CS 766/QIC 820 Theory of Quantum Information (Fall 2011)

Lecture 13: Majorization for real vectors and Hermitian operators

This lecture discusses the notion of *majorization* and some of its connections to quantum information. The main application of majorization that we will see in this course will come in a later lecture when we study *Nielsen's theorem*, which precisely characterizes when it is possible for two parties to transform one pure state into another by means of local operations and classical communication. There are other interesting applications of the notion, however, and a few of them will be discussed in this lecture.

13.1 Doubly stochastic operators

Let Σ be a finite, nonempty set, and for the sake of this discussion let us focus on the real vector space $\mathbb{R}^{\Sigma}$. An operator $A \in \mathrm{L}\left(\mathbb{R}^{\Sigma}\right)$ acting on this vector space is said to be *stochastic* if

1. $A(a,b) \geq 0$ for each $(a,b) \in \Sigma \times \Sigma$, and
2. $\sum_{a\in\Sigma} A(a,b) = 1$ for each $b \in \Sigma$.

This condition is equivalent to requiring that $A e_b$ is a probability vector for each $b \in \Sigma$, or equivalently that A maps probability vectors to probability vectors.

An operator $A \in \mathrm{L}\left(\mathbb{R}^{\Sigma}\right)$ is *doubly stochastic* if it is the case that both A and A^{T} (or, equivalently, A and A^{*}) are stochastic. In other words, when viewed as a matrix, every row and every column of A forms a probability vector:

1. $A(a,b) \geq 0$ for each $(a,b) \in \Sigma \times \Sigma$,
2. $\sum_{a\in\Sigma} A(a,b) = 1$ for each $b \in \Sigma$, and
3. $\sum_{b\in\Sigma} A(a,b) = 1$ for each $a \in \Sigma$.

Next, let us write $\mathrm{Sym}(\Sigma)$ to denote the set of one-to-one and onto functions of the form $\pi : \Sigma \rightarrow \Sigma$ (or, in other words, the *permutations* of Σ). For each $\pi \in \mathrm{Sym}(\Sigma)$ we define an operator $V_\pi \in \mathrm{L}\left(\mathbb{R}^{\Sigma}\right)$ as

$$V_\pi(a,b) = \begin{cases} 1 & \text{if } a = \pi(b) \\ 0 & \text{otherwise} \end{cases}$$

for every $(a,b) \in \Sigma \times \Sigma$. Equivalently, V_π is the operator defined by $V_\pi e_b = e_{\pi(b)}$ for each $b \in \Sigma$. Such an operator is called a *permutation operator*.

It is clear that every permutation operator is doubly stochastic, and that the set of doubly stochastic operators is a convex set. The following famous theorem establishes that the doubly stochastic operators are, in fact, given by the convex hull of the permutation operators.

Theorem 13.1 (The Birkhoff–von Neumann theorem)**.** *Let Σ be a finite, nonempty set and let $A \in \mathrm{L}\left(\mathbb{R}^{\Sigma}\right)$ be a linear operator on $\mathbb{R}^{\Sigma}$. It holds that A is a doubly stochastic operator if and only if there exists a probability vector $p \in \mathbb{R}^{\mathrm{Sym}(\Sigma)}$ such that*

$$A = \sum_{\pi \in \mathrm{Sym}(\Sigma)} p(\pi) V_{\pi}.$$

Proof. The Krein-Milman theorem states that every compact, convex set is equal to the convex hull of its extreme points. As the set of doubly stochastic operators is compact and convex, the theorem will therefore follow if we prove that every extreme point in this set is a permutation operator.

