Exact derivatives in Quantum Circuits

Adrián Pérez Salinas[1, 2]

[1] *Barcelona Supercomputing Center, Barcelona, Spain*
[2] *Universitat de Barcelona, Barcelona, Spain*

In this brief document I will review a method to compute exact derivatives in a given parameterized quantum circuit by measuring observables for the circuit itself whose parameters were shifted. This method allows for a slightly more efficient calculation of gradients and second-order derivatives. In addition, the method is much more resilient to statistical noise than numerical derivatives. Both features provide an advantage when dealing with classical optimization procedures for quantum circuits.

I. INTRODUCTION

Many classical optimization methods make use of gradients to drive the quest for the parameters that minimize some cost function. Gradient-based recipes are nowadays broadly used in classical computation due to its success and its efficiency. A prominent example is the Stochastic Gradient Descent (SGD) algorithm that underlies at the core of most Artificial Intelligence implementations, in particular for Neural Networks. Although there are alternatives to gradient-based optimizers, these have proved their accuracy and speed for solving interesting problems.

Derivatives are classically computed usually in two different ways, namely an exact and a numerical calculation. Exact computations are useful when the derivative of interest has an analytical form that can be easily introduced in a computer. For instance, one of the main strengths of SGD is due to the existence of these formula for many Neural Networks. This manner is fast and exact, but not always feasible as it depends on the function we are interested in to minimize. On the other hand, the chance to compute numerical derivatives is always present. A standard procedure to do so is known as the finite differences method. In this method, we compute the derivative in a given point by evaluating the function in some close points and combining the results, *e. g.*: $\partial_x f(x) \sim (2\varepsilon)^{-1}(f(x+\varepsilon) - f(x-\varepsilon))$. This way of computing a derivative is approximate and sensitive to the behaviour of the function around the point of interest.

In these notes we explore exact and approximate methods to compute derivatives of a certain observable in a parameterized quantum circuit. We show that numerical derivatives may suffer large errors due to the inherent statistical noise of quantum measurements. We explain also a method to compute exact derivatives in the case where the quantum circuit is constructed out of single qubit gates. This method is further deployed in Ref. [1]

II. NUMERICAL DERIVATIVES FOR QUANTUM COMPUTERS

Let us consider the standard finite differences method to estimate derivatives

$$\frac{df(x)}{dx} \sim \frac{f(x+h) - f(x-h)}{2h}, \tag{1}$$

where the approximation error is

$$\left| \frac{df(x)}{dx} - \frac{f(x+h) - f(x-h)}{2h} \right| \approx \mathcal{O}(f^{(3)}(c)h^2), \tag{2}$$

where $f^{(3)}(c)$ is the third derivative of f evaluated at some point $x - h \leq c \leq x + h$. There are higher order finite differences recipes that are useful in the cases where the absolute values of the derivative are small, but we restrict the discusion to this method since the arguments hold in general.

We now extend this method to quantum circuits. The role of $f(x)$ is now played by the measurement of some observable M. This observable is defined by the formula

$$M(\theta) = \langle 0 | U^\dagger(\theta) M U^\dagger(\theta) | 0 \rangle , \tag{3}$$

and the approximated value of $\langle M(\theta) \rangle$ provided by some measurement performed with N shots is

$$\tilde{M}(\theta) = M(\theta) \pm \mathcal{O}(\frac{1}{\sqrt{N}}). \tag{4}$$

Having these two errors in mind it is straightforward to see that

$$\left|\frac{dM(\theta)}{d\theta} - \frac{d\tilde{M}(\theta)}{d\theta}\right| = \mathcal{O}(h^2) + \mathcal{O}(N^{-1/2}h^{-1}). \tag{5}$$

This error analysis reveals that it is no longer useful to decrease the value of h to small quantities, as in the classical case, since now a larger source of error emerges from the noise of quantum measurement. We cannot get rid of this error since it is inherent to the quantum mechanical nature of quantum devices. Just to get a feeling on how much error happens using numerical approximations, we see that the value of h at which the minimum error is achieved for $N = 10^5$ shots is $h \sim 0.15$, which is not enough to precisely approximate derivatives when their values approach zero. The obvious way to decrease the amount of error is to average the observables over more shots, that is using larger Ns. However, this approach is costly and resource consuming.

Notice that this analysis does not take into account the error of the quantum circuit itself. We have assumed perfect circuits where the only source of uncertainty comes from the sampling noise. This scenario is an optimistic one since current devices cannot perform any computation with a perfect fidelity. However, addressing this problem is not a cornerstone to explain the methods proposed in this notes. Therefore we leave this issue for future work.

III. ANALYTICAL DERIVATIVES FOR QUANTUM COMPUTERS

We are now convinced about the unfeasibility of numerical derivatives. We deal now with analytical derivatives to be applied in quantum circuits. From Ref. [1] we can extract the following result. Let M be an observable of some quantum circuit depending on a set of parameters θ. Then, the gradient can be computed by means of the formula

$$\partial_i M = \frac{\partial M(\theta)}{d\theta} = r\left(M(\theta + se_i) - M(\theta - se_i)\right), \tag{6}$$

where r and s are real numbers and e_i is a vector with as many components as θ where all the elements are 0 except for the i-th element, which is 1. The values of s and r are determined by the operation affected by θ_i. In particular for standard single-qubit rotation gates $r = 1/2$ and $s = \pi/2$.

