Quantum information theory (MAT4430) **Spring 2021**

Lecture 7: Measuring distances between quantum states

Lecturer: Alexander Müller-Hermes

To analyze protocols in quantum information theory, we need to introduce some quantities measuring the distance between quantum states and quantum channels. To measure the distance between quantum states we will often use the *trace distance* based on the Schatten 1-norm and the *fidelity*. Another distance measure is the *quantum relative entropy*, which we will introduce later.

1 Positivity of block operators and operator inequalities

Recall, that we discussed the following lemma in the exercises:

Lemma 1.1 (Positivity of a block matrix)**.** *Let $\mathcal{H}$ denote a complex Euclidean space. For $X \in B(\mathcal{H})$, we have*

$$\begin{pmatrix} \mathbb{1}_{\mathcal{H}} & X^\dagger \\ X & \mathbb{1}_{\mathcal{H}} \end{pmatrix} \in B(\mathcal{H} \oplus \mathcal{H})^+,$$

if and only if $\|X\|_\infty \leqslant 1$.

Proof. By the singular value decomposition there are unitaries $U, V \in \mathcal{U}(\mathcal{H})$ and an operator $S \in B(\mathcal{H})$, that is diagonal in the computational basis with positive diagonal entries, such that $X = USV$. We can then verify that

$$\begin{pmatrix} \mathbb{1}_{\mathcal{H}} & X^\dagger \\ X & \mathbb{1}_{\mathcal{H}} \end{pmatrix} = \begin{pmatrix} V^\dagger & 0 \\ 0 & U \end{pmatrix} \begin{pmatrix} \mathbb{1}_{\mathcal{H}} & S \\ S & \mathbb{1}_{\mathcal{H}} \end{pmatrix} \begin{pmatrix} V & 0 \\ 0 & U^\dagger \end{pmatrix}.$$

This operator is positive if and only if

$$\begin{pmatrix} \mathbb{1}_{\mathcal{H}} & S \\ S & \mathbb{1}_{\mathcal{H}} \end{pmatrix} \in B(\mathcal{H} \oplus \mathcal{H})^+.$$

Finally, we note that there exists a unitary $W \in \mathcal{U}(\mathcal{H} \oplus \mathcal{H})$ such that

$$W \begin{pmatrix} \mathbb{1}_{\mathcal{H}} & S \\ S & \mathbb{1}_{\mathcal{H}} \end{pmatrix} W^\dagger = \begin{pmatrix} 1 & s_1 \\ s_1 & 1 \end{pmatrix} \oplus \begin{pmatrix} 1 & s_2 \\ s_2 & 1 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 1 & s_{\text{dim}(\mathcal{H})} \\ s_{\text{dim}(\mathcal{H})} & 1 \end{pmatrix},$$

where $s_i \in \mathbb{R}^+$ are the diagonal entries of S. This operator is positive if and only if $\max_i s_i = \|X\|_\infty \leqslant 1$, and the proof is finished. □

A consequence of this lemma is the following lemma:

Lemma 1.2 (Positivity of a block matrix)**.** *Let $\mathcal{H}$ denote a complex Euclidean space. For $X, Y \in B(\mathcal{H})^+$ and $Z \in B(H)$, we have*

$$\begin{pmatrix} X & Z^\dagger \\ Z & Y \end{pmatrix} \in B(\mathcal{H} \oplus \mathcal{H})^+,$$

if and only if there exists a $K \in B(H)$ satisfying $\|K\|_\infty \leqslant 1$ and $Z = Y^{\frac{1}{2}} K X^{\frac{1}{2}}$.

Proof. Note that the cone $B(\mathcal{H}\oplus\mathcal{H})^+$ is closed, and therefore we have

$$\begin{pmatrix} X & Z^\dagger \\ Z & Y \end{pmatrix} \in B(\mathcal{H}\oplus\mathcal{H})^+,$$

if and only if

$$\begin{pmatrix} X+\epsilon \mathbb{1}_{\mathbb{C}} & Z^\dagger \\ Z & Y+\epsilon \mathbb{1}_{\mathbb{C}} \end{pmatrix} \in B(\mathcal{H}\oplus\mathcal{H})^+,$$

for all $\epsilon > 0$. Thus, it is sufficient to consider the case where X and Y are invertible. In this case, we have

$$\begin{pmatrix} X^{-\frac{1}{2}} & 0 \\ 0 & Y^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} X & Z^\dagger \\ Z & Y \end{pmatrix} \begin{pmatrix} X^{-\frac{1}{2}} & 0 \\ 0 & Y^{-\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \mathbb{1}_{\mathcal{H}} & X^{-\frac{1}{2}} Z^\dagger Y^{-\frac{1}{2}} \\ Y^{-\frac{1}{2}} Z X^{-\frac{1}{2}} & \mathbb{1}_{\mathcal{H}} \end{pmatrix}.$$