To this end, let us consider any doubly stochastic operator A that is not a permutation operator. Our goal is to prove that A is not an extreme point in the set of doubly stochastic operators. Given that A is doubly stochastic but not a permutation operator, there must exist at least one pair $(a_1, b_1) \in \Sigma \times \Sigma$ such that $A(a_1, b_1) \in (0,1)$. As $\sum_b A(a_1, b) = 1$ and $A(a_1, b_1) \in (0,1)$, we conclude that there must exist $b_2 \neq b_1$ such that $A(a_1, b_2) \in (0,1)$. Applying similar reasoning, but to the first index rather than the second, there must exist $a_2 \neq a_1$ such that $A(a_2, b_2) \in (0,1)$. This argument may repeated, alternating between the first and second indices (i.e., between rows and columns), until eventually a closed loop of even length is formed that alternates between horizontal and vertical moves among the entries of A. (Of course a loop must eventually be formed, given that there are only finitely many entries in the matrix A, and an odd length loop can be avoided by an appropriate choice for the entry that closes the loop.) This process is illustrated in Figure 13.1, where the loop is indicated by the dotted lines.

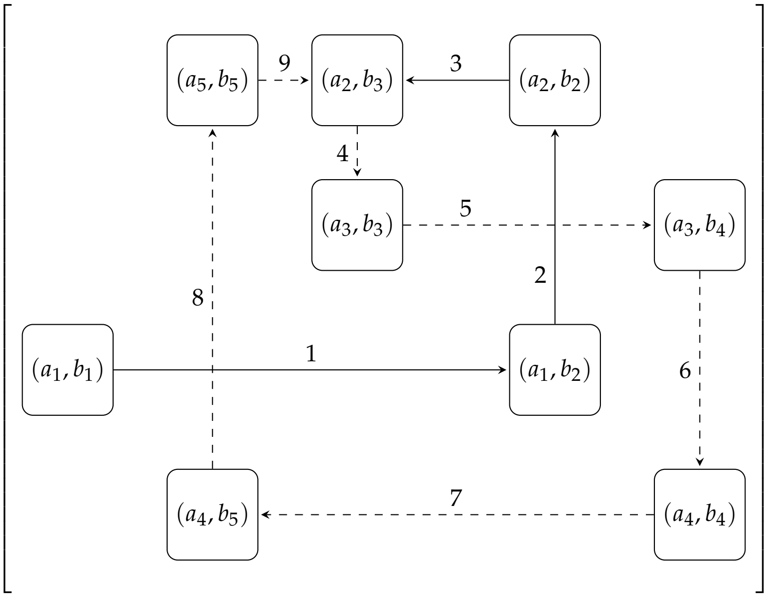

Figure 13.1: An example of a closed loop consisting of entries of A that are contained in the interval $(0,1)$.

Now, let $\varepsilon \in (0,1)$ be equal to the minimum value over the entries in the closed loop, and define B to be the operator obtained by setting each entry in the closed loop to be $\pm\varepsilon$, alternating sign along the entries as suggested in Figure 13.2. All of the other entries in B are set to 0.

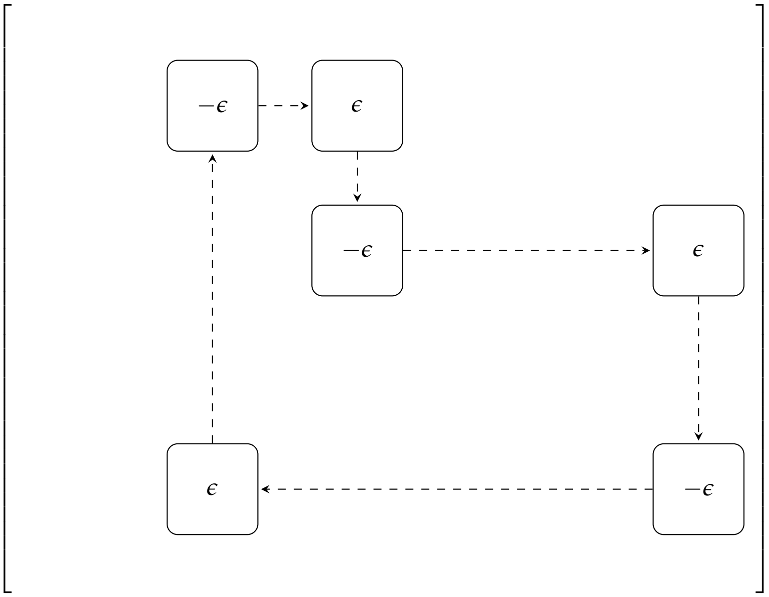

Figure 13.2: The operator B. All entries besides those indicated are 0.

