Investigation of shot Noise for SU(2) gates for variational quantumcomputing

Yassin Ouali, Supervisor: Manuel Rudolph

December 16, 2024

Abstract

This brief introduction explores variational quantum computing using a single qubit. We examine two parametrizations of a general rotation: The SU(2) gates and the "ZYZ" decomposition. Our findings reveal that the "ZYZ" decomposition is computationally faster than the SU(2) approach.

Additionally, we give an introduction to shot noise and analyze the impact on these two decompositions, highlighting how noise behaves under varying conditions. We also discuss how switching between parametrization landscapes influences the system and examine its interplay with shot noise. Notably, our analysis indicates that shot noise amplifies divergently under specific circumstances.

Contents

1 Intruduction

Variational quantum computing is a field of quantum computing where one does not try to build a pure quantum computer, but rather tries to use benefits of both worlds from ordinary silicon based computers and quantum computers. In this field one tries to define an algorithm on a quantum computer where some classical parameters are included to optimize the algorithm. In this work we will work with cost functions of the type:

$$C(\theta) = \langle\psi|U^{\dagger}(\theta)BU(\theta)|\psi\rangle$$

Here $|\Psi\rangle$ is the prepared input state for the Quantum algorithm, B is the Obersvable one want to for example minimize and $U(\theta)$ is a unitary transformation controled by a classical paramter θ. What makes this algorithm variational, is that this parameter is just a real number which is an input for the circuit. One can now imagine to measure this cost function multiple times in order to optimize θ, so that this expectation value gets minimized. In the case of $B =$ Hamiltonian H, this task is equivalent to trying to find the groundstate of a quantum mechanical system. In the beginning of the following work we will start our consideration on a single qubit system. A unitary transformation with a free paramter then may be given by a rotation, or multiple rotations. If one tries to reach the hole blochsphere to ensure that one finds the global minimum of a system one uses a general rotation which is given by the elements of the Group $\mathbb{SU}(2)$. One example to Parametrize the group are the euler rotations given by

$$U(\boldsymbol{\theta}) = R_Z(\theta_3)\, R_Y(\theta_2)\, R_Z(\theta_1) \tag{1}$$

where $R_A(\theta) = \exp\{i\theta\underline{A}\}$ is a rotation generated by the Pauli matrices $\underline{X}, \underline{Y}, \underline{Z}$, thus a rotation around the corresponding axis. Those gates are also called Pauli gates. In quantumcircuit notation for such a kind of measurement

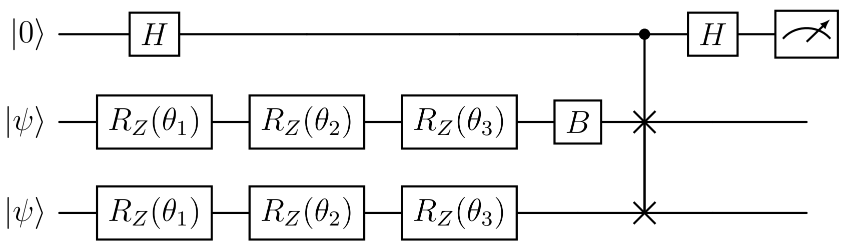

Figure 1: Setup for a variational quantum circuit where pauli gates on a single qubit are varied and the cost function is measuered via the Swap test.

In this work different parametrisations of the $\mathbb{SU}(2)$ to create the unitary transformation are compared in terms of theire convergenz speed to a lokal minimum and shot noise due to the measurement of the cost function.

2 Different Parametrisations of $\mathbb{SU}(2)$

Imagine triying to build a one qubit variational quantum circuit. What Parametrization of $\mathbb{SU}(2)$ am I gonna to choose? Different paramtrizations lead to different speeds of convergence, when somoane uses gradient descent for example. In the following part we are going to follow [3]. In this paper it is advertised, that $\mathbb{SU}(N)$-Gates are most suitable for optimizing such a circuit. A $\mathbb{SU}(N)$-Gates is defined via the Pauli Monomials:

$$\mathcal{P}^{N_{qubits}} = \left\{ i \left(\sigma_1 \otimes ... \otimes \sigma_{N_{qubits}} \right) \right\} / \left\{ I_{N_{qubits}} \right\}$$

where $\sigma_i \in \{\underline{I}, \underline{X}, \underline{Y}, \underline{Z}\}$ and $I_{N_{qubits}} = iI^{\otimes N_{qubits}}$. With whom one creates parametrized gates via:

$$U(\boldsymbol{\phi}) = \exp\{A(\boldsymbol{\phi})\}, \quad A(\boldsymbol{\phi}) = \sum_m \phi_m G_m$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, ..., \phi_{N^2-1})$ and $\{G_m\} \in \mathcal{P}^{N_{qubits}}$. This means for $N_{qubits} = 1$, the Gate is given by:

$$U(\boldsymbol{\phi}) = \exp\{i\left(\phi_1 \underline{X} + \phi_2 \underline{Y} + \phi_3 \underline{Z}\right)\} = \exp\{i\boldsymbol{\phi} \cdot \boldsymbol{\sigma}\}$$

2.1 Hint 1: Unbiased optimization landscape

Looking at the Pauli-Gate one recognizes that each direction is parametrized equally. Meaning a change $\delta\phi$ in the coordinate ϕ_1 has an equally long rotation to the same change in the coordinate ϕ_2. Lets map the ZYZ-Decomposition of eq. (1) into the one of the Pauli-Gates. One direction of the map can be done analytically. The calculation is treated in Appendix A and results in:

$$\boldsymbol{\phi} = \frac{\arccos\left(\cos\left(\theta_2\right)\cos\left(\theta_1+\theta_3\right)\right)}{\sqrt{1-\cos^2\left(\theta_2\right)\cos^2\left(\theta_1+\theta_3\right)}} \begin{pmatrix} \sin\left(\theta_2\right)\sin\left(\theta_1-\theta_3\right) \\ \sin\left(\theta_2\right)\cos\left(\theta_1-\theta_3\right) \\ \cos\left(\theta_2\right)\sin\left(\theta_1+\theta_3\right) \end{pmatrix}$$

If we now define a speed of the gates, via the path length covered by a rotation around the Bloch sphere and want to fix the speed to for example $\sqrt{2^{N_{qubits}}} = \sqrt{2}$ one has to fix the norm of the vectors $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ to

$$1 = \phi_1^2 + \phi_2^2 + \phi_3^2 = |\theta_1| + |\theta_2| + |\theta_3|$$

Equipped with this one can compare the to maps by plotting such a hypersurface seen in figure 2. On can clearly see that the ZYZ-decomposition has different magnitues regarding different directions, so that the landscape is biased. This is one advantage for the Pauli gates.

