Notes for Barren Plateaus in Quantum Neural Network Training Landscapes

Steven White

2023-04-22

Abstract

This paper discusses the barren plateau problem in quantum neural network training landscapes. This note is mainly focused on the introduction and the methods of calculation of this paper.

1 Introduction

Many experimental proposals for noisy intermediate scale quantum devices involve training a parameterized quantum circuit with a classical optimization loop. Such hybrid quantumclassical algorithms are popular for applications in quantum simulation, optimization, and machine learning. Due to its simplicity and hardware efficiency, random circuits are often proposed as initial guesses for exploring the space of quantum states. We show that the exponential dimension of Hilbert space and the gradient estimation complexity make this choice unsuitable for hybrid quantum-classical algorithms run on more than a few qubits. Specifically, we show that for a wide class of reasonable parameterized quantum circuits, the probability that the gradient along any reasonable direction is non-zero to some fixed precision is exponentially small as a function of the number of qubits. We argue that this is related to the 2-design characteristic of random circuits, and that solutions to this problem must be studied.[1]

2 Methods

Assume a quantum circuit $U(\theta)$ with L parameters $\theta = (\theta_1, \theta_2, \cdots, \theta_L)$ which can be expressed as

$$U(\theta) = \prod_{l=1}^{L} U_l(\theta_l) W_l \tag{1}$$

where $U_l(\theta_l) = \exp(-i\theta_l V_l)$, and V_l is a Hermitian matrix. And a cost function $E(\theta)$ which can be expressed as

$$E(\theta) = \langle 0 | U(\theta)^\dagger H U(\theta) | 0 \rangle \tag{2}$$

where H is a Hermitian operator.

We introduce the notations $U_- = \prod_{l=0}^{k-1} U_l(\theta_l)W_l$ (I dont know why this prod begin from k = 0 ?)and $U_+ = \prod_{l=k}^{L} U_l(\theta_l)W_l$. Thus, we can calculate that the gradient of $E(\theta)$ is

$$\begin{aligned}
\partial_k E &\equiv \frac{\partial E(\theta)}{\partial \theta_k} && (3)\\
&= \langle 0|\frac{\partial U(\theta)}{\partial \theta_k}^\dagger HU(\theta)|0\rangle + \langle 0|U(\theta)^\dagger H\frac{\partial U(\theta)}{\partial \theta_k}|0\rangle && (4)\\
&= \langle 0|\frac{\partial (U_-U_+)}{\partial \theta_k}^\dagger HU(\theta)|0\rangle + \langle 0|U(\theta)^\dagger H\frac{\partial (U_-U_+)}{\partial \theta_k}|0\rangle && (5)\\
&= \langle 0|\frac{\partial U_+^\dagger}{\partial \theta_k}U_-HU_-U_+|0\rangle + \langle 0|U_+^\dagger U_-^\dagger HU_-\frac{\partial U_+}{\partial \theta_k}|0\rangle && (6)
\end{aligned}$$

Then we calculate the gradient of $U_+^\dagger$ and U_+, we can get

$$\begin{aligned}
\frac{\partial U_+}{\partial \theta_k} &= \frac{\partial U_k}{\partial \theta_k}W_k \prod_{l=k+1}^{L} U_l(\theta_l)W_l && (7)\\
&= iV_kU_kW_k \prod_{l=k+1}^{L} U_l(\theta_l)W_l && (8)\\
&= iV_kU_+ && (9)
\end{aligned}$$

and in the same way, we can get

$$\frac{\partial U_+^\dagger}{\partial \theta_k} = iU_+^\dagger V_k \tag{10}$$

Move the i to the front, we can get

$$\begin{aligned}
\frac{\partial E(\theta)}{\partial \theta_k} &= \langle 0|\frac{\partial U_+^\dagger}{\partial \theta_k}U_-HU_-U_+|0\rangle + \langle 0|U_+^\dagger U_-^\dagger HU_-\frac{\partial U_+}{\partial \theta_k}|0\rangle && (11)\\
&= i\langle 0|U_+^\dagger V_kU_-HU_-U_+|0\rangle + i\langle 0|U_+^\dagger U_-^\dagger HU_-V_kU_+|0\rangle && (12)\\
&= i\langle 0|U_+^\dagger [V_k, U_-^\dagger HU_-]U_+|0\rangle && (13)
\end{aligned}$$

and we introduce the notation V as V_k here. Thus, we can get

$$\frac{\partial E(\theta)}{\partial \theta_k} = i\langle 0|U_+^\dagger [V, U_-^\dagger HU_-]U_+|0\rangle \tag{14}$$

as the equation 3 in the paper.(I dont know if the author calculate the wrong answer here ?)

