Freie Universität Berlin
Tutorials on Quantum Information Theory
Summer Term 2018

Solution to Problem 3 on Sheet 4
Schatten norms

Discussed in Tutorial: 24/05/2018 J. Eisert, D. Hangleiter, I. Roth

3. **Schatten p-norms**
On the last excercise sheet we have studied the ℓ_p-norms on vector spaces. The ℓ_p-norms have important cousins on matrix spaces, the Schatten p-norms. As they are important distant measures in quantum information, we study there different definitions and properties in this excerice.

One way to introduce the Schatten p-norm with $p \in [1, \infty)$ for a matrix $A \in \mathbb{C}^{n\times n}$ is

$$\|A\|_p := (\mathrm{Tr}\,[|A|^p])^{\frac{1}{p}}\,, \tag{1}$$

where $|A| := \sqrt{A^\dagger A}$ is the matrix absolute value. Furthermore, the case $p = \infty$ is defined as the limit $\|A\|_\infty = \lim_{p\to\infty} \|A\|_p$.

These norms are related to the ℓ_p-norms of the eigenvalues (or more generally the singular values) of A.

a) Let A be a Hermitian matrix and let $\lambda = (\lambda_1, \ldots, \lambda_n)$ be the vector of its eigenvalues. Show that

$$\|A\|_p = \|\lambda\|_{\ell_p} \tag{2}$$

for all p.

Solution: Let $A = U\Lambda U^\dagger$ be the eigenvalue decomposition of A. Recall that for Hermitian A it holds that $\sqrt{A^\dagger A} = U\,\mathrm{diag}(|\lambda_1|, \ldots, |\lambda_n|)U^\dagger$ as can be easily checked by squaring the equation. Then,

$$\|A\|_p^p = \mathrm{Tr}\left[(A^\dagger A)^{\frac{p}{2}}\right] = \mathrm{Tr}\left[U\Lambda^p U^\dagger\right] = \mathrm{Tr}\,\Lambda^p = \sum_i \lambda_i^p = \|\lambda\|_{\ell_p}^p. \tag{3}$$

With this characterisation we have also established that the Schatten p-norms are invariant under unitary transformations.

b) Give the statement and proof for the Hölder inequality for Schatten p-norms.

Solution: There are different matrix version of Hölder's inequality: Let $1 \geq p \geq \infty$ and q such that $\frac{1}{p} + \frac{1}{q} = 1$, then
Matrix Hölder I:

$$\|A^\dagger B\|_1 \leq \|A\|_p \|B\|_q. \tag{4}$$

Matrix Hölder II:

$$\left|\mathrm{Tr}\left[A^\dagger B\right]\right| \leq \|A\|_p \|B\|_q. \tag{5}$$

Just as a side remark, there is even the more general version that holds for every unitarily invariant norm $\|\cdot\|$.
Matrix Hölder III:

$$\|A^\dagger B\| \leq \|(A^\dagger)^p\|^{\frac{1}{p}} \|B^q\|^{\frac{1}{q}}. \tag{6}$$

The idea of the proof is to reduce the Matrix Hölder inequality to Hölder's inequality for ℓ_p-norms on vectors. To this end, we have to first establish the following inquality, the so-called von Neumann inequality. *Lemma "von Neumann-inequality":* Let A and B be two matrices and let $s(A)$ and $s(B)$ be the vector of singular values of A and B, respectively, ordered decreasingly. Then it holds that

$$|\operatorname{Tr}[AB]| \leq \operatorname{Tr}|AB| \leq \sum_i s_i(A)s_i(B). \tag{7}$$

The proof of this inequality is actually much more tricky than I have anticipated. If someone has a simpler or more elementary proof I am very keen to read it. (Seriously, please write me an email.) Some elegant ways to proof it is make use of the theory of majorisation and doubly stochastic matrices. To give you a flavor, here is a sketch of the argument following Bhatia's book on matrix analysis:

Proof sketch, the majorisation way: One can show that $s(AB)$ is weakly majorised[1] by the vector of the element-wise product $s(A)s(B)$, which is a stronger statement that implies the von Neumann inequality. For a matrix A, we denote by $\wedge^k A$ the linear map $A^{\otimes k}$ restricted to act only on vectors in $(\mathbb{C}^n)^{\otimes k}$ that are anti-symmetric under permuting the tensor factors. Using the min-max characterisation of singular values it is easy to see that $\| \wedge^k A\|_1 = \prod_{i=1}^k s_i$. Using the sub-multiplicativity of the spectral norm (see below) one concludes that $\prod_{i=1}^k s_i(AB) = \| \wedge^k AB\|_1 \leq \| \wedge^k A\| \| \wedge^k B\| = \prod_{i=1}^k s_i(A)s_i(B)$. A function $\Phi : \mathbb{C}^n \to \mathbb{C}^m$ that preserves weak majorisation, i.e. x is weakly majorised by y implies that $\Phi(x)$ is weakly majorised by $\Phi(y)$, is called strongly isotone. The function acting as the exponential function on each component of a vector can be shown to be strongly isotone. Thus, we have for the component-wise logarithm that $\log x$ weakly majorised by $\log y$ implies that x is weakliy majorised by y. Using the preservation of majorisation when dropping the logarithm allows to conclude the desired statement that $s(AB)$ is weakly majorised by $s(A)s(B)$ from the relate statement about the partial products.

