# Exercise Sheet 3 for the 'Quantum Machine Learning' Course

TA: Subadra Echeverria,[1] TA: Felipe Choy,[1] and Teacher: M. Cerezo[2]

[1]*Escuela de Ciencias Físicas y Matemáticas, Universidad de San Carlos de Guatemala, Guatemala 01004, Guatemala*
[2]*Information Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

This third set of exercises is aimed at introducing useful quantum circuit results that will be used throughout the course.

## I. Useful Quantum Circuits

*Many quantum circuits are build from basic components in a Lego-like fashion, where simple components are combined to evaluate non-trivial and complex quantities.*

Let us fist consider the so-called *Hadamard test* circuit in Fig. 1(a). This circuit allows us to compute the expectation value of a unitary $U$ over a quantum state $|\psi\rangle$.

**Exercise 1** (Hadamard test). *Show that the probability of measuring the first qubit in the $|0\rangle$ state is*

$$p(0) = \frac{1}{2}(1 + Re[\langle\psi|U|\psi\rangle]) \,. \tag{1}$$

Note that Eq. (1) is valid for general $n$ qubit states $|\psi\rangle$ and for general unitaries $U$ acting on $n$-qubits.

In particular if $U = Z \otimes Z$, note that we can use the Hadamard test to measure $\langle Z \otimes Z\rangle$. From the Exercise Sheet 2 we know that given a two qubit state $|\psi\rangle$, we can also compute the expectation value of the $Z \otimes Z$ operator by measuring in the computational basis and then computing $\langle Z \otimes Z\rangle = p(00) + p(11) - p(10) - p(01)$. Moreover, given $N$ measurement shots, we know that the uncertainty in estimating each of those probabilities is $1/\sqrt{N}$.

**Exercise 2** (Error analysis). *Show that, given a budget of $N$ measurement shots, using the Hadamard test to compute $\langle Z \otimes Z\rangle$ leads to a smaller error than simply measuring the two qubit state in the computational basis and computing $\langle Z \otimes Z\rangle = p(00) + p(11) - p(10) - p(01)$. [TIP: use the error propagation formula.]*

Note that while using the Hadamard test leads to smaller errors, there is a trade-off as here we need the use of an additional qubit, plus being able to implement controlled operations.

Let us now introduce a useful quantum gate known as the *SWAP* gate. Explicitly, the SWAP gate is given by

$$\text{SWAP} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

**Exercise 3** (Useful equality). *Explicitly show via matrix multiplication that the SWAP operator con be obtained via the quantum circuit in Fig. 1(b).*

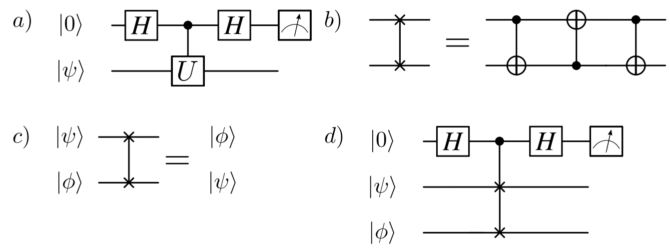


FIG. 1. **Useful quantum circuits**. a) Circuit for the Hadamard test. b) Decomposition of the SWAP gate into 3 CNOTs. c) Effect of the SWAP test on a product state. d) Circuit for the SWAP test.

**Exercise 4** (Useful equality). *Let $|\psi\rangle = \sum_{i=0}^{1} c_i|i\rangle$ and $|\phi\rangle = \sum_{j=0}^{1} b_i|j\rangle$ be two single-qubit states. Prove that, as shown in Fig. 1(c), we have*

$$SWAP|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle. \tag{3}$$

*[TIP: First, explicitly prove via matrix multiplication that $SWAP|i\rangle \otimes |j\rangle = |j\rangle \otimes |i\rangle$].*

Many times, we combine quantum circuits to perform useful and non-trivial operations. For instance, we can combine the Hadamard test with the SWAP operator in the so-called SWAP test shown in Fig. 1(d).

**Exercise 5** (SWAP test). *Show that the SWAP test circuit in Fig. 1(d) can be used to compute $|\langle\phi|\psi\rangle|^2$.*

## II. Cost functions

*Cost functions encode in their extrema the solution to an optimization task.*

Consider a cost function of the form

$$C(\boldsymbol{\theta}) = \mathrm{Tr}[U(\boldsymbol{\theta})\rho U^\dagger(\boldsymbol{\theta})O] \,, \tag{4}$$

where $\rho$ is an input quantum state acting on a Hilbert space $\mathcal{H}$, $U(\boldsymbol{\theta})$ a parametrized quantum circuit, and $O$ a Hermitian operator. Without loss of generality, we can express $U(\boldsymbol{\theta})$ as

$$U(\boldsymbol{\theta}) = \prod_{j=1}^{L} e^{-i\theta_j\sigma_\mu^j/2} W_j \tag{5}$$

$$= e^{-i\theta_L\sigma_\mu^L/2} W_L \dots e^{-i\theta_1\sigma_\mu^1/2} W_1 \tag{6}$$

with $\sigma_\mu^j \in \{X,Y,Z\}$ a Pauli matrix, and $W_j$ an unparametrized fixed unitary. Here, the set of trainable parameters are given by $\boldsymbol{\theta} = (\theta_1, \dots, \theta_L)$.

Being able to exactly compute the partial derivatives of $C(\boldsymbol{\theta})$ with respect of the trainable parameters allows us to use optimization methods such as gradient descent. In what follows we will show that one can exactly compute the partial derivatives $\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k}$ on a quantum computer for any $\theta_k$.

