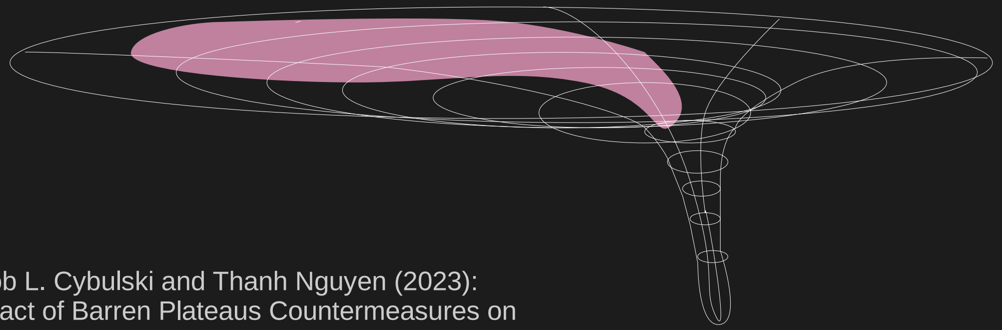


Jacob L. Cybulski and Thanh Nguyen (2023):
"Impact of Barren Plateaus Countermeasures on the Quantum Neural Network Capacity to Learn."
Quantum Information Processing, 22, 442.

Introduction
Variational quantum models (VQM)
Emergence of barren plateaus (BP) in VQM training
BP countermeasures
Impact of BP countermeasures on the capacity to learn
Measures of the capacity to learn
Experimental results
Summary, reflections and questions

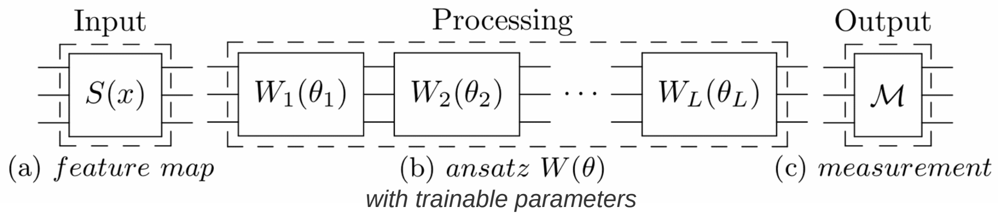
(Brief)

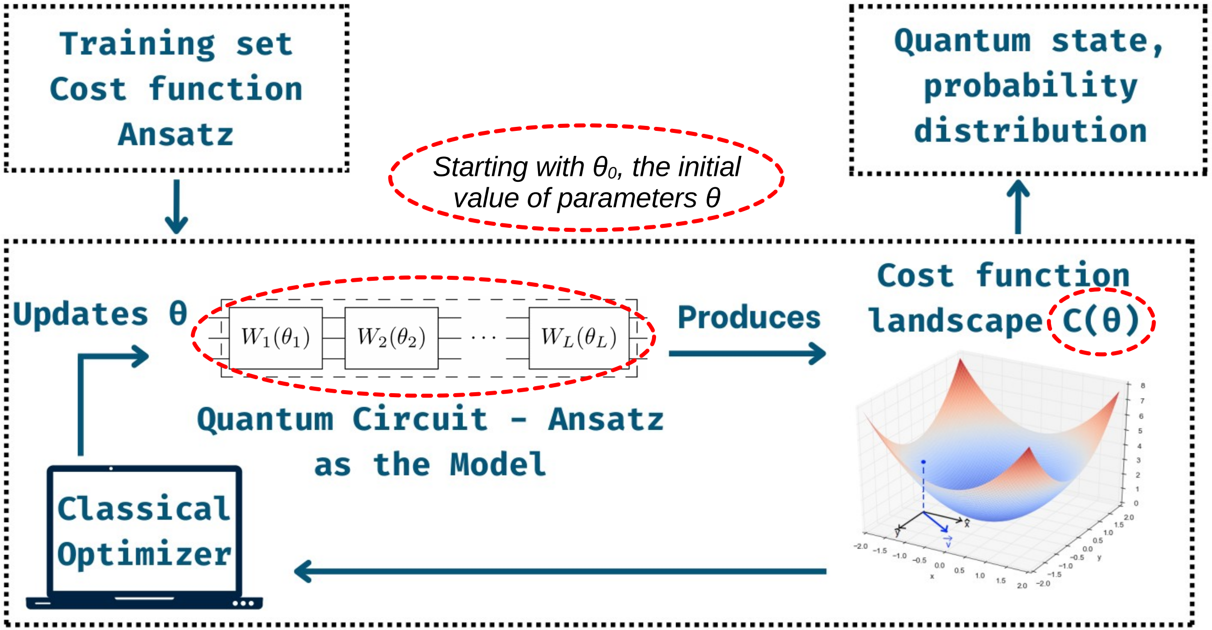
Strategies for dealing with barren plateaus in training quantum machine learning models

