Chapter 7

Norms and Distance Measures

7.1 Schatten Norm

We introduce the following family of "norms".

Definition 7.1. *For* $p \in \mathbb{R} \setminus \{0\}$ *and* $R \in \mathcal{L}(\mathcal{H})$*, we define the* **Schatten** p**-norm** *of* R *as*

$$\|R\|_p := \mathrm{tr}\big(|R|^p\big)^{\frac{1}{p}} .$$

This extends to

$$\|R\|_\infty := \lambda_{\max}(|R|) ,$$

where $\lambda_{\max}(|R|)$ *is the largest eigenvalue of* $|R|$*.*

In other words, the Schatten p-norm equals the standard p-norm of the **singular values** of R, i.e., the eigenvalues of $|R|$. In particular, for a normal or Hermitian $R \in \mathcal{L}(\mathcal{H})$ with spectrum $\{\lambda_1, \ldots, \lambda_d\}$, we have

$$\|R\|_p = \Big(\sum_i |\lambda_i|^p\Big)^{\frac{1}{p}} \quad \text{if } p < \infty, \text{ and} \quad \|R\|_\infty = \max\{|\lambda_1|, \ldots, |\lambda_d|\} .$$

The Schatten norm is a norm for $p \geq 1$, as we will see soon, and only for $p \geq 1$, but it will be convenient to have the definition at hand also for $0 < p < 1$, and even for (strictly) *negative* p; we refer to Section 0.4 for the definition of $|R|^p$, in particular in case of a negative p where $|R|^p$ is then defined by means of the pseudo-inverse.

The Schatten ∞-norm coincides with the **operator norm**, and $\|\cdot\|_2$ is the norm induced by the Hilbert-Schmidt/Frobenius inner product, and thus known as **Hilbert-Schmidt** or **Frobenius norm**. Most relevant for us is the case $p = 1$, in which case the norm is also called the **trace norm** and denoted as $\|\cdot\|_{tr}$.

The following of **Hölder inequality** (and its reverse) for the Schatten p-norm is a generalization of the corresponding inequality for the standard p-norm.

Theorem 7.1 (Hölder Inequality)**.** *Let* $L, R \in \mathcal{L}(\mathcal{H})$*, and let* $p, q \in \mathbb{R} \cup \{\infty\}$ *with the property that* $\frac{1}{p} + \frac{1}{q} = 1$*. If* $p, q \geq 1$ *then*

$$|\mathrm{tr}(LR)| \leq \|L\|_p \cdot \|R\|_q .$$

In reverse, if $0 < p < 1$ *and* $q < 0$*, and if* $L, R \geq 0$ *with* $\mathrm{supp}(L) \subseteq \mathrm{supp}(R)$*, then*

$$\mathrm{tr}(LR) \geq \|L\|_p \cdot \|R\|_q .$$

Proof. For simplicity, we restrict to normal L and R. Let $R = \sum_i \lambda_i |e_i\rangle\langle e_i|$ be the spectral decomposition of R. Then, by Hölder inequality for the (standard) p-norm, we have

$$|\mathrm{tr}(LR)| = \Big|\sum_i \lambda_i \langle e_i|L|e_i\rangle\Big| \leq \Big(\sum_i |\langle e_i|L|e_i\rangle|^p\Big)^{\frac{1}{p}} \Big(\sum_i |\lambda_i|^q\Big)^{\frac{1}{q}} = \Big(\sum_i |\langle e_i|L|e_i\rangle|^p\Big)^{\frac{1}{p}} \|R\|_q \,.$$

with obvious modifications if, say, $p = 1$ and $q = \infty$. If, furthermore, $L = \sum_j \mu_j |f_j\rangle\langle f_j|$ is the spectral decomposition of L, we see that

$$\sum_i |\langle e_i|L|e_i\rangle|^p \leq \sum_i \Big(\sum_j |\mu_j||\langle e_i|f_j\rangle|^2\Big)^p \leq \sum_i \sum_j |\langle e_i|f_j\rangle|^2 |\mu_j|^p = \sum_j |\mu_j|^q = \|L\|_p^p \,,$$

where the second inequality is Jensen's inequality (Proposition B.5), noting that raising to power $p \geq 1$ is a convex function, and that $\sum_j |\langle e_i|f_j\rangle|^2 = \sum_j \langle e_i|f_j\rangle\langle f_j|e_i\rangle = \langle e_i|e_i\rangle = 1$.