First, show that:

**Exercise 6** (Useful equality)**.** *Let $\rho$ be an arbitrary quantum state acting on $\mathcal{H}$, and let $\sigma_k$ be a Pauli matrix. Show that*

$$i[\sigma_k, \rho] = e^{i\pi\sigma_k/4}\rho e^{-i\pi\sigma_k/4} - e^{-i\pi\sigma_k/4}\rho e^{i\pi\sigma_k/4} \,. \tag{7}$$

*[Tip: Recall that $e^{i\theta\sigma_k} = \cos(\theta)\, I + i \sin(\theta)\, \sigma_k$].*

**Exercise 7** (Useful equality 2)**.** *Show that*

$$\frac{\partial (e^{-i\theta\sigma})}{\partial \theta} = -i\sigma e^{-i\theta\sigma} \,. \tag{8}$$

**Exercise 8** (Partial derivative of $U(\boldsymbol{\theta})$)**.** *Let $\theta_k$ be a trainable parameter in $\boldsymbol{\theta}$. Show that*

$$\frac{\partial U(\boldsymbol{\theta})}{\partial \theta_k} = \frac{-i}{2} U_L(\boldsymbol{\theta}_L)\sigma_\mu^k \mathrm{U}_R(\boldsymbol{\theta}_R) \,, \tag{9}$$

*with*

$$U_L(\boldsymbol{\theta}_L) = \prod_{j>k} e^{-i\theta_j \sigma_\mu^j 2} W_j \,, \tag{10}$$

$$U_R(\boldsymbol{\theta}_R) = \prod_{j\leqslant k} e^{-i\theta_j \sigma_\mu^j 2} W_j \,, \tag{11}$$

*and where $\boldsymbol{\theta}_R = (\theta_1, \dots, \theta_k)$, and $\boldsymbol{\theta}_R = (\theta_{k+1}, \dots, \theta_L)$. Note that such that $U(\boldsymbol{\theta}) = U_L(\boldsymbol{\theta}_L) U_R(\boldsymbol{\theta}_R)$. [TIP: Recall that given three parametrized matrices $A(x)$, $B(y)$ and $C(z)$, we have $\frac{\partial A(x)B(y)C(z)}{\partial y} = A(x)\frac{\partial B(y)}{\partial y}C(z)$.]*

Using the results in the previous two exercises prove that

**Exercise 9** (Parameter shift rule)**.** *Let $\theta_k$ be a trainable parameter in $\boldsymbol{\theta}$, and let $C(\boldsymbol{\theta})$ be the cost function in Eq. (4). Show that*

$$\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k} = \frac{1}{2}(C(\boldsymbol{\theta}_k^+) - C(\boldsymbol{\theta}_k^-)) \,, \tag{12}$$

*with $\boldsymbol{\theta}_k^\pm = (\theta_1, \dots, \theta_k \pm \frac{\pi}{2}, \dots, \theta_L)$. [TIP: Use the fact that:*

$$\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k} = \frac{\partial \mathrm{Tr}[U(\boldsymbol{\theta})\rho U^\dagger(\boldsymbol{\theta}) O]}{\partial \theta_k} \tag{13}$$

$$= \mathrm{Tr}[\frac{\partial U(\boldsymbol{\theta})}{\partial \theta_k}\rho U^\dagger(\boldsymbol{\theta}) O] + \mathrm{Tr}[U(\boldsymbol{\theta})\rho \frac{\partial U^\dagger(\boldsymbol{\theta})}{\partial \theta_k} O] \tag{14}$$

*where in the second equality we have use the linearity of the partial derivative, and the derivative of a product.]*

Equation (12) is known as the *parameter shift-rule*, and is one of the most surprising result of quantum computing as it implies that we can *exactly* compute the gradient of the cost function by evaluating it at two sets of *shifted* parameters.

## III. Quantum Noise

*Thus far we have modeled quantum systems as being closed. However, in reality, quantum systems always interact with their environment leading to decoherence and loss of information due to uncontrolled system-environment interactions.*

The quantum channel formalism is a general tool for describing the evolution of quantum systems in a wide variety of circumstances, including quantum systems interacting with their environment.

Here, we recall that a quantum channel $\mathcal{E}$ is a map that takes a quantum states, and outputs a quantum state. That is

$$\mathcal{E}(\rho) = \rho' \,. \tag{15}$$

For instance, applying a unitary $U$ is a map, where $\mathcal{E}_U(\rho) = U\rho U^\dagger$, and so is performing a measurement, $\mathcal{E}_m(\rho) = M_m \rho M_m^\dagger$.

One of the most widely used noisy channels is the global depolarizing channel, which acts on an $n$-qubit state as

$$\mathcal{D}(\rho) = p\rho + (1-p)\mathbb{1}/2^n \,, \tag{16}$$

where $p \in [0,1]$ is the noise probability.

**Exercise 10** (Noisy states are quantum states)**.** *Show that if $\rho$ is an $n$-qubit quantum state, then $\rho' = \mathcal{D}(\rho)$ is also a quantum state. That is, show that $\mathrm{Tr}[\rho'] = 1$ and $\rho' \geqslant 0$. [TIP: recall that since $\rho$ is a quantum state, then, $\rho \geqslant 0$. Moreover, we recall that $\rho \geqslant 0$ implies that for any $|x\rangle$, we have that $\langle x|\rho|x\rangle \geqslant 0$. ]*

**Exercise 11** (Noisy states are mixed)**.** *Show that even if $\rho$ is a pure quantum state, the state $\rho' = \mathcal{D}(\rho)$ is mixed. That is, prove that $\mathrm{Tr}[(\rho')^2] \leqslant 1$.*

**Exercise 12** (Multiple noisy channels)**.** *Show that if $L$ global depolarizing channels act on $\rho$, then the output state is*

$$\rho' = p^L \rho + (1-p^L)\mathbb{1}/2^n \,. \tag{17}$$