Probability distributions consistent with a mixed state

M. A. Nielsen*
Department of Physics, MC 12-33, California Institute of Technology, Pasadena, California 91125
(Recieved 9 September 1999; published 16 October 2000)

A density matrix ρ may be represented in many different ways as a mixture of pure states, $\rho = \Sigma_i p_i |\psi_i\rangle\langle\psi_i|$. This paper characterizes the class of probability distributions (p_i) that may appear in such a decomposition, for a fixed density matrix ρ. Several illustrative applications of this result to quantum mechanics and quantum information theory are given.

PACS number(s): 03.67.−a, 03.65.−Bz

I. INTRODUCTION

The density matrix was introduced [1,2] as a means of describing a quantum system when the state of the system is not completely known. In particular, if the state of the system is $|\psi_i\rangle$ with probability p_i, then the density matrix is defined by

$$\rho \equiv \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{1.1}$$

For a fixed density matrix it is natural to ask what class of ensembles $\{p_i, |\psi_i\rangle\}$ gives rise to that density matrix? This problem was addressed by Schrödinger [3], whose results have been extended by Jaynes [4], and by Hughston, Jozsa, and Wootters [5]. The result of these investigations, the *classification theorem for ensembles*, has been of considerable utility in quantum statistical mechanics, quantum information theory, quantum computation, and quantum error correction.

In this paper we use the classification theorem for ensembles to obtain an explicit classification of probability distributions (p_i) such that there exist pure states $|\psi_i\rangle$ satisfying $\rho = \Sigma_i p_i |\psi_i\rangle\langle\psi_i|$, for some fixed density matrix ρ. This is done in Sec. II. Section III illustrates the result with several simple applications to quantum mechanics and quantum information theory. Section IV concludes the paper.

II. PROBABILITY DISTRIBUTIONS CONSISTENT WITH A MIXED STATE

To state and prove our results we need to introduce some notions from the theory of *majorization* [6–8]. Majorization is an area of mathematics concerned with the problem of comparing two vectors to determine which is more ''disordered.'' Suppose x and y are two d-dimensional real vectors. Then we say x is *majorized* by y, written $x \prec y$, if

$$\sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow} \tag{2.1}$$

for $k = 1, \ldots, d-1$, with strict equality required when $k = d$. The $^{\downarrow}$ notation indicates that the vector components are to be ordered into decreasing order. The usual interpretation is that x is more ''disordered'' or ''mixed'' than y. When x and y are probability distributions it can be shown that $x \prec y$ implies many quantities commonly used as measures of disorder, such as the Shannon entropy, are never lower for x than for y.

There is a close relation between unitary matrices and majorization. Any matrix D whose components may be written in the form $D_{ij} = |u_{ij}|^2$ for some unitary matrix $u = (u_{ij})$ is said to be *unitary stochastic*. The following theorem [9] connects the unitary stochastic matrices to majorization.

Theorem 1: Let x and y be d-dimensional vectors. Then $x \prec y$ if and only if there exists unitary-stochastic D such that $x = Dy$. The proof of this theorem [9] is constructive in nature. That is, given $x \prec y$ it is possible to explicitly construct a unitary matrix $u = (u_{ij})$ such that $x = Dy$ where $(D_{ij}) = (|u_{ij}|^2)$. Indeed, even more is true—for the forward implication in Theorem 1 it turns out to be sufficient to consider only orthogonal matrices u, that is, real matrices satisfying $uu^T = u^T u = I$, where ^{T} is the transpose operation. The corresponding matrix $D_{ij} = u_{ij}^2$ is known as an *orthostochastic* matrix. Note that the expression u_{ij}^2 indicates the square of the ijth component of the matrix u, not the ijth component of u^2. The Appendix to this paper gives an outline of the construction needed for the reverse implication in Theorem 1, somewhat different from the proof in [9].