For the reverse part, since $\mathrm{supp}(L) \subseteq \mathrm{supp}(R)$, we may restrict $\mathcal{H}$ to $\mathrm{supp}(R)$. We first show the reverse Hölder inequality for the (standard) p-norm, i.e., for *commuting* L and R. Noting that $\frac{p-1}{p} = \frac{1}{q}$, this follows from

$$\|L\|_p^p = \mathrm{tr}(L^p R^p R^{-p}) \leq \|L^p R^p\|_{1/p} \, \|R^{-p}\|_{1/(1-p)} = \mathrm{tr}(LR)^p \, \|R\|_q^{-p} \,.$$

Then, for non-commuting L and R, with the same spectral decompositions as above — but now it is ensured that $\lambda_i > 0$ and $\mu_j \geq 0$ — we get

$$\mathrm{tr}(LR) = \sum_i \lambda_i \langle e_i|L|e_i\rangle \geq \Big(\sum_i \langle e_i|L|e_i\rangle^p\Big)^{\frac{1}{p}} \Big(\sum_i \lambda_i^q\Big)^{\frac{1}{q}} = \Big(\sum_i \langle e_i|L|e_i\rangle^p\Big)^{\frac{1}{p}} \|R^{-1}\|_{-q}^{-1} \,.$$

and

$$\sum_i \langle e_i|L|e_i\rangle^p = \sum_i \Big(\sum_j \mu_j |\langle e_i|f_j\rangle|^2\Big)^p \geq \sum_i \sum_j |\langle e_i|f_j\rangle|^2 \mu_j^p = \sum_j \mu_j^p = \|L\|_p^p \,,$$

now noting that raising to power $p < 1$ is a concave function. □

The following in particular implies that the p-Schatten norm satisfies the triangle inequality when $p \geq 1$, and so indeed is a norm in this range for p.

Proposition 7.2. *For any $L \in \mathcal{L}(\mathcal{H})$, and with $p, q \geq 1$ so that $\frac{1}{p} + \frac{1}{q} = 1$,*

$$\|L\|_p = \max_{\|R\|_q = 1} |\mathrm{tr}(LR)|$$

where the max *is over $R \in \mathcal{L}(\mathcal{H})$ with $\|R\|_q = 1$, but may be restricted to $R \geq 0$ if $L \geq 0$.*

We obtain a similar statement when $p < 1$ and $L > 0$, but then with a min instead of max.

Proof. Again, for simplicity, we assume L to be normal. Then, setting $R := \|L\|_p^{1-p} L^{p-1}$, we see that

$$\|R\|_q^q = \|L\|_p^{(1-p)q} \, \mathrm{tr}\big(L^{(p-1)q}\big) = \|L\|_p^{-p} \, \mathrm{tr}(L^p) = 1$$

and

$$\mathrm{tr}(LR) = \|L\|_p^{1-p} \, \mathrm{tr}\big(L^p\big) = \|L\|_p \,.$$

Thus, Hölder inequality becomes an equality, and the claim follows. □

Given that for $L, R > 0$: $|\mathrm{tr}(LR)| = |\mathrm{tr}(R^{1/2} L R^{1/2})| = \mathrm{tr}(R^{1/2} L R^{1/2})$, we obtain the following.

Corollary 7.3. *For any $p \neq 1$ and $0 \leq L \leq M \in \mathcal{L}(\mathcal{H})$, it holds that $\|L\|_p \leq \|M\|_p$.*

Applying the logarithm to Hölder's inequality and using the concavity of the logarithm to argue that $\frac{1}{p}\log\mathrm{tr}(|L|^p)+\frac{1}{q}\log\mathrm{tr}(|R|^q) \leq \log\bigl(\frac{1}{p}\mathrm{tr}(|L|^p)+\frac{1}{q}\mathrm{tr}(|R|^q)\bigr)$, we obtain the following "operator trace version" of **Young's inequality**.

Corollary 7.4 (Young's inequality)**.** *Let $p, q \geq 1$ with $\frac{1}{p}+\frac{1}{q}=1$. Then, for any $L, R \in \mathcal{L}(\mathcal{H})$:*

$$|\mathrm{tr}(LR)| \leq \frac{\mathrm{tr}(|L|^p)}{p}+\frac{\mathrm{tr}(|R|^q)}{q}\,,$$

with equality if $|L|^p = |R|^q$.

