Exercise 1. *Distinguishing channels*

The setting is the following: With equal probabilities you are given either the identity channel I on some finite alphabet $\mathcal{X}$ or an arbitrary channel W on the same alphabet, without knowing which. In terms of conditional probability distributions the channels are described by

$$I(x\,|\,x') = \delta_{xx'} \quad \textit{and} \quad W(x\,|\,x')\,, \tag{1}$$

for $x, x' \in \mathcal{X}$. You are allowed to use the given channel once, possibly with a stochastic (randomized) input, and then asked which channel was used.

(a) The error probability of a channel W is defined as $P_{\text{error}}(W) := \max_{x\in\mathcal{X}}\big(1 - W(x\,|\,x)\big)$. Argue why this is a sensible definition.

(b) Using properties of the trace distance of probability distributions previously derived in Exercise Sheet 1, Exercise 1(c), show that the probability of guessing the channel correctly in the above scenario is

$$P_{\text{guess}}(I \text{ vs. } W) = \frac{1}{2}\big(1 + P_{\text{error}}(W)\big). \tag{2}$$

Solution.

(a) For some specific $x \in \mathcal{X}$ the error probability of W is the probability that the output differs from the input which is equal to 1 minus the probability that the output equals the input, i.e. $P_{\text{error}}(W \text{ on } x) = 1 - W(x\,|\,x)$. The ‘overall’ error probability of W is then just the maximum error probability of W on the inputs,

$$P_{\text{error}}(W) = \max_{x\in\mathcal{X}} P_{\text{error}}(W \text{ on } x) = \max_{x\in\mathcal{X}}\big(1 - W(x\,|\,x)\big). \tag{S.1}$$

Notice, however, that a high error probability according to this definition does not mean that the channel is useless! This way of defining the error probability merely compares the channel to the identity channel.

(b) According to the setting, the only thing we will have access to at the end is the output of the channel, so what we will be confronted with is the problem of distinguishing two probability distributions from one sample. This is exactly the setting considered in the mentioned exercise in Exercise Sheet 1. Furthermore, we have not specified the input distribution. Calling the chosen input RV X, distributed P_X, and denoting the output RV if W is applied by $X' := W(X)$, distributed $P_{X'}$, we see that

$$P_{\text{guess}}(I \text{ vs. } W) = \max_{P_X} P_{\text{guess}}(X \text{ vs. } X') = \max_{P_X} \frac{1}{2}\big(1 + \delta(P_X, P_{X'})\big). \tag{S.2}$$

We can express $P_{X'}$ in terms of P_X and $W(\cdot\,|\,\cdot)$ as

$$P_{X'}(x') = \sum_{x\in\mathcal{X}} P_X(x) W(x'\,|\,x) \tag{S.3}$$

for $x' \in \mathcal{X}$. Together with the definition of the trace distance $\delta(\cdot,\cdot)$ therefore:

$$\begin{aligned}
P_{\text{guess}}(I \text{ vs. } W) &= \frac{1}{2}\big(1 + \max_{P_X} \delta(P_X, P_{X'})\big) \\
&= \frac{1}{2}\big(1 + \max_{P_X} \frac{1}{2} \sum_{x\in\mathcal{X}} |P_X(x) - P_{X'}(x)|\,\big) \\
&= \frac{1}{2}\big(1 + \frac{1}{2} \max_{P_X} \sum_{x\in\mathcal{X}} \Big|\sum_{x'\in\mathcal{X}} P_X(x')\,[\delta_{xx'} - W(x\,|\,x')\,]\Big|\,\big)
\end{aligned} \tag{S.4}$$

A few observations are in order. Consider the function g of P_X,

$$g(P_X) := \sum_{x\in\mathcal{X}} \Big|\sum_{x'\in\mathcal{X}} P_X(x')\,[\delta_{xx'} - W(x\,|\,x')\,]\Big|. \tag{S.5}$$

This is the function we want to maximise in (S.4). The absolute value $|\cdot|$ is a convex function. In g this is applied to a linear function of P_X, $\sum_{x'\in\mathcal{X}} P_X(x')\,[\delta_{xx'} - W(x\,|\,x')\,]$. Finally, g is a linear combination of such terms with positive coefficients (here all coefficients are 1). Hence, g is a convex function on the space of probability distributions.