2.2 Hint 2: Most efficient Gradient.

Without clear calculation, the paper stresses that the gradient of the Pauli Gates point directly to the minimum whereas the gradient of the ZYZ-decomposition could be influenced vie unstable points on the bloch sphere. This argument is shown in figure 3

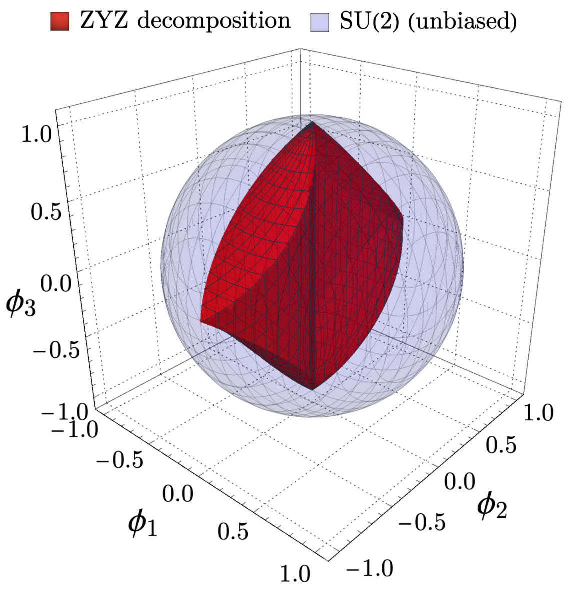

Figure 2: Comparison of unitary gates with same gate speed. The blue circle are all the parametrisations possible for a pauli gate with gatespeed $\sqrt{2}$. The red surface are all possible parametrisations via the ZYZ-decomposition with the same gatespeed.

3 Numerical Comparision

Lets now start with some numerics. We start with vanilla gradient descent. The gradient is calculated numerically via the implemented grad function of jax.

even tough this is just a random hamiltonian, it depicts several properties which happen quite often in numerical situations:

- The norm of the gradients is much smaller for the Pauli Gates than for the ZYZ-decomposition.
- The Pauli gates are slower than the ZYZ-decomposition!

Lets make a statistical analysis. The difference of the cost function averaged over 50 random hamilitonians and 10 random initial values is plotted in figure 5.

This result is very suprising. Besides its possible advantage, the optimization in the ZYZ-decomposition is up to 6% better and apart from the first 10 steps always better. Why is that. Let's have a look at the bloch sphere. The different optimizations can be seen in figure 6. One sees that the gradient of the Pauli gates is much smaller and doesn't walk on a geodesic! Rather the ZYZ-decomposed gate goes along a geodesic.

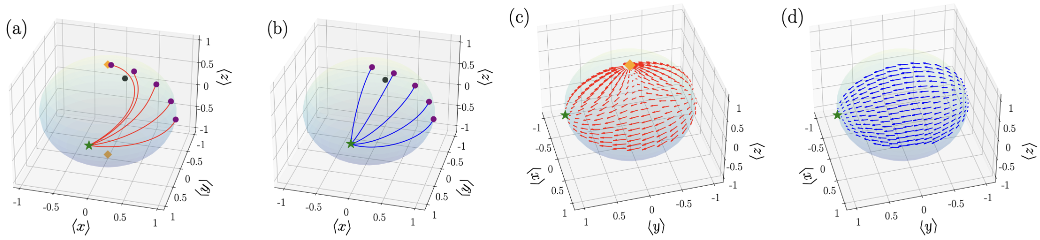

Figure 3: Comparison of the update of circuit parameters from various initial parameters acting on the initial state $\rho = |0\rangle\langle 0|$. The training paths are depicted on the Bloch sphere for: (a) parameterized single-qubit rotations using the ZYZ ansatz; and (b) the SU(N) gate. The purple dots represent initial states generated by applying $U(\boldsymbol{\theta}_0)$ with $\boldsymbol{\theta}_0 = (0, a, 0)$, where $a \in \{\pi, \pi/2, 2\pi, 3\pi, \pi/4\}$, to ρ. The unstable equilibrium points are shown as orange diamonds, located at $(0, 0, 1)$ and $(0, 0, -1)$, while the black point at $(0, 1, 0)$ corresponds to the maximum of the cost function. (c) displays the gradient vector field for the decomposed ZYZ ansatz. The vector field for the SU(2) gate is shown in (d).

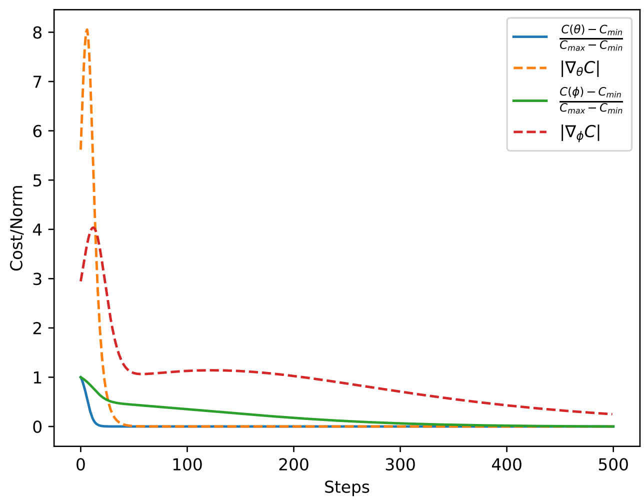

Figure 4: Vanilla gradient descent on a random Hamiltonian with the Pauli gates parametrized by $\boldsymbol{\phi}$ and the ZYZ-decomposition $\boldsymbol{\theta}$ and the the size of their respective gradient.

4 Shot Noise

The learn a bit more about the quantum nature of the system lets investigate shot noise. Shot noise is the phenomena that arises, when someone actually implements the Cost function and measures it to calculate the gradient or wnats to know the value of the costfunction. Due to the quantum nature of things one will get a distribution rather than a single value. The mean value is often of main interrest. To estimate the mean value one constructs estimators. One measurement of the system is called a shot.

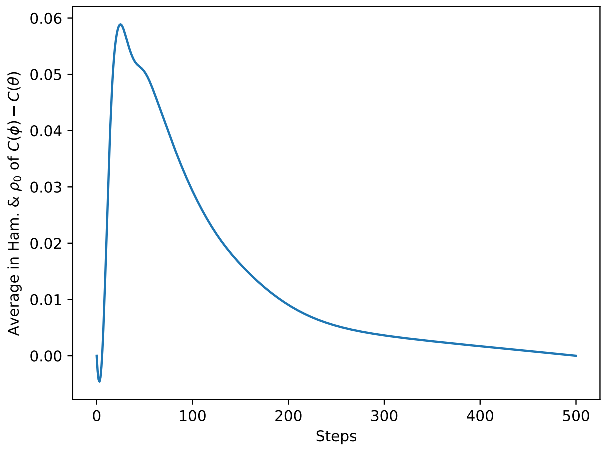

Figure 5: Averaged difference in Cost by Pauli gates and ZYZ-Decomposition. It is averaged over 50 random hamiltonians and 10 random initial values with a learning rate $\eta = 10^{-2}$.