The second result we need is the classification theorem for ensembles [3–5]:

Theorem 2: Let ρ be a density matrix. Then $\{p_i, |\psi_i\rangle\}$ is an ensemble for ρ if and only if there exists a unitary matrix $u = (u_{ij})$ such that

$$\sqrt{p_i}|\psi_i\rangle = \sum_j u_{ij}|e_j\rangle, \tag{2.2}$$

where $|e_j\rangle$ are eigenvectors of ρ normalized so that $\lambda_j^{\rho} = \langle e_j|e_j\rangle$ are the corresponding eigenvalues.

In the statement of Theorem 2 it is understood that there may be more elements in the ensemble $\{p_i, |\psi_i\rangle\}$ than there are eigenvectors $|e_j\rangle$. When this is the case one appends extra zero vectors to the list of eigenvectors, until the number of elements in the two lists matches. Combining Theorem 1 and Theorem 2 in an appropriate way gives the following

*Electronic address: mnielsen@theory.caltech.edu

classification theorem for the class of probability distributions consistent with a given density matrix:

Theorem 3: Suppose ρ is a density matrix. Let (p_i) be a probability distribution. Then there exist normalized quantum states $|\psi_i\rangle$ such that

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{2.3}$$

if and only if $(p_i) \prec \lambda^\rho$, where λ^ρ is the vector of eigenvalues of ρ.

In the statement of Theorem 3 it is understood that if the vector (p_i) contains more elements than the vector λ^ρ, then one should append sufficiently many zeros to λ^ρ that the two vectors be of the same length.

Proof of Theorem 3: Suppose there exists a set of states $|\psi_i\rangle$ such that $\rho = \Sigma_i p_i |\psi_i\rangle\langle\psi_i|$. By Theorem 2 Eq. (2.2) must hold. Multiplying Eq. (2.2) by its adjoint gives

$$p_i = \sum_{jk} u_{ik}^* u_{ij} \lambda_j^\rho \delta_{jk}, \tag{2.4}$$

which simplifies to

$$p_i = \sum_j |u_{ij}|^2 \lambda_j^\rho. \tag{2.5}$$

Setting $D_{ij} \equiv |u_{ij}|^2$, we have $(p_i) = D\lambda^\rho$ for unitary stochastic D, and by Theorem 1, $(p_i) \prec \lambda^\rho$.

Conversely, if $(p_i) \prec \lambda^\rho$ then by Theorem 1 we can find unitary u such that Eq. (2.5) is satisfied. Now define states $|\psi_i\rangle$ by Eq. (2.2); since u_{ij}, p_i, and $|e_j\rangle$ are known this equation determines the $|\psi_i\rangle$ uniquely. By Theorem 2 we need only check that these are properly normalized pure states to complete the proof. Multiplying the definition of $|\psi_i\rangle$, Eq. (2.2), by its adjoint gives

$$p_i \langle\psi_i|\psi_i\rangle = \sum_{jk} u_{ij} u_{ik}^* \langle e_k|e_j\rangle \tag{2.6}$$

$$= \sum_j |u_{ij}|^2 \lambda_j^\rho \tag{2.7}$$

$$= p_i, \tag{2.8}$$

where the last step follows from the choice of u to satisfy Eq. (2.5). It follows that $|\psi_i\rangle$ is a normalized pure state. Q.E.D.

Theorem 3 is the central result of this paper. Many elements of the proof are already implicit in the paper of Hughston, Jozsa, and Wootters [5], however, they do not explicitly draw the connection with majorization. The forward implication has been proved by Uhlmann [10], who conjectured but did not find an explicit construction for the reverse implication.

III. APPLICATIONS

The remaining sections of this paper demonstrate several illustrative applications of Theorem 3 to elementary quantum mechanics and quantum information theory.

A. Uniform ensembles exist for any density matrix

As our first application of Theorem 3, suppose d is the rank of ρ, and that $m \geq d$. Then it is easy to verify that $(1/m, 1/m, \ldots, 1/m) \prec \lambda^\rho$, and therefore there exist pure states $|\psi_1\rangle, \ldots, |\psi_m\rangle$ such that ρ is an equal mixture of these states with probability $1/m$,

$$\rho = \sum_i \frac{|\psi_i\rangle\langle\psi_i|}{m}. \tag{3.1}$$

Indeed, if we choose $m \geq d$ where d is the dimension of the underlying space, then for any ρ there exists a set of states such that Eq. (3.1) holds. *A priori* it is not at all obvious that such a set of pure states should exist for *any* density matrix ρ, however Theorem 3 guarantees that this is indeed the case: any density matrix may be regarded as the result of picking uniformly at random from some ensemble of pure states.