By Lemma 1.1, this matrix is positive if and only if $K = Y^{-\frac{1}{2}} Z X^{-\frac{1}{2}}$ satisfies $\|K\|_\infty \leqslant 1$. Since the completely positive map Ad_M for

$$M = \begin{pmatrix} X^{-\frac{1}{2}} & 0 \\ 0 & Y^{-\frac{1}{2}} \end{pmatrix},$$

is invertible with completely positive (and in particular positive) inverse, the statement of the lemma follows. □

We will also need the following operator inequality, which we will prove in the exercises:

Theorem 1.3 (Choi's inequality)**.** *For any positive and invertible operator $X \in B(\mathbb{C}^{d_A})^+$ and any positive and unital map $P : B\left(\mathbb{C}^{d_A}\right) \to B\left(\mathbb{C}^{d_B}\right)$ we have*

$$P(X)^{-1} \leqslant P(X^{-1}),$$

where we used the Moore-Penrose pseudoinverse.

2 Schatten p-norms

The most important norms in quantum information theory are the Schatten norms, i.e., the non-commutative analogues of the l_p-norms. In particular, the Schatten 1-norm, also known as the *trace norm*, and the Schatten ∞-norm, which you know as the *operator norm*, are ubiquitous.

2.1 Basic definition and properties

For a complex Euclidean space $\mathcal{H}$, an operator $X \in B(\mathcal{H})$ and $p \in [1,\infty)$ the *Schatten p-norm* is given by

$$\|X\|_p = \mathrm{Tr}\left[|X|^p\right]^{1/p},$$

where $|X| = \sqrt{X^*X}$ denotes the unique positive square root of the positive operator X^*X. It can be shown that

$$\|X\|_\infty := \lim_{p\to\infty} \|X\|_p,$$

coincides with the usual operator norm $\|X\|$. For $p = 2$, we recover the Hilbert-Schmidt norm $\|X\|_2 = \|X\|_{HS}$ induced by the Hilbert-Schmidt inner product on $B(\mathcal{H})$. A useful alternative expression of the Schatten norms is given in terms of the singular values $s_1(X), \dots, s_R(X)$ of the operator $X \in B(\mathcal{H})$, and we have

$$\|X\|_p = \left(\sum_{i=1}^{R} s_i(X)^p\right)^{\frac{1}{p}} ,$$

which coincides with the l_p-norms of the vector of singular values. The Schatten norms behave mostly like the l_p-norms, and you might want to verify the following standard facts:

- We have the ordering $\|\cdot\|_\infty \leqslant \|\cdot\|_p \leqslant \|\cdot\|_1$ for any $p \in [1,\infty]$.
- The norms $\|\cdot\|_p$ and $\|\cdot\|_q$ for $1/p + 1/q = 1$ are dual norms with respect to the Hilbert-Schmidt inner product.
- We have the Hölder's inequality

  $$\|XY\|_1 \leqslant \|X\|_p \|Y\|_q,$$

  for any operators $X, Y \in B(\mathcal{H})$ and any $p, q \in [1,\infty)$ satisfying $1/p + 1/q = 1$.

2.2 Fun fact: Extreme points of the unit balls B_∞ and B_1

We will start with the following easy theorem:

Theorem 2.1. *Let $\mathcal{H}$ denote a complex Euclidean space and consider the unit ball*

$$B_1 = \{X \in B(\mathcal{H}) \ : \ \|X\|_1 \leqslant 1\}.$$

Then, we have

$$\text{Ext}\,(B_1) = \{|v\rangle\langle w| \in B(\mathcal{H}) \ : \ \langle v|v\rangle = \langle w|w\rangle = 1\}.$$

Proof. By the singular value decomposition any operator $X \in B_1$ can be written as

$$X = \sum_{i=1}^{\dim(\mathcal{H})} s_i |v_i\rangle\langle w_i|,$$

such that $s_i \geqslant 0$ for all $i \in \{1, \dots, \dim(\mathcal{H})\}$ and $\sum_{i=1}^{\dim(\mathcal{H})} s_i = 1$. This shows that the extreme points of B_1 are a subset of the set of rank-1 operators $|v\rangle\langle w| \in B(\mathcal{H})$ satisfying $\langle v|v\rangle = \langle w|w\rangle = 1$. Finally, it is clear that any such rank-1 operators is extremal since

$$|v\rangle\langle w| = (1-\lambda)X_1 + \lambda X_2$$

for $X_1, X_2 \in B_1$ and $\lambda \in (0,1)$ implies that $X_1 = X_2 = |v\rangle\langle w|$. □

The next theorem determines the extreme points of the $\|\cdot\|_\infty$-unit ball.

