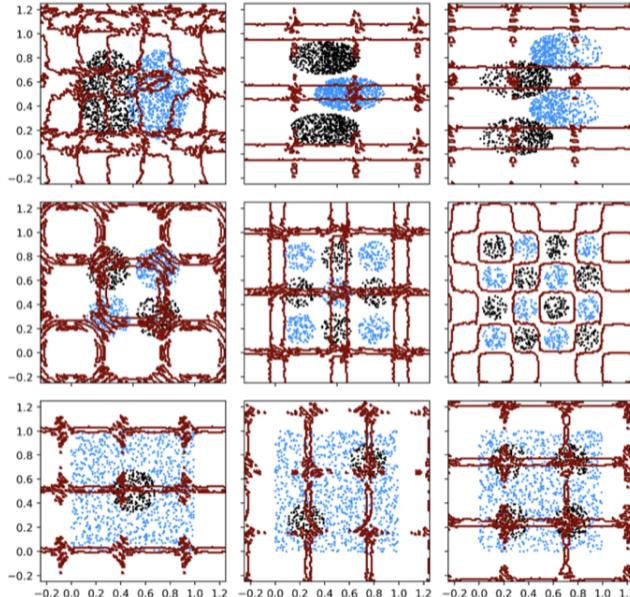


Fig. 25 Decision boundary for circuit 1 (2 layers) optimized on our dataset

5.3.3 Circuit 16 with 1 layer

Circuit 16, consisting of sets of R_x , R_z and conditional R_z gates again is only able to effectively leverage the R_x gates. Because of this, the overall performance of 68% is roughly equal to the performance of circuit 1 with 1

layer. This can also be seen in the convergence plot in Figure 26 and the decision boundary plot in Figure 27, which appear similar to the Figures for circuit 1 with 1 layer. The only notable difference is the vertical orientation versus the horizontal orientation of the decision boundary for circuit 2b.

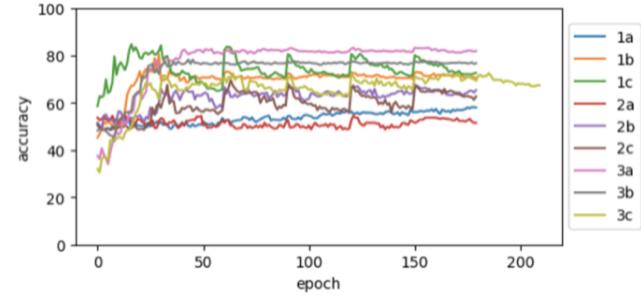


Fig. 26 Convergence for circuit 16 (1 layer)

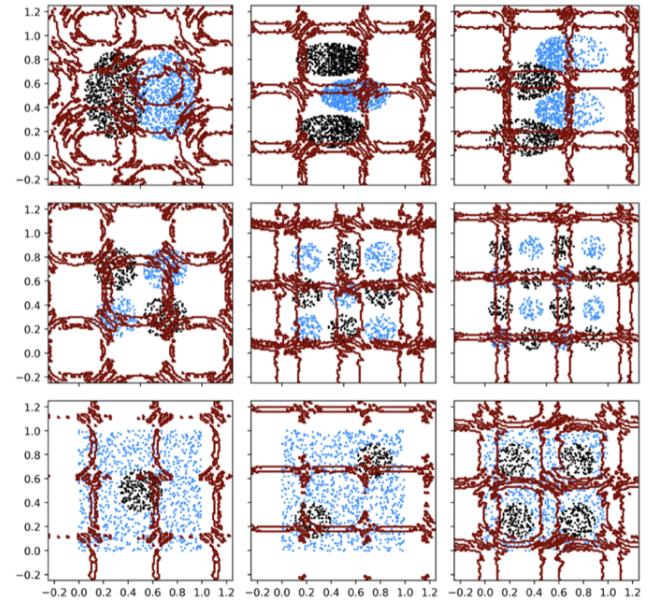


Fig. 27 Decision boundary for circuit 16 (1 layer) optimized on our dataset

5.3.4 Circuit 16 with 2 layers

Circuit 16 with 2 layers can leverage the effects of the R_z , conditional R_z and the extra R_x gates. This translates in an increased overall average accuracy of 84%, reached by steadily improving over the epochs as shown in Figure 28. The decision boundary plots, as visualised in Figure 29, show clearly fitted decision boundaries along the samples of the data. Wave patterns try to fit the data in dataset 1c and 2b. Dataset 2b is fitted