B. Schur-convex functions of ensemble probabilities

A second application of Theorem 3 relates functions of the eigenvalues of ρ to functions of the probabilities (p_i). The theory of *isotone functions* [6] is concerned with functions which preserve the majorization order. More specifically, the *Schur-convex functions* are real-valued functions f such that $x \prec y$ implies $f(x) \leq f(y)$. Examples of Schur-convex functions include $f(x) \equiv \Sigma_i x_i \ln(x_i)$, $f(x) \equiv \Sigma_i x_i^k$ (for any constant $k \geq 1$), $f(x) \equiv -\Pi_i x_i$, and $f(x) \equiv -x_1^\downarrow$. More examples and a characterization of the Schur-convex functions may be found in [7,6]. Each such Schur-convex function gives rise to an inequality relating the vector of probabilities (p_i) in Eq (2.3) to the vector λ^ρ. For example, we see from the Schur convexity of $\Sigma_i x_i \ln(x_i)$ the useful inequality that $H(p_i) \geq S(\rho)$, where $H(\,)$ is the Shannon entropy, and $S(\,)$ is the von Neumann entropy. (This result was obtained by Lanford and Robinson [11] using different techniques.) In general, any Schur-convex function will give rise to a similar inequality relating (p_i) and λ^ρ. A similar property related to convex functions has previously been noted (see the review [12] for an overview, as well as the original Refs. [10,13–16]), however, those results are a special case [7] of the more general result given here based upon Schur-convex functions. The earlier results may be obtained by noting that if $f(x)$ is convex then the map $(p_i) \to \Sigma_i f(p_i)$ is Schur convex.

C. Representation of bipartite pure states

A third application of Theorem 3 gives us insight into the properties of pure states of bipartite systems. We state the result formally as follows:

Corollary 4: Suppose $|\psi\rangle$ is a pure state of a composite system AB with Schmidt decomposition [17]

$$|\psi\rangle = \sum_i \sqrt{p_i} |i_A\rangle |i_B\rangle. \tag{3.2}$$

Then given a probability distribution (q_i) there exists an orthonormal basis $|i'_A\rangle$ for system A and corresponding pure states $|\psi_i\rangle$ of system B such that

$$|\psi\rangle = \sum_i \sqrt{q_i}|i'_A\rangle|\psi_i\rangle \tag{3.3}$$

if and only if $(q_i) \prec (p_i)$.

In the statement of Corollary 4 it is understood that if (q_i) contains more terms than (p_i) then the former vector should be extended by adding extra zeros. In the case where the number of terms in (q_i) exceeds the number of dimensions of A's Hilbert space, A's Hilbert space must be extended so its dimension matches the number of terms in (q_i).

Proof of Corollary 4: To prove the forward implication, note that tracing out system A in Eqs. (3.2) and (3.3) gives $\Sigma_i p_i |i_B\rangle\langle i_B| = \Sigma_i q_i |\psi_i\rangle\langle\psi_i|$, and thus by Theorem 3, $(q_i) \prec (p_i)$. Conversely, suppose $|\psi\rangle$ has Schmidt decomposition given by Eq. (3.2), and that $(q_i) \prec (p_i)$. Let ρ be the reduced density matrix of system B when A is traced out,

$$\rho = \mathrm{tr}_A(|\psi\rangle\langle\psi|) = \sum_i p_i |i_B\rangle\langle i_B|. \tag{3.4}$$

By Theorem 3, $\rho = \Sigma_i q_i |\psi_i\rangle\langle\psi_i|$ for some set of pure states $|\psi_i\rangle$. The state $|\phi\rangle$ defined by

$$|\phi\rangle \equiv \sum_i \sqrt{q_i}|i_A\rangle|\psi_i\rangle \tag{3.5}$$

is a purification of ρ, that is, a pure state of system AB such that when system A is traced out, $\mathrm{tr}_A(|\phi\rangle\langle\phi|) = \rho$. Thus $|\psi\rangle$ and $|\phi\rangle$ are both purifications of ρ. It can easily be shown [5] that there exists a unitary matrix U acting on system A such that $U|\phi\rangle = |\psi\rangle$. Defining $|i'_A\rangle \equiv U|i_A\rangle$ we see that

$$|\psi\rangle = \sum_i \sqrt{q_i}|i'_A\rangle|\psi_i\rangle, \tag{3.6}$$

as claimed. Q.E.D.