Theorem 2.2. *Let $\mathcal{H}$ denote a complex Euclidean space and consider the unit ball*

$$B_\infty = \{X \in B(\mathcal{H}) \ : \ \|X\|_\infty \leqslant 1\}.$$

Then, we have

$$\text{Ext}\,(B_\infty) = \mathcal{U}(\mathcal{H}),$$

i.e., the extreme points are the unitary operators $U \in \mathcal{U}(\mathcal{H})$.

Proof. Consider a $X \in B(\mathcal{H})$ satisfying $\|X\|_\infty \leqslant 1$. By the singular value decomposition, we have $X = UDV$ for unitary operators $U, V \in \mathcal{U}(\mathcal{H})$ and an operator $D \in B(\mathcal{H})$ diagonal in the computational basis such that $D_{ii} = s_i \in [0,1]$ for each $i \in \{1, \dots, \dim(\mathcal{H})\}$. Observe, that every $s \in [0,1]$ can be written as

$$s = \frac{1}{2}\left(e^{it} + e^{-it}\right)$$

for some $t \in \mathbb{R}$. Using this decomposition for all singular values shows that

$$X = UDV = \frac{1}{2}U\left(D_1 + D_2\right)V = \frac{1}{2}\left(U_1 + U_2\right),$$

for unitary operators $D_1, D_2 \in \mathcal{U}(\mathcal{H})$ diagonal in the computational basis, and unitary operators $U_1, U_2 \in \mathcal{U}(\mathcal{H})$ obtained as $U_1 = UD_1V$ and $U_2 = UD_2V$. We conclude that each contraction can be written as a convex combination of two unitary operators!

Clearly, $\|U\|_\infty = 1$ for any unitary operator $U \in \mathcal{U}(\mathcal{H})$. Consider now a unitary operator $U \in \mathcal{U}(\mathcal{H})$, and assume that there are $X_1, X_2 \in B_\infty \setminus \{0\}$ satisfying

$$U = pX_1 + (1-p)X_2, \tag{1}$$

for some $p \in (0,1)$. By the singular value decomposition, we have $X_1 = VSW$ for $V, W \in \mathcal{U}(\mathcal{H})$ and an operator $S \in B_\infty$ diagonal in the computational basis and containing the singular values $s_1, \dots, s_d$ of X_1 on its diagonal. We define $U' = V^\dagger U W^\dagger \in \mathcal{U}(\mathcal{H})$ and $X_2' = V^\dagger X_2 W^\dagger \in B_\infty$. Then, we have

$$U' = pD + (1-p)X_2'.$$

Consider now a normalized eigenvector $|v\rangle \in \mathcal{H}$ of the unitary operator U'. We find that

$$1 = |\langle v|U'|v\rangle| \leqslant p|\langle v|D|v\rangle| + (1-p)|\langle v|X'|v\rangle| \leqslant p\sum_{i=1}^{d} s_i|v_i|^2 + (1-p) \leqslant 1,$$

where $d = \dim(\mathcal{H})$ and $v_i \in \mathbb{C}$ are the entries of $|v\rangle$ in the computational basis. This implies that

$$\sum_{i=1}^{d} s_i|v_i|^2 = \sum_{i=1}^{d} |v_i|^2 = 1,$$

and therefore $s_i = 1$ for each $i \in \{1, \dots, d\}$. By repeating the same argument as above, we conclude that X_1 and X_2 in (1) are unitary operators. In this case, we can compute

$$\mathbb{1}_{\mathcal{H}} = U^\dagger U = p^2\mathbb{1}_{\mathcal{H}} + (1-p)^2\mathbb{1}_{cH} + p(1-p)\left(X_1^\dagger X_2 + X_2^\dagger X_1\right),$$

and using that $p(1-p) \neq 0$ we find that

$$\frac{1}{2}\left(X_1^\dagger X_2 + X_2^\dagger X_1\right) = \mathbb{1}_{\mathcal{H}}.$$