Again, we obtain a similar statement when $0 < p < 1$ and $L, R > 0$ with $\mathrm{supp}(L) \subseteq \mathrm{supp}(R)$, but with the inequality reversed.

7.2 Trace Distance

Since the density operator uniquely describes the behavior of a quantum system, two systems whose respective states are given by the *same* density operator behave in exactly the same way. Now we want to be able to say that if the density operators of two states are *close* then the states behave *similarly* and are hard to distinguish. For measuring the closeness of two density operators, the distance induced by the trace norm turns out the be the right choice.

Definition 7.2. *The* **trace distance** *of $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ is defined as $\delta(\rho,\sigma) := \frac{1}{2}\|\rho-\sigma\|_{tr}$.*[1]

In case of *classical* density operators, the trace distance coincides with the statistical distance (see Definition A.1) of the two distributions: if $\rho_X = \sum_x P_X(x)|x\rangle\langle x|$ and $\rho_Y = \sum_x P_Y(x)|x\rangle\langle x|$ in $\mathcal{D}(\mathcal{X})$ are the respective density operator representations of random variables X and Y, then $\delta(\rho_X,\rho_Y) = \delta(P_X,P_Y)$.

The following, in combination with Lemma A.1 or Corollary A.2, implies that two states that are close are hard to distinguish. In other words, if $\delta(\rho,\sigma)$ is small then the two states behave very much the same way, i.e., they are hard to distinguish.

Theorem 7.5. *Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, and let $\mathbf{M} = \{M_i\}_i \in \mathcal{M}eas_I(\mathcal{H})$ be a measurement. Also, set $P(i) = \mathrm{tr}(M_i\rho M_i^\dagger)$ and $Q(i) = \mathrm{tr}(M_i\sigma M_i^\dagger)$ for every $i \in I$. Then:*

$$\delta(P,Q) \leq \delta(\rho,\sigma)\,.$$

Proof. Note that

$$\delta(P,Q) = \frac{1}{2}\sum_i \bigl|P(i)-Q(i)\bigr| = \frac{1}{2}\sum_i \bigl|\mathrm{tr}\bigl(M_i\rho M_i^\dagger\bigr)-\mathrm{tr}\bigl(M_i\sigma M_i^\dagger\bigr)\bigr| = \frac{1}{2}\sum_i \bigl|\mathrm{tr}\bigl(M_i(\rho-\sigma)M_i^\dagger\bigr)\bigr|\,.$$

By considering the spectral decomposition $\rho - \sigma = \sum_j \lambda_j |j\rangle\langle j|$, it follows that the above equals

$$= \frac{1}{2}\sum_i \Bigl|\sum_j \lambda_j\,\mathrm{tr}\bigl(M_i|j\rangle\langle j|M_i^\dagger\bigr)\Bigr| \leq \frac{1}{2}\sum_i\sum_j |\lambda_j|\bigl|\mathrm{tr}\bigl(M_i|j\rangle\langle j|M_i^\dagger\bigr)\bigr| = \frac{1}{2}\sum_{i,j}|\lambda_j||\langle j|M_i^\dagger M_i|j\rangle|$$
$$= \frac{1}{2}\sum_{i,j}|\lambda_j|\langle j|M_i^\dagger M_i|j\rangle = \frac{1}{2}\sum_j|\lambda_j|\langle j|\mathbb{I}|j\rangle = \frac{1}{2}\sum_j|\lambda_j| = \delta(\rho,\sigma)\,,$$

which shows the claimed inequality. □

[1]The factor $\frac{1}{2}$ is for normalization purposes: it ensures that $\delta(\rho,\sigma) \leq 1$.

This is actually a special case of the following result, which states that the distance between two states can only decrease when manipulating the states.