Jacob L. Cybulski[1] and Thanh Nguyen[2]

[1]*School of IT, SEBE, Deakin University, Australia*
[2]*Institute of R&D, Duy Tan University, Da Nang City, Vietnam.*

VQM Training
and Barren Plateaus





Variational Quantum Model (VQM)
Variational Quantum Algorithm (VQA)

A typical hybrid process of VQM training, where a classical optimiser varies trainable parameters to optimise a quantum model to support some analytic task.

The process may be affected by

Barren plateaus (BPs)

which are large “flat” areas in the quantum model’s cost landscape, which impede the effective model optimisation.

The phenomenon is related to the problem of vanishing gradients in classical neural network training

What causes barren plateaus?

Other factors, such as data, optimiser, noise, etc.

Barren Plateaus

The true nature of BPs is commonly idealised and in consequence misunderstood.

BPs do not just “exist” -
BPs emerge in the model training

BPs are not nice and flat -
BPs may be rough and bumpy

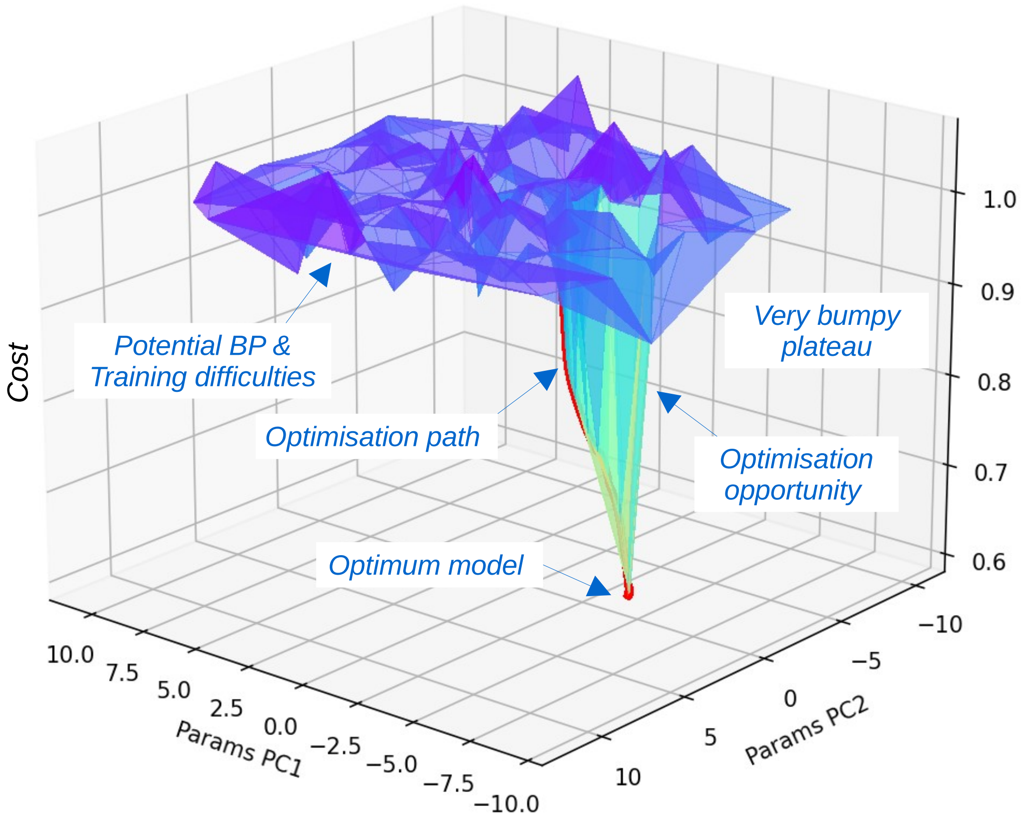
BPs presence is often declared when facing difficulties in model training -
BPs are only one reason for this to happen

