Freie Universität Berlin
Tutorials on Quantum Information Theory
Summer Term 2018

Problem Sheet 3
POVM, impossibilities and ℓ_p-norms

Discussed in Tutorial: 17/05/2018 J. Eisert, D. Hangleiter, I. Roth

1. **Encoding classical bits.** On the last excercise sheet we introduced the description of quantum measurements with the help of POVMs. We want to use this formulation to study the following question:

   Let $\mathcal{H}$ be a d-dimensional Hilbert space. Our aim is to encode n classical bits into the space of quantum states $\mathcal{D}(\mathcal{H})$. To this end, we choose a set of 2^n states $\{\rho_i\}_{i\in\{0,1\}^n} \subset \mathcal{D}(\mathcal{H})$, each state corresponding to a bit string. To decode the bit string we have to make a measurement described by a POVM $\{F_i\}_{i\in\{0,1\}^n}$, where the bit string is the outcome.

   How many classical bits can be encoded and decoded in a d-dimensional quantum system in this way?

   Consider a source that outputs the bit string $x \in \{0,1\}^n$ with probability $p(x)$.

   a) Define the success probability of the decoding procedure.

   Solution: $\mathrm{Tr}[\rho_i F_i]$ should be maximal (1) for each i. The total success probability is then the expectation of that with respect to p, i.e., $\sum_x p(x)\,\mathrm{Tr}[\rho_x F_x]$

   b) Show that for $p(x) = 2^{-n}$ the success probability is bounded by $2^{-n}d$.
   (*Hint:* Argue that $\mathbb{1} \succcurlyeq \rho_i$ for all i and show that for $A \succcurlyeq 0$ and $B \succcurlyeq C$ it holds that $\mathrm{Tr}(AB) \geq \mathrm{Tr}(AC)$ as a starting point.)

   Solution: Clearly $\mathbb{1} - \rho = U(\mathbb{1} - \Lambda)U^\dagger$, where U diagonalises ρ. But since ρ is a quantum state with eigenvalues smaller than one, $\mathbb{1} - \Lambda$ has only nonnegative entries, hence the claim $\mathbb{1} \succcurlyeq \rho_i$ for all i. If $A \succcurlyeq 0$ and $B - C \succcurlyeq 0$, then $\mathrm{Tr}\,AB - \mathrm{Tr}\,AC = \mathrm{Tr}(A(B-C)) \geq 0$. Thus, $\mathrm{Tr}(AB) \geq \mathrm{Tr}(AC)$.

   Hence, we have

   $$\sum_x p(x)\,\mathrm{Tr}[\rho_x F_x] = 2^{-n}\sum_i \mathrm{Tr}[\rho_i F_i] \leq 2^{-n}\sum_i \mathrm{Tr}[F_i] = 2^{-n}\,\mathrm{Tr}\,\mathbb{1} = 2^{-n}d \quad (1)$$

   and the claim follows.

   c) What does this imply?

   Solution: One cannot encode more than $\log_2 d$ bits in a d-dimensional quantum system.

2. **Impossible machines – no cloning.**

   In this problem we will re-derive the impossibility results that you have seen in the lecture but now directly using the structure of quantum theory.

   Show that there does not exist a unitary map on two copies of a Hilbert space $\mathcal{H}$ which acts in the following way:

   $$\forall\,|\psi\rangle \in \mathcal{H} :\; U\,|\psi\rangle\,|0\rangle = \mathrm{e}^{\mathrm{i}\phi(\psi)}\,|\psi\rangle\,|\psi\rangle\ .$$

Solution: Assume this was the case for $|\psi\rangle$ and $|\phi\rangle$ with $|\psi\rangle \neq \mathrm{e}^{\mathrm{i}\alpha}|\phi\rangle$ for any α.