We assume that for any gradient $\partial_k E$ defined above, the circuit implements $U(\theta)$ is sufficiently deep and random such that either U_- , U_+ or both match the

Haar distribution up to the second moment, and U_- and U_+ are independent of each other.

The target is to find the minimum of $E(\theta)$, and we can use the gradient descent algorithm to find the minimum. Therefore, we will calculate the expectation value of $\partial_k E(\theta)$ and the variance of $\partial_k E(\theta)$.

First, we calculate the expectation value of $E(\theta)$:

$$\langle \partial_k E \rangle = \int \mathrm{d}U \, p(U) \partial_k E \tag{15}$$

$$= \int \mathrm{d}U \, p(U) i \langle 0 | U_+^\dagger [V, U_-^\dagger H U_-] U_+ | 0 \rangle \tag{16}$$

Beacuse U_+ and U_- are independent of each other, we can get

$$p(U)\mathrm{d}U = p(U_+)p(U_-)\mathrm{d}U_+ \mathrm{d}U_- \tag{17}$$

which allow us to calculate $\langle E(\theta) \rangle$:

$$\langle \partial_k E \rangle = \int \mathrm{d}U \, p(U) \partial_k E \tag{18}$$

$$= \int p(U_+)p(U_-)\mathrm{d}U_+ \mathrm{d}U_- i \langle 0 | U_+^\dagger [V, U_-^\dagger H U_-] U_+ | 0 \rangle \tag{19}$$

$$= \int p(U_+)\mathrm{d}U_+ i \langle 0 | U_+^\dagger [V, \int p(U_-)\mathrm{d}U_- U_-^\dagger H U_-] U_+ | 0 \rangle \tag{20}$$

$$= i \int p(U_+)\mathrm{d}U_+ \, Tr\{ \langle 0 | U_+^\dagger [V, \int p(U_-)\mathrm{d}U_- U_-^\dagger H U_-] U_+ | 0 \rangle \} \tag{21}$$

$$= i \int p(U_+)\mathrm{d}U_+ \, Tr\{ U_+ | 0 \rangle \langle 0 | U_+^\dagger \times [V, \int p(U_-)\mathrm{d}U_- U_-^\dagger H U_-] \} \tag{22}$$

$$= i \int p(U_+)\mathrm{d}U_+ \, Tr\{ \rho_+ \times [V, \int p(U_-)\mathrm{d}U_- U_-^\dagger H U_-] \} \tag{23}$$

where we introduce $\rho_+ = U_+ | 0 \rangle \langle 0 | U_+^\dagger$.

We introduce elementwise formula here which calculate the integration over the unitary group with respect to the Haar measure, which up to the first moment can be expressed as

$$\int \mathrm{d}\mu(U) U_{ij} U_{km}^\dagger = \int \mathrm{d}\mu(U) U_{ij} U_{mk}^\star = \frac{\delta_{im}\delta_{jk}}{N} \tag{24}$$

which up to the second moment can be expressed as

$$\begin{aligned} \int \mathrm{d}\mu(U) U_{ij} U_{ql} U_{km}^\dagger U_{np}^\dagger = & \frac{\delta_{ik}\delta_{qn}\delta_{jm}\delta_{lp} + \delta_{in}\delta_{qk}\delta_{jp}\delta_{lm}}{N^2 - 1} \\ & - \frac{\delta_{ik}\delta_{qn}\delta_{jp}\delta_{lm} + \delta_{in}\delta_{qk}\delta_{jm}\delta_{lp}}{N(N^2 - 1)} \end{aligned} \tag{25}$$