BPs presence does not mean the model is bad –
BPs just make the model difficult to train

There exist many well-known causes of BPs -
There are many well-known BP countermeasures

BP strategies should not be applied as a precaution -
BP countermeasures can make your model worse!
So apply them with caution...

A 3D projection of a sample cost landscape for a simple quantum model of 114 parameters, i.e. its parameter space is 114 dimensional.



Approaches to BPs
Conducted experiments

Can BP mitigation strategies be detrimental to the model's capacity to learn ?

- *Focus on ansatz depth*

  Many qubits / layers / parameters invite BPs
- *Focus on cost*

  Global cost functions (measure all qubits) promote BPs
- *Focus on parameters initialisation*

  Random parameters encourage BPs
- *Focus on the current practice (baseline)*

  large circuits +
  randomly initialised +
  measured with a global cost function

Method / Aspect	**#0: Generic - No BP strategy**	**#1: Shallow circuit with Local cost**	**#2: Layerwise learning**	**#3: Identity blocks**
			Both search for the best initialisation, subsequent training as in other methods	
Ansatz depth	Any	Shallow	Any	Any
Cost Function	Global	Local	Global	Global
Initial Params	Randomised	Randomised	Restricted	Restricted
Reported BP Effect	Unpredictable	Eliminated [16]	Avoided [14]	Avoided [17]

Conducted Experiments

[14] **Skolik**, A., McClean, J. R., Mohseni, M., van der Smagt, P. & Leib, M. Layerwise learning for quantum neural networks. *Quantum Machine Intelligence* 3, 1–11 (2021).

[16] **Cerezo**, M., Sone, A., Volkoff, T., Cincio, L. & Coles, P. J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications* 12, 1791 (2021).

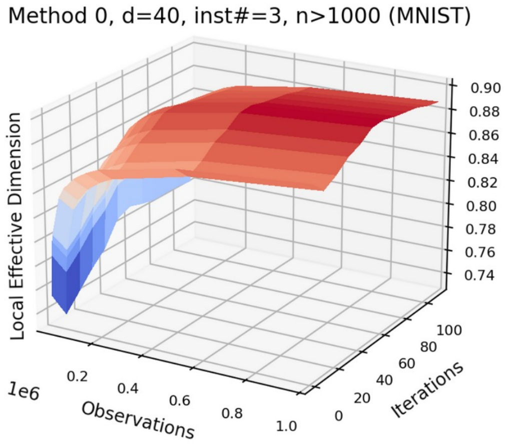
[17] **Grant**, E., Wossnig, L., Ostaszewski, M. & Benedetti, M. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum* 3, 214 (2019).