Now, when maximising a convex function we know that the maximiser lies on the boundary of the domain of the function. In this case, the boundary of probability distributions are the deterministic distributions,

$$P_X(x) = \delta_{xx^\star}\,, \quad \text{for some } x^\star \in \mathcal{X}. \tag{S.6}$$

Therefore we can replace the maximization in (S.4) over probability distributions with a maximization over all possible deterministic distributions and find

$$\begin{aligned}
P_{\text{guess}}(I \text{ vs. } W) &= \frac{1}{2}\big(1 + \frac{1}{2} \max_{x^\star} \sum_{x\in\mathcal{X}} \Big|\sum_{x'\in\mathcal{X}} \delta_{x'x^\star}\,[\delta_{xx'} - W(x\,|\,x')\,]\Big|\,\big) \\
&= \frac{1}{2}\big(1 + \frac{1}{2} \max_{x^\star} \sum_{x\in\mathcal{X}} \big|\delta_{xx^\star} - W(x\,|\,x^\star)\big|\,\big) \\
&= \frac{1}{2}\big(1 + \frac{1}{2} \max_{x^\star} \Big[1 - W(x^\star\,|\,x^\star) + \sum_{x\neq x^\star} W(x\,|\,x^\star)\Big]\big) \\
&= \frac{1}{2}\big(1 + \max_{x^\star} \,[1 - W(x^\star\,|\,x^\star)]\,\big) \\
&= \frac{1}{2}\big(1 + P_{\text{error}}(W)\big),
\end{aligned} \tag{S.7}$$

where we used that $W(\cdot\,|\,\cdot)$ is a conditional probability distribution in the fourth line.

Exercise 2. *Source coding*

In the lecture we have seen a channel coding theorem with a more or less tricky derivation involving interesting ideas which were needed to get to the final result. In this exercise we will consider source coding, the challenge of which is to compress some input described by a RV X to another RV Y. One can think of this as the task to find the optimal channel $P_{Y|X}$ such that all information contained in X can be retrieved from Y, but the alphabet $\mathcal{Y}$ of Y should be as small as possible. Instead of phrasing the problem in a technical way, we are here taking a simpler and more conceptual approach.

Consider a k-bit string described by the RV X distributed P_X. We would like to compress this to a l-bit string described by a RV Y distributed P_Y. The goal is to find the minimal l such that Y contains (almost) all information from X.

(a) *Suppose you are not interested in many but just in a single use of the compressed source. Furthermore you will not tolerate any errors. How small can l be chosen in this case?*

(b) *If you allow for some small error ε, what is the answer now?*

(c) *Often one is interested in many (independent) uses of the source. Using asymptotic equipartition results for max-entropies, show that for N i.i.d. uses of the source the optimum compression requires $NH(X)_{P_X}$ bits in the limit $N \to \infty$.*

This is called Shannon's source coding theorem. Interestingly, using single-shot quantities and results about asymptotic equipartition this result becomes almost trivial.

Solution.

(a) If no errors are tolerated all one can do is to map X in a $1-1$ fashion to Y such that every possible value of X has exactly one corresponding value in Y (by 'value' here we basically mean a specific bit string). Hence, optimally the alphabet of Y is as large as the support of X. In terms if bits this is

$$l = \log \operatorname{supp} P_X = H_{\max}(X)_{P_X}, \tag{S.8}$$

where log is the logarithm w.r.t. base 2.

(b) As encountered in Exercise Sheet 2, Exercise 1, allowing for small errors ε suggests that we should use the smooth max-entropy instead of the standard max-entropy. Indeed, by definition of the smooth max-entropy we find that the optimal number of bits to encode X is in this case given by

$$l = \min_{Q_X \in \mathcal{B}^\varepsilon(P_X)} H_{\max}(X)_{Q_X} = H^\varepsilon_{\max}(X)_{P_X}. \tag{S.9}$$

(c) The AEP for the classical smooth max-entropy says

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{1}{N} H^\varepsilon_{\max}(X^N)_{P_X^{\times N}} = H(X)_{P_X}. \tag{S.10}$$

According to (b), in N independent uses the compressed data finds space on $H^\varepsilon_{\max}(X^N)_{P_X^{\times N}}$ bits, while the error is given by ε. The AEP now relates this asymptotically to $NH(X)_{P_X} \sim H^\varepsilon_{\max}(X^N)_{P_X^{\times N}}$ for large N, which concludes the proof.