Using the elementwise formula, we can get

$$M = \int \mathrm{d}\mu(U) U H U^\dagger \tag{26}$$
$$= \int \mathrm{d}\mu(U) \sum_{i,j,k,l} U_{ij} H_{jk} U^\dagger_{kl} \tag{27}$$
$$= \sum_{i,j,k,l} H_{jk} \int \mathrm{d}\mu(U) U_{ij} U^\dagger_{kl} \tag{28}$$
$$= \sum_{i,j,k,l} H_{jk} \int \mathrm{d}\mu(U) U_{ij} U^\star_{lk} \tag{29}$$
$$= \sum_{i,j,k,l} H_{jk} \frac{\delta_{il}\delta_{jk}}{N} \tag{30}$$
$$= \frac{1}{N} \mathrm{Tr}\{H\} I \tag{31}$$

for who up to second moment we can use the same way which we will show in the process of derivation.

Move back to the derivation of the expectation value of $E(\theta)$ which have been expressed as

$$\langle \partial_k E \rangle = \int \mathrm{d}U \, p(U) \partial_k E \tag{32}$$
$$= i \int p(U_+) \mathrm{d}U_+ \, Tr\{\rho_+ \times [V, \int p(U_-) \mathrm{d}U_- U_-^\dagger H U_-]\} \tag{33}$$

According to our assumption that either U_- , U_+ or both match the Haar distribution up to the second moment (of course they match the Haar distribution up to the first moment).

Case 1: If U_+ is 1-design

$$\langle \partial_k E \rangle = i \mathrm{Tr}\{ \int p(U_+) \mathrm{d}U_+ \, \rho_+ \times [V, \int p(U_-) \mathrm{d}U_- U_-^\dagger H U_-]\} \tag{34}$$
$$= i \mathrm{Tr}\{ \int \mu(U_+) \, \rho_- \times [V, \int p(U_-) \mathrm{d}U_- U_-^\dagger H U_-]\} \tag{35}$$
$$= i \mathrm{Tr}\{ \frac{1}{N} \mathrm{Tr}\{O\} I \times [V, \int p(U_-) \mathrm{d}U_- U_-^\dagger H U_-]\} \tag{36}$$
$$= \frac{i}{N} \mathrm{Tr}\{O\} \mathrm{Tr}\{[V, \int p(U_-) \mathrm{d}U_- U_-^\dagger H U_-]\} \tag{37}$$
$$= 0 \tag{38}$$

where we use the fact that the trace of a commutator is zero.

Case 2: If U_- is 1-design

$$\langle \partial_k E \rangle = i \int p(U_+) \mathrm{d}U_+ \, Tr\{\rho_+ \times [V, \int \mu(U_-) U_-^\dagger H U_-]\} \tag{39}$$

$$= i \int p(U_+) \mathrm{d}U_+ \, Tr\{\rho_+ \times [V, \frac{1}{N} \mathrm{Tr}\{H\} I]\} \tag{40}$$

$$= \frac{i}{N} \mathrm{Tr}\{H\} \int p(U_+) \mathrm{d}U_+ \, Tr\{\rho_- \times [V, I]\} \tag{41}$$

$$= 0 \tag{42}$$

So we can say that the expectation value of the gradient of $E(\theta)$ is zero, which means that the gradient of $E(\theta)$ is unbiased.

The variance of the gradient of $E(\theta)$ is

$$\mathrm{Var}(\partial_k E) = \langle (\partial_k E)^2 \rangle - \langle \partial_k E \rangle^2 \tag{43}$$

$$= \langle (\partial_k E)^2 \rangle \tag{44}$$

$$= \int \mathrm{d}U \, p(U) (\partial_k E)^2 \tag{45}$$

$$= \int \mathrm{d}U \, p(U) (i \langle 0 | U_+^\dagger [V, U_-^\dagger H U_-] U_+ | 0 \rangle)^2 \tag{46}$$

$$= \int \mathrm{d}U \, p(U) \mathrm{Tr}\{(i \langle 0 | U_+^\dagger [V, U_-^\dagger H U_-] U_+ | 0 \rangle)^2\} \tag{47}$$

$$= -\int \mathrm{d}U \, p(U) \mathrm{Tr}\{U_+ |0\rangle\langle 0| U_+^\dagger [V, U_-^\dagger H U_-] U_+ |0\rangle\langle 0| U_+^\dagger [V, U_-^\dagger H U_-]\} \tag{48}$$

Case 1: If U_+ is 2-design

$$\mathrm{Var}(\partial_k E) = -\int \mathrm{d}U \, p(U) \mathrm{Tr}\{U_+ O U_+^\dagger A U_+ O U_+^\dagger A\} \tag{49}$$

where we introduce the notation that $A = [V, U_-^\dagger H U_-]$ and $O = |0\rangle\langle 0|$.