Capacity to Learn
And how it was used

Measures of quantum model "capacity to learn"

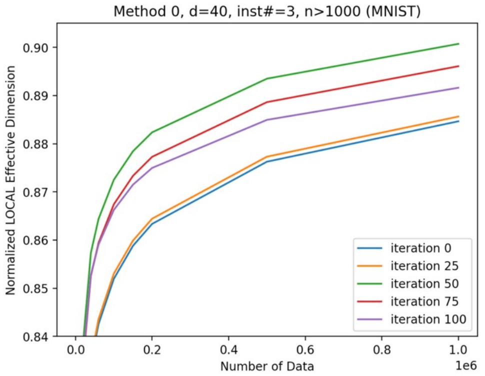
global effective dimension (GED)
(Abbas et al 2020 arXiv, then Nature 2021)
Combines Vapnik and Chervonenkis Effective Dimension and Fisher Information Matrix as a static, probabilistic measure of the model complexity as the geometry of its entire parameter space (via gradients)

local effective dimension (LED)
(Abbas et al 2021 arXiv)
Dynamic geometric measure of the model complexity in training, derived from GED, but affected by data distribution and optimisation algorithm



Each BP method was tested using a QNN classifier.
Each classifier had 10 instances.
Each instance was trained on Iris and MNIST data.

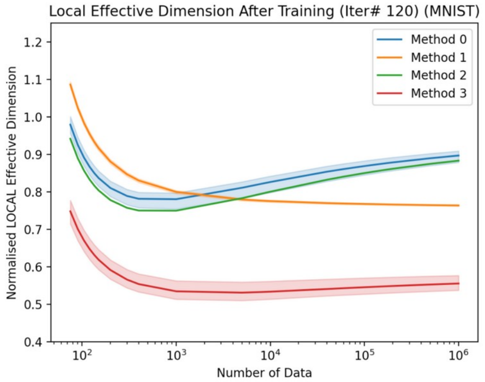
GED and LED were calculated in the process.



The LED results were depicted in 3D.

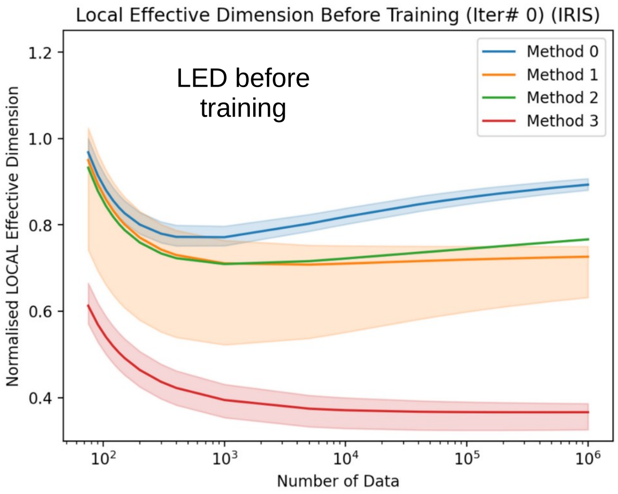
To simplify LED 3D visualisation we 2D plot only the selected data slices.

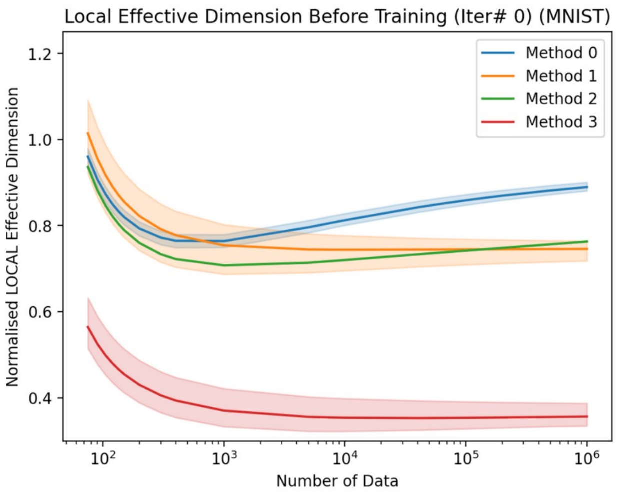
QNN instances were plotted with their mean and variance bands, and x in log-scale.