Exercise 3. *Getting used to the Bloch representation*

In this exercise we will see that any density operator of a qubit can be written as

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{r} \cdot \vec{\sigma}), \tag{3}$$

where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector of Pauli matrices and $\vec{r} = (r_x, r_y, r_z) \in \mathbb{R}^3, |\vec{r}| \leq 1$ is the so-called Bloch vector that gives us the position of a point in a unit ball. The surface of that ball is usually known as the Bloch sphere. Thereby we go through a few properties of density matrices describing states of a qubit using the Bloch representation. This way of expressing qubit states is very convenient and frequently used in quantum information theory.

Hint: *The (anti-)commutation relations for Pauli matrices will be helpful.*

(a) Using Eq. (3):

- *(i) Find and draw in the ball the Bloch vectors of a fully mixed state and the pure states that form three bases, $\{|\uparrow\rangle, |\downarrow\rangle\}$, $\{|+\rangle, |-\rangle\}$ and $\{|\circlearrowleft\rangle, |\circlearrowright\rangle\}$.*
- *(ii) Find and diagonalise the states represented by Bloch vectors $\vec{r}_1 = (\frac{1}{2}, 0, 0)$ and $\vec{r}_2 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$.*

(b) Show that the operator ρ defined in Eq. (3) is a valid density operator for any vector $\vec{r}$ with $|\vec{r}| \leq 1$ by proving it fulfils the following properties:

- *(i) Hermiticity: $\rho = \rho^\dagger$.*
- *(ii) Positivity: $\rho \geq 0$.*
- *(iii) Normalisation: $\mathrm{tr}(\rho) = 1$.*

(c) Now do the converse: show that any qubit density operator may be written in the form of Eq. (3).

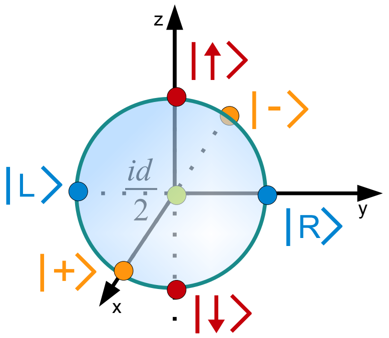
(d) Check that the surface of the ball is formed by all the pure states, i.e.

$$\rho \textit{ pure } \Leftrightarrow \rho = \frac{1}{2}(\mathbb{1} + \vec{r} \cdot \vec{\sigma}) \textit{ with } |\vec{r}| = 1. \tag{4}$$

Solution.

(a) (i)

state	**density matrix**	**Bloch vector**	**in the figure**
$\frac{\mathbb{1}}{2}$	$\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$(0, 0, 0)$	green
$\vert 0\rangle$	$\frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$	$(0, 0, 1)$	red
$\vert 1\rangle$	$\frac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$	$(0, 0, -1)$	red



state	**density matrix**	**Bloch vector**	**in the figure**
$\vert +\rangle$	$\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$(1, 0, 0)$	yellow
$\vert -\rangle$	$\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$	$(-1, 0, 0)$	yellow
$\vert \circlearrowright\rangle$	$\frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$	$(0, 1, 0)$	blue: $\vert R\rangle$
$\vert \circlearrowleft\rangle$	$\frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$	$(0, -1, 0)$	blue: $\vert L\rangle$

(ii) We have

$$\begin{aligned}\rho_1 &= \frac{1}{2}\left[\mathbb{1} + \left(\frac{1}{2}, 0, 0\right)\cdot(\sigma_x, \sigma_y, \sigma_z)\right] \\ &= \frac{1}{2}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] \\ &= \frac{1}{4}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad \Rightarrow \quad \text{Eigenvalues: } \left\{\frac{1}{4}, \frac{3}{4}\right\},\end{aligned} \tag{S.11}$$

$$\begin{aligned}\rho_2 &= \frac{1}{2}\left[\mathbb{1} + \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)\cdot(\sigma_x, \sigma_y, \sigma_z)\right] \\ &= \frac{1}{2}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right] \\ &= \frac{1}{2\sqrt{2}}\begin{pmatrix} \sqrt{2}+1 & 1 \\ 1 & \sqrt{2}-1 \end{pmatrix} \qquad \Rightarrow \quad \text{Eigenvalues: } \{0, 1\}.\end{aligned} \tag{S.12}$$

The first Bloch vector lies inside the ball ($|\vec{r}_1 = \frac{1}{4}|$), and the state that it represents is mixed. The Bloch vector of the second state is on the surface of the sphere, and that state is pure.