Let us consider the scalar product between two such vectors

$$\begin{aligned}\langle\varphi \mid \psi\rangle &= \langle 0|\langle\varphi| U^\dagger U |\psi\rangle |0\rangle \\ &= \mathrm{e}^{\mathrm{i}(\phi(\psi)-\phi(\varphi))}\langle\varphi|\langle\varphi|\psi\rangle|\psi\rangle \\ &= \langle\varphi \mid \psi\rangle^2 \mathrm{e}^{\mathrm{i}(\phi(\psi)-\phi(\varphi))}.\end{aligned}$$

Taking absolute values on both sides shows that $\langle\varphi \mid \psi\rangle$ can only be 0 or 1, so it cannot be the case that U clones arbitrary states.

3. **ℓ_p-norms**

In quantum information we deal with a handful of different matrix spaces such as the set of quantum states and in the near future also quantum channels. For quantitative statements we have to equip these spaces with distance measures. Depending on the application and context different distance measures have the desired operational meaning.

A prominent role is played by the so called *Schatten p-norms*. But to set the stage we have to first familiarise ourself with their analogons on vector spaces, namely *ℓ_p-norms*. For $1 \leq p < \infty$ the ℓ_p-norm on the complex vector space $\mathbb{C}^n$ is defined as

$$\|\cdot\|_{\ell_p} : x \mapsto \|x\|_{\ell_p} \coloneqq \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}},$$

and the ℓ_∞-norm as

$$\|\cdot\|_{\ell_\infty} : x \mapsto \|x\|_{\ell_\infty} \coloneqq \lim_{p\to\infty} \|x\|_{\ell_p}.$$

We will now characterise the function $\|\cdot\|_{\ell_p}$ and derive important properties. We begin with an explicit expression for the ℓ_∞-norm.

a) Show that $\|x\|_{\ell_\infty} = \max_{1\leq i\leq n} |x_i|$.

Solution: We assume w.l.o.g. that $|x_1| = \max_i |x_i|$

$$\begin{aligned}\|x\|_{\ell_\infty} = \lim_{p\to\infty} = \lim_{p\to\infty}\left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} &\qquad (2)\\ = |x_1| \lim_{p\to\infty}\left[1 + \sum_{i=2}^n \frac{|x_i|^p}{|x_1|^p}\right]^{\frac{1}{p}} &\qquad (3)\\ = |x_1|. &\qquad (4)\end{aligned}$$

For all of what follows the notion of a convex function will be important. Let $D \subset \mathbb{R}$ be a convex set. We say that a function $f : D \to \mathbb{R}$ is *convex* if

$$f\left(\sum_i a_i x_i\right) \leq \sum_i a_i f(x_i),$$

for all $x_i \in D$ and $a_i \geq 0, i = 1, \ldots, m$ such that $\sum_i a_i = 1$.

b) Show that any twice continuously differentiable function on an open interval is convex if and only if its second derivative is everywhere nonnegative.

Solution: First, observe that the definition above is equivalent to requiring that for $\lambda \in (0,1)$ and $x, y \in D$ it holds that $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. The definition follows directly by applying this result repeatedly.

Convex $\Leftarrow f'' \geq 0$
For any $x, y \in (\alpha, \beta)$, $x < y$, and $\lambda \in (0,1)$, we set $z = \lambda x + (1-\lambda)y$. Assume $f''(x)$ be non-negative on (α, β), correspondingly $f'(x)$ is non-decreasing on (α, β). Now,

$$
\begin{aligned}
f(z) &= \lambda f(z) + (1-\lambda) f(z) && (5)\\
&= \lambda \int_x^z f'(t)\,\mathrm{d}t + \lambda f(x) + (1-\lambda)\int_y^z f'(t)\,\mathrm{d}t + (1-\lambda) f(y) && (6)\\
&\leq \lambda f(x) + (1-\lambda) f(y) + \lambda f'(z)(z-x) + (1-\lambda) f'(z)(z-y) && (7)\\
&= \lambda f(x) + (1-\lambda) f(y) + f'(z)\left[z - (\lambda x + (1-\lambda)y)\right] && (8)\\
&= \lambda f(x) + (1-\lambda) f(y). && (9)
\end{aligned}
$$

