Evaluation of derivatives using approximate generalized parameter shift rule

Vytautas Abramavicius, Evan Philip, Kaonan Micadei, Mario Dagrada, Vincent Elfving and Panagiotis Barkoutsos

Pasqal

Derivative calculations for quantum circuit outputs with respect to variable circuit parameters (input or variational parameters) is essential to implement a breadth of quantum algorithms, especially in the field of Quantum Machine Learning (QML). Notable examples include algorithms like Differentiable Quantum Circuit (DQC) [1] and Quantum Extremal Learning (QEL) [2]. Analytical or automatic differentiation (AD) methods are usually favored over numerical differentiation methods, *e.g.* based on finite differences, even in classical machine learning due to their robustness against errors. Quantum computation exhibits fundamental shot noise, therefore numerical differentiation is error prone and unfeasible on noisy intermediate-scale quantum (NISQ) hardware with limited shot budget. On quantum hardware where qubit interactions can be turned off (superconducting or ion-trapped), control parameters are rotation angles in single-qubit gates. In these settings, parameter shift rule (PSR) [3, 4] has become the *de facto* method for calculating derivatives. It has also been experimentally demonstrated to be more robust against noise [5].

Usual PSR is valid where parameters control gates that can be described by involutary and idempotent generators having spectrum consisting of values $\pm\lambda$ after removing possible duplicates. However, not all parametrized quantum circuits can be represented by such generators. Notably, on neutral atom platforms such as Pasqal's, any non-trivial configuration for a parameterized quantum circuit exhibits a background Rydberg interaction between atoms that cannot be turned off. In such cases, generators governing the dynamics of the quantum circuit are neither involutary nor idempotent, thus one has to rely on the generalized parameter shift rule (GPSR) [6] to calculate derivatives.

According to [6], for a given parameter x, one can define a function $f(x)$ as quantum expectation of a Hermitian cost operator $\hat{C}$

$$f\left(x\right) = \left\langle \psi \right| \hat{U}^{\dagger}(x) \hat{C} \hat{U}(x) \left| \psi \right\rangle , \tag{1}$$

where $|\psi\rangle$ is some initial state and $\hat{U}(x) = \exp\left(-\frac{i}{2} x \hat{G}\right)$ is a unitary evolution operator governed by the Hermitian generator $\hat{G}$. The derivative of $f(x)$ is given by

$$\frac{\mathrm{d}f\left(x\right)}{\mathrm{d}x} = \sum_{s=1}^{S} \Delta_s R_s, \tag{2}$$

where S is the number of unique non-zero spectral gaps $\{\Delta_s\}_{s=1}^{S}$ of generator $\hat{G}$. Parameters R_s necessary for derivative evaluation can be obtained by solving the following system of linear equations:

$$\begin{cases} F_1 &= 4 \sum\limits_{s=1}^{S} \sin\left(\frac{\delta_1 \Delta_s}{2}\right) R_s, \\ F_2 &= 4 \sum\limits_{s=1}^{S} \sin\left(\frac{\delta_2 \Delta_s}{2}\right) R_s, \\ & \ldots \\ F_S &= 4 \sum\limits_{s=1}^{S} \sin\left(\frac{\delta_S \Delta_s}{2}\right) R_s. \end{cases} \tag{3}$$

Here, we denoted $F_s = f(x + \delta_s) - f(x - \delta_s)$ as the difference between values of function $f(x)$ with argument shifted by $\pm\delta_s$.

The derivative expression in Eq. (2) is free from any approximations, hence mathematically exact, and referred to as the GPSR. It provides a method to calculate the exact derivative with respect to parameters represented by an arbitrary generator with a rich spectrum of eigenvalues. Moreover, it remains robust against noise by tailoring the parameter shift values δ_s to minimize the variance of the derivative estimation. Since the generator of a N-qubit quantum circuit has 2^N eigenvalues and (at most) $S = \frac{2^N(2^N-1)}{2}$ unique spectral gaps, GPSR requires making measurements at $2S$ different values $x \pm \delta_s$ of the parameter to estimate the derivative at any point x. For instance, derivative calculations of a $N = 5$ qubit circuit on neutral atom hardware cannot be performed with ordinary PSR, and its generalized version has to applied. However, even with a 1000-shot budget, merely a single estimate at each of the $2S = 992$ shifted points $x \pm \delta_s$ demanded by GPSR will be possible, rendering it practically useless for the calculation of the derivative.