D. Communication cost of entanglement transformation

Corollary 4 can be used to give insight into a recent result in the study of entanglement transformation [18]. Suppose Alice and Bob are in possession of an entangled pure state $|\psi\rangle$. They wish to transform this state into another pure state $|\phi\rangle$, with the restriction that they may only use local operations on their respective systems, together with a possibly unlimited amount of classical communication. It was shown in [18] that the transformation can be made if and only if $\lambda_\psi \prec \lambda_\phi$, where λ_ψ denotes the vector of eigenvalues of the reduced density matrix of Alice's system when the joint Alice-Bob system is in the state $|\psi\rangle$, and λ_ϕ is defined similarly for the state $|\phi\rangle$.

To see how Corollary 4 applies in this context, suppose $|\psi\rangle$ and $|\phi\rangle$ are bipartite states with Schmidt decompositions

$$|\psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle, \tag{3.7}$$

$$|\phi\rangle = \sum_i \sqrt{q_i}|i\rangle|i\rangle, \tag{3.8}$$

where without loss of generality we may assume the two states have the same Schmidt bases, since local unitary transformations can be used to interconvert between different Schmidt bases. Note that $\lambda_\psi = (p_i)$ and $\lambda_\phi = (q_i)$. Suppose that $\lambda_\psi = (p_i) \prec \lambda_\phi = (q_i)$. By Corollary 4, and ignoring unimportant local unitary transformations, it is possible to write $|\psi\rangle$ and $|\phi\rangle$ in the form

$$|\psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle, \tag{3.9}$$

$$|\phi\rangle = \sum_i \sqrt{p_i}|i\rangle|\psi_i\rangle, \tag{3.10}$$

for some set of pure states $|\psi_i\rangle$. This form makes it quite plausible that the state $|\psi\rangle$ can be transformed into the state $|\phi\rangle$ by local operations and classical communication: all that needs to be done is for Bob to transform $|i\rangle$ into $|\psi_i\rangle$ in such a way as to preserve coherence between different terms in the sum.

I have not found a general method utilizing this fact to transform $|\psi\rangle$ into $|\phi\rangle$. However, it will now be shown how Corollary 4 can be applied successfully in the special case where $|\psi\rangle$ is a maximally entangled state of a d-dimensional system with a $d' \geqslant d$-dimensional system,

$$|\psi\rangle = \sum_i \frac{|i\rangle|i\rangle}{\sqrt{d}}. \tag{3.11}$$

The new proof has the feature that it is *exponentially more efficient* from the point of view of classical communication than the protocol described in [18]. The argument runs as follows. By Corollary 4 we can find pure states $|\phi_i\rangle$ such that

$$|\phi\rangle = \sum_i \frac{|i\rangle|\phi_i\rangle}{\sqrt{d}}, \tag{3.12}$$

up to local unitary transformations. Define an operator on Bob's system,

$$F \equiv \sum_i |\phi_i\rangle\langle i|. \tag{3.13}$$

Ideally, we would apply F to the system B taking $|\psi\rangle$ directly to $|\phi\rangle$. This does not work because F is not unitary. Instead, we use F to define a quantum measurement with essentially the same effect. Define

$$E \equiv \frac{F}{\sqrt{\mathrm{tr}(F^\dagger F)}}. \tag{3.14}$$

Let $|0\rangle, \ldots, |d-1\rangle$ be the Schmidt basis for Bob's system. Define operators X and Z by

$$X|j\rangle \equiv |j \oplus 1\rangle; \quad Z|j\rangle \equiv \omega^j |j\rangle, \tag{3.15}$$