Convex $\Rightarrow f'' \geq 0$
Conversely, assume that $f''(x)$ is negative somewhere, than by continuity there exist a subinterval (α', β') where f' is decreasing everywhere. Choosing $x, y \in (\alpha', \beta')$ and $\lambda \in (0,1)$ implies that $\int_x^z f'(t)\,\mathrm{d}t > f'(z)(z-x)$ and $\int_y^z f'(t)\,\mathrm{d}t > f'(z)(z-y)$. Thus, the same calculation as above yields $f(z) > \lambda f(x) + (1-\lambda) f(y)$ establishing a contradiction to f being convex.

Alternatively: Assume $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ holds. Then, the following is also true

$$
\begin{aligned}
f(x) &= f\left(\frac{1}{2}(x+h) + \frac{1}{2}(x-h)\right) && (10)\\
&\leq \frac{1}{2} f(x+h) + \frac{1}{2} f(x-h) && (11)\\
\Leftrightarrow 0 &\leq f(x+h) + f(x-h) - 2f(x). && (12)
\end{aligned}
$$

This is exactly the term one encounters in the Taylor expansion of the second derivative

$$
f''(x) = \lim_{h \to \infty} \frac{f(x+h) + f(x-h) + 2f(x)}{h^2}. \tag{13}
$$

c) Show that $|\cdot|^p$ is a convex function for $p \geq 1$.

Solution: Applying the criterion of the last excercise, it is obvious that the function $(0, \infty) \to \mathbb{R}$ $x \mapsto x^p$ with $p \geq 1$ is convex. Consider $x, y \in \mathbb{R}$, $x, y \neq 0$ and $\lambda \in (0,1)$ then

$$
\begin{aligned}
|\lambda x + (1-\lambda)y|^p &\leq (\lambda |x| + (1-\lambda)|y|)^p && (14)\\
&\leq \lambda |x|^p + (1-\lambda)|y|^p. && (15)
\end{aligned}
$$

We will now use this fact to show that $\|\cdot\|_{\ell_p}$ is a norm (positive definite, absolutely homogeneous, subadditive aka triangle inequality).

d) Argue that $\|\cdot\|_{\ell_p}$ is positive definite and absolutely homogeneous for $1 \leq p < \infty$ and $p = \infty$.

Solution: Clear ;)

That was easy. Now comes the hard part; we have to show that the norms satisfy the triangle inequality, i.e.

$$\|x+y\|_{\ell_p} \leq \|x\|_{\ell_p} + \|y\|_{\ell_p}. \tag{16}$$

In fact, the triangle inequality for ℓ_p-norms has even its own name, *Minkowski inequality*. A clever way to prove this inequality is to normalise the right hand side, introduce normalised vectors and then use the convexity of $|\cdot|^p$.

e) Argue that it is sufficient to consider the case $\|x\|_{\ell_p} = \lambda$ and $\|y\|_{\ell_p} = (1-\lambda)$ with $\lambda \in (0,1)$ in order to prove the Minkowski inequality.

Solution: Let $\tilde{x}, \tilde{y} \in \mathbb{R}$ then by absolute homogeneity of $\|\cdot\|_{\ell_p}$ we have

$$\|\tilde{x}+\tilde{y}\|_{\ell_p} \leq \|\tilde{x}\|_{\ell_p} + \|\tilde{y}\|_{\ell_p} \tag{17}$$

if and only if

$$\|\frac{1}{s}\tilde{x} + \frac{1}{s}\tilde{y}\|_{\ell_p} \leq 1 \tag{18}$$

with $s \coloneqq \|\tilde{x}\|_{\ell_p} + \|\tilde{y}\|_{\ell_q}$. Choosing $x = \frac{1}{s}\tilde{x}$ and $y = \frac{1}{s}\tilde{y}$ yields the situation above.

f) Show the Minkowski inequality for the ℓ_p-norms when $1 \leq p < \infty$.