To solve this exponential scaling problem, we propose aGPSR (approximate GPSR) as a method of estimating derivative of a function spawned by an arbitrary generator having a non-trivial spectrum of eigenvalues in a

limited shot budget setting. The foundation of this method is the observation that only a small fraction of spectral gaps is sufficient to accurately estimate the value of derivative in Eq. (2). Expressions for aGPSR are similar to Eqs. (2) and (3) with the only difference that the effective set of used spectral gaps is an arbitrary collection $\{\gamma_k\}_{k=1}^K$, instead of the actual gap spectrum $\{\Delta_s\}_{s=1}^S$ of generator $\hat{G}$. This leads to a truncated system in Eq. (3) containing only $K \ll S$ equations. The final expression of the derivative in this case, thus reads

$$\left.\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|_K = \sum_{k=1}^{K} \gamma_k R(\{\gamma_k\}_{k=1}^K, \{\delta_k\}_{k=1}^K)_k = \sum_{s=1}^{S} (\Delta_s + Q_K\left(\Delta_s, \{\gamma_k\}_{k=1}^K, \{\delta_k\}_{k=1}^K\right)) R_s. \tag{4}$$

where $R(\{\gamma_k\}_{k=1}^K, \{\delta_k\}_{k=1}^K)_k$ are values obtained by solving the truncated linear system in Eq. (3). δ_k and γ_k are shifts and gap values that can be selected with different strategies for optimal performance. This allows aGPSR to accurately estimate derivatives even in cases where the full generator spectrum is unknown, since γ_k can be selected independently. One of the most obvious yet robust selection strategies is to draw distinct γ_k values from $k \in (0, \cdots, K]$. We observed that as long as the majority of actual spectral gaps Δ_s fall in the range $[0, K]$, the accuracy of derivative estimation is very high.

The second equality in Eq. (4) shows that aGPSR derivative estimates can be written in a form closely resembling the original exact derivative expression in Eq. (2) but contains deviations Q_K from the original spectral gap values Δ_s. These deviations become zero in the limit of $K = S$, when aGPSR turns into full mathematically exact GPSR. Thus, accuracy of the approximation used in aGPSR can be assessed by investigating the behavior of the deviation function $Q_K\left(\Delta, \{\gamma_k\}_{k=1}^K, \{\delta_k\}_{k=1}^K\right)$ when Δ is varied, as depicted in Fig. 1. Here, we can see that the larger K values, the smaller the deviation function for a wider range of spectral gaps Δ. This confirms the observation that by increasing K, we can accurately estimate a larger portion of spectral gaps of the generator, thus obtaining a more accurate estimate for the derivative. For instance, if the spectral gaps of an arbitrary generator are mainly distributed in the range $[0, 5]$, they are approximated with a relative error $\frac{Q_K}{\Delta} \sim 10^{-2}$ for $K = 4$. The derivative estimation will already be very accurate in this case.

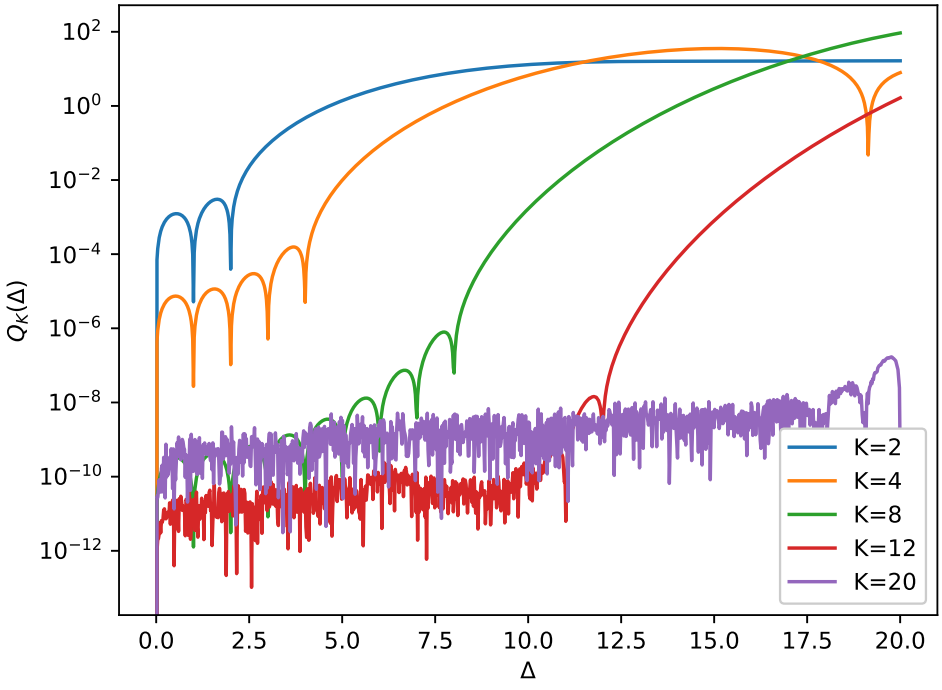

Figure 1: Spectral gap error function Q_K dependence on spectral gap value Δ for different values of K.