But an operator of the form $2\mathbb{1}_{\mathcal{H}} - W$ for a unitary $W \in \mathcal{U}(\mathcal{H})$ is only itself a unitary operator if $W = \mathbb{1}_{\mathcal{H}}$, since the eigenvalues of $2\mathbb{1}_{\mathcal{H}} - W$ have modulus strictly larger than 1 when $W \in \mathcal{U}(\mathcal{H}) \setminus \{\mathbb{1}_{cH}\}$. We conclude that $X_1^\dagger X_2 = \mathbb{1}_{\mathcal{H}}$ and thus $X_1 = X_2 = U$. This means that U is an extreme point of B_∞. □

2.3 Induced norms and the Russo-Dye Theorem

Inspired by the operator norm, we can use the Schatten p-norms to define norms on the space of linear maps $L : B(\mathcal{H}_A) \rightarrow B(\mathcal{H}_B)$. We will write

$$\|L\|_{\alpha\rightarrow\beta} = \sup_{x\in B(\mathcal{H}_A)} \frac{\|L(x)\|_\beta}{\|x\|_\alpha},$$

for any linear map $L : B(\mathcal{H}_A) \rightarrow B(\mathcal{H}_B)$ and $\alpha, \beta \in [1,\infty]$. These norms have many nice properties which they inherent from the Schatten p-norms, or get by specializing general properties of operator norms. For example we have the following properties, which you might want to prove yourself:

- For any linear map $L : B(\mathcal{H}_A) \rightarrow B(\mathcal{H}_B)$ and all $\alpha, \beta \in [1,\infty]$ we have
$$\|L\|_{\alpha\rightarrow\beta} = \|L^*\|_{\beta'\rightarrow\alpha'},$$
where $L^* : B(\mathcal{H}_B) \rightarrow B(\mathcal{H}_A)$ is the adjoint operator with respect to the Hilbert-Schmidt inner product, and $1/\alpha + 1/\alpha' = 1 = 1/\beta + 1/\beta'$.
- For linear maps $L_1 : B(\mathcal{H}_A) \rightarrow B(\mathcal{H}_B)$ and $L_2 : B(\mathcal{H}_B) \rightarrow B(\mathcal{H}_C)$ we have
$$\|L_2 \circ L_1\|_{\alpha\rightarrow\gamma} \leqslant \|L_2\|_{\beta\rightarrow\gamma}\|L_1\|_{\alpha\rightarrow\beta},$$
for any $\alpha, \beta, \gamma \in [1,\infty]$.

In quantum information theory, we will mostly use the two special cases of $\alpha = \beta = 1$ and $\alpha = \beta = \infty$. These norms are closely related to the trace norm and the operator norm, and they behave particularly nicely when applied to positive maps:

Theorem 2.3 (Russo-Dye)**.** *For any positive map $P : B(\mathcal{H}_A) \rightarrow B(\mathcal{H}_B)$, we have*

$$\|P\|_{1\rightarrow 1} = \|P^*\|_{\infty\rightarrow\infty} = \|P^*(\mathbb{1}_{\mathcal{H}_B})\|_\infty .$$

In particular, $\|P\|_{1\rightarrow 1} = 1$ if P is positive and trace-preserving.

Proof of Theorem 2.3. Using the duality of the norms $\|\cdot\|_{1\rightarrow 1}$ and $\|\cdot\|_{\infty\rightarrow\infty}$, it will be enough to show that

$$\|P\|_{\infty\rightarrow\infty} := \sup_{X\in B_\infty} \|P(X)\|_\infty = \|P(\mathbb{1}_{\mathcal{H}_A})\|_\infty .$$

for any positive map $P : B(\mathcal{H}_A) \rightarrow B(\mathcal{H}_B)$. Since B_∞ is convex and compact, the supremum is attained in an extreme point of B_∞, and by Theorem 2.2 we know that these are the unitary operators. Therefore, we have $\|P\|_{\infty\rightarrow\infty} = \|P(U)\|_\infty$ for some unitary operator $U \in \mathcal{U}(\mathcal{H}_A)$. Note that

$$M_U = \begin{pmatrix} \mathbb{1}_{\mathcal{H}_A} & U^\dagger \\ U & \mathbb{1}_{\mathcal{H}_A} \end{pmatrix} = \sum_{i=1}^{d_A} \begin{pmatrix} 1 & \overline{\lambda_i} \\ \lambda_i & 1 \end{pmatrix} \otimes |v_i\rangle\langle v_i|,$$

where $U = \sum_{i=1}^{d} \lambda_i |v_i\rangle\langle v_i|$ is the spectral decomposition of U. Since $|\lambda_i| \leqslant 1$, we find that M_U is separable, and that

$$(\mathrm{id}_2 \otimes P)(M_U) = \begin{pmatrix} P(\mathbb{1}_{\mathcal{H}_A}) & P(U^\dagger) \\ P(U) & P(\mathbb{1}_{\mathcal{H}_A}) \end{pmatrix} \geqslant 0.$$

By Lemma 1.2, we have

$$P(U) = P(\mathbb{1}_{\mathcal{H}_A})^{\frac{1}{2}} K P(\mathbb{1}_{\mathcal{H}_A})^{\frac{1}{2}},$$

for some $K \in B(\mathcal{H}_A)$ satisfying $\|K\|_\infty \leqslant 1$. We conclude that

$$\|P(U)\|_\infty \leqslant \|P(\mathbb{1}_{\mathcal{H}_A})^{\frac{1}{2}}\|_\infty^2 \|K\|_\infty \leqslant \|P(\mathbb{1}_{\mathcal{H}_A})\|_\infty,$$

since $\|X^{\frac{1}{2}}\|_\infty = \|X\|_\infty^{\frac{1}{2}}$ for any positive operator X.

□

The Russo-Dye theorem is sometimes stated in an alternative form, which we point out for completeness:

Corollary 2.4. *For any positive map $P : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$, we have*

$$\|P\|_{1\to 1} = \max\{\|P(|v\rangle\langle v|)\|_1 \ : \ |v\rangle \in \mathcal{H}_A, \langle v|v\rangle = 1\}.$$

Proof. Since P is positive, we have

$$\|P^*(\mathbb{1}_{\mathcal{H}_B})\|_\infty = \max\{\langle v|P^*(\mathbb{1}_{\mathcal{H}_B})|v\rangle \ : \ |v\rangle \in \mathcal{H}_A, \langle v|v\rangle = 1\}.$$

For any $|v\rangle \in \mathcal{H}_A$, we have

$$\langle v|P^*(\mathbb{1}_{\mathcal{H}_B})|v\rangle = \mathrm{Tr}\left[P(|v\rangle\langle v|)\right] = \|P(|v\rangle\langle v|)\|_1,$$

since $P(|v\rangle\langle v|) \in B(\mathcal{H}_B)^+$. □

Another important corollary of the Russo-Dye theorem is the following:

Corollary 2.5. *The trace norm is contractive under quantum channels, i.e., we have*

$$\|T(X)\|_1 \leqslant \|X\|_1$$

for every quantum channel $T : B(\mathcal{H}_A) \to B(\mathcal{H}_B)$ and any $X \in B(\mathcal{H}_A)$.

3 The trace distance

When using the trace norm to measure the distance between quantum states $\rho, \sigma \in D(\mathcal{H})$, we will often speak about the *trace distance* $\|\rho - \sigma\|_1$. Note that $\|\rho - \sigma\|_1 \leqslant 2$ with equality if and only if $\mathrm{supp}(\rho) \perp \mathrm{supp}(\sigma)$. Some authors define the trace distance as the norm distance with a factor $1/2$ such that the maximum distance of two quantum states is 1 rather than 2. We will occasionally do so as well. The following lemma contains two useful properties of the trace norm, which we prove in the exercises:

Lemma 3.1 (Some properties of the trace norms). *Consider an operator $X \in B(\mathcal{H})$ and normalized vectors $|\psi\rangle, |\phi\rangle \in \mathcal{H}$. We have:*

1. $\|X\|_1 = \sup\{|\langle U, X\rangle_{HS}| \ : \ U \in \mathcal{U}(\mathcal{H})\}$.
2. $\|a|\psi\rangle\langle\psi| - b|\phi\rangle\langle\phi|\|_1 = \sqrt{(a+b)^2 - 4ab|\langle\psi|\phi\rangle|}$ *for any $a, b \in \mathbb{R}^+$.*

3.1 Operational interpretation: Quantum state discrimination

The trace distance between quantum states is the quantum analogue of the statistical distance between probability distributions $p, q \in \mathcal{P}(\Sigma)$ given by

$$\|p - q\|_1 = \sum_{x\in\Sigma} |p(x) - q(x)|.$$

As the statistical distance quantifies the maximum probability of discriminating between p and q in a one-shot setting, it is not surprising that the trace-distance can be interpreted in the same way. Consider the following scenario:

Scenario: State discrimination . Two researchers Alice and Bob are given the following task visualized in Figure 1. Alice has a device with two buttons labeled “0” and “1”. After pressing the button “0” the device emits a particle in quantum state $\rho_0 \in D(\mathcal{H})$, and after pressing the button “1” the device emits a particle in quantum state $\rho_1 \in D(\mathcal{H})$. Bob catches the emitted particle and measures it using some POVM. Then, he tries to guess whether Alice pressed button 0 or button 1. Assume that Alice presses button “0” with probability $\lambda \in [0,1]$ and button “1” with probability $1-\lambda$, then, what is the optimal success probability of Bob’s guess?

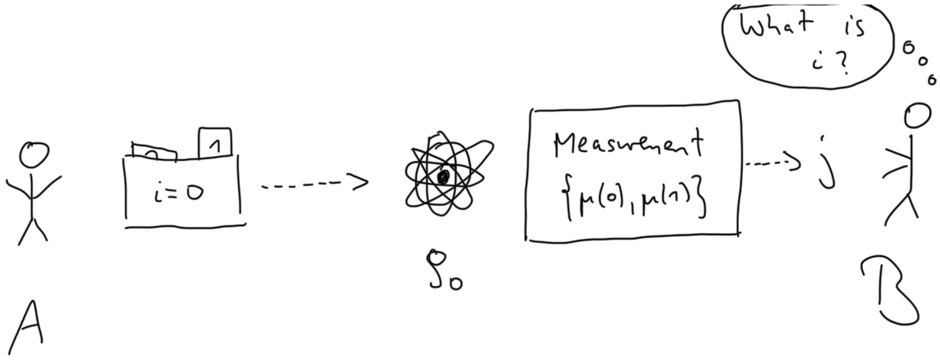

Figure 1: Alice and Bob at work.

Let us consider the case of some POVM $\mu : \{0,1\} \to B(\mathcal{H})^+$, which Bob could measure, i.e., the operators $\mu(0)$ and $\mu(1)$ are positive semidefinite and add up to $\mathbb{1}_{\mathcal{H}}$. Furthermore, we assume (without loss of generality) that the outcomes “0” and “1” of this measurement determine exactly whether he guesses that Alice pressed the corresponding button. Using the POVM formalism, we compute the success probability as

$$p_{\text{succ}}(\mu) = \lambda \langle \mu(0), \rho_0 \rangle_{HS} + (1-\lambda)\langle \mu(1), \rho_1 \rangle_{HS}.$$

How large can this probability be? The following theorem gives an upper bound and shows how to achieve it:

Theorem 3.2 (Holevo-Helstrom)**.** *Let $\mathcal{H}$ denote a complex Euclidean space and $\rho_0, \rho_1 \in D(\mathcal{H})$ a pair of quantum states. For any $\lambda \in [0,1]$ and any POVM $\mu : \{0,1\} \to B(\mathcal{H})^+$, we have*

$$\lambda \langle \mu(0), \rho_0 \rangle_{HS} + (1-\lambda)\langle \mu(1), \rho_1 \rangle_{HS} \leqslant \frac{1}{2} + \frac{1}{2}\|\lambda\rho_0 - (1-\lambda)\rho_1\|_1.$$

The upper bound is achieved by the PVM $\mu_{opt} : \{0,1\} \to \mathrm{Proj}(\mathcal{H})$ such that $\mu(0)$ is the projector onto $supp\left((\lambda\rho_0 - (1-\lambda)\rho_1)^+\right)$ (where $\cdot^+$ denotes the positive part in the Jordan-Hahn decomposition[1]).

Proof. Any binary POVM $\mu : \{0,1\} \to B(\mathcal{H})^+$ can be written as

$$\mu(0) = \frac{\mathbb{1}_{\mathcal{H}} + X}{2} \quad \text{and} \quad \mu(1) = \frac{\mathbb{1}_{\mathcal{H}} - X}{2},$$

for some contraction $X \in B(\mathcal{H})$ satisfying $\|X\|_\infty \leqslant 1$. Inserting this decomposition into the formula for the success probability shows that

$$\lambda \langle \mu(0), \rho_0 \rangle_{HS} + (1-\lambda)\langle \mu(1), \rho_1 \rangle_{HS} = \frac{1}{2} + \frac{1}{2}\langle X, \lambda\rho_0 - (1-\lambda)\rho_1 \rangle_{HS}.$$

[1] For $H \in B(\mathcal{H})_{sa}$ with Jordan-Hahn decomposition $H = X_1 - X_2$ with $X_1, X_2 \in B(\mathcal{H})^+$ we have $H^+ = X_1$.

Now, we note that

$$\langle X, \lambda\rho_0 - (1-\lambda)\rho_1\rangle_{HS} \leqslant \sup_{U\in\mathcal{U}(\mathcal{H})} |\langle U, \lambda\rho_0 - (1-\lambda)\rho_1\rangle_{HS}| = \|\lambda\rho_0 - (1-\lambda)\rho_1\|_1,$$

where we used the fact that the unitaries are the extreme points of the $\|\cdot\|_\infty$-unit ball and Lemma 3.1. Since the set of unitaries $\mathcal{U}(\mathcal{H})$ is compact, the supremum is attained. The optimal unitary U_{opt} is the one flipping the signs of the negative eigenvalues of the operator $\lambda\rho_0 - (1-\lambda)\rho_1$ and we see that

$$\mu_{\text{opt}}(0) = \frac{\mathbb{1}_{\mathcal{H}} + U_{\text{opt}}}{2},$$

which is the projector onto $\operatorname{supp}\left((\lambda\rho_0 - (1-\lambda)\rho_1)^+\right)$. □

Scenario: Discriminating states from an ensemble . In the previous scenario, Bob received a quantum state ρ_0 with probability λ and a state ρ_1 with probability $1-\lambda$. Let us now generalize this scenario to more than two quantum states: Again, Alice is in the position of a device with $n \in \mathbb{N}$ buttons. After pressing a button "i", the device emits a particle in quantum state ρ_i from some set $\{\rho_1, \ldots, \rho_n\} \subset D(\mathcal{H})$ of quantum states. Again, Bob catches the particle, measures it using a POVM $\mu : \{1, \ldots, n\} \to B(\mathcal{H})^+$ and guesses that Alice pressed button j if he received that outcome. Assuming that Alice presses the buttons according to a probability distribution $p \in \mathcal{P}\{1, \ldots, n\}$, what is the optimal success probability of Bob's guess?

Again, given a particular POVM $\mu : \{1, \ldots, n\} \to B(\mathcal{H})^+$, we can express the success probability by

$$p_{\text{succ}}(\mu) = \sum_{i=1}^{n} p_i \langle \mu(i), \rho_i\rangle.$$

In the following, we will denote by $\operatorname{opt}\left(\{p_i, \rho_i\}_{i=1}^n\right)$ the optimal success probability achievable by Bob's measurement, when Alice chooses the states ρ_i with probability p_i. The collection of probabilities and corresponding operators $\{p_i, \rho_i\}_{i=1}^n$ is also called an *ensemble of quantum states*.

Contrary to the previous scenario, it is more difficult to analyze this expression. In particular, there is no easy closed formula for the optimal guessing probability. Instead, the optimal guessing probability can be expressed as a particular convex optimization problem known as a *semidefinite program* (or SDP for short). Such optimization problems can be solved efficiently (i.e., their runtime scales at most polynomial in the size of the problem and the inverse of the desired accuracy). We will not go into the details of this, but instead prove the following remarkable fact, that there actually is a very simple measurement Bob can do, which achieves a pretty good success probability:

Theorem 3.3 (Barnum and Knill's pretty good measurement)**.** *Consider a set $\{\rho_1, \ldots, \rho_n\} \subset D(\mathcal{H})$ of quantum states on a complex Euclidean space $\mathcal{H}$. For any probability distribution $p \in \mathcal{P}_n$, we can define a POVM $\mu : \{1, \ldots, n\} \to B(\mathcal{H})^+$ by*

$$\mu(i) = \rho^{-\frac{1}{2}} p_i \rho_i \rho^{-\frac{1}{2}} + \frac{1}{n}\Pi_{ker(\rho)},$$

where $\rho = \sum_{i=1}^n p_i\rho_i$, and ρ^{-1} denotes the Moore-Penrose pseudo-inverse. This POVM satisfies the inequality

$$\sum_{i=1}^{n} p_i \langle \mu(i), \rho_i\rangle_{HS} \geqslant \left(opt\left(\{p_i, \rho_i\}_{i=1}^n\right)\right)^2.$$