without showing aliases in the symmetrically counter positions along the horizontal and vertical axis. Still, dataset 1a seems difficult to fit, and dataset 2a is still fitted with a decision boundary equal to the embedding circuit.

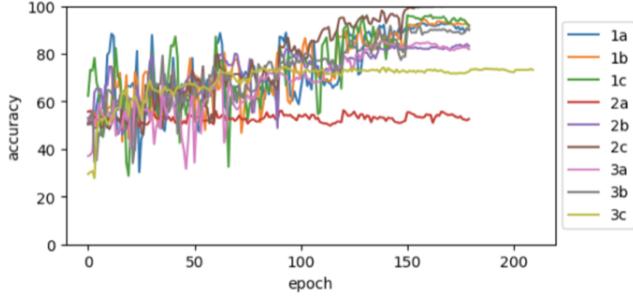


Fig. 28 Convergence for circuit 16 (2 layers)

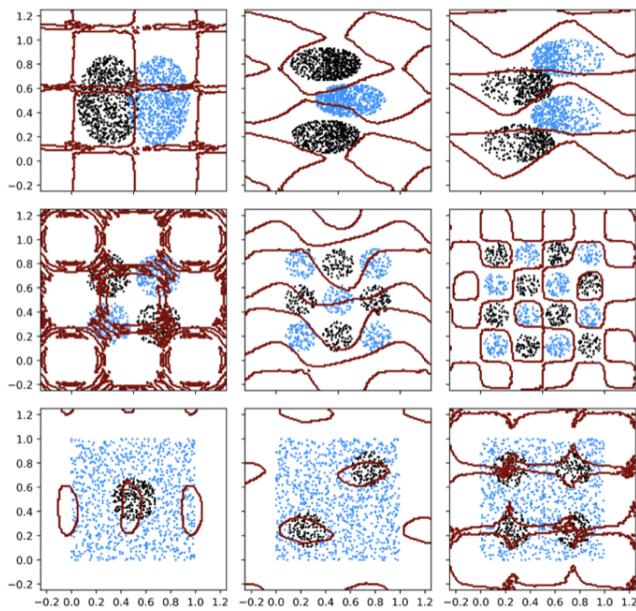


Fig. 29 Decision boundary for circuit 16 (2 layers) optimized on our dataset

5.3.5 Circuit 19 with 1 layer

Circuit 19 with 1 layer has an average classification accuracy of 79%, with convergence being reached early, as shown in Figure 30. The difference with circuit 16 with 1 layer is clearly present, even though this circuit contains the same number of rotational and conditional rotational gates. This strengthens the point that rotations in the z axis at the end of a circuit have no effect. The difference to circuit 1 is also significant: adding

a conditional rotational gate raises the accuracy with 10%-point, while adding a full second layer of "standard" rotational gates only raises the accuracy with 5%-point. Finally, as a general remark to the circuit characteristics, we see a lot of curvature and breaking of the simple pattern in the decision boundary shown in 31. Still, datasets 2a, 3b and 3c prove hard to find a fit to.