Accuracy of aGPSR and its suitability for calculating derivatives on quantum systems exhibiting arbitrary qubit interactions can be illustrated by considering the neutral-atom generator:

$$\hat{G} = \sum_{i=1}^{N} \left(\hat{\sigma}_i^x + \frac{2}{\Omega} \sum_{j<i}^{N} J_{ij} \hat{n}_i \hat{n}_j \right). \tag{5}$$

Here, Ω is the drive amplitude of the laser pulse, J_{ij} is the interaction strength between atoms i and j and

$\hat{n_i} = \frac{\hat{\sigma}_i^z + \hat{I}_i}{2}$ the number operator. The differentiation parameter in this case is $x = \Omega t$. In Fig. 2 we can see the derivative calculated for a 6-qubit system with 2×3 qubit layout and parameters of generator $\hat{G}$ such that Ω is significantly smaller than the characteristic interaction strength J. In this case, the dynamics of the system are mainly determined by the always-on interaction between the qubits whose generator cannot be represented as single-qubit operations. Ordinary PSR will definitely fail to accurately estimate the derivative in such system as confirmed in Fig. 2.

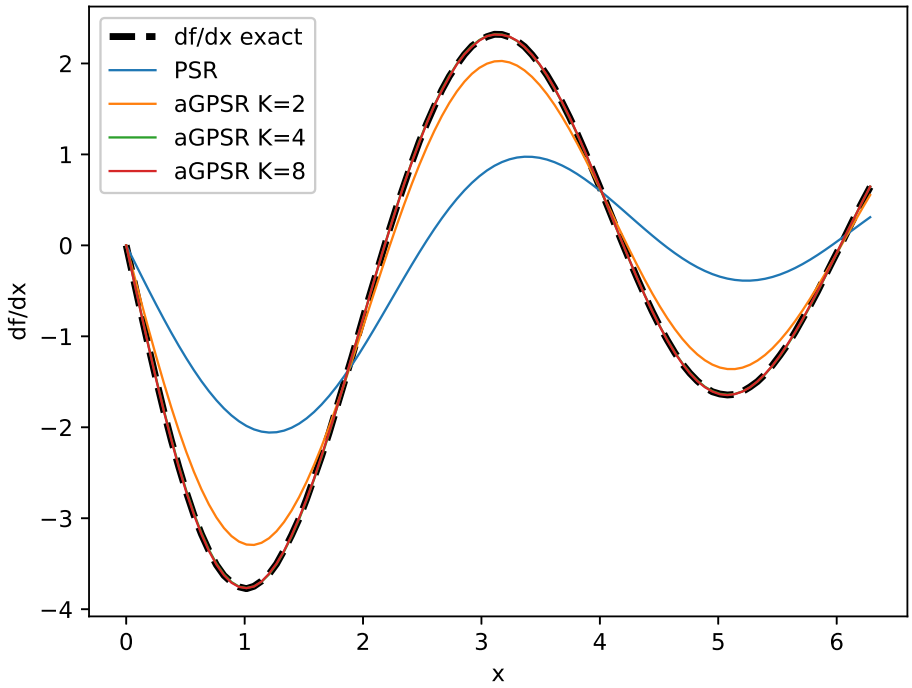

Figure 2: Comparison of derivatives $\frac{\mathrm{d}f}{\mathrm{d}x}$ calculated for different K values. The ratio of the drive amplitude with the interaction strength is $\frac{J}{\Omega} \gg 1$.

Here, we can observe that PSR with a single gap to calculate the derivative is quite at variance with respect to the exact result. On the other hand, it is evident that aGPSR method using at least $K = 4$ terms already coincides perfectly with the exact derivative curve. To emphasize the considerable benefit of aGPSR in computation requirements, we must note that full GPSR for such a 6-qubit system would require solving a large linear system in Eq. (3) with $S = 1952$ equations instead of $K = 4$.

References

[1] Oleksandr Kyriienko, Annie E Paine, and Vincent E Elfving. "Solving nonlinear differential equations with differentiable quantum circuits". In: *Physical Review A* 103.5 (2021), p. 052416.

[2] Savvas Varsamopoulos et al. *Quantum Extremal Learning*. 2022. arXiv: `2205.02807 [quant-ph]`.

[3] K. Mitarai et al. "Quantum circuit learning". In: *Phys. Rev. A* 98 (3 2018), p. 032309.

[4] Maria Schuld et al. "Evaluating analytic gradients on quantum hardware". In: *Phys. Rev. A* 99 (3 2019), p. 032331.

[5] Yuki Ishiyama et al. "Noise-robust optimization of quantum machine learning models for polymer properties using a simulator and validated on the IonQ quantum computer". In: *Scientific Reports* 12.1 (2022), p. 19003.

[6] Oleksandr Kyriienko and Vincent E Elfving. "Generalized quantum circuit differentiation rules". In: *Physical Review A* 104.5 (2021), p. 052417.