Finally, consider the operators $A + B$ and $A - B$. As A is doubly stochastic and the row and column sums of B are all 0, we have that both $A + B$ and $A - B$ also have row and column sums equal to 1. As ε was chosen to be no larger than the smallest entry within the chosen closed loop, none of the entries of $A + B$ or $A - B$ are negative, and therefore $A - B$ and $A + B$ are doubly stochastic. As B is non-zero, we have that $A + B$ and $A - B$ are distinct. Thus, we have that

$$A = \frac{1}{2}(A + B) + \frac{1}{2}(A - B)$$

is a proper convex combination of doubly stochastic operators, and is therefore not an extreme point in the set of doubly stochastic operators. This is what we needed to prove, and so we are done. □

13.2 Majorization for real vectors

We will now define what it means for one real vector to *majorize* another, and we will discuss two alternate characterizations of this notion. As usual we take Σ to be a finite, nonempty set, and as in the previous section we will focus on the real vector space $\mathbb{R}^{\Sigma}$. The definition is as follows: for $u, v \in \mathbb{R}^{\Sigma}$, we say that u *majorizes* v if there exists a doubly stochastic operator A such that $v = Au$. We denote this relation as $v \prec u$ (or $u \succ v$ if it is convenient to switch the ordering).

By the Birkhoff–von Neumann theorem, this definition can intuitively be interpreted as saying that $v \prec u$ if and only if there is a way to "randomly shuffle" the entries of u to obtain v, where by "randomly shuffle" it is meant that one averages in some way a collection of vectors that is obtained by permuting the entries of u to obtain v. Informally speaking, the relation $v \prec u$ means that u is "more ordered" than v, because we can get from u to v by randomizing the order of the vector indices.

An alternate characterization of majorization (which is in fact more frequently taken as the definition) is based on a condition on various sums of the entries of the vectors involved. To state the condition more precisely, let us introduce the following notation. For every vector $u \in \mathbb{R}^{\Sigma}$ and for $n = |\Sigma|$, we write

$$s(u) = (s_1(u), \ldots, s_n(u))$$

to denote the vector obtained by *sorting* the entries of u from largest to smallest. In other words, we have

$$\{u(a) \,:\, a \in \Sigma\} = \{s_1(u), \ldots, s_n(u)\},$$

where the equality considers the two sides of the equation to be multisets, and moreover

$$s_1(u) \geq \cdots \geq s_n(u).$$

The characterization is given by the equivalence of the first and second items in the following theorem. (The equivalence of the third item in the theorem gives a third characterization that is closely related to the definition, and will turn out to be useful later in the lecture.)

Theorem 13.2. *Let Σ be a finite, non-empty set, and let $u, v \in \mathbb{R}^{\Sigma}$. The following are equivalent.*

1. *$v \prec u$.*
2. *For $n = |\Sigma|$ and for $1 \leq k < n$ we have*

$$\sum_{j=1}^{k} s_j(v) \leq \sum_{j=1}^{k} s_j(u), \tag{13.1}$$

and moreover

$$\sum_{j=1}^{n} s_j(v) = \sum_{j=1}^{n} s_j(u). \tag{13.2}$$

3. *There exists a unitary operator $U \in \mathrm{L}\left(\mathbb{C}^{\Sigma}\right)$ such that $v = Au$, where $A \in \mathrm{L}\left(\mathbb{R}^{\Sigma}\right)$ is the operator defined by $A(a,b) = |U(a,b)|^2$ for all $(a,b) \in \Sigma \times \Sigma$.*

Proof. First let us prove that item 1 implies item 2. Assume that A is a doubly stochastic operator satisfying $v = Au$. It is clear that the condition (13.2) holds, as doubly stochastic operators preserve the sum of the entries in any vector, and so it remains to prove the condition (13.1) for $1 \leq k < n$.