where $\oplus$ denotes addition modulo d, and ω is a dth root of unity. Define unitary operators $U_{s,t}$ by

$$U_{s,t} \equiv X^s Z^t. \tag{3.16}$$

The indices s and t are integers in the range 0 to $d-1$. By checking on an operator basis and applying linearity it is easily verified that for any Hermitian A,

$$\sum_{st} U_{s,t}^\dagger A U_{s,t} = \mathrm{tr}(A) I. \tag{3.17}$$

Therefore, defining $E_{s,t} \equiv E U_{s,t}$ gives

$$\sum_{st} E_{s,t}^\dagger E_{s,t} = I. \tag{3.18}$$

The set $\{E_{s,t}\}$ therefore defines a generalized measurement on Bob's system with d^2 outcomes. Suppose Bob performs this measurement. If he obtains the result (s,t) then the state of the system after the measurement is

$$\sum_i \frac{\omega^{it} |i\rangle |\phi_{i \oplus s}\rangle}{\sqrt{d}}. \tag{3.19}$$

Bob sends the measurement result to Alice, which requires $\lceil 2 \log_2 d \rceil$ bits of communication, and then Alice performs $X^s Z^{-t}$ (where X and Z are now defined with respect to Alice's Schmidt basis) on her system, giving the state

$$\sum_i \frac{|i \oplus s\rangle |\phi_{i \oplus s}\rangle}{\sqrt{d}}, \tag{3.20}$$

which is just $|\phi\rangle$.

This protocol for entanglement transformation requires only $\lceil 2 \log_2(d) \rceil$ bits of communication, compared with the protocol in [18], which required $d-1$. Another method [19] for achieving this result is as follows: Alice prepares locally a system $A'B'$ in a copy of $|\phi\rangle$. She then uses the shared maximal entanglement $|\psi\rangle$ with Bob to teleport [20] system B' to Bob, creating the desired state $|\phi\rangle$. Again, this protocol requires $\lceil 2 \log_2(d) \rceil$ bits of communication.

The present approach is interesting, in that it does not require knowledge of the teleportation protocol in order to succeed. Moreover, the method used strongly suggests that it may be possible to *always* perform the transformation using $O(\log_2 d)$ bits of communication, even when $|\psi\rangle$ is not maximally entangled, a result that does not appear obvious from the teleportation protocol. A method for doing so has recently been found using different methods, and will be reported elsewhere.

IV. CONCLUSION

The results reported here answer a fundamental question about the nature of the density matrix as a representation for ensembles of pure states, and give some elementary applications of this result to quantum mechanics and quantum information theory. I expect that the connection revealed here between majorization and ensembles of pure states will be of considerable use in future investigations of fundamental properties of quantum systems.

Note added in proof. Recently I learned that Theorem 3 was obtained by Ruskai in unpublished work (1993).

ACKNOWLEDGMENTS

I thank Sumit Daftuar and Andrew Landahl for pointing out some glitches in earlier versions of this work, and Armin Uhlmann for discussions on majorization. This work was supported by DARPA through the Quantum Information and Computing Institute (QUIC) administered through the ARO.

APPENDIX: UNITARY-STOCHASTIC MATRICES AND MAJORIZATION

In this appendix we outline the constructive steps in the proof of Theorem 1. To begin, we first take a slight detour connecting majorization with a class of matrices known as *T transforms*.

By definition, a T transform is a matrix which acts as the identity on all but two dimensions, where it has the form

$$T = \begin{bmatrix} t & 1-t \\ 1-t & t \end{bmatrix}, \tag{A1}$$

for some parameter t, $0 \leqslant t \leqslant 1$. The following result connects majorization and T transforms [7]:

Theorem 5: If $x \prec y$ there exists a finite set of T transforms $T_1, T_2, \ldots, T_n$ such that $x = T_1 T_2 \cdots T_n y$.

The converse of Theorem 5 is also true [7], but will not be needed. For convenience we provide details of the construction of the sequence $T_1, \ldots, T_n$ here.