This allows to compute the gradient for all parameters θ defining the computation in a straightforward manner. The number of function evaluations to be performed for obtaining the gradient is twice the number of single-qubit rotation gates around the axis X, Y, Z. That is, a R_z gate will need 2 evaluations, while a $U_3 = R_z R_y R_z$ gate will need 6 of them.

In order to get a fair comparison, we analyze the error commited using this analytical method. On the one hand, there is no approximation error since the output of this function is the exact derivative. On the other hand, we cannot get rid of the sampling noise, and thus

$$\left|\partial_i M - \partial_i \tilde{M}\right| = \mathcal{O}(N^{-1/2}). \tag{7}$$

Notice that this method removes the problem of the step size h since it is not needed anymore.

A simple example

Let us review now a simple example. We have the state

$$|\psi\rangle = R_z(\beta)R_y(\alpha)\,|0\rangle\,, \tag{8}$$

and we want to study the observable

$$Z = \langle\psi|\, Z\,|\psi\rangle \tag{9}$$

It is inmediate to compute the gradient of Z as

$$\nabla Z = \begin{pmatrix}\partial_\alpha Z\\ \partial_\beta Z\end{pmatrix} = \\ \frac{1}{2}\begin{pmatrix}\langle 0|\, R_y(-\alpha-\pi/2)R_z(-\beta)ZR_z(\beta)R_y(\alpha+\pi/2)\,|0\rangle - \langle 0|\, R_y(-\alpha+\pi/2)R_z(-\beta)ZR_z(\beta)R_y(\alpha-\pi/2)\,|0\rangle\\ \langle 0|\, R_y(-\alpha)R_z(-\beta-\pi/2)ZR_z(\beta+\pi/2)R_y(\alpha)\,|0\rangle - \langle 0|\, R_y(-\alpha)R_z(-\beta+\pi/2)ZR_z(\beta-\pi/2)R_y(\alpha)\,|0\rangle\end{pmatrix} \tag{10}$$

Using this quantity we can recall a classical optimizer to look for optimal values (α, β) minimizing the measured observable.

IV. ANALYTICAL GRADIENTS FOR RE-UPLOADING

Let us work now with analytical gradients applied to algorithms in which the re-uploading principle is present. This re-uploading approach consists of introducing classical data as a parameter for a quantum circuit as in

$$\theta_i \leftarrow w_i x + b_i, \tag{11}$$

what allows for a great flexibility for quantum circuits. Let us assume that the problem we want to solve is to create a circuit depending on x and some parameters θ such that the observable M encodes a function as

$$M(x,\theta) = \langle 0| U^\dagger(x,\theta) M U(x,\theta) |0\rangle \approx f(x), \tag{12}$$

for a given domain $\mathcal{D}$ of x. We can drive an optimization to find a proper set of θ by defining the cost function

$$\chi^2 = \frac{1}{2\mathcal{D}} \int_{\mathcal{D}} dx \left(M(x,\theta) - f(x)\right)^2. \tag{13}$$

The gradient of χ^2 can be easily obtained with previous formulae and the chain rule. We are interested in

$$\frac{\partial \chi^2}{\partial w_i}; \qquad \frac{\partial \chi^2}{\partial b_i}. \tag{14}$$

The chain rule ensures that

$$\frac{\partial \chi^2}{\partial w_i} = \frac{\partial \chi^2}{\partial \theta_i} x \tag{15}$$

$$\frac{\partial \chi^2}{\partial b_i} = \frac{\partial \chi^2}{\partial \theta_i}. \tag{16}$$

On the other hand,

$$\frac{\partial \chi^2}{\partial \theta_i} = \frac{1}{\mathcal{D}} \int_{\mathcal{D}} dx \left(M(x,\theta) - f(x)\right) \partial_i M(x,\theta), \tag{17}$$

with $\partial_i M(x,\theta)$ the previously defined derivative. Combining both results it is clear that we can obtain the gradient of the quantum circuit with respect to both $\{w\}$ and $\{b\}$ sets of parameters by just applying the exact derivative method to the $\{\theta\}$ parameters.

$$\frac{\partial \chi^2}{\partial w_i} = \frac{1}{\mathcal{D}} \int_{\mathcal{D}} dx \left(M(x,\theta) - f(x)\right) \partial_i M(x,\theta) x \tag{18}$$

$$\frac{\partial \chi^2}{\partial b_i} = \frac{1}{\mathcal{D}} \int_{\mathcal{D}} dx \left(M(x,\theta) - f(x)\right) \partial_i M(x,\theta). \tag{19}$$

Since $\{w\}$ and $\{b\}$ outnumber $\{\theta\}$, this method is more efficient since it allows to re-use the same quantum function evaluation for different parameters.

V. DISCUSSION

The method explained hereby allows to obtain exact gradients for quantum circuits. This method is resilient against sampling noise in contradistinction to the numerical approaches. In runtimes efficiencies are similar since it is necessary to perform two function evaluations per parameter.

In the re-uploading case, there exists a manner to exploit the inner structure of the parameters to obtain a gradient for all components without performing as many derivative calculations. This method takes advantage of the chain rule. The gain in runtime is not substantially different, but rather a prefactor.

---

[1] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3), Mar 2019.