Solutions Problem Sheet 8: Shot noise and convergence inequalities

Class problems

1. The parameter shift rule can be used to compute gradients for parameterized quantum circuits. Specifically, given a cost of the form $C(\boldsymbol{\theta}) := \mathrm{Tr}[U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger M]$ the general parameter shift rule can be stated as

$$\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k} = \frac{1}{2\sin(\alpha)}\left(\mathrm{Tr}[U(\boldsymbol{\theta}_+)\rho U(\boldsymbol{\theta}_+)^\dagger M] - \mathrm{Tr}[U(\boldsymbol{\theta}_-)\rho U(\boldsymbol{\theta}_-)^\dagger M]\right) \tag{1}$$

where $\boldsymbol{\theta}_\pm = \boldsymbol{\theta} \pm \alpha \boldsymbol{e_k}$. Here $\boldsymbol{e_k}$ is a vector having 1 as its kth element and 0 otherwise. What value of α is best to use to minimize the effect of shot noise?

Answer: The variance of the gradient estimate is given by

$$\mathrm{Var}\left(\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k}\right) = \frac{\mathrm{Var}(C(\boldsymbol{\theta}_+)) + \mathrm{Var}(C(\boldsymbol{\theta}_-))}{4\sin^2(\alpha)} \, .$$

Indeed $\boldsymbol{\theta}_\pm$ depends on α, but when the variance $\mathrm{Var}(C(\boldsymbol{\theta}_\pm))$ has a negligible dependence on α the optimal value is $\alpha = \pi/2$ (i.e. value that minimises the gradient variance.) In practice, it is quite common to use $\alpha = \pi/2$. See https://link.aps.org/accepted/10.1103/PhysRevA.103.012405 for example (or other papers cited there for more details).

2. Quantum generative modelling is an active area of research in quantum machine learning. The aim is to use samples from a target distribution $p(\boldsymbol{x})$ to learn a quantum model $q(\boldsymbol{x})$ of $p(\boldsymbol{x})$ which can be used to generate new samples.

Suppose the quantum model probabilities are computed by preparing a quantum state $|\psi_{\boldsymbol{\theta}}\rangle$ and measuring it in the computational basis. That is $q_{\boldsymbol{\theta}}(\boldsymbol{x}) = |\langle \boldsymbol{x}|\psi_{\boldsymbol{\theta}}\rangle|^2$. In practise, this will be done using a finite number of shots.

A number of different loss functions can be used to evaluate the similarity between the target and model distribution.

Which of the following losses are unbiased with a finite number of shots?

KL divergence (KLD)

$$\mathcal{L}^{\mathrm{KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right) \, . \tag{2}$$

The classical fidelity,

$$\mathcal{L}^{\mathrm{CF}}(\boldsymbol{\theta}) = 1 - \sum_{\boldsymbol{x}\in\mathcal{X}} \sqrt{p(\boldsymbol{x}) q_{\boldsymbol{\theta}}(\boldsymbol{x})} \, . \tag{3}$$

The quantum fidelity,

$$\mathcal{L}^{\mathrm{QF}}(\boldsymbol{\theta}) = 1 - \left|\sum_{\boldsymbol{x}\in\mathcal{X}} \sqrt{p(\boldsymbol{x}) q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right|^2 \, . \tag{4}$$

The total variation distance,

$$\mathcal{L}^{\mathrm{TVD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} |p(\boldsymbol{x}) - q_{\boldsymbol{\theta}}(\boldsymbol{x})| \, . \tag{5}$$

The squared Euclidean distance,

$$\mathcal{L}^{\mathrm{SED}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} (p(\boldsymbol{x}) - q_{\boldsymbol{\theta}}(\boldsymbol{x}))^2 \, . \tag{6}$$

Answer: The quantum fidelity and the squared Euclidean distance are unbiased with a finite number of shots as they can be rewritten as expectations of some operators in some states. We define the state $|\phi\rangle$ such that $p(\boldsymbol{x}) = |\langle \boldsymbol{x}|\psi\rangle|^2$. The quantum fidelity loss is nothing but the expectation of operator $\mathbb{1} - |\phi\rangle\langle\phi|$ in state $|\psi_{\boldsymbol{\theta}}\rangle$. For the square Euclidean distance, we have to compute terms $\sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x})^2$, $\sum_{\boldsymbol{x}\in\mathcal{X}} q_{\boldsymbol{\theta}}(\boldsymbol{x})^2$ and $\sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x})q_{\boldsymbol{\theta}}(\boldsymbol{x})$ that can be written as expectations of operator $\sum_{\boldsymbol{x}\in\mathcal{X}} |\boldsymbol{x}\rangle\langle\boldsymbol{x}|\otimes|\boldsymbol{x}\rangle\langle\boldsymbol{x}|$ in states $|\phi\rangle\otimes|\phi\rangle$, $|\psi_{\boldsymbol{\theta}}\rangle\otimes|\psi_{\boldsymbol{\theta}}\rangle$ and $|\phi\rangle\otimes|\psi_{\boldsymbol{\theta}}\rangle$ respectively. So, the square Euclidean distance can be obtain by unbiased estimation of these 3 terms. Feel free to look at one of our paper on NPJ `https://www.nature.com/articles/s41534-024-00902-0#citeas`.

3. Suppose you want to compute the Hilbert-Schmidt norm $||\rho - \sigma||_2^2 = \mathrm{Tr}[\rho^2 + \sigma^2 - 2\rho\sigma]$ between two mixed states ρ and σ.

   a) Show that $\mathrm{Tr}[\rho\sigma]$ can be measured on a quantum computer via performing a SWAP measurement. (Assume both ρ and σ are given to you.)

   b) Show that $\mathrm{Tr}[\rho\sigma]$ can also be measured using a generalization of the Loschmidt echo test. (Assume ρ is given to you but you know a circuit to prepare σ - note there are different ways in which the circuit to prepare σ could be given to you).

   c) Which measurement method converges more efficiently?

   Answer: Everything is well explained in the *Quantum Mixed State Compiling* paper (`https://arxiv.org/abs/2209.00528`). You can consider the eigen-decomposition of both ρ and σ and compute all fidelity terms between the eigenstates of σ and the eigenstates of ρ (pure states). For n qubits states, there are 2^n eigenstates for both (if full rank) so in total there are 2^{2n} fidelity to compute. Then, the answer is the same as in the standard setting with pure states discussed in lecture, but the number of shots required is multiplied by 2^{2n} (which is bad).