Theorem 7.6. *For any CPTP map* $\mathfrak{T} \in \mathcal{L}\big(\mathcal{L}(A), \mathcal{L}(A')\big)$ *and for all* $\sigma, \rho \in \mathcal{D}(A)$

$$\delta\big(\mathfrak{T}(\rho), \mathfrak{T}(\sigma)\big) \leq \delta(\rho, \sigma) \, .$$

Proof. It is sufficient to show that $\delta(\rho \otimes \tau, \sigma \otimes \tau) = \delta(\rho, \sigma)$ for all $\rho, \sigma \in \mathcal{D}(A)$ and $\tau \in \mathcal{D}(B)$, and that $\delta(\rho, \sigma) \geq \delta\big(\mathrm{tr}_A(\rho), \mathrm{tr}_A(\sigma)\big)$ for all $\rho, \sigma \in \mathcal{D}(AB)$; in combination with the Stinespring representation (Theorem 6.6) and the obvious invariance of δ under unitary transformations, the claim then follows. The first claim is easy to verify, and we leave it as an exercise. For the second claim, let us consider the spectral decomposition $\rho_{AB} - \sigma_{AB} = \sum_i \lambda_i |i\rangle\langle i|$ of $\rho_{AB} - \sigma_{AB}$. Then, we see that

$$\begin{aligned} \delta(\rho_A, \sigma_A) &= \frac{1}{2}\big\|\mathrm{tr}_B(\rho_{AB}) - \mathrm{tr}_B(\sigma_{AB})\big\|_{tr} = \frac{1}{2}\big\|\mathrm{tr}_B(\rho_{AB} - \sigma_{AB})\big\|_{tr} \\ &\leq \frac{1}{2}\sum_i |\lambda_i| \, \big\|\mathrm{tr}_B(|i\rangle\langle i|)\big\|_{tr} = \frac{1}{2}\sum_i |\lambda_i| = \delta(\rho_{AB}, \sigma_{AB}) \end{aligned}$$

where we used triangle inequality and the fact that $\mathrm{tr}_B(|i\rangle\langle i|)$ is a density operator and thus has trace norm 1. □

We conclude the section with a couple of useful results regarding the trace distance. In case of *pure* states, the trace distance is determined by their inner product:

Lemma 7.7. *For* $|\varphi\rangle, |\psi\rangle \in \mathcal{S}(\mathcal{H})$*:* $\delta\big(|\varphi\rangle\langle\varphi|, |\psi\rangle\langle\psi|\big) = \sqrt{1 - |\langle\varphi|\psi\rangle|^2}$.

$|\langle\varphi|\psi\rangle|$ is referred to as the **fidelity** of the states $|\varphi\rangle$ and $|\psi\rangle$. For general mixed states ρ and σ, the fidelity is given by $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_{tr}$, and for classical states it coincides with the so-called **Bhattacharyya coefficient**.[2]

Proof. We can choose an orthonormal basis $\{|0\rangle, |1\rangle, \cdots, |d-1\rangle\}$ of $\mathcal{H}$ with $|\varphi\rangle = \omega|0\rangle$ and $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ such that α_0 and α_1 are real (and $|\omega| = 1$ and $|\alpha_0|^2 + |\alpha_1|^2 = \alpha_0^2 + \alpha_1^2 = 1$). As both sides of the equation to be proven are invariant under multiplying $|\varphi\rangle$ with $\bar{\omega}$, we may assume without loss of generality that $\omega = 1$. It follows that $1 - |\langle\varphi|\psi\rangle|^2 = 1 - \alpha_0^2 = \alpha_1^2$ and thus $\sqrt{1 - |\langle\varphi|\psi\rangle|^2} = |\alpha_1|$. On the other hand, when expressing $|\varphi\rangle\langle\varphi| - |\psi\rangle\langle\psi|$ in this basis we see that

$$\delta\big(|\varphi\rangle\langle\varphi|, |\psi\rangle\langle\psi|\big) = \frac{1}{2}\mathrm{tr}\big(\big||\varphi\rangle\langle\varphi| - |\psi\rangle\langle\psi|\big|\big) = \frac{1}{2}\mathrm{tr}\left(\begin{vmatrix} 1-\alpha_0^2 & -\alpha_0\alpha_1 \\ -\alpha_0\alpha_1 & -\alpha_1^2 \end{vmatrix}\right) = \frac{1}{2}\mathrm{tr}\left(\begin{vmatrix} \alpha_1^2 & -\alpha_0\alpha_1 \\ -\alpha_0\alpha_1 & -\alpha_1^2 \end{vmatrix}\right)$$

where the matrix in the right-hand-side expression has eigenvalues $\pm\alpha_1$ (which can easily be seen by computing its characteristic polynomial). It follows that $\delta\big(|\varphi\rangle\langle\varphi|, |\psi\rangle\langle\psi|\big) = |\alpha_1|$. □

Finally, in case of two hybrid states with *the same* classical part, the trace distance coincides with the expectation over the classical part; the proof is given as an exercise.