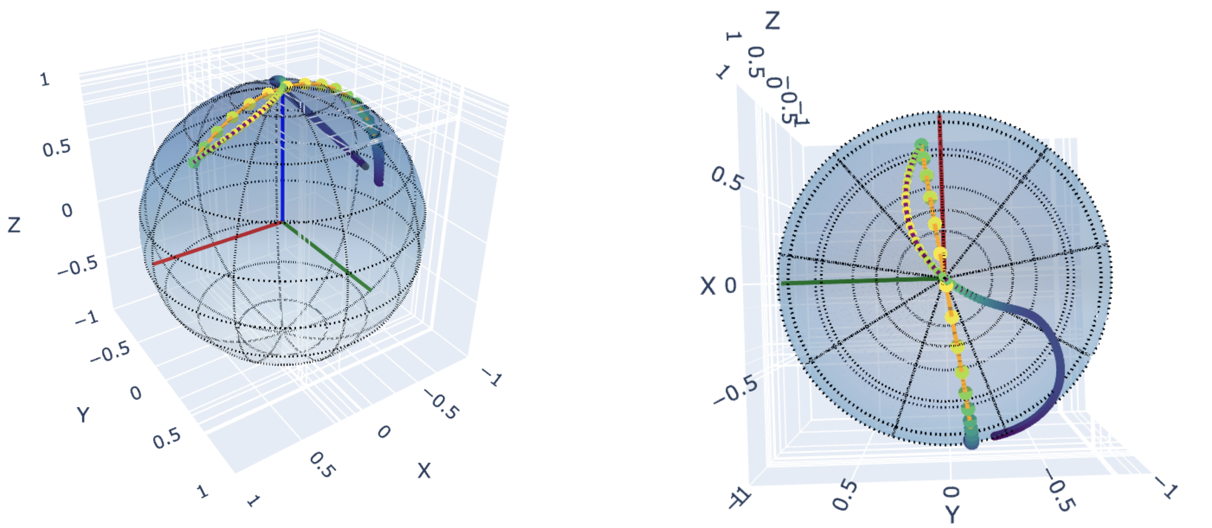

Figure 6: Vanilla gradient descent on the Bloch sphere. The blue line corresponds to the Pauli-Gate and the orange line corresponds to the ZYZ-decomposition. The dots correspont to each step and the color for the size of the gradient. They are normalized to 1 and get from large yellow to small blue.

4.1 Basics – Estimators

Let's consider an observable

$$M = \sum_{M=1}^{N} \lambda_M \left|\lambda_M\right\rangle \left\langle\lambda_M\right| \quad \Big| \text{ with } \left|\left\langle\lambda_M|\lambda_M\right\rangle\right|^2 = \delta_{ij} \text{ and } \lambda_M \in \mathbb{R} \tag{2}$$

and a density matrix in the corresponding hilbertspace ρ, where we want to know the expectation value. To get an estimation for it, one prepares copies of said state and measures N times to get an estimation. This motivates the Estimator With $\lambda_{M(s)}$ beeing the eigenvalue measured on the s^{th} shot. This is essentially a sum of stochastically independent events. The

$$\boxed{\hat{M}_N = \frac{1}{N}\sum_{s=1}^{N} \lambda_{M(s)}}$$

Expectation value is

$$\mathbb{E}\left[\hat{M}_N\right] = \frac{1}{N}\sum_{s=1}^{N}\mathbb{E}\left[\lambda_{M(s)}\right] = \frac{1}{N}\sum_{s=1}^{N}\sum_{M=1}^{N}\lambda_M\, p(\lambda_M) = \frac{1}{N}\sum_{s=1}^{N}\mathrm{Tr}\left[M\rho\right] = \mathrm{Tr}\left[M\rho\right] = \mathbb{E}\left[M\right]$$

An estimator where the expectation values matches with the corresponding observable is called an unbiased estimator. The average of the estimator for finite N is equal to the time average. Let's have a look at the variance:

$$\begin{aligned}
Var\left[\hat{M}\right] &= \mathbb{E}\left[\hat{M}_N^2\right] - \mathbb{E}\left[\hat{M}_N\right]^2 \\
\mathbb{E}\left[\hat{M}_N^2\right] &= \frac{1}{N^2}\mathbb{E}\left[\sum_{s,t=1}^{N}\lambda_{M(s)}\lambda_{M(t)}\right] = \frac{1}{N^2}\left(\sum_{s=1}^{N}\mathbb{E}\left[\lambda_{M(s)}^2\right] + \sum_{s\neq t=1}^{N}\mathbb{E}\left[\lambda_{M(s)}\lambda_{M(t)}\right]\right) \\
&= \frac{1}{N^2}\left(\sum_{s=1}^{N}\mathrm{Tr}\left[\rho M^2\right] + \sum_{s\neq t=1}^{N}\mathbb{E}\left[\lambda_{M(s)}\right]\mathbb{E}\left[\lambda_{M(t)}\right]\right) \\
&= \frac{1}{N}\mathrm{Tr}\left[\rho M^2\right] + \frac{1}{N^2}\left(N^2+N\right)\mathrm{Tr}\left[M\rho\right]^2 \\
\Rightarrow Var\left[\hat{M}\right] &= \frac{1}{N}\left(\mathrm{Tr}\left[\rho M^2\right] - \mathrm{Tr}\left[\rho M\right]^2\right) = \frac{1}{N}Var\left[M\right]
\end{aligned}$$

Hence we define an estimate precision ϵ

$$\epsilon = \sqrt{\frac{Var\left[M\right]}{N}} \quad \Leftrightarrow \quad N = \frac{Var\left[M\right]}{\epsilon^2}$$

4.2 Examples: Subtleties of Shot Noise

Consider two different ways of computing the fidelity between to states

1. Loschmidt Echo circuit

$$|\langle\psi|\phi\rangle|^2 = \langle\psi| R_\phi |0\rangle\langle 0| R_\phi^\dagger |\psi\rangle = \langle\psi| R_\phi M_A R_\phi^\dagger |\psi\rangle$$

with the variance

$$\begin{aligned}
Var\left[\hat{M}_A\right] &= \frac{Var_{|\psi\rangle}(M_A)}{N} = \frac{\langle\psi| R_\phi \left(|0\rangle\langle 0|\right)^2 R_\phi^\dagger |\psi\rangle - |\langle\psi|\phi\rangle|^4}{N} \\
&= \frac{|\langle\psi|\phi\rangle|^2 - |\langle\psi|\phi\rangle|^4}{N}
\end{aligned}$$

2. Swap Test

$$\begin{aligned}
|\langle\psi|\phi\rangle|^2 = \langle\psi|\otimes\langle\phi|\,\mathrm{SWAP}\,|\psi\rangle\otimes|\phi\rangle &= \sum_{i,j,i',j'}\psi_i^*\phi_j^*\psi_{i'}\phi_{j'}\langle ij|\,\mathrm{SWAP}\,|i'j'\rangle \\
= \sum_{i,j,i',j'}\psi_i^*\phi_j^*\psi_{i'}\phi_{j'}\delta_{ij'}\delta_{i'j} &= \sum_{ij}\psi_i^*\psi_j\phi_j^*\phi_i = |\langle\psi|\phi\rangle|^2
\end{aligned}$$

with $\mathrm{SWAP} = M_B$ and the variance

$$\begin{aligned}
Var\left[\hat{M}_B\right] &= \frac{Var_{|\psi\rangle}(M_B)}{N} = \frac{\langle\psi|\otimes\langle\phi|\,\mathrm{SWAP}^2\,|\psi\rangle\otimes|\phi\rangle - |\langle\psi|\phi\rangle|^4}{N} \\
&= \frac{1 - |\langle\psi|\phi\rangle|^4}{N}
\end{aligned}$$

Due to the fact that $|\langle\psi|\phi\rangle|^2 \leq 1$ it follows that $Var\left[\hat{M}_A\right] \leq Var\left[\hat{M}_B\right]$ so that Loschmidt Echo has less shot noise.

4.3 Parameter Shift Rule

This derivation follows [2]. Let us consider a circuit decribed in figure 7. This is a multiple parameter quantum circuit, where in general multiple parameter are variied. But we are only going to varii one parameter. Therefore the circuit containing a transformation before the gate containing θ_i is taken into account by describing the input state as $|\psi\rangle$ and the circuit afterwards by defining a new obervable B.