Proof of Theorem 5: The result is proved by induction on d, the dimension of the vector space x and y live in. For notational convenience we assume that the components of x and y have been ordered into decreasing order; if this is not the case then one can easily reduce to this case by insertion of appropriate transposition matrices (which are T transforms). The result is clear when $d=2$, so let us assume the result is true for arbitrary d, and try to prove it for $(d+1)$-dimensional x and y.

Choose k such that $y_k \leqslant x_1 \leqslant y_{k-1}$. Such a k is guaranteed to exist because $x \prec y$ implies that $x_1 \leqslant y_1$ and $x_1 \geqslant x_{d+1} \geqslant y_{d+1}$. Choose t such that

$$x_1 = t y_1 + (1-t) y_k. \tag{A2}$$

Now define z to be the result of applying a T transform T with parameter t to the first and kth components of y, so that

$$z = Ty \tag{A3}$$

$$= (x_1, y'), \tag{A4}$$

where

$$y' \equiv (y_2, \ldots, y_{k-1}, (1-t)y_1 + ty_k, y_{k+1}, \ldots, y_{d+1}). \tag{A5}$$

Define $x' \equiv (x_2, x_3, \ldots, x_{d+1})$. It is not difficult to verify that $x' \prec y'$ (see [7] for details), and thus by the inductive hypothesis, $x' = T_1 \cdots T_r y'$ for some sequence of T transforms in d dimensions. But these T transforms can equally well be regarded as T transforms on $(d+1)$ dimensions by acting as the identity on the first dimension, and thus $x = T_1 \cdots T_r T y$, that is, x can be obtained from y by a finite sequence of T transforms, as we set out to show. Q.E.D.

Note that the inductive step of the proof of Theorem 5 can immediately be converted into an iterative procedure for constructing the matrices $T_1, \ldots, T_n$, and also implies that $n = d-1$ in a d-dimensional space. The proof of Theorem 1, which we now give, is also inductive in nature, and is easily converted into an iterative procedure for constructing an orthogonal matrix $u = (u_{ij})$ such that D defined by $D_{ij} \equiv u_{ij}^2$ satisfies Theorem 1. Note again the convention that expressions like u_{ij}^2 represent the square of the real number u_{ij}, not the ijth component of the matrix u^2.

To prove Theorem 1 we use the decomposition $x = T_1 T_2 \cdots T_n y$ from the proof of Theorem 5. The strategy is to use induction on n to prove that $T_1 T_2 \cdots T_n = (W_{ij}^2)$ for some orthogonal matrix W. Suppose $n = 1$. Omitting components on which T_1 acts as the identity, we have

$$T_1 = \begin{bmatrix} t & 1-t \\ 1-t & t \end{bmatrix} \tag{A6}$$

for some t, $0 \leqslant t \leqslant 1$. Define a unitary matrix U to act as the identity on all components on which T_1 acts as the identity, and as

$$U \equiv \begin{bmatrix} \sqrt{t} & -\sqrt{1-t} \\ \sqrt{1-t} & \sqrt{t} \end{bmatrix}, \tag{A7}$$

on the components where T_1 acts nontrivially. It is clear that $T_1 = (U_{ij}^2)$, as required.

To do the inductive step, suppose that products of n T transforms of the form used in the proof of Theorem 5 are orthostochastic, and consider the product $T_1 T_2 \cdots T_{n+1}$. We assume T_{n+2-k} acts on components k and component $d_k > k$, as per the proof of Theorem 5. Let P be the permutation matrix which transposes components 2 and d_1. (The following proof is more transparent if one assumes that $d_1 = 2$, and drops all reference to P, which is a technical device to make certain equations more compact.) Then

$$PT_{n+1}P = \begin{bmatrix} t & 1-t & 0 \\ 1-t & t & 0 \\ 0 & 0 & I_{d-2} \end{bmatrix}, \tag{A8}$$

where I_{d-2} is the $d-2$ by $d-2$ identity matrix. Furthermore, let us define a $d-1$ by $d-1$ matrix Δ by

$$T_1 T_2 \cdots T_n = \begin{bmatrix} 1 & 0 \\ 0 & \Delta \end{bmatrix}. \tag{A9}$$