Then we can expand the trace of $U_+ O U_+^\dagger A U_+ O U_+^\dagger A$

$$Var(\partial_k E) = -\int \mathrm{d}U \, p(U) \mathrm{Tr}\{U_+ O U_+^\dagger A U_+ O U_+^\dagger A\} \tag{50}$$

$$= -\int \mathrm{d}U \, p(U) \mathrm{Tr}\{\sum U_{ij} O_{jk} U_{kl}^\dagger A_{lm} U_{mn} O_{no} U_{op}^\dagger A_{pq}\} \tag{51}$$

beacuse $\mathrm{Tr}\{M\} = \sum M_{ii}$, we can get

$$Var(\partial_k E) = -\int \mathrm{d}U\, p(U) \sum_{i,j,k,l,m,n,o,p} U_{ij} O_{jk} U^{\dagger}_{kl} A_{lm} U_{mn} O_{no} U^{\dagger}_{op} A_{pi} \tag{52}$$

$$= -\int \mathrm{d}U\, p(U) \sum_{i,j,k,l,m,n,o,p} U_{ij} U_{mn} U^{\dagger}_{kl} U^{\dagger}_{op} O_{jk} O_{no} A_{lm} A_{pi} \tag{53}$$

$$= -\int \mathrm{d}U\, p(U) \sum_{i,j,k,l,m,n,o,p} U_{ij} U_{mn} U^{\star}_{lk} U^{\star}_{po} O_{jk} O_{no} A_{lm} A_{pi} \tag{54}$$

$$= -\int \mathrm{d}U_-\, p(U_-) \sum O_{jk} O_{no} A_{lm} A_{pi} \int \mathrm{d}U_+\, p(U_+) U_{ij} U_{mn} U^{\star}_{lk} U^{\star}_{po} \tag{55}$$

Then we can use the formula(25) which we have introduced before. We will devide the expression into four parts. The first part is

$$\sum \delta_{il}\delta_{mp}\delta_{jk}\delta_{no} O_{jk} A_{lm} O_{no} A_{pi} = \sum O_{jj} O_{nn} A_{im} A_{mi} = \mathrm{Tr}\{A^2\}\mathrm{Tr}^2\{O\} \tag{56}$$

calculate other three parts in the same way, we can get

$$\sum \delta_{ip}\delta_{ml}\delta_{jo}\delta_{nk} O_{jk} O_{no} A_{lm} A_{pi} = \sum A_{ii} A_{mm} O_{ok} O_{ko} = \mathrm{Tr}\{O^2\}\mathrm{Tr}^2\{A\} \tag{57}$$

$$\sum \delta_{il}\delta_{mp}\delta_{jo}\delta_{nk} O_{jk} O_{no} A_{lm} A_{pi} = \sum A_{im} A_{mi} O_{ok} O_{ko} = \mathrm{Tr}\{O^2\}\mathrm{Tr}\{A^2\} \tag{58}$$

$$\sum \delta_{ip}\delta_{ml}\delta_{jk}\delta_{no} O_{jk} A_{lm} O_{no} A_{pi} = \sum O_{jj} O_{nn} A_{ii} A_{mm} = \mathrm{Tr}^2\{A\}\mathrm{Tr}^2\{O\} \tag{59}$$

take all the four parts back to the expression of $Var(\partial_k E)$, we can get

$$Var(\partial_k E) = -\int \mathrm{d}U_-\, p(U_-) \sum O_{jk} O_{no} A_{lm} A_{pi} \int \mathrm{d}U_+\, p(U_+) U_{ij} U_{mn} U^{\star}_{lk} U^{\star}_{po} \tag{60}$$

$$= -\int \mathrm{d}U_-\, p(U_-)\Big(\frac{1}{N^2-1}(\mathrm{Tr}\{A^2\}\mathrm{Tr}^2\{O\} + \mathrm{Tr}\{O^2\}\mathrm{Tr}^2\{A\})$$
$$- \frac{1}{N(N^2-1)}(\mathrm{Tr}\{O^2\}\mathrm{Tr}\{A^2\} + \mathrm{Tr}^2\{A\}\mathrm{Tr}^2\{O\})\Big) \tag{61}$$