To do this, let us first consider the effect of an arbitrary doubly stochastic operator $B \in \mathrm{L}\left(\mathbb{R}^{\Sigma}\right)$ on a vector of the form

$$e_{\Gamma} = \sum_{a \in \Gamma} e_a$$

where $\Gamma \subseteq \Sigma$. The vector e_{Γ} is the *characteristic vector* of the subset $\Gamma \subseteq \Sigma$. The resulting vector Be_{Γ} is a convex combination of permutations of e_{Γ}, or in other words is a convex combination of characteristic vectors of sets having size $|\Gamma|$. The sum of the entries of Be_{Γ} is therefore $|\Gamma|$, and each entry must lie in the interval $[0,1]$. For any set $\Delta \subseteq \Sigma$ with $|\Delta| = |\Gamma|$, the vector $e_{\Delta} - Be_{\Gamma}$ therefore has entries summing to 0, and satisfies $(e_{\Delta} - Be_{\Gamma})(a) \geq 0$ for every $a \in \Delta$ and $(e_{\Delta} - Be_{\Gamma})(a) \leq 0$ for every $a \notin \Delta$.

Now, for each value $1 \leq k < n$, define $\Delta_k, \Gamma_k \subseteq \Sigma$ to be the subsets indexing the k largest entries of u and v, respectively. In other words,

$$\sum_{j=1}^{k} s_j(u) = \sum_{a \in \Delta_k} u(a) = \langle e_{\Delta_k}, u \rangle \qquad \text{and} \qquad \sum_{j=1}^{k} s_j(v) = \sum_{a \in \Gamma_k} v(a) = \langle e_{\Gamma_k}, v \rangle\,.$$

We now see that

$$\sum_{i=1}^{k} s_i(u) - \sum_{i=1}^{k} s_i(v) = \langle e_{\Delta_k}, u \rangle - \langle e_{\Gamma_k}, v \rangle = \langle e_{\Delta_k}, u \rangle - \langle e_{\Gamma_k}, Au \rangle = \langle e_{\Delta_k} - A^* e_{\Gamma_k}, u \rangle .$$

This quantity in turn may be expressed as

$$\langle e_{\Delta_k} - A^* e_{\Gamma_k}, u \rangle = \sum_{a \in \Delta_k} \alpha_a \, u(a) - \sum_{a \notin \Delta_k} \alpha_a \, u(a),$$

where

$$\alpha_a = \begin{cases} (e_{\Delta_k} - A^* e_{\Gamma_k})(a) & \text{if } a \in \Delta_k \\ -(e_{\Delta_k} - A^* e_{\Gamma_k})(a) & \text{if } a \notin \Delta_k. \end{cases}$$

As argued above, we have $\alpha_a \geq 0$ for each $a \in \Sigma$ and $\sum_{a \in \Delta_k} \alpha_a = \sum_{a \notin \Delta_k} \alpha_a$. By the choice of Δ_k we have $u(a) \geq u(b)$ for all choices of $a \in \Delta_k$ and $b \notin \Delta_k$, and therefore

$$\sum_{a \in \Delta_k} \alpha_a \, u(a) \geq \sum_{a \notin \Delta_k} \alpha_a \, u(a).$$

This is equivalent to (13.1) as required.

Next we will prove item 2 implies item 3, by induction on $|\Sigma|$. The case $|\Sigma| = 1$ is trivial, so let us consider the case that $|\Sigma| \geq 2$. Assume for simplicity that $\Sigma = \{1, \ldots, n\}$, that $u = (u_1, \ldots, u_n)$ for $u_1 \geq \cdots \geq u_n$, and that $v = (v_1, \ldots, v_n)$ for $v_1 \geq \cdots \geq v_n$. This causes no loss of generality because the majorization relationship is invariant under renaming and independently reordering the indices of the vectors under consideration. Let us also identify the operators U and A we wish to construct with $n \times n$ matrices having entries denoted $U_{i,j}$ and $A_{i,j}$.