Solution: We write $x = \lambda\hat{x}$ and $y = (1-\lambda)\hat{y}$ with $\|\hat{x}\|_{\ell_p} = \|\hat{y}\|_{\ell_p} = 1$. Then,

$$\begin{aligned}
\|x+y\|_{\ell_p}^p &= \|\lambda\hat{x} + (1-\lambda)\hat{y}\|_{\ell_p} && (19)\\
&= \sum_i |\lambda\hat{x}_i + (1-\lambda)\hat{y}_i|^p && (20)\\
&\leq \sum_i [\lambda|\hat{x}_i|^p + (1-\lambda)|\hat{y}_i|^p] && (21)\\
&= \lambda\|\hat{x}\|_{\ell_p} + (1-\lambda)\|\hat{y}\|_{\ell_p} && (22)\\
&= \lambda + (1-\lambda) = 1. && (23)
\end{aligned}$$

A crucial property of the ℓ_p-norms is Hölder's inequality. It generalises the Cauchy-Schwarz inequality, which is its special case for $p = 2$. Let $\langle\cdot,\cdot\rangle$ be the Euclidean inner product on $\mathbb{C}^n$, i.e. $\langle x, y\rangle = \sum_{i=1}^n \bar{x}_i y_i$ with $\bar{\cdot}$ denoting the complex conjugate. Hölder's inequality reads

$$|\langle x, y\rangle| \leq \|x\|_{\ell_p}\|y\|_{\ell_q}, \text{ where } \frac{1}{p} + \frac{1}{q} = 1.$$

Like for the proof of Minkowski's inequality, it will be useful to use normalised vectors in the proof of Hölder's inequality. Furthermore, we will need to first establish the *arithmetic-geometric mean inequality*

$$\prod_{i=1}^n x_i^{a_i} \leq \sum_{i=1}^n a_i x_i \text{ if } x_i \geq 0,\ a_i \geq 0,\ \sum_i a_i = 1. \tag{24}$$

g) Show that $-\log$ is a convex function and use this to show the arithmetic-geometric mean inequality, Eq. (24).

Solution: The function $-\log$ is twice continously differential and $(-\log x)'' = (-1/x)' = 1/x^2 > 0$ for $x \in \mathbb{R}$ and, thus, convex.

Then,

$$-\log\left[\prod_i x_i^{a_i}\right] = -\sum_i a_i \log x_i \geq -\log\left[\sum_i a_i x_i\right]. \tag{25}$$

By monotonicity of the logarithm, this implies Eq. (24).

h) Now, prove Hölder's inequality for $1 < p < \infty$.

Solution: Again, by absolute homogeneity of the norms and bilinearity of the scalar product, it is sufficient to consider the case $\|x\|_{\ell_p} = \|y\|_{\ell_q} = 1$.

$$|\langle x, y\rangle| = \sum_i |x_i||y_i| = \sum_i (|x_i|^p)^{1/p}(|y_i|^q)^{1/q}) \tag{26}$$

$$\leq \sum_i (\frac{1}{p}|x_i|^p + \frac{1}{q}|y_i|^q) = \frac{1}{p}\|x\|_{\ell_p}^p + \frac{1}{q}\|y\|_{\ell q}^q \tag{27}$$

$$= \frac{1}{p} + \frac{1}{q} = 1. \tag{28}$$

i) Finally, prove Hölder's inequality for $p = 1$.

Solution: $|\langle x, y\rangle| = \sum_i |x_i||y_i| \leq \max_i\{|x_i|\} \sum_i |y_i| = \|x\|_{\ell_\infty}\|y\|_{\ell_1}$.