where we have known that $A = [V, U_-^{\dagger} H U_-]$, $O = |0\rangle\langle 0|$. In order to consistent with the paper, we will use the notation $H_u = U_-^{\dagger} H U_-$, take them all back to the expression of $Var(\partial_k E)$, we can get

$$Var(\partial_k E) = -\int \mathrm{d}U_-\, p(U_-)(\frac{1}{N^2-1}(\mathrm{Tr}\{[V,H_u]^2\}\mathrm{Tr}^2\{O\} + \mathrm{Tr}\{O^2\}\mathrm{Tr}^2\{[V,H_u]\}) - \frac{1}{N(N^2-1)}(\mathrm{Tr}\{O^2\}\mathrm{Tr}\{[V,H_u]^2\} + \mathrm{Tr}^2\{[V,H_u]\}\mathrm{Tr}^2\{O\})) \tag{62}$$

$$= -\frac{1}{N^2-1}(\mathrm{Tr}\{O\}\langle\mathrm{Tr}\{[V,H_u]^2\}\rangle + \mathrm{Tr}\{O^2\}\langle\mathrm{Tr}^2\{[V,H_u]\}\rangle) + \frac{1}{N(N^2-1)}(\mathrm{Tr}\{O^2\}\langle\mathrm{Tr}\{[V,H_u]^2\}\rangle + \mathrm{Tr}^2\{O\}\langle\mathrm{Tr}^2\{[V,H_u]\}\rangle) \tag{63}$$

$$= -\frac{1}{2^{2n}-1}(\mathrm{Tr}\{O\}\langle\mathrm{Tr}\{[V,H_u]^2\}\rangle + \mathrm{Tr}\{O^2\}\langle\mathrm{Tr}^2\{[V,H_u]\}\rangle) + \frac{1}{2^n(2^{2n}-1)}(\mathrm{Tr}\{O^2\}\langle\mathrm{Tr}\{[V,H_u]^2\}\rangle + \mathrm{Tr}^2\{O\}\langle\mathrm{Tr}^2\{[V,H_u]\}\rangle) \tag{64}$$

where $\langle\cdot\rangle$ means the average over the Haar measure.

Case 2: If U_- is 2-design

$$Var(\partial_k E) = -\int \mathrm{d}U\, p(U)\mathrm{Tr}\{U_+|0\rangle\langle 0|U_+^\dagger[V,U_-^\dagger HU_-]U_+|0\rangle\langle 0|U_+^\dagger[V,U_-^\dagger HU_-]\} \tag{65}$$

$$= -\int \mathrm{d}U\, p(U)\mathrm{Tr}\{\rho_+[V,U_-^\dagger HU_-]\rho_+[V,U_-^\dagger HU_-]\} \tag{66}$$

$$= -\int \mathrm{d}U\, p(U)\mathrm{Tr}\{\rho_+(VU_-^\dagger HU_- - U_-^\dagger HU_-V)\rho_+(VU_-^\dagger HU_- - U_-^\dagger HU_-V)\} \tag{67}$$

$$= -\int \mathrm{d}U\, p(U)\mathrm{Tr}\{(\rho_+VU_-^\dagger HU_-\rho_+VU_-^\dagger HU_-) - (\rho_+VU_-^\dagger HU_-\rho_+U_-^\dagger HU_-V) - (\rho_+U_-^\dagger HU_-V\rho_+VU_-^\dagger HU_-) + (\rho_+U_-^\dagger HU_-V\rho_+U_-^\dagger HU_-V)\}$$

where $\rho_+ = U_+|0\rangle\langle 0|U_+^\dagger$. We can see that each term is similar to the expression of $Var(\partial_k E)$ in case 1, so we can express each part as

$$Part\, i = -\int \mathrm{d}U\, p(U) \sum_{i,j,k,l,m,n,o,p} U_{ij}A_{jk}U^\dagger_{kl}B_{lm}U_{mn}C_{no}U^\dagger_{op}D_{pi} \tag{68}$$