Now, assuming item 2 holds, we must have that $u_1 \geq v_1 \geq u_k$ for some choice of $k \in \{1, \ldots, n\}$. Take k to be minimal among all such indices. If it is the case that $k = 1$ then $u_1 = v_1$; and by setting $x = (u_2, \ldots, x_n)$ and $y = (v_2, \ldots, v_n)$ we conclude from the induction hypothesis that there exists an $(n-1) \times (n-1)$ unitary matrix X so that the doubly stochastic matrix B defined by $B_{i,j} = \left| X_{i,j} \right|^2$ satisfies $y = Bx$. By taking U to be the $n \times n$ unitary matrix

$$U = \begin{pmatrix} 1 & 0 \\ 0 & X \end{pmatrix}$$

and letting A be defined by $A_{i,j} = \left| U_{i,j} \right|^2$, we have that $v = Au$ as is required.

The more difficult case is where $k \geq 2$. Let $\lambda \in [0,1]$ satisfy $v_1 = \lambda u_1 + (1 - \lambda) u_k$, and define W to be the $n \times n$ unitary matrix determined by the following equations:

$$\begin{aligned} We_1 &= \sqrt{\lambda} e_1 - \sqrt{1 - \lambda} e_k \\ We_k &= \sqrt{1 - \lambda} e_1 + \sqrt{\lambda} e_k \\ We_j &= e_j \qquad (\text{for } j \notin \{1, k\}). \end{aligned}$$

The action of W on the span of $\{e_1, e_k\}$ is described by this matrix:

$$\begin{pmatrix} \sqrt{\lambda} & \sqrt{1 - \lambda} \\ -\sqrt{1 - \lambda} & \sqrt{\lambda} \end{pmatrix}.$$

Notice that the $n \times n$ doubly stochastic matrix D given by $D_{i,j} = \left|W_{i,j}\right|^2$ may be written

$$D = \lambda \mathbb{1} + (1-\lambda) V_{(1\ k)},$$

where $(1\ k) \in S_n$ denotes the permutation that swaps 1 and k, leaving every other element of $\{1, \ldots, n\}$ fixed.

Next, define $(n-1)$-dimensional vectors

$$\begin{aligned}
x &= (u_2, \ldots, u_{k-1}, \lambda u_k + (1-\lambda) u_1, u_{k+1}, \ldots, u_n) \\
y &= (v_2, \ldots, v_n).
\end{aligned}$$

We will index these vectors as $x = (x_2, \ldots, x_n)$ and $y = (y_2, \ldots, y_n)$ for clarity. For $1 \leq l \leq k-1$ we clearly have

$$\sum_{j=2}^{l} y_j = \sum_{j=2}^{l} v_j \leq \sum_{j=2}^{l} u_j = \sum_{j=2}^{l} x_j$$

given that $v_n \leq \cdots \leq v_1 \leq u_{k-1} \leq \cdots \leq u_1$. For $k \leq l \leq n$, we have

$$\sum_{j=2}^{l} x_j = \sum_{j=1}^{l} u_i - (\lambda u_1 + (1-\lambda) u_k) \geq \sum_{j=2}^{l} v_j.$$

Thus, we may again apply the induction hypothesis to obtain an $(n-1) \times (n-1)$ unitary matrix X such that, for B the doubly stochastic matrix defined by $B_{i,j} = \left|X_{i,j}\right|^2$, we have $y = Bx$.

Now define

$$U = \begin{pmatrix} 1 & 0 \\ 0 & X \end{pmatrix} W.$$

This is a unitary matrix, and to complete the proof it suffices to prove that the doubly stochastic matrix A defined by $A_{i,j} = \left|U_{i,j}\right|^2$ satisfies $v = Au$. We have the following entries of A:

$$\begin{aligned}
A_{1,1} &= \left|U_{1,1}\right|^2 = \lambda, & A_{i,1} &= (1-\lambda)\left|X_{i-1,k-1}\right|^2 = (1-\lambda) B_{i-1,k-1}, \\
A_{1,k} &= \left|U_{1,k}\right|^2 = 1-\lambda, & A_{i,k} &= \lambda \left|X_{i-1,k-1}\right|^2 = \lambda B_{i-1,k-1}, \\
A_{1,j} &= 0, & A_{i,j} &= \left|X_{i-1,j-1}\right|^2 = B_{i-1,j-1},
\end{aligned}$$

Where i and j range over all choices of indices with $i, j \notin \{1, k\}$. From these equations we see that

$$A = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix} D,$$

which satisfies $v = Au$ as required.