Proof. By positivity, we have $\ker(\rho) \subseteq \ker(\rho_i)$ and hence $\mathrm{im}(\rho_i) \subseteq \mathrm{im}(\rho)$ for any $i \in \{1, \ldots, n\}$. For any $X \in B(\mathcal{H})^+$, we have

$$\langle X, \rho_i \rangle_{HS} = \langle X, P_{\mathrm{im}(\rho)} \rho_i P_{\mathrm{im}(\rho)} \rangle_{HS} = \langle \rho^{\frac{1}{4}} X \rho^{\frac{1}{4}}, \rho^{-\frac{1}{4}} \rho_i \rho^{-\frac{1}{4}} \rangle_{HS},$$

where ρ^{-1} denotes the Moore-Penrose pseudo-inverse. Applying the Cauchy-Schwarz inequality, we find that

$$\langle X, \rho_i \rangle_{HS} \leqslant \|\rho^{\frac{1}{4}} X \rho^{\frac{1}{4}}\|_2 \|\rho^{-\frac{1}{4}} \rho_i \rho^{-\frac{1}{4}}\|_2. \tag{2}$$

Consider now a POVM $\nu : \{1, \ldots, n\} \to B(\mathcal{H})^+$. We have

$$\begin{aligned} \sum_{i=1}^n p_i \langle \nu(i), \rho_i \rangle_{HS} &\leqslant \sum_{i=1}^n \|\rho^{\frac{1}{4}} \nu(i) \rho^{\frac{1}{4}}\|_2 \|\rho^{-\frac{1}{4}} p_i \rho_i \rho^{-\frac{1}{4}}\|_2 \\ &\leqslant \left(\sum_{i=1}^n \|\rho^{\frac{1}{4}} \nu(i) \rho^{\frac{1}{4}}\|_2^2 \right)^{1/2} \left(\sum_{i=1}^n \|\rho^{-\frac{1}{4}} p_i \rho_i \rho^{-\frac{1}{4}}\|_2^2 \right)^{1/2}, \end{aligned}$$

where we used (2) for the first inequality and the Cauchy-Schwarz inequality for vectors in the second inequality. Using that $0 \leqslant \nu(i) \leqslant \mathbb{1}_{\mathcal{H}}$, we find that

$$\|\rho^{\frac{1}{4}} \nu(i) \rho^{\frac{1}{4}}\|_2^2 = \mathrm{Tr}\left[\nu(i) \rho^{\frac{1}{2}} \nu(i) \rho^{\frac{1}{2}} \right] \leqslant \mathrm{Tr}\left[\rho \nu(i) \right],$$

and therefore

$$\sum_{i=1}^n \|\rho^{\frac{1}{4}} \nu(i) \rho^{\frac{1}{4}}\|_2^2 \leqslant \sum_{i=1}^n \mathrm{Tr}\left[\rho \nu(i) \right] = \mathrm{Tr}\left[\rho \right] = 1.$$

We conclude that

$$\sum_{i=1}^n p_i \langle \nu(i), \rho_i \rangle_{HS} \leqslant \left(\sum_{i=1}^n \|\rho^{-\frac{1}{4}} p_i \rho_i \rho^{-\frac{1}{4}}\|_2^2 \right)^{1/2},$$

for any POVM $\nu : \{1, \ldots, n\} \to B(\mathcal{H})^+$. Consider now the pretty good measurement $\mu\{1, \ldots, n\} \to B(\mathcal{H})^+$ defined in the statement of the theorem. We can compute

$$p_i \langle \mu(i), \rho_i \rangle_{HS} = \langle \rho^{-\frac{1}{4}} p_i \rho_i \rho^{-\frac{1}{4}}, \rho^{-\frac{1}{4}} p_i \rho_i \rho^{-\frac{1}{4}} \rangle_{HS} = \|\rho^{-\frac{1}{4}} p_i \rho_i \rho^{-\frac{1}{4}}\|_2^2.$$

Finally, if $\nu : \{1, \ldots, n\} \to B(\mathcal{H})^+$ is an optimal measurement, then we conclude that

$$\mathrm{opt}\left(\{p_i, \rho_i\}_{i=1}^n \right) = \sum_{i=1}^n p_i \langle \nu(i), \rho_i \rangle_{HS} \leqslant \left(\sum_{i=1}^n \|\rho^{-\frac{1}{4}} p_i \rho_i \rho^{-\frac{1}{4}}\|_2^2 \right)^{1/2} = \left(\sum_{i=1}^n p_i \langle \nu(i), \rho_i \rangle_{HS} \right)^{1/2},$$

which finishes the proof.

□