Slightly more direct proofs using doubly stochastic matrices were worked out by Mirsky. A more elementary proof was given R. D. Grigorieff in a note in '92. You can find it on his webpage.

Proof of matrix Hölder I & II.

With the help of the von Neumann inequality, it is easy to reduce matrix Hölder to the standard Hölder inequality for vectors:

$$\operatorname{Tr}|AB| \leq |\langle s(A) \mid s(B)\rangle| \leq \|s(A)\|_{\ell_p}\|s(B)\|_{\ell_q} = \|A\|_p\|B\|_q. \tag{8}$$

The second version follows from the first version by showing that

$$\left|\operatorname{Tr}\left[A^\dagger B\right]\right| \leq \operatorname{Tr}\left|A^\dagger B\right|. \tag{9}$$

The most important Schatten p-norms have other interesting expressions:

c) Show that the Schatten 2-norm or Frobenius norm fulfils

$$\|A\|_2^2 = \sum_{i,j=1}^n |A_{ij}|^2. \tag{10}$$

[1] A n-dimensional vector x is weakly majorising a n-dimensional vector y if $\sum_{i=1}^k x_i \leq \sum_{i=1}^k y_i$ for all $k \in \{1, 2, \ldots, n\}$.

Solution:

$$\|A\|_2^2 = \operatorname{Tr} A^\dagger A = \sum_{i,j} \bar{A}_{ji} A_{ji} = \sum_{i,j} |A_{ij}|^2. \tag{11}$$

In general, one can define the operator norms induced by the ℓ_p-norms:

$$\|A\|_{\ell_p \to \ell_q} = \sup_{\|x\|_{\ell_p}=1} \|Ax\|_{\ell_q}. \tag{12}$$

d) What is the Schatten p-norm equal to $\|\cdot\|_{\ell_2\to\ell_2}$?

Solution:

$$\|A\|_{\ell_2\to\ell_2} = \sup_{\|x\|_{\ell_2}=1} \|Ax\|_{\ell_2} = \sup_{\|x\|_{\ell_2}=1} \sqrt{\langle x, A^\dagger A\, x\rangle} \tag{13}$$

$$= \sqrt{\lambda_{\mathsf{max}}(A^\dagger A)} = |\lambda_{\mathsf{max}}(A)|. \tag{14}$$

Another important properties of Schatten p-norms is *sub-multiplicativity*, $\|AB\|_p \leq \|A\|_p\|B\|_p$ for all p and $A, B \in \mathbb{C}^{n\times n}$. Sometimes the term *matrix norm* is exclusively used for sub-multiplicative norms on matrix spaces.

e) Show the sub-multiplicativity of the Schatten p-norms.

Solution: Using the min-max principle of the Rayleigh quotient, we first establish that $|\lambda_i(AB)| \leq \|A\|_\infty |\lambda_i(B)|$. Proof:

$$|\lambda_i(AB)| = \min_{U, \dim U = k} \max_{x\in U, \|x\|_{\ell_2}=1} |\langle x, ABx\rangle| \tag{15}$$

$$\leq \min_{U, \dim U = k} \max_{x\in U, \|x\|_{\ell_2}=1} \|A\|_\infty |\langle x, Bx\rangle| \tag{16}$$

$$= \|A\|_\infty |\lambda_i(B)|, \tag{17}$$

where we have used that $|\langle x, Ay\rangle| \leq \|A\|_\infty |\langle x, y\rangle|$, which follows from the operator norm definition of the spectral norm.

Now we have

$$\|AB\|_p = \left[\sum_i |\lambda_i(AB)|^p\right]^{\frac{1}{p}} \leq \|A\|_\infty \left[\sum_i |\lambda_i(B)|^p\right]^{\frac{1}{p}} \tag{18}$$

$$= \|A\|_\infty \|B\|_p \leq \|A\|_p \|B\|_p. \tag{19}$$

In the last step we have used the ordering of the p-norms inherited by the ordering of the ℓ_p-norms, in particular $\|A\|_\infty \leq \|A\|_p$ for all $p < \infty$.