(b) (i) *Hermiticity:* $\rho = \rho^\dagger$. All Pauli matrices are Hermitian and the vector $\vec{r}$ is real, so the result comes from direct application of Eq. (3).

(ii) *Positivity:* $\rho \geq 0$. The general form of a state given by Eq. (3) is

$$\rho = \frac{1}{2}\begin{pmatrix} 1 + r_z & r_x - i r_y \\ r_x + i r_y & 1 - r_z \end{pmatrix} \quad \Rightarrow \quad \text{Eigenvalues: } \left\{\frac{1-|\vec{r}|}{2}, \frac{1+|\vec{r}|}{2}\right\}. \tag{S.13}$$

Since $0 \leq |\vec{r}| \leq 1$, the eigenvalues are non negative.

(iii) *Normalisation:* $\mathrm{tr}(\rho) = 1$. From Eq. (S.13) we have that

$$\mathrm{tr}(\rho) = (1 + r_z) + (1 - r_z) = 1. \tag{S.14}$$

(c) One can always expand an operator A in an orthonormal basis $\{e_i\}_i$ as $A = \sum_i (A, e_i) e_i$, where the inner product (A, B) is defined as $\mathrm{tr}(A^* B)$. The three Pauli matrices and the identity form a basis for 2×2 matrices, $\mathcal{B}$. However, this basis is not normalized. A normalized basis would be

$$\mathcal{B}' = \left\{\frac{\sigma_x}{\sqrt{2}}, \frac{\sigma_y}{\sqrt{2}}, \frac{\sigma_z}{\sqrt{2}}, \frac{\mathbb{1}}{\sqrt{2}}\right\}. \tag{S.15}$$

We can expand any 2×2 matrix in this basis, and in particular any two-level density operator:

$$\begin{aligned}\rho &= \mathrm{tr}(\rho \mathbb{1})\ \frac{\mathbb{1}}{2} + \sum_i \mathrm{tr}(\rho \sigma_i)\ \frac{\sigma_i}{2} \\ &= \frac{\mathbb{1}}{2} + \frac{1}{2}\Big(\mathrm{tr}(\rho\sigma_x), \mathrm{tr}(\rho\sigma_y), \mathrm{tr}(\rho\sigma_z)\Big)\cdot\Big(\sigma_x, \sigma_y, \sigma_z\Big) \\ &= \frac{\mathbb{1}}{2}\,(r_x, r_y, r_z)\cdot(\sigma_x, \sigma_y, \sigma_z)\,,\end{aligned} \tag{S.16}$$

where we used the property of density operators $\mathrm{tr}(\rho) = 1$. To obtain the bound $|\vec{r}| \leq 1$ we use the fact that for any density operator $\mathrm{tr}(\rho^2) \leq 1$ (because all eigenvalues $\lambda_j \leq 1$ and $\sum \lambda_j = 1$) and get

$$\begin{aligned}
1 &\geq \mathrm{tr}(\rho^2) \\
&= \mathrm{tr}\Big(\left[\frac{\mathbb{1}}{2} + \sum_i r_i \frac{\sigma_i}{2} \right] \left[\frac{\mathbb{1}}{2} + \sum_i r_i \frac{\sigma_i}{2} \right] \Big) \\
&= \frac{1}{4}\mathrm{tr}\Big(\left[1 + \sum_i {r_i}^2 \right] \mathbb{1} \Big) \qquad \text{(because } \mathcal{B}' \text{ is an orthonormal basis)} \\
&= \frac{1}{2}\Big(1 + \sum_i {r_i}^2 \Big) \\
&\Rightarrow 1 \geq \sum_i {r_i}^2 .
\end{aligned} \tag{S.17}$$

(d) For pure states, $\mathrm{tr}(\rho^2) = 1$ and we can replace all "$\geq$" with "$=$" above, obtaining $|\vec{r}| = 1$.