The final step in the proof is to observe that item 3 implies item 1, which is trivial given that the operator A determined by item 3 must be doubly stochastic. □

13.3 Majorization for Hermitian operators

We will now define an analogous notion of majorization for Hermitian operators. For Hermitian operators $A, B \in \operatorname{Herm}(\mathcal{X})$ we say that A majorizes B, which we express as $B \prec A$ or $A \succ B$, if there exists a mixed unitary channel $\Phi \in \mathrm{C}(\mathcal{X})$ such that

$$B = \Phi(A).$$

Inspiration for this definition partly comes from the Birkhoff–von Neumann theorem, along with the intuitive idea that randomizing the entries of a real vector is analogous to randomizing the choice of an orthonormal basis for a Hermitian operator.

The following theorem gives an alternate characterization of this relationship that also connects it with majorization for real vectors.

Theorem 13.3. *Let $\mathcal{X}$ be a complex Euclidean space and let $A, B \in \operatorname{Herm}(\mathcal{X})$. It holds that $B \prec A$ if and only if $\lambda(B) \prec \lambda(A)$.*

Proof. Let $n = \dim(\mathcal{X})$. By the Spectral theorem, we may write

$$B = \sum_{j=1}^{n} \lambda_j(B) u_j u_j^* \quad \text{and} \quad A = \sum_{j=1}^{n} \lambda_j(A) v_j v_j^*$$

for orthonormal bases $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_n\}$ of $\mathcal{X}$.

Let us first assume that $\lambda(B) \prec \lambda(A)$. This implies there exist a probability vector $p \in \mathbb{R}^{S_n}$ such that

$$\lambda_j(B) = \sum_{\pi \in S_n} p(\pi) \lambda_{\pi(j)}(A)$$

for $1 \leq j \leq n$. For each permutation $\pi \in S_n$, define a unitary operator

$$U_\pi = \sum_{j=1}^{n} u_j v_{\pi(j)}^*.$$

It holds that

$$\sum_{\pi \in S_n} p(\pi) U_\pi A U_\pi^* = \sum_{j=1}^{n} \sum_{\pi \in S_n} p(\pi) \lambda_{\pi(j)}(A)\, u_j u_j^* = B.$$

Suppose on the other hand that there exists a probability vector $(p_1, \ldots, p_m)$ and unitary operators $U_1, \ldots, U_m$ so that

$$B = \sum_{i=1}^{m} p_i U_i A U_i^*.$$

By considering the spectral decompositions above, we have

$$\lambda_j(B) = \sum_{i=1}^{m} p_i u_j^* U_i A U_i^* u_j = \sum_{i=1}^{m} \sum_{k=1}^{n} p_i \left| u_j^* U_i v_k \right|^2 \lambda_k(A).$$

Define an $n \times n$ matrix D as

$$D_{j,k} = \sum_{i=1}^{m} p_i \left| u_j^* U_i v_k \right|^2.$$

It holds that D is doubly stochastic and satisfies $D\lambda(A) = \lambda(B)$. Therefore $\lambda(B) \prec \lambda(A)$ as required. □

13.4 Applications

Finally, we will note a few applications of the facts we have proved about majorization.

13.4.1 Entropy, norms, and majorization

We begin with two simple facts, one relating entropy with majorization, and the other relating Schatten p-norms to majorization.

Proposition 13.4. *Suppose that $\rho, \xi \in \mathrm{D}(\mathcal{X})$ satisfy $\rho \succ \xi$. It holds that $S(\rho) \leq S(\xi)$.*

Proof. The proposition follows from fact that for every density operator ρ and every mixed unitary operation Φ, we have $S(\rho) \leq S(\Phi(\rho))$ by the concavity of the von Neumann entropy. □

Note that we could equally well have first observed that $H(p) \leq H(q)$ for probability vectors p and q for which $p \succ q$, and then applied Theorem 13.3.

Proposition 13.5. *Suppose that $A, B \in \mathrm{Herm}(\mathcal{X})$ satisfy $A \succ B$. For every $p \in [1, \infty]$, it holds that $\|A\|_p \geq \|B\|_p$.*

Proof. As $A \succ B$, there exists a mixed unitary operator

$$\Phi(X) = \sum_{a \in \Gamma} q(a) U_a X U_a^*$$

such that $B = \Phi(A)$. It holds that

$$\|B\|_p = \|\Phi(A)\|_p = \left\| \sum_{a \in \Gamma} q(a) U_a A U_a^* \right\|_p \leq \sum_{a \in \Gamma} q(a) \|U_a A U_a^*\|_p = \sum_{a \in \Gamma} q(a) \|A\|_p = \|A\|_p,$$

where the inequality is by the triangle inequality, and the third equality holds by the unitary invariance of Schatten p-norms. □

13.4.2 A theorem of Schur relating diagonal entries to eigenvalues

The second application of majorization concerns a relationship between the diagonal entries of an operator and its eigenvalues, which is attributed to Issai Schur. First, a simple lemma relating to dephasing channels (discussed in Lecture 6) is required.

Lemma 13.6. *Let $\mathcal{X}$ be a complex Euclidean space and let $\{x_a : a \in \Sigma\}$ be any orthonormal basis of $\mathcal{X}$. The channel*

$$\Phi(A) = \sum_{a \in \Sigma} (x_a^* A x_a) \, x_a x_a^*$$

is mixed unitary.

Proof. We will assume that $\Sigma = \mathbb{Z}_n$ for some $n \geq 1$. This assumption causes no loss of generality, because neither the ordering of elements in Σ nor their specific names have any bearing on the statement of the lemma.

First consider the standard basis $\{e_a : a \in \mathbb{Z}_n\}$, and define a mixed unitary channel

$$\Delta(A) = \frac{1}{n} \sum_{a \in \mathbb{Z}_n} Z^a A (Z^a)^*,$$

as was done in Lecture 6. We have

$$\Delta(E_{b,c}) = \frac{1}{n} \sum_{a \in \mathbb{Z}_n} \omega^{a(b-c)} E_{b,c} = \begin{cases} E_{b,b} & \text{if } b = c \\ 0 & \text{if } b \neq c, \end{cases}$$

and therefore

$$\Delta(A) = \sum_{a \in \mathbb{Z}_n} (e_a^* A e_a) E_{a,a}.$$

For $U \in \mathrm{U}(\mathcal{X})$ defined as

$$U = \sum_{a \in \mathbb{Z}_n} x_a e_a^*$$

it follows that

$$\frac{1}{n} \sum_{a \in \mathbb{Z}_n} (UZ^a U^*) \, A \, (UZ^a U^*)^* = U\Delta(U^* A U) U^* = \Phi(A).$$

The mapping Φ is therefore mixed unitary as required. □

Theorem 13.7 (Schur). *Let $\mathcal{X}$ be a complex Euclidean space, let $A \in \mathrm{Herm}(\mathcal{X})$ be a Hermitian operator, and let $\{x_a : a \in \Sigma\}$ be an orthonormal basis of $\mathcal{X}$. For $v \in \mathbb{R}^\Sigma$ defined as $v(a) = x_a^* A x_a$ for each $a \in \Sigma$, it holds that $v \prec \lambda(A)$.*

Proof. Immediate from Lemma 13.6 and Theorem 13.3. □

Notice that this theorem implies that the probability distribution arising from any *complete* projective measurement of a density operator ρ must have Shannon entropy at least $S(\rho)$.

It is natural to ask if the converse of Theorem 13.7 holds. That is, given a Hermitian operator $A \in \mathrm{Herm}(\mathcal{X})$, for $\mathcal{X} = \mathbb{C}^\Sigma$, and a vector $v \in \mathbb{R}^\Sigma$ such that $\lambda(A) \succ v$, does there necessarily exist an orthonormal basis $\{x_a : a \in \Sigma\}$ of $\mathcal{X}$ such that $v(a) = \langle x_a x_a^*, A \rangle$ for each $a \in \Sigma$? The answer is "yes," as the following theorem states.