More generally, for a norm $\|\cdot\|$ on $\mathbb{C}^d$ one can define its dual norm $\|\cdot\|^*$ as

$$\|x\|^* := \sup_{y \in \mathbb{C}^d, \|y\|=1} |\langle x, y\rangle|. \tag{29}$$

j) Show that for every norm $\|\cdot\|$ on $\mathbb{C}^d$ it holds:

$$|\langle x, y\rangle| \leq \|x\|\|y\|^* \tag{30}$$

for all $x, y \in \mathbb{C}^d$.

k) Show that the dual norm $\|\cdot\|_{\ell_p}^*$ of the ℓ_p-norm $\|\cdot\|_{\ell_p}$ is the ℓ_q-norm $\|\cdot\|_{\ell_q}$ with $\frac{1}{p} + \frac{1}{q} = 1$.

Solution: By Hölder's inequality, we have

$$\|x\|_{\ell_q} \geq \sup_{\|y\|_{\ell_p}=1} |\langle x, y\rangle|. \tag{31}$$

Again by absolute homogeneity we can assume $\|x\|_{\ell_q} = 1$. Setting $y_i = |x_i|^{q/p+1}/x_i$ for all nonzero x_i and 0 else one checks that for $1/p + 1/q = 1$ the inequality is saturated establishing the claim:

$$|\langle x, y\rangle| = \left|\sum_i x_i \frac{|x_i|^{q/p+1}}{x_i}\right| = \sum_i |x_i|^{q/p+1} = \sum_i |x_i|^q = \|x\|_{\ell_q}^q = 1$$

Finally, we will show another convenient property of the ℓ_p norms.

l) Show that the ℓ_p norms are ordered in the sense that

$$\|x\|_{\ell_p} \leq \|x\|_{\ell_q}, \text{ for } q \leq p.$$

Solution: Consider $x \in \mathbb{C}^n$ and define $\hat{x} = x/\|x\|_{\ell_q}$. In particular, we have $\hat{x}_i \leq 1$ for all i. Then, $\|x\|_{\ell_p}^p = \|x\|_{\ell_q}^p \sum_i |\hat{x}_i|^p \leq \|x\|_{\ell_q}^p \sum_i |\hat{x}_i|^q = \|x\|_{\ell_q}^p \|x\|_{\ell_q}^q = \|x\|_{\ell_q}^p$.

4. **Non-uniqueness of the decomposition of mixed states.**

Consider two macroscopically different preparation schemes of a large number of polarised photons:

Preparation A. For each photon we toss a fair coin. Depending on whether we get head or tail, we prepare the photon to have either vertical or horizontal *linear* polarisation.

Preparation B. For each photon we toss a fair coin. Depending on whether we get head or tail, we prepare the photon to have either left-handed or right-handed *circular* polarisation.

We are given a large number of photons which all were prepared by the same scheme.

a) Argue that having only access to the photons we can not distinguish which of the preparation scheme was used.

Solution: Both preparations give rise to the same quantum state, namely, the maximally mixed state. Hence, there is no measurement that distinguishes the two preparations.

b) Argue that if it were possible to distinguish such types of preparations by measuring the photon, locality would be violated.

Solution: Protocol: EPR setting with Bell state

Bob chooses a measurement setting, X or Z and measures his half of the state.

Then, the state reads

$$\rho_A = \mathrm{Tr}[|\psi\rangle\langle\psi| P_1] + \mathrm{Tr}[|\psi\rangle\langle\psi| P_2], \tag{32}$$

where $P_{1,2}$ are either $|+\rangle\langle+|, |-\rangle\langle-|$ or $|0\rangle\langle 0|, |1\rangle\langle 1|$.

Depending on which measurement setting Bob chooses, the state on Alice's side reads $\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$ or $\frac{1}{2}(|+\rangle\langle+| + |-\rangle\langle-|)$.

If Alice had a way of distinguishing the two mixtures, they could have communicated a bit encoded as $\{X, Z\}$.