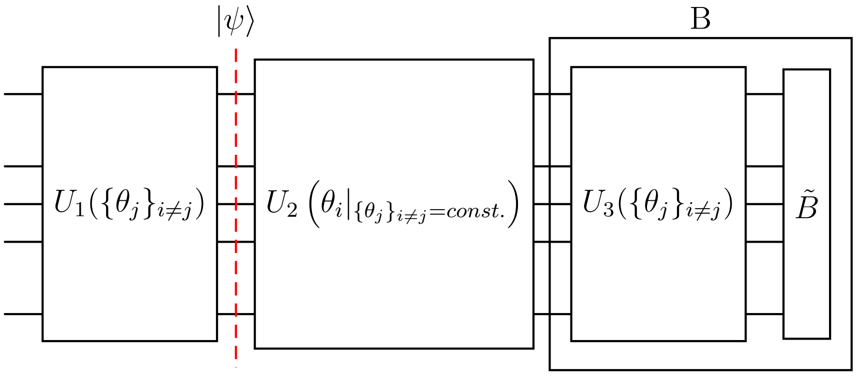

Figure 7: Variational Quantum Circuit where one parameter is varied.

The goal is to find the minimal expectation Value for B in dependence of $U_2(\theta_i) = U(\theta)$. Therfore we define a cost function:

$$C(\theta) = \langle\psi|U^\dagger(\theta)BU(\theta)|\psi\rangle \tag{3}$$

Lets say U is an unitary operation, we can fiddle around with the eigenvalue equation:

$$U(\theta)\,|u(\theta)\rangle = u_j(\theta)\,|u_j(\theta)\rangle \quad \Leftrightarrow \quad \langle u_j(\theta)|\,U(\theta)^\dagger = \langle u_j(\theta)|\,\overline{u_j(\theta)}$$
$$\Rightarrow \langle u_j(\theta)|\,U(\theta)^\dagger U(\theta)\,|u_j(\theta)\rangle = 1 = |u_j(\theta)|^2$$

So that every eigenvalue is on the unit circle in the complex plane If we now assume that the parametrized gate is generated by

$$U(\theta) = e^{i\theta G}$$

the eigenvalues can be parametrized via $u(\theta)_j = e^{i\theta\omega_j}$, where $j \in [d]$ and $[d] := 1, ..., d$ with d beeing the dimension. Lets express (3) via the eigenfunctions of U:

$$|\psi\rangle = \sum_{j=1}^{d} \psi_j\,|u_j\rangle \quad B = \sum_{i,j=1}^{d} |u_j\rangle\,\langle u_j|\,B\,|u_i\rangle\,\langle u_i| \quad U(\theta)$$
$$\Rightarrow C(\theta) = \sum_{j,k=1}^{d} \overline{\psi_j e^{i\omega_j\theta}} b_{jk}\psi_k e^{i\omega_k\theta} \quad \Big| \sum_{j,k=1}^{d} A_{jk} = \sum_{j,k=1;j<k}^{d} A_{kj} + A_{kj}^T + \sum_{k}^{d} A_{kk}$$
$$= \sum_{j,k=1;j<k}^{d} \overline{\psi_j} b_j k\psi_k e^{i(\omega_k-\omega_j)\theta} + \psi_j\overline{b_j k\psi_k e^{i(\omega_k-\omega_j)\theta}} + \sum_{k=1}^{d} |\psi_k|^2 b_{kk}$$

We now collect all terms independent of θ into coefficients $c_{jk} = \overline{\psi}_j b_{jk} \psi_k$ and introduce R unique positive differences $\{\Omega_l\}_{l\in[R]} := \{\omega_k - \omega_j | j,k \in [d], \omega_k > \omega_j\}$. Introducing a new index

$$(k,j) \mapsto l(k,j) \text{ with } l(j,k) = l(j',k') \Leftrightarrow \omega_k - \omega_j = \omega_{k'} - \omega_{j'}$$

therfore $c_l = c_{l(j,k)}$ is well defined und we can write the expectation value as a finite term Fourier Series:

$$C(\theta) = a_0 + \sum_{l=1}^{R} c_l e^{i\Omega_l \theta} + \overline{c_l} e^{-i\Omega_l \theta} \quad \Big| \; a_0 = \sum_{k=1}^{d} |\psi_k|^2 b_{kk} \tag{4}$$

$$= a_0 + \sum_{l=1}^{R} a_l \cos \Omega_l \theta + b_l \sin \Omega_l \theta \quad \Big| \; c_l =: \frac{1}{2}(a_l - b_l) \, \forall l \in [R]; \; a_l, b_l \in \mathbf{R} \tag{5}$$

Lets simplify our situation to figure 8, where our variational paramter enters only in one Pauli gate.

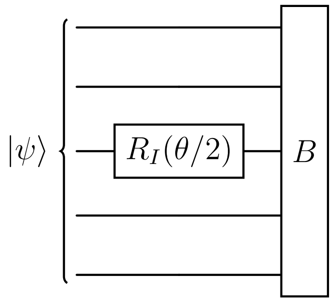

Figure 8: Variational Quantum Circuit where a pauli gate on a random qubit is varied, $I \in \{X, Y, Z\}$.

The Eigenvalues of the rotation matrices are then (I believe I didnt check):

$$\text{Eigenvalues}[\exp\{i\theta X/2\}] = e^{\pm i\theta/2} \Rightarrow \Omega_l = \Omega = \frac{1}{2} - \left(-\frac{1}{2}\right) = 1$$

Therfore Equation (5) reduces to:

$$C(\theta) = a_0 + a\cos\theta + b\sin\theta$$

Since we are interrested in minimizing the function we calculate its derivative regarding following identities (vie addition theorems)

$$\frac{d\cos(\theta)}{d\theta} = \frac{\cos(\theta+s) - \cos(\theta-s)}{2\sin(s)}$$
$$\frac{d\sin(\theta)}{d\theta} = \frac{\sin(\theta+s) - \sin(\theta-s)}{2\sin(s)}$$

with $s \neq k\pi, \; k \in \mathbb{Z}$ We get:

$$\frac{dC(\theta)}{d\theta} = \frac{C(\theta+s) - C(\theta-s)}{2\sin(s)} \tag{6}$$

Notice that there is no approximation here. This formula is exact and has a free parameter in it! This parameter will be crucial in the next section when considering short noise. This notion of a derivative is the one of a partial derivative, so that this rule can be lifted to:

$$\frac{\partial C(\theta)}{\partial \theta_k} = \frac{C(\theta + s\mathbf{e}_k) - C(\theta - s\mathbf{e}_k)}{2\sin(s)}$$

Looking back at the estimator in section 2 we defined beforehand, our cost function is just an expectation value of the operator B for a given State $U(\theta)|\psi\rangle$. The following derivation follows [1]. Lets rewrite the estimator like

$$\begin{aligned} \hat{B}_N(\theta) &= \mathbb{E}\left[\hat{B}_N(\theta)\right] + \hat{B}_N(\theta) - \mathbb{E}\left[\hat{B}_N(\theta)\right] \quad \Big| \, \mathbb{E}\left[\hat{B}_N(\theta)\right] = C(\theta) \\ &= C(\theta) + \hat{\epsilon}(\theta) = \hat{C} \end{aligned}$$

where $\hat{\epsilon}$ is a random variable with:

$$\mathbb{E}\left[\hat{\epsilon}\right] = 0 \quad \mathrm{Var}\left[\hat{\epsilon}\right] = \mathrm{Var}\left[\hat{B}_N(\theta)\right] = \frac{\mathrm{Var}\left[B(\theta)\right]}{N}$$