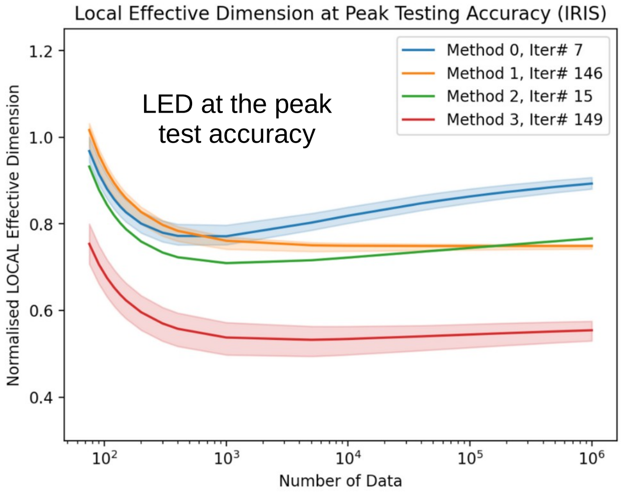


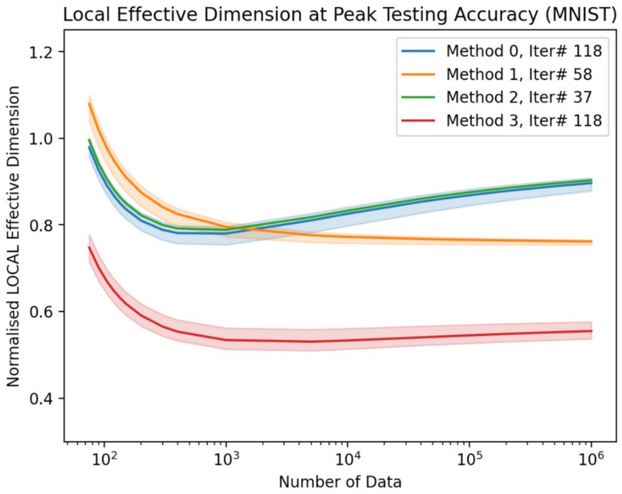
LED Results

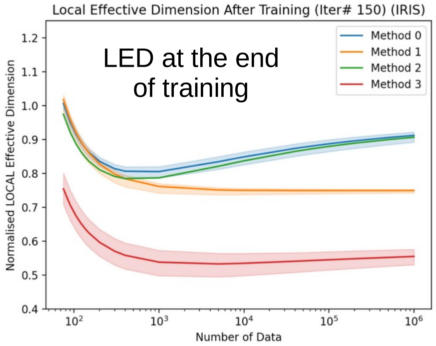
- LED evolves with the training
- LED scales with test scores
- LED variance reduces with training
- GED is a good predictor of the LED at the peak model test score
- Training volatility leads to anomalies











Observe that for IRIS method #2, LED improves past the peak in testing accuracy, due to high volatility in training.

Insights

Avoids BPs while its accuracy and capacity to learn is not worse than ignoring BPs

- Ignoring BPs (baseline)
  Had the best accuracy and LED

  However, the prior studies warn such models are BP prone, hence we recommend to use this approach with a great deal of caution!

- Models with shallow circuits and local cost
  Converged earliest in training and reduced their large initial variance quickest

  However, the model’s small overall capacity was insufficient to learn complex data

- The layerwise parameter initialisation
  Had excellent accuracy scores and consistently high LED

  However, it was computationally expensive

- Identity blocks parameter initialisation
  Had high accuracy scores

  However, they had worst LED in all tests

  Note that authors warn of the method’s sensitivity to the number and depth of identity blocks

- Critical insight

  With the exception of method #0 all tested BP countermeasures altered the model ansatz!

- Other insights
  LED is more influenced by the model structure and its optimisation than data

Thank You!

And some more following...

Questions?

- In this work, we compared four methods associated with four distinct BP mitigation strategies.
- We found that each approach to dealing with BPs had a different impact on model training and the model capacity to learn.
- We demonstrated usefulness of measuring the effective dimension of quantum models to assess the evolution of their learning capacity.
- Our experimental approach to studying variational quantum models provided valuable insights into their development.