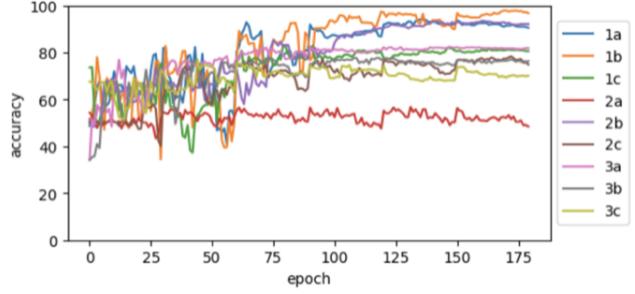


Fig. 30 Convergence for circuit 19 (1 layer)

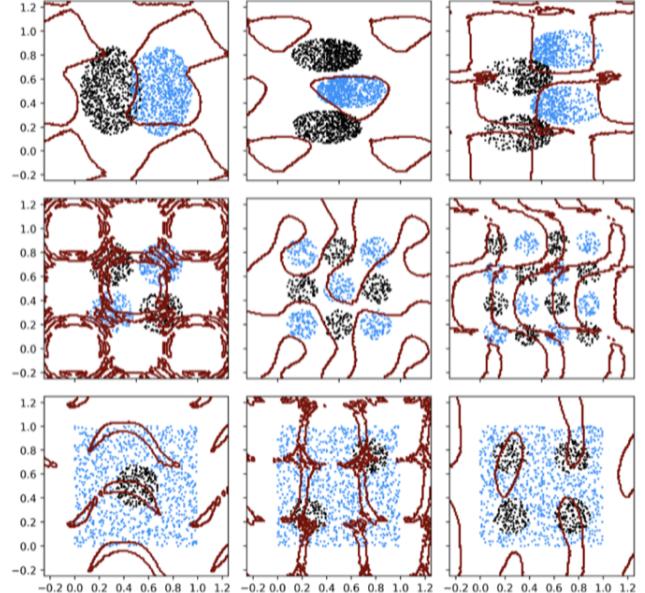


Fig. 31 Decision boundary for circuit 19 (1 layer) optimized on our dataset

5.3.6 Circuit 19 with 2 layers

Circuit 19 has the most active gates and achieves the best overall accuracy of 90%. The advancements are again reached in later epochs, as shown in Figure 32. The decision boundaries are also the cleanest, as shown in Figure 33. It is able to handle data that is not horizontally nor vertically symmetrical, such as dataset 3b.

It also picks out the details in dataset 2b. It is even able to make a fit for dataset 2a, something none of the other circuits were capable off. The overall increase in average accuracy by adding a layer is 11%-point.

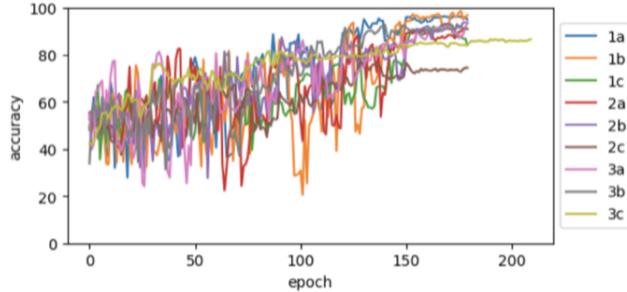


Fig. 32 Convergence for circuit 19 (2 layers)

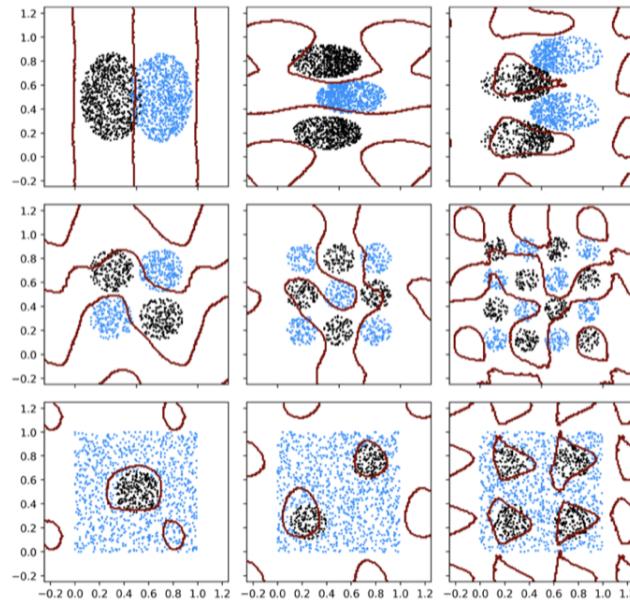


Fig. 33 Decision boundary for circuit 19 (2 layers)