Lemma 7.8. *Let* $\rho_{XE}, \tilde{\rho}_{XE} \in \mathcal{D}(\mathcal{X} \otimes \mathcal{H}_E)$ *be hybrid states with classical* X *(w.r.t. the same basis) and such that* $\rho_X = \tilde{\rho}_X$ *(i.e.,* $P_X = P'_X$*). Then,*

$$\delta\big(\rho_{XE}, \tilde{\rho}_{XE}\big) = \sum_x P_X(x) \, \delta\big(\rho_{E|X=x}, \tilde{\rho}_{E|X=x}\big) \, .$$

[2]In some literature, the fidelity is defined as the *square* of the above, i.e., as $\|\sqrt{\rho}\sqrt{\sigma}\|_{tr}^2$ respectively $|\langle\varphi|\psi\rangle|^2$.

7.3 The Gentle-Measurement Lemma

Performing a measurement on a state *disturbs* the state in general. For instance, measuring the qubit $|0\rangle$ in the Hadamard basis results in the post-measurement state $|+\rangle$ or $|-\rangle$. The exception is when we measure the state in a basis that *contains the state*. Like, measuring $|0\rangle$ in the computational basis results gives measurement outcome 0 with certainty, and the post-measurement state is still $|0\rangle$.

This suggests the following: if a measurement is such that one particular outcome occurs with probability close to 1 then the corresponding post-measurement state must be close to the original. The following shows that this intuition is indeed true, for arbitrary projective measurements. It is *not* true for general measurements because a general measurement may "twist" the state before or after the measurement. For example, $\{U|0\rangle\langle 0|, U|1\rangle\langle 1|\}$ is a measurement for any $U \in \mathcal{U}(\mathbb{C}^2)$, as can easily be verified, and has a definite outcome when applied to the state $|0\rangle$, but the corresponding post-measurement state is $U|0\rangle$.

Proposition 7.9. *Let $\rho \in \mathcal{D}(\mathcal{H})$, and let $\Pi_0 \in \mathcal{L}(\mathcal{H})$ be a projection (which we think of being part of a projective measurement). We set $p_0 = \mathrm{tr}(\Pi_0 \rho)$ and $\rho^0 = \frac{1}{p_0}\Pi_0 \rho \Pi_0$. Then*

$$\delta(\rho, \rho^0) \le \sqrt{1 - p_0}\,,$$

with equality if ρ is pure.

Proof. First, consider a pure $\rho = |\varphi\rangle\langle\varphi|$. We then have that $p_0 = \langle\varphi|\Pi_0|\varphi\rangle$ and $\rho^0 = |\varphi^0\rangle\langle\varphi^0|$ with $|\varphi^0\rangle = \frac{1}{\sqrt{p_0}}\Pi_0|\varphi\rangle$, and thus, by Lemma 7.7,

$$\delta\big(|\varphi\rangle\langle\varphi|, |\varphi^0\rangle\langle\varphi^0|\big) = \sqrt{1 - |\langle\varphi|\varphi^0\rangle|^2} = \sqrt{1 - \tfrac{1}{p_0}|\langle\varphi|\Pi_0|\varphi\rangle|^2} = \sqrt{1 - p_0}\,.$$

For a mixed state $\rho = \sum_\ell \varepsilon_\ell |\varphi_\ell\rangle\langle\varphi_\ell|$, we can consider its purification $|\varphi\rangle = \sum_\ell \sqrt{\varepsilon_\ell}|\varphi_\ell\rangle|\ell\rangle$. We then have that $p_0 = \langle\varphi|(\Pi_0 \otimes \mathbb{I})|\varphi\rangle$, and a straightforward calculation then shows that $|\varphi^0\rangle = \frac{1}{\sqrt{p_0}}(\Pi_0 \otimes \mathbb{I})|\varphi\rangle$ is a purification of $\rho^0 = \frac{1}{p_0}\Pi_0\rho\Pi_0$, and therefore we have

$$\delta(\rho, \rho^0) \le \delta\big(|\varphi\rangle\langle\varphi|, |\varphi^0\rangle\langle\varphi^0|\big) = \sqrt{1 - p_0}\,,$$

which was to be proven. □