$$= -\int \mathrm{d}U\, p(U) \sum_{i,j,k,l,m,n,o,p} U_{ij}U_{mn}U^\star_{lk}U^\star_{po}A_{jk}B_{lm}C_{no}D_{pi} \tag{69}$$

$$= -\int \mathrm{d}U_-\, p(U_-)\sum A_{jk}B_{lm}C_{no}D_{pi}\int \mathrm{d}U_+\, p(U_+)U_{ij}U_{mn}U^\star_{lk}U^\star_{po} \tag{70}$$

and we can use the same method to calculate the integral of $\int \mathrm{d}U_+\, p(U_+)U_{ij}U_{mn}U^\star_{lk}U^\star_{po}$ as the case 1. But its form is not AUOU+AUOU+, so we can't use the case 1 directly

$$\sum \delta_{il}\delta_{mp}\delta_{jk}\delta_{no}A_{jk}B_{lm}C_{no}D_{pi} = \sum A_{jj}C_{nn}B_{im}D_{mi} = \mathrm{Tr}\{A\}\mathrm{Tr}\{C\}\mathrm{Tr}\{BD\} \tag{71}$$

$$\sum \delta_{ip}\delta_{ml}\delta_{jo}\delta_{nk}A_{jk}C_{no}B_{lm}D_{pi} = \sum D_{ii}B_{mm}A_{ok}C_{ko} = \mathrm{Tr}\{B\}\mathrm{Tr}\{D\}\mathrm{Tr}\{AC\} \tag{72}$$

$$\sum \delta_{il}\delta_{mp}\delta_{jo}\delta_{nk}A_{jk}C_{no}B_{lm}D_{pi} = \sum B_{im}D_{mi}A_{ok}C_{ko} = \mathrm{Tr}\{BD\}\mathrm{Tr}\{AC\} \tag{73}$$

$$\sum \delta_{ip}\delta_{ml}\delta_{jk}\delta_{no}A_{jk}B_{lm}C_{no}D_{pi} = \sum A_{jj}C_{nn}D_{ii}B_{mm} = \mathrm{Tr}\{A\}\mathrm{Tr}\{B\}\mathrm{Tr}\{C\}\mathrm{Tr}\{D\} \tag{74}$$

so we can get

$$\begin{aligned} Part\, i &= -\int \mathrm{d}U_-\, p(U_-) \sum A_{jk}B_{lm}C_{no}D_{pi} \int \mathrm{d}U_+\, p(U_+) U_{ij}U_{mn}U^\star_{lk}U^\star_{po} \quad (75) \\ &= -\int \mathrm{d}U_+\, p(U_+)\Big(\frac{1}{N^2-1}(\mathrm{Tr}\{A\}\mathrm{Tr}\{C\}\mathrm{Tr}\{BD\} + \mathrm{Tr}\{B\}\mathrm{Tr}\{D\}\mathrm{Tr}\{AC\}) \\ &\quad - \frac{1}{N(N^2-1)}(\mathrm{Tr}\{BD\}\mathrm{Tr}\{AC\} + \mathrm{Tr}\{A\}\mathrm{Tr}\{B\}\mathrm{Tr}\{C\}\mathrm{Tr}\{D\})\Big) \quad (76) \end{aligned}$$

where

$$\begin{cases} A_1 = \rho_+ V \\ B_1 = H \\ C_1 = \rho_+ V \\ D_1 = H \end{cases} \quad \begin{cases} A_2 = \rho_+ \\ B_2 = H \\ C_2 = V\rho_+ V \\ D_2 = H \end{cases} \quad \begin{cases} A_3 = V\rho_+ V \\ B_3 = H \\ C_3 = \rho_+ \\ D_3 = H \end{cases} \quad \begin{cases} A_4 = V\rho_+ \\ B_4 = H \\ C_4 = V\rho_+ \\ D_4 = H \end{cases}$$