5.3.7 Summary of the classification performance of the circuits

In the previous subsections, we have seen the following:

- Any R_z or conditional R_z gates that are not followed by a rotation along a different axis do not contribute to changes in the decision boundary
- Rotational gates do not change the overall pattern, but only straighten out or curve the decision boundary
- Conditional rotational gates are able to create more complex decision boundaries, such as lines oriented

along the diagonal, circles and trapezoids. They also allow to break patterns that single qubit gates cannot break

- The combination of rotational gates and conditional rotational gates provide further advancement of the average classification accuracy of a circuit
- Concatenating layers provide advancement in the average classification accuracy of a circuit
- Highest average accuracy achieved by a quantum circuit is 90%, which is still 9%-point below the average classification accuracy of the 2-layer Neural Network.

5.4 Relation between classification performance, expressibility and entangling capability of the circuits

The entangling capabilities of the circuits are plotted against the average classification performance on the left in Figure 34. Here, we can see no direct or inverse relation. However, on a coarse level one could say that lower entangled circuits perform less, medium entangled circuits perform average and the highly entangled circuit performed best.

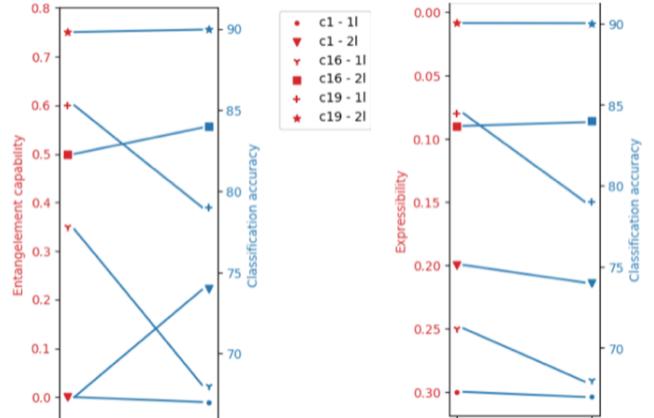


Fig. 34 Left: Entanglement capability versus classification accuracy. Right: Expressibility versus classification accuracy

The expressibility of every circuit against the average classification performance is plotted on the right in Figure 34. Given that a circuit with a lower expressibility number is better able to explore the Hilbert space, we have inverted the axis. We can see that the relative ordering is preserved for all circuits, except for a minor overlap between circuit 19 with 1 layer and circuit 16 with 1 layer, with an estimated expressibility of 0.8 and 0.9 respectively. Given that the expressibility ranges from 0.009 to 0.3, we find that these points do lay very close together and do not steer away from the general trend.

Table 2 Evaluation results — Details

Setup		Performance									
Type	Alg.	1a	1b	1c	2a	2b	2c	3a	3b	3c	Average
Classical	NN - 1l	96%	100%	100%	92%	94%	82%	100%	98%	70%	92%
Classical	NN - 2l	96%	100%	100%	100%	99%	98%	100%	99%	97%	99%
Hybrid	c1 - 1l	57%	74%	72%	64%	63%	64%	82%	77%	63%	69%
Hybrid	c1 - 2l	59%	81%	80%	53%	69%	100%	80%	78%	62%	74%
Hybrid	c16 - 1l	58%	71%	73%	52%	65%	63%	82%	77%	67%	68%
Hybrid	c16 - 2l	90%	91%	92%	53%	83%	100%	82%	90%	73%	84%
Hybrid	c19 - 1l	91%	97%	81%	48%	92%	75%	82%	76%	70%	79%
Hybrid	c19 - 2l	95%	97%	84%	91%	94%	74%	93%	91%	87%	90%

We have considered combining the entanglement and expressibility in a single quantification. However, when doing so linearly, regardless of weights used, it does not improve the relative ordering. This holds both for the cross happening between circuit 16 with 2 layers and circuit 19 with 1 layer for both descriptors, nor does it improve the cross in entanglement capability for circuit 1 with 2 layers and circuit 16 with 1 layer.