By the inductive hypothesis there is a $d-1$ by $d-1$ orthogonal matrix U_{ij} such that $\Delta_{ij} = U_{ij}^2$. Define a new matrix U' by interchanging the role of the first and (d_1-1)th coordinates in U, $U' = P' U P'$, where P' transposes the first and (d_1-1)th coordinates, and similarly define Δ' by $\Delta' \equiv P' \Delta P'$. Then $\Delta'_{ij} = U'^2_{ij}$. Also we have

$$PT_1 T_2 \cdots T_n P = \begin{bmatrix} 1 & 0 \\ 0 & \Delta' \end{bmatrix}. \tag{A10}$$

Multiplying the previous equation by $PT_{n+1}P$ gives, from Eq. (A8) and the identity $P^2 = I$,

$$PT_1 T_2 \cdots T_{n+1} P = \begin{bmatrix} t & 1-t & 0 \\ (1-t)\vec{\delta} & t\vec{\delta} & \tilde{\Delta}, \end{bmatrix}, \tag{A11}$$

where $\vec{\delta}$ is the first column of Δ', and $\tilde{\Delta}$ is the $d-2$ by $d-1$ matrix that results when the first column of Δ' is removed. Let $\tilde{U}$ denote the $d-2$ by $d-1$ matrix that results when the first column of U' is removed, and let $\vec{u}$ denote the first column of U'. Define a d by d matrix V by

$$V \equiv \begin{bmatrix} \sqrt{t} & -\sqrt{1-t} & 0 \\ \sqrt{1-t}\vec{u} & \sqrt{t}\vec{u} & \tilde{U} \end{bmatrix}. \tag{A12}$$

We claim that V is an orthogonal matrix. To see this we need to show that the columns of V are of unit length and orthogonal. The length of the first column is

$$\sqrt{t + (1-t)\vec{u} \cdot \vec{u}} = \sqrt{1} = 1. \tag{A13}$$

A similar calculation shows that the second column is of unit length. The remaining columns are all of unit length since they are all columns of the unitary matrix U'. Simple algebra along similar lines can be used to check that the correct orthogonality relations between columns of V are satisfied. Observe that $PT_1 T_2 \cdots T_{n+1} P = (V_{ij}^2)$, so if we define $W \equiv PVP$, we see that W is an orthogonal matrix such that $T_1 T_2 \cdots T_{n+1} = (W_{ij}^2)$, which completes the induction.

[1] L. Landau, Z. Phys. **45**, 430 (1927).

[2] J. von Neumann, Göttinger Nachrichten, 245 (1927).

[3] E. Schrödinger, Proc. Cambridge Philos. Soc. **32**, 446 (1936).

[4] E.T. Jaynes, Phys. Rev. **108**, 171 (1957).

[5] L.P. Hughston, R. Jozsa, and W.K. Wootters, Phys. Lett. A **183**, 14 (1993).

[6] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications* (Academic, New York, 1979).

[7] R. Bhatia, *Matrix Analysis* (Springer-Verlag, New York, 1997).

[8] P. M. Alberti and A. Uhlmann, *Stochasticity and Partial Order: Doubly Stochastic Maps and Unitary Mixing* (Kluwer,

Dordrecht, 1982).
[9] A. Horn, Am. J. Math. **76**, 620 (1954).
[10] A. Uhlmann, Rep. Math. Phys. **1**, 147 (1970).
[11] O.E. Lanford and D. Robinson, J. Math. Phys. **9**, 1120 (1968).
[12] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).
[13] A. Uhlmann, Wiss. Z. Karl-Marx-Univ. Leipzig **20**, 633 (1971).
[14] A. Uhlmann, Wiss. Z. Karl-Marx-Univ. Leipzig **21**, 421 (1972).
[15] A. Uhlmann, Wiss. Z. Karl-Marx-Univ. Leipzig **22**, 139 (1973).
[16] A. Wehrl, Rev. Mod. Phys. **6**, 15 (1974).
[17] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Dordrecht, 1993).
[18] M.A. Nielsen, Phys. