Problem Sheet 8: Shot noise and convergence inequalities

All the questions in this week's problem sheets are based around real research questions that I've come across in the last couple of years.

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? (You shouldn't find an exact optimal value of α, but rather try to get an intuitive answer with the terms you are able to deal with.)

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,

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

3. (Assessed problem 2023) 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?

Assessed Problem

This question will compare the effect of shot noise for three approaches to estimating $\mathrm{Tr}[\rho H]$, i.e., the energy of a state ρ with Hamiltonian $H = \sum_{i=1}^{M} \alpha_i P_i$ where the P_i are Pauli operators.

1. First suppose you measure each Pauli term individually using N/M shots for each Pauli term. Write down an estimator for this measurement strategy. Show that the estimator is unbiased and compute its variance. Find an upper bound for the variance of the estimator in terms of a norm of the vector $\boldsymbol{\alpha}$. Carefully list any assumptions about N and M needed in this calculation.

   (5 marks)

2. Next suppose you instead use $N_i = p_i N$ shots for each Pauli term where $p_i = |\alpha_i| / \sum_i |\alpha_i|$. Why intuitively might this be a better strategy than the previous one?

   Write down an estimator for this measurement strategy. Show that the estimator is unbiased and compute its variance. Find an upper bound for the variance of the estimator in terms of a norm of the vector $\boldsymbol{\alpha}$. Carefully list any assumptions about N needed in this calculation.

   (5 marks)

3. Discuss which of these strategies is best for estimating $\mathrm{Tr}[\rho H]$ if only taking into account shot noise? Construct some example Hamiltonians to illustrate your arguments.

   (5 marks)

4. Finally, suppose you instead have a source of classical randomness and with probability $p_i = |\alpha_i| / \sum_i |\alpha_i|$ measure Pauli P_i.

   Again, write down an estimator for this measurement strategy. Show that the estimator is unbiased and compute its variance. And find an upper bound for the estimator in terms of a norm of the vector $\boldsymbol{\alpha}$.

   (5 marks)

   Discuss how this method compares to the others? When might you choose to use it?

   (2 marks)

5. Suggest two other methods you might use to estimate $\mathrm{Tr}[\rho H]$. What are the advantages / disadvantages of these methods? (3 marks)