so we can get

$$
\begin{aligned}
Var(\partial_k E) =& -\int \mathrm{d}U_+\, p(U_+) \sum\Big(\frac{1}{N^2-1}(\mathrm{Tr}\{A\}\mathrm{Tr}\{C\}\mathrm{Tr}\{BD\} + \mathrm{Tr}\{B\}\mathrm{Tr}\{D\}\mathrm{Tr}\{AC\}) \\
& - \frac{1}{N(N^2-1)}(\mathrm{Tr}\{BD\}\mathrm{Tr}\{AC\} + \mathrm{Tr}\{A\}\mathrm{Tr}\{B\}\mathrm{Tr}\{C\}\mathrm{Tr}\{D\})\Big) \quad (77) \\
=& -\int \mathrm{d}U_+\, p(U_+)\Big(\frac{2}{N^2-1}(\mathrm{Tr}^2\{\rho_+ V\}\mathrm{Tr}\{H^2\} + \mathrm{Tr}^2\{H\}\mathrm{Tr}\{\rho_+ V \rho_+ V\} \\
& - \mathrm{Tr}\{\rho_+\}\mathrm{Tr}\{V\rho_+ V\}\mathrm{Tr}\{H^2\} - \mathrm{Tr}^2\{H\}\mathrm{Tr}\{\rho_+ V\rho_+ V\}) \\
& - \frac{2}{N(N^2-1)}(\mathrm{Tr}\{H^2\}\mathrm{Tr}\{\rho_+ V\rho_+ V\} + \mathrm{Tr}^2\{\rho_+ V\}\mathrm{Tr}^2\{H\} \\
& - \mathrm{Tr}\{H^2\}\mathrm{Tr}\{\rho_+ V\rho_+ V\} - \mathrm{Tr}\{\rho_+\}\mathrm{Tr}^2\{H\}\mathrm{Tr}\{V\rho_+ V\})\Big) \quad (78) \\
=& -\int \mathrm{d}U_+\, p(U_+)\Big(\frac{2}{2^{2n}-1}(\mathrm{Tr}^2\{\rho_+ V\}\mathrm{Tr}\{H^2\} + \mathrm{Tr}^2\{H\}\mathrm{Tr}\{\rho_+ V \rho_+ V\} \\
& - \mathrm{Tr}\{\rho_+\}\mathrm{Tr}\{V\rho_+ V\}\mathrm{Tr}\{H^2\} - \mathrm{Tr}^2\{H\}\mathrm{Tr}\{\rho_+ V\rho_+ V\}) \\
& - \frac{2}{2^n(2^{2n}-1)}(\mathrm{Tr}\{H^2\}\mathrm{Tr}\{\rho_+ V\rho_+ V\} + \mathrm{Tr}^2\{\rho_+ V\}\mathrm{Tr}^2\{H\} \\
& - \mathrm{Tr}\{H^2\}\mathrm{Tr}\{\rho_+ V\rho_+ V\} - \mathrm{Tr}\{\rho_+\}\mathrm{Tr}^2\{H\}\mathrm{Tr}\{V\rho_+ V\})\Big) \quad (79)
\end{aligned}
$$

so the result is

$$
\begin{aligned}
Var(\partial_k E) =& -\Big(\frac{2}{2^{2n}-1}(\mathrm{Tr}\{H^2\}\langle\mathrm{Tr}^2\{\rho_+ V\}\rangle + \mathrm{Tr}^2\{H\}\langle\mathrm{Tr}\{\rho_+ V\rho_+ V\}\rangle \\
& - \mathrm{Tr}\{H^2\}\langle\mathrm{Tr}\{\rho_+\}\mathrm{Tr}\{V\rho_+ V\}\rangle - \mathrm{Tr}^2\{H\}\langle\mathrm{Tr}\{\rho_+ V\rho_+ V\}\rangle) \\
& - \frac{2}{2^n(2^{2n}-1)}(\mathrm{Tr}\{H^2\}\langle\mathrm{Tr}\{\rho_+ V\rho_+ V\}\rangle + \mathrm{Tr}^2\{H\}\langle\mathrm{Tr}^2\{\rho_+ V\}\rangle \\
& - \mathrm{Tr}\{H^2\}\langle\mathrm{Tr}\{\rho_+ V\rho_+ V\}\rangle - \mathrm{Tr}^2\{H\}\langle\mathrm{Tr}\{\rho_+\}\mathrm{Tr}\{V\rho_+ V\}\rangle)\Big) \quad (80)
\end{aligned}
$$

In both 2 cases, the exponential decay of the gradient as a function of the number of qubits is evident.

References

[1] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):4812, 2018.