Inserting this into equation (6) we get:

$$\frac{\partial \hat{C}}{\partial \theta}(s) = \frac{C(\theta + s) - C(\theta - s)}{2\sin(s)} + \frac{\hat{\epsilon}(\theta + s) - \hat{\epsilon}(\theta - s)}{2\sin(s)}$$

Therfore we have:

$$\begin{aligned} \mathbb{E}\left[\frac{\partial \hat{C}}{\partial \theta}(s)\right] &= \frac{\partial C}{\partial \theta}(s) \\ \mathbb{V}\left[\frac{\partial \hat{C}}{\partial \theta}(s)\right] &= \mathbb{V}\left[\frac{\hat{\epsilon}(\theta + s) - \hat{\epsilon}(\theta - s)}{2\sin(s)}\right] = \ast \end{aligned}$$

We now calculate for two independent random variables, e.g. two shots at postion $\theta - s$ and $\theta + s$

$$\begin{aligned} \mathbb{V}\left[\alpha X + \beta Y\right] &= \mathbb{E}\left[(\alpha X + \beta Y)^2\right] - \mathbb{E}\left[\alpha X + \beta Y\right]^2 \\ &= \alpha^2 \mathbb{E}\left[X^2\right] + \beta^2 \mathbb{E}\left[Y^2\right] + 2\alpha\beta \mathbb{E}\left[XY\right] \\ &- \alpha^2 \mathbb{E}\left[X\right]^2 - \beta^2 \mathbb{E}\left[Y\right]^2 - 2\alpha\beta \mathbb{E}\left[X\right] \mathbb{E}\left[X\right] \quad \Big| \, \mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right] \mathbb{E}\left[Y\right] \\ &= \alpha^2 \mathbb{V}\left[X\right] + \beta^2 \mathbb{V}\left[Y\right] \end{aligned}$$

We also postulate before going further:

Assumption 1: The variance of the measured observable depends weakly on the parameter shift, such that $\mathbb{V}\left[B(\theta + s)\right] + \mathbb{V}\left[B(\theta - s)\right] \approx 2\mathbb{V}\left[B(\theta)\right]$

Therefore:

$$\begin{aligned} \ast &= \frac{\mathbb{V}\left[\hat{\epsilon}(\theta + s)\right] + \mathbb{V}\left[\hat{\epsilon}(\theta - s)\right]}{4\sin^2(s)} \\ &= \frac{\mathbb{V}\left[B(\theta + s)\right] + \mathbb{V}\left[B(\theta - s)\right]}{4N\sin^2(s)} \\ &\approx \frac{\mathbb{V}\left[B(\theta)\right]}{2N\sin^2(s)} \end{aligned}$$

And there we have it. The optimal parameter for the parameter shift to reduce noise on a weakly dependend observable is $s = \pm\frac{\pi}{2}$. Let's test this parameter shift rule numerically. In figure 9 we calculate the gradient for a one particle hamiltonian and have Z-Rotation as the unitary transformation. It can be clearly seen that the gradient decreases while s approaches $\pi/2$.

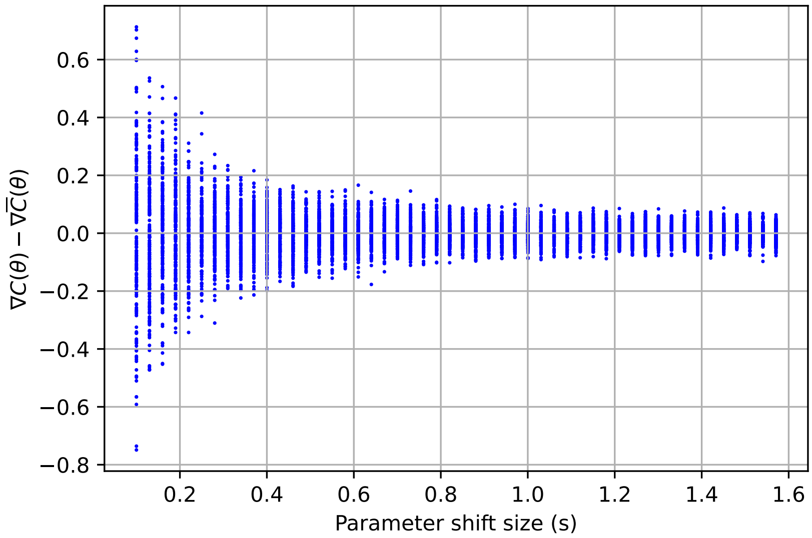

Figure 9: Numercial verification of the parameter shift rule in regards of the dependency of the variance to the shift Paramter s. 200 shots are done per s to estimate the gradient. The distribution of $C(\theta)$ is approximated as Gaussion centered at this value with a standart diviation of $\frac{\pi}{8}$.

4.4 Effect of shot noise on the Pauli- vs. ZYZ-Gates

Let's set the problem of the measurement of the cost function aside and look at what effect the shot noise would have when a standart diviation in the gradient exists. For that we repeat the experiment done in figure 5 but rather than plotting the averaged difference we plot the averaged variance. This can be seen in figure 10. The mean of this plot is ≈ 0 so that there is no difference in the decompositions.

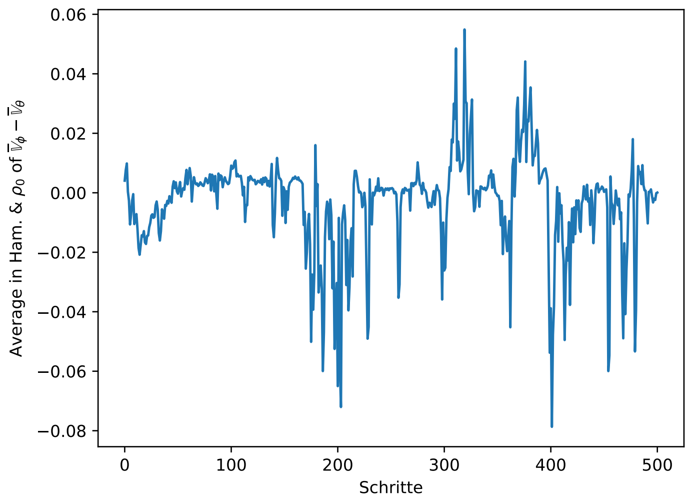

Figure 10: Averaged Variance difference in Cost by Pauli gates and ZYZ-Decomposition. It is averaged over 50 random hamiltonians and 10 random initial values with a learning rate $\eta = 10^{-2}$ over the normed cost function.