6 Conclusion

We previously set out to answer how to structure quantum circuits by investigating the effects of individual gates. We have seen that rotational gates can straighten out or curve our decision boundary, whereas conditional rotational gates can break larger symmetrical patterns. We have also seen that concatenating layers increases accuracy. From this, we can conclude that the common approach to layer combinations of rotational gates and conditional rotational gates provides the best accuracy. We also asked if any relation exists between expressibility, entangling capability and classification for variational quantum circuits. We have seen that, for our experiments on a limited number of circuits, a coarse-grained relationship exists between entangling capability and classification accuracy, as well as a more fine-grained correlation between expressibility and classification accuracy. Future research will need to be performed to quantify this relation. In both cases, for accuracy and expressibility, it should be taken into consideration in which basis the circuit is going to be measured, as several of our circuit consisted out of gates that would perform rotations around the axis in which a measurement would follow.

References

- Abadi Mea (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. URL <http://tensorflow.org/>, software available from tensorflow.org
- Abraham H (2019) Qiskit: An open-source framework for quantum computing. DOI 10.5281/zenodo.2562110
- Benedetti M, Lloyd E, Sack S, Fiorentini M (2019) Parameterized quantum circuits as machine learning models DOI 10.1088/2058-9565/ab4eb5, arXiv: 1906.07682
- Denil M, Freitas ND (2011) Toward the implementation of a quantum rbm.
- Du Y, Hsieh MH, Liu T, Tao D (2018) The expressive power of parameterized quantum circuits. arXiv: 1810.11922
- Farhi E, Neven H (2018) Classification with quantum neural networks on near term processors. arXiv: 1802.06002
- Glasser I, Sweke R, Pancotti N, Eisert J, Cirac JI (2019) Expressive power of tensor-network factorizations for probabilistic modeling, with applications from hidden markov models to quantum machine learning arXiv:1907.03741
- Havlicek V, Croles AD, Temme K, Harrow AW, Kandala A, Chow JM, Gambetta JM (2018) Supervised learning with quantum enhanced feature spaces DOI 10.1038/s41586-019-0980-2, arXiv:1804.11326
- Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow JM, Gambetta JM (2017) Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets.
- Killoran N, Bromley TR, Arrazola JM, Schuld M, Quesada N, Lloyd S (2019) Continuous-variable quantum neural networks.
- Kingma D, Ba J (2014) Adam: A method for stochastic optimization.
- Krizhevsky IS Alex, Hinton GE (2012) Imagenet classification with deep convolutional neural networks.
- Kullback S, Leibler RA (1951) On information and sufficiency.
- Li L, Jafarpour B, Mohammad-Khaninezhad MR (2013) A simultaneous perturbation stochastic approximation algorithm for coupled well placement and control optimization under geologic uncertainty.

- Meyer DA, Wallach NR (2002) Global entanglement in multiparticle systems.
- Purushothaman G, Karayiannis NB (1997) Quantum neural networks (qnns): inherently fuzzy feedforward neural networks. *IEEE Transactions on neural networks* 8, no 3: 679-693
- Rebentrost P, Mohseni M, Lloyd S (2014) Quantum support vector machine for big data classification. *Physical review letters* 113:13: 130503
- Schuld M, Killoran N (2019) Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4), p040504
- Schuld M, Bocharov A, Svore K, Wiebe N (2018) Circuit-centric quantum classifiers. [arXiv:1804.00633](https://arxiv.org/abs/1804.00633)
- Sim S, Johnson PD, Aspuru-Guzik A (2019) Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms DOI 10.1002/qute.201900070, [arXiv:1905.10876](https://arxiv.org/abs/1905.10876)
- Stoudenmire EM, Schwab DJ (2016) Supervised learning with quantum-inspired tensor networks [arXiv: 1605.05775](https://arxiv.org/abs/1605.05775)
- Xiang ZL, Sahel Ashhab JQY, Nori F (2013) Hybrid quantum circuits: Superconducting circuits interacting with other quantum systems.