Theorem 13.8. *Suppose Σ is a finite, nonempty set, let $\mathcal{X} = \mathbb{C}^\Sigma$, and suppose that $A \in \mathrm{Herm}(\mathcal{X})$ and $v \in \mathbb{R}^\Sigma$ satisfy $v \prec \lambda(A)$. There exists an orthonormal basis $\{x_a : a \in \Sigma\}$ of $\mathcal{X}$ such that $v(a) = x_a^* A x_a$ for each $a \in \Sigma$.*

Proof. Let

$$A = \sum_{a \in \Sigma} w(a) u_a u_a^*$$

be a spectral decomposition of A. The assumptions of the theorem imply, by Theorem 13.2, that there exists a unitary operator U such that, for D defined by

$$D(a,b) = |U(a,b)|^2$$

for $(a,b) \in \Sigma \times \Sigma$, we have $v = Dw$.

Define

$$V = \sum_{a \in \Sigma} e_a u_a^*$$

and let $x_a = V^* U^* V u_a$ for each $a \in \Sigma$. It holds that

$$x_a^* A x_a = \sum_b |U(a,b)|^2 \, w(b) = (Dw)(a) = v(a),$$

which proves the theorem. □

Theorems 13.7 and 13.8 are sometimes together referred to as the *Schur-Horn theorem*.

13.4.3 Density operators consistent with a given probability vector

Finally, we will prove a characterization of precisely which probability vectors are consistent with a given density operator, meaning that the density operator could have arisen from a random choice of pure states according to the distribution described by the probability vector.

Theorem 13.9. *Let $X = \mathbb{C}^{\Sigma}$ for Σ a finite, nonempty set, and suppose that a density operator $\rho \in \mathrm{D}(\mathcal{X})$ and a probability vector $p \in \mathbb{R}^{\Sigma}$ are given. There exist (not necessarily orthogonal) unit vectors $\{u_a : a \in \Sigma\}$ in $\mathcal{X}$ such that*

$$\rho = \sum_{a \in \Sigma} p(a) u_a u_a^*$$

if and only if $p \prec \lambda(\rho)$.

Proof. Assume first that $p \prec \lambda(\rho)$. By Theorem 13.8 we have that there exists an orthonormal basis $\{x_a : a \in \Sigma\}$ of $\mathcal{X}$ with the property that $\langle x_a x_a^*, \rho \rangle = p(a)$ for each $a \in \Sigma$. Let $y_a = \sqrt{\rho} x_a$ for each $a \in \Sigma$. It holds that

$$\|y_a\|^2 = \langle \sqrt{\rho} x_a, \sqrt{\rho} x_a \rangle = x_a^* \rho x_a = p(a).$$

Define

$$u_a = \begin{cases} \frac{y_a}{\|y_a\|} & \text{if } y_a \neq 0 \\ z & \text{if } y_a = 0, \end{cases}$$

where $z \in \mathcal{X}$ is an arbitrary unit vector. We have

$$\sum_{a \in \Sigma} p(a) u_a u_a^* = \sum_{a \in \Sigma} y_a y_a^* = \sum_{a \in \Sigma} \sqrt{\rho} x_a x_a^* \sqrt{\rho} = \rho$$

as required.

Suppose, on the other hand, that

$$\rho = \sum_{a \in \Sigma} p(a) u_a u_a^*$$

for some collection $\{u_a : a \in \Sigma\}$ of unit vectors. Define $A \in \mathrm{L}(\mathcal{X})$ as

$$A = \sum_{a \in \Sigma} \sqrt{p(a)} u_a e_a^*,$$

and note that $AA^* = \rho$. It holds that

$$A^* A = \sum_{a,b \in \Sigma} \sqrt{p(a)p(b)} \langle u_a, u_b \rangle E_{a,b},$$

so $e_a^* A^* A e_a = p(a)$. By Theorem 13.7 this implies $\lambda(A^* A) \succ p$. As $\lambda(A^* A) = \lambda(AA^*)$, the theorem is proved. □