5 Mapping between landscapes

Let us now consider the following case. Physicaly we could implement the decomposed gate. We can measure our gradients in this parametrization, but as [3] shows the gradient descent in this parameter space is not going to be optimal[1]. We would rather do the gradient descent in the basis of Pauli Matrices. But we have a Map which maps the coordinates! Therefore it is possible to do the translation. The Algorithm is then:

Algorithm 1 Optimized Gradient Descent

Input: $\nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta})$ at step t via PMSR

Calculate $\nabla_{\boldsymbol{\Phi}} C(\boldsymbol{\theta})$ at step t via chain rule

Correct $\boldsymbol{\phi}(t+1) = \boldsymbol{\phi}(t) + \eta\, \nabla_{\boldsymbol{\Phi}} C(\boldsymbol{\theta})$ with step size η

Find $\boldsymbol{\theta}(\boldsymbol{\phi}(t+1)) = \boldsymbol{\theta}(t+1)$

As we see in the first step we have to calculate the gradient in respect to other coordinates. Lets remind our self, we have a coordinate transformation:

$$\boldsymbol{\phi} = \mathbf{f}(\boldsymbol{\theta}) \Leftrightarrow \boldsymbol{\theta} = \mathbf{f}^{-1}(\boldsymbol{\phi})$$

Known for us is $\mathbf{f}$ and we don't know the functional form of $\mathbf{f}^{-1}$. We now use the chain rule to get:

$$\frac{\partial C}{\partial \phi_i} = \frac{\partial C}{\partial \theta_l}\frac{\partial \theta_l}{\partial \phi_i} = \frac{\partial C}{\partial \theta_l}\frac{\partial f_l^{-1}}{\partial \phi_i} \quad \Big| \text{ with Einstein summation} \tag{7}$$

The first term was found before via the parameter shift rule. The latter needs a bit more consideration since we don't know $\mathbf{f}^{-1}$. The exact calculation of the gradient is done in B.

5.1 The effect of shot noise in the new gradient

If we now use equation (7) and implement shot noise in the measurement of the cost-function we get:

$$\begin{aligned}
\frac{\partial C}{\partial \phi_i} &= \frac{\partial C}{\partial \theta_1}\frac{\partial \theta_1}{\partial \phi_i} + \frac{\partial C}{\partial \theta_2}\frac{\partial \theta_2}{\partial \phi_i} + \frac{\partial C}{\partial \theta_3}\frac{\partial \theta_3}{\partial \phi_i} \\
\Rightarrow \mathbb{V}\left[\frac{\partial \hat{C}}{\partial \phi_i}\right] &\approx \left(\frac{\partial \theta_1}{\partial \phi_i}\right)^2 \frac{\mathbb{V}[B(\theta_1)]}{2N} + \left(\frac{\partial \theta_2}{\partial \phi_i}\right)^2 \frac{\mathbb{V}[B(\theta_2)]}{2N} + \left(\frac{\partial \theta_3}{\partial \phi_i}\right)^2 \frac{\mathbb{V}[B(\theta_3)]}{2N} \\
&\geq \left(\left(\frac{\partial \theta_1}{\partial \phi_i}\right)^2 + \left(\frac{\partial \theta_2}{\partial \phi_i}\right)^2 + \left(\frac{\partial \theta_3}{\partial \phi_i}\right)^2\right) \min_{i\in\{1,2,3\}}\left\{\frac{\mathbb{V}[B(\theta_i)]}{2N}\right\} \\
&= |\nabla_{\phi_i}\boldsymbol{\theta}|^2 \min_{i\in\{1,2,3\}}\left\{\frac{\mathbb{V}[B(\theta_i)]}{2N}\right\}
\end{aligned}$$

Therefore the shot noise is depending on the magnitude of the Gradient. Lets investigate

$$h(\boldsymbol{\theta}) = \log\left(|\nabla_{\phi_1}\boldsymbol{\theta}|^2\right)$$

It is plotted in figure 11.

We see that there are multiple poles. Therefore the shot noise diverges for some angles.

[1] We have shown the oppsosite in the investigation before. This result is currently discussed with Dr. Wiersema

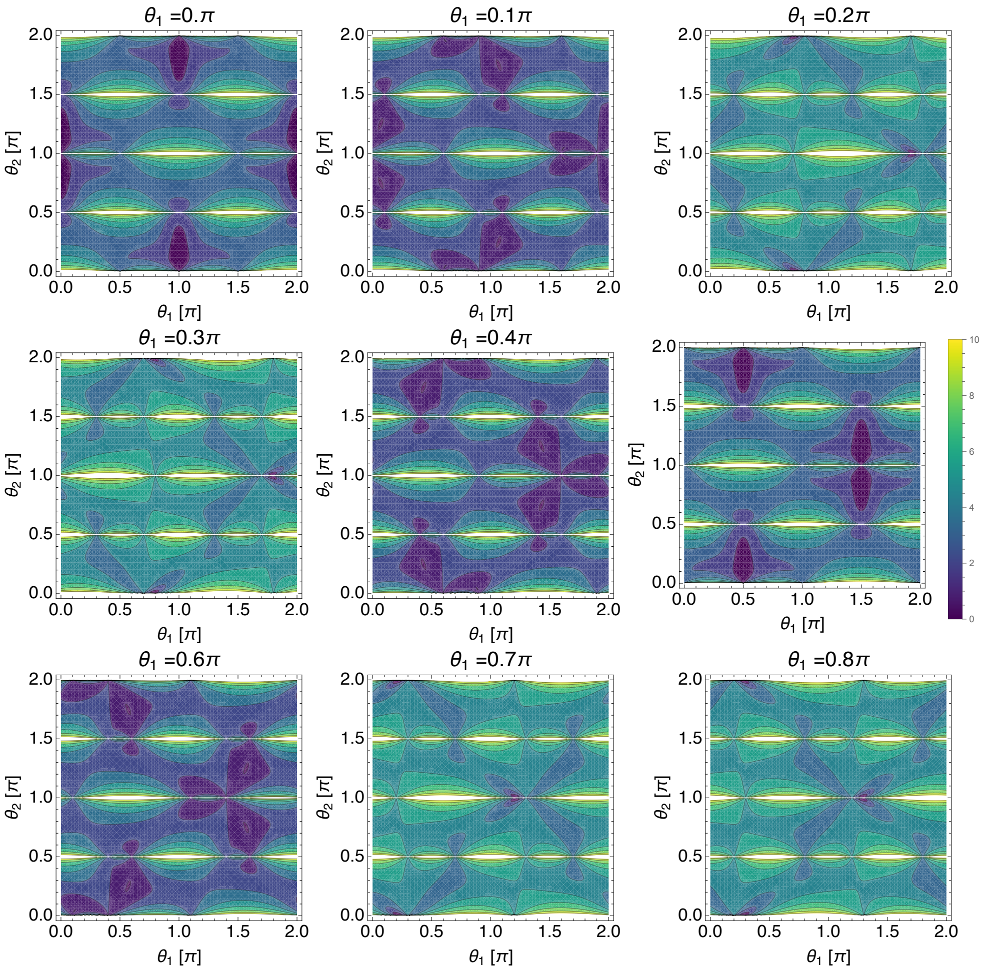

Figure 11: Plot of $h(\boldsymbol{\theta})$ in crosssections in θ_3.

