An Algorithmic Framework for Local Approximations of Parametrized Quantum Circuits via RKHS Methods

Lars Simon[1], Holger Eble[2], Hagen-Henrik Kowalski[3], and Manuel Radons[4]

[1,2,3,4]Bundesdruckerei GmbH, firstname.lastname@bdr.de

We describe a reproducing kernel Hilbert space (RKHS) containing the set of functions computed by variational quantum algorithms (VQAs) and develop an algorithmic framework which exploits the RKHS structure to construct local approximations in parameter space. After outlining the general method, we describe three applications.

Method

We consider functions of the form $f\colon \mathbb{R}^m \to \mathbb{R}$, $\theta \mapsto \langle\psi(\theta)|\mathcal{M}|\psi(\theta)\rangle$, where $\mathcal{M} \in \mathbb{C}^{2^n\times 2^n}$ is an observable and $|\psi(\theta)\rangle = C_{m+1}R_m(\theta_m)C_m\cdots R_1(\theta_1)C_1|0\rangle^{\otimes n}$, where $C_1,\dots,C_{m+1}$ are n-qubit unitaries and each $R_j(\theta_j) = \exp\left(-i\frac{\theta_j}{2}G_j\right)$ is a rotation, where $G_j \in \mathbb{C}^{2^n\times 2^n}$ is Hermitian with set of Eigenvalues $\{-1,1\}$ (e.g., G_j could be a non-identity Pauli string).

The function f is contained in H, where H is the set of all functions $g\colon \mathbb{R}^m \to \mathbb{R}$ of the form $g(z) = \sum_{\omega\in\{-1,0,1\}^m} c_\omega e^{i\omega^\intercal z}$ with $\overline{c_{-\omega}} = c_\omega \in \mathbb{C}$ for all ω.

RKHS Structure When equipping H with the (real) inner product given by

$$\langle g_1, g_2\rangle_H = \int_{[-\pi,\pi]^m} g_1(z)g_2(z)\mathrm{d}z,\ g_1, g_2 \in H,$$

it turns out that H carries the structure of an RKHS with kernel $K = \left(\frac{3}{2\pi}\right)^m \tilde{K}$, where

$$\tilde{K}(x,z) = \prod_{j=1}^m \frac{1+2\cos(x_j - z_j)}{3},\ x, z \in \mathbb{R}^m.$$

Approximations of f If $p_1,\dots,p_D \in \mathbb{R}^m$ are points in parameter space, then the linear system of equations

$$\left(\tilde{K}(p_i,p_j)\right)_{1\le i,j\le D}\cdot\tilde{\eta} = (f(p_1),\dots,f(p_D))^\intercal$$

has at least one solution $\tilde{\eta}\in\mathbb{R}^D$. We then define a classical approximation $\tilde{f}\colon \mathbb{R}^m\to\mathbb{R}$ to f, given by

$$\tilde{f}(\theta) = \sum_{j=1}^D \tilde{\eta}_j\tilde{K}(p_j,\theta).$$

This approximation is natural in the sense that $\tilde{f}$ coincides with f on $\{p_1,\dots,p_D\}$ and $\tilde{f}$ is the orthogonal projection (wrt. $\langle\cdot,\cdot\rangle_H$) of f onto the subspace of H spanned by $K(p_1,\cdot),\dots,K(p_D,\cdot)$.

Applications

Classical Surrogates [1] Given a point $\theta_0\in\mathbb{R}^m$ in parameter space, we build a local approximation $\tilde{f}$ of f around θ_0 by choosing $p_1 = \theta_0 + q_1,\dots,p_D = \theta_0 + q_D$ to be the elements of the set $\theta_0 + \left\{-\frac{\pi}{2},0,\frac{\pi}{2}\right\}^m$ with the property that at most L entries of the q_j are nonzero, where $0\le L\le m$. Here, L is a hyperparameter controlling the order of the approximation. Computing this approximation does not require more circuit evaluations than computing the L-th order Taylor polynomial using the parameter-shift rules. We give experimental and theoretical evidence that our approximation outperforms the L-th order Taylor polynomial.

Denoising Gradient Descent [2] When carrying out a gradient descent step using the parameter-shift rules at a point $\theta\in\mathbb{R}^m$ in parameter space, a quantum device is used to evaluate f at the $2m$ coordinate-wise shifts by $\pm\frac{\pi}{2}$ at θ. To counter the resulting noise, we carry out a gradient descent step wrt. a local approximation $\tilde{f}$ (instead of wrt. f) which is constructed to mitigate effects of the former, by also using the $\pm\frac{\pi}{2}$-shifts from the previous $\ell-1$ iterations in combination with regularization of the kernel matrix. Here, $\ell\ge 1$ is a hyperparameter. We give experimental evidence that our algorithm outperforms gradient descent in the presence of both shot noise and (simulated) quantum hardware noise, without requiring more circuit evaluations.

Optimization Problem for VQAs [3] When aiming to minimize f, a common technique is to compute a local approximation around a given point $\theta_0\in\mathbb{R}^m$ in parameter space and subsequently perform one (or several) optimization step(s) with respect to this approximation. Examples of this are gradient descent via the parameter-shift rules (linear approximation) and quantum analytic descent (trigonometric approximation coinciding with f up to second order). We use this technique with the L-th order approximation $\tilde{f}$ around θ_0 described above. However, $\pm\frac{\pi}{2}$-shifts are replaced by $\pm\frac{2\pi}{3}$-shifts, since this makes the kernel matrix trivial to invert and hence significantly reduces the classical computational overhead. We compare this to gradient descent (for $L=1$) and to quantum analytic descent (for $L=2$).

[1] Lars Simon, Holger Eble, Hagen-Henrik Kowalski, and Manuel Radons, "Interpolating Parametrized Quantum Circuits using Blackbox Queries," *SIAM Journal on Scientific Computing*, To appear, 2024, arXiv:2310.04396.

[2] Lars Simon, Holger Eble, Hagen-Henrik Kowalski, and Manuel Radons, "Denoising Gradient Descent in Variational Quantum Algorithms," 2024, arXiv:2403.03826.

[3] In preparation.