A The mapping between the Pauli- and ZYZ-decomposition

We now consider the decomposed $\mathbb{SU}(2)$ gate $V(\boldsymbol{\theta}) = R_Z(\theta_3)\, R_Y(\theta_2)\, R_Z(\theta_1)$. The gate on which the optimization advantage maybe comes to light is $V(\boldsymbol{\theta}) = \exp\{i\boldsymbol{\phi}\cdot\boldsymbol{\sigma}\}$. Where one has to find $\boldsymbol{\phi}(\boldsymbol{\theta})$. To get this we just compare the two equations

$$\exp\{i\theta_1\underline{Z}\}\exp\{i\theta_2\underline{Y}\}\exp\{i\theta_3\underline{Z}\} = \exp\{i\boldsymbol{\phi}\cdot\boldsymbol{\sigma}\}$$

Lets investigate the lefthand side. Since we know that for any Pauli Matrix $\underline{H}$ it holds $\underline{H}^2 = \mathbf{1}$ we can calculate the exponetial expilitly:

$$\exp\{i\theta_i\underline{H}\} = \cos(\theta_i)\,\mathbf{1} + i\sin(\theta_i)\,\underline{H}$$

Therefore:

$$
\begin{aligned}
\text{LHS} &= \left(\cos\left(\theta_1\right)\mathbf{1} + i\sin\left(\theta_1\right)\underline{Z}\right)\left(\cos\left(\theta_2\right)\mathbf{1} + i\sin\left(\theta_2\right)\underline{Y}\right)\left(\cos\left(\theta_3\right)\mathbf{1} + i\sin\left(\theta_3\right)\underline{Z}\right) \quad \Big| \ \sin = s \ ... \\
&= \left(c_1c_2\mathbf{1} - s_1s_2\underline{ZY} + is_1c_2\underline{Z} + ic_1s_2\underline{Y}\right)\left(c_3\mathbf{1} + is_3\underline{Z}\right) \quad \Big| \ \underline{ZY} = -i\underline{X} \quad \underline{Y}\,\underline{Z} = i\underline{X} \quad \underline{ZY}\,\underline{Z} = -\underline{Y} \\
&= c_1c_2c_3\mathbf{1} + is_1s_2c_3\underline{X} + is_1c_2c_3\underline{Z} + ic_1s_2c_3\underline{Y} \\
&\quad + ic_1c_2s_3\underline{Z} + is_1s_2s_3\underline{Y} - s_1c_2s_3\mathbf{1} - ic_1s_2s_3\underline{X} \\
&= \left(c_1c_2c_3 - s_1c_2s_3\right)\mathbf{1} + i\begin{pmatrix} s_1s_2c_3 - c_1s_2s_3 \\ c_1s_2c_3 + s_1s_2s_3 \\ s_1c_2c_3 + c_1c_2s_3 \end{pmatrix} \cdot \begin{pmatrix} \underline{X} \\ \underline{Y} \\ \underline{Z} \end{pmatrix} \\
&= \left(c_2c_{1+3}\right)\mathbf{1} + i\begin{pmatrix} s_2s_{1-3} \\ s_2c_{1-3} \\ c_2s_{1+3} \end{pmatrix} \cdot \begin{pmatrix} \underline{X} \\ \underline{Y} \\ \underline{Z} \end{pmatrix} = \eta\mathbf{1} + i\begin{pmatrix} x \\ y \\ x \end{pmatrix} \cdot \begin{pmatrix} \underline{X} \\ \underline{Y} \\ \underline{Z} \end{pmatrix}
\end{aligned}
$$

In general one can find the parametrisation to an arbitrary number of type of $\mathbb{SU}(2)$ gates, by finding η, x, y, z. We now have to look at the RHS. First we notice:

$$
i\boldsymbol{\phi} \cdot \boldsymbol{\theta} = i\begin{pmatrix} \phi_3 & (\phi_1 - i\phi_2) \\ (\phi_1 + i\phi_2) & -\phi_3 \end{pmatrix} = \begin{pmatrix} i\phi_3 & (\phi_2 + i\phi_1) \\ -(\phi_2 - i\phi_1) & -i\phi_3 \end{pmatrix}
$$

And then use mathematica to get, $|\boldsymbol{\phi}| = \phi \quad si_\phi = \frac{\sin(\phi)}{\phi}$:

$$
e^{i\boldsymbol{\phi}\cdot\boldsymbol{\sigma}} = \begin{pmatrix} c_\phi + i\phi_3 si_\phi & \left(i\phi_1 + \phi_2\right) si_\phi \\ i\left(\phi_1 + i\phi_2\right) si_\phi & c_\phi - i\phi_3 si_\phi \end{pmatrix} = c_\phi\mathbf{1} + isi_\phi\boldsymbol{\phi} \cdot \boldsymbol{\sigma}
$$

Since we found representations in an orthonormal basis, we just have to compare the coefficients, which results in four equations:

$$
\begin{aligned}
c_\phi &= \eta \\
si_\phi\phi_1 &= x \\
si_\phi\phi_2 &= y \\
si_\phi\phi_3 &= z
\end{aligned}
$$

This results in

$$
\begin{aligned}
&\boldsymbol{\phi} = \frac{\arccos\left(\eta\right)}{\sqrt{1 - \cos\left(\eta^2\right)}}\begin{pmatrix} x \\ y \\ z \end{pmatrix} \\
\Leftrightarrow \quad &\boldsymbol{\phi} = \frac{\arccos\left(\cos\left(\theta_2\right)\cos\left(\theta_1 + \theta_3\right)\right)}{\sqrt{1 - \cos^2\left(\theta_2\right)\cos^2\left(\theta_1 + \theta_3\right)}}\begin{pmatrix} \sin\left(\theta_2\right)\sin\left(\theta_1 - \theta_3\right) \\ \sin\left(\theta_2\right)\cos\left(\theta_1 - \theta_3\right) \\ \cos\left(\theta_2\right)\sin\left(\theta_1 + \theta_3\right) \end{pmatrix} = \mathbf{f}\left(\boldsymbol{\theta}\right)
\end{aligned}
$$

B Calculation of the gradient

Continuing the calculation, we want to find the multivariable derivative of the inverse function. Luckily we can calculate via it the inverse function theorem:

$$\frac{\partial\left(f_1^{-1}, f_2^{-1}, f_3^{-1}\right)}{\partial\left(\phi_1, \phi_2, \phi_3\right)}(\boldsymbol{\phi}) = \left(\frac{\partial\left(f_1, f_2, f_3\right)}{\partial\left(\theta_1, \theta_2, \theta_3\right)}(\boldsymbol{\theta})\right)^{-1} \quad \Big|\, \boldsymbol{\phi} = \mathbf{f}(\boldsymbol{\theta})$$

$$= \begin{pmatrix} \frac{\partial f_1}{\partial\theta_1} & \frac{\partial f_1}{\partial\theta_2} & \frac{\partial f_1}{\partial\theta_3} \\ \frac{\partial f_2}{\partial\theta_1} & \frac{\partial f_2}{\partial\theta_2} & \frac{\partial f_2}{\partial\theta_3} \\ \frac{\partial f_3}{\partial\theta_1} & \frac{\partial f_3}{\partial\theta_2} & \frac{\partial f_3}{\partial\theta_3} \end{pmatrix}^{-1} = \begin{pmatrix} \partial_{\theta_1} f_1 & \partial_{\theta_2} f_1 & \partial_{\theta_3} f_1 \\ \partial_{\theta_1} f_2 & \partial_{\theta_2} f_2 & \partial_{\theta_3} f_2 \\ \partial_{\theta_1} f_3 & \partial_{\theta_2} f_3 & \partial_{\theta_3} f_3 \end{pmatrix}^{-1} = \maltese$$

We now use the formula:

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

$$\Rightarrow \quad A^{-1} = \frac{1}{a(ei - fh) - b(di - fg) + c(dh - eg)} \begin{pmatrix} ei - fh & -(bi - ch) & bf - ce \\ -(di - fg) & ai - cg & -(af - cd) \\ dh - eg & -(ah - bg) & ae - bd \end{pmatrix}$$

So that we get

$$\maltese = \frac{1}{\partial_{\theta_1} f_1(\partial_{\theta_2} f_2 \partial_{\theta_3} f_3 - \partial_{\theta_3} f_2 \partial_{\theta_2} f_3) - \partial_{\theta_2} f_1(\partial_{\theta_1} f_2 \partial_{\theta_3} f_3 - \partial_{\theta_3} f_2 \partial_{\theta_1} f_3) + \partial_{\theta_3} f_1(\partial_{\theta_1} f_2 \partial_{\theta_2} f_3 - \partial_{\theta_2} f_2 \partial_{\theta_1} f_3)}$$
$$\times \begin{pmatrix} \partial_{\theta_2} f_2 \partial_{\theta_3} f_3 - \partial_{\theta_3} f_2 \partial_{\theta_2} f_3 & -(\partial_{\theta_2} f_1 \partial_{\theta_3} f_3 - \partial_{\theta_3} f_1 \partial_{\theta_2} f_3) & \partial_{\theta_2} f_1 \partial_{\theta_3} f_2 - \partial_{\theta_3} f_1 \partial_{\theta_2} f_2 \\ -(\partial_{\theta_1} f_2 \partial_{\theta_3} f_3 - \partial_{\theta_3} f_2 \partial_{\theta_1} f_3) & \partial_{\theta_1} f_1 \partial_{\theta_3} f_3 - \partial_{\theta_3} f_1 \partial_{\theta_1} f_3 & -(\partial_{\theta_1} f_1 \partial_{\theta_3} f_2 - \partial_{\theta_3} f_1 \partial_{\theta_1} f_2) \\ \partial_{\theta_1} f_2 \partial_{\theta_2} f_3 - \partial_{\theta_2} f_2 \partial_{\theta_1} f_3 & -(\partial_{\theta_1} f_1 \partial_{\theta_2} f_3 - \partial_{\theta_2} f_1 \partial_{\theta_1} f_3) & \partial_{\theta_1} f_1 \partial_{\theta_2} f_2 - \partial_{\theta_2} f_1 \partial_{\theta_1} f_2 \end{pmatrix}$$

It looks even more horrifying when I type it into mathematica:

$$\frac{\partial\boldsymbol{\theta}}{\partial\phi_1} = \Bigg(\frac{\csc(2\theta2)\left(\sin^2(\theta2)\sin(\theta1+\theta3)\left(\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3)) - \cos(\theta2)\cos(\theta1+\theta3)\sqrt{1-\cos^2(\theta2)\cos^2(\theta1+\theta3)}\right)\left(\cos(\theta2)(\sin(2\theta1)+\sin(2\theta3))\sqrt{}\right.\right.}{\csc(2\theta2)\left(1-\cos^2(\theta2)\cos^2(\theta1+\theta3)\right)\left(\cos(\theta2)\sqrt{1-\cos^2(\theta2)\cos^2(\theta1+\theta3)}\left(\frac{1}{8}\sin^2(2\theta2)\sin(2(\theta1+\theta3))\sin(3\theta1+\theta3)+\frac{1}{8}\sin^2(2\theta2)\sin(2(\theta1+\theta3\right.\right.}$$

$$\frac{\partial\boldsymbol{\theta}}{\partial\phi_2} = \Bigg(-\frac{256\tan(\theta2)\left(1-\cos^2(\theta2)\cos^2(\theta1+\theta3)\right)^{3/2}\left(\sin(\theta1+\theta3)\cos(\theta1-\theta3)\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3))\sqrt{1-\cos^2(\theta2)\cos^2(\theta1+\theta3)}+\cos(\theta2)\left(-\sin(\right.\right.}{\cos^{-1}(\cos(\theta2)\cos(\theta1+}$$
$$-\frac{512\cos(\theta1-\theta3)\left(\cos^2(\theta2)\cos^2(\theta1+\theta3)-1\right)^2}{\cos^{-1}(\cos(\theta2)\cos(\theta1+}$$
$$-\frac{256\tan(\theta2)\left(1-\cos^2(\theta2)\cos^2(\theta1+\theta3)\right)^{3/2}\left(\sin(\theta1+\theta3)\cos(\theta1-\theta3)\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3))\sqrt{1-\cos^2(\theta2)\cos^2(\theta1+\theta3)}+\cos(\theta2)\left(\sin(\right.\right.}{\cos^{-1}(\cos(\theta2)\cos(\theta1+}$$

$$\frac{\partial\boldsymbol{\theta}}{\partial\phi_3} = \Bigg(-\frac{4\left(\sin^2(\theta1+\theta3)\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3))+\sin(\theta2)\tan(\theta2)\cos(\theta1+\theta3)\sqrt{1-\cos^2(\theta2)\cos^2(\theta1+\theta3)}\right)}{\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3))(\cos(2(\theta1-\theta2+\theta3))+\cos(2(\theta1+\theta2+\theta3))+2\cos(2(\theta1+\theta3))+2\cos(2\theta2)-6)}$$
$$-\frac{128\sin^2(\theta2)\cos(\theta2)\csc(2\theta2)\sin(\theta1+\theta3)\left(\cos^2(\theta2)\cos^2(\theta1+\theta3)-1\right)\left(\cos(\theta2)\cos(\theta1+\theta3)\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3))-\sqrt{1-\cos^2(\theta2)\cos^2(\theta\right.}{\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3))(\cos(2(\theta1-\theta2+\theta3))+\cos(2(\theta1+\theta2+\theta3))+2\cos(2(\theta1+\theta3))+2\cos(2\theta2)-6)^2}$$
$$-\frac{4\left(\sin^2(\theta1+\theta3)\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3))+\sin(\theta2)\tan(\theta2)\cos(\theta1+\theta3)\sqrt{1-\cos^2(\theta2)\cos^2(\theta1+\theta3)}\right)}{\cos^{-1}(\cos(\theta2)\cos(\theta1+\theta3))(\cos(2(\theta1-\theta2+\theta3))+\cos(2(\theta1+\theta2+\theta3))+2\cos(2(\theta1+\theta3))+2\cos(2\theta2)-6)}$$

The full expressions can be seen in the mathematica notebook. Suprisingly there are some symmetries:

$$\frac{\partial\theta_1}{\partial\phi_2} = \frac{\partial\theta_3}{\partial\phi_2} \quad \frac{\partial\theta_1}{\partial\phi_3} = \frac{\partial\theta_3}{\partial\phi_3}$$

With this expressions the calculation of the gradient is done.

References

[1] Andrea Mari, Thomas R. Bromley, and Nathan Killoran. Estimating the gradient and higher-order derivatives on quantum hardware. *Phys. Rev. A*, 103:012405, Jan 2021.

[2] David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. General parameter-shift rules for quantum gradients. *Quantum*, 6:677, March 2022.

[3] Roeland Wiersema, Dylan Lewis, David Wierichs, Juan Carrasquilla, and Nathan Killoran. Here comes the SU(N): multivariate quantum gates and gradients. *Quantum*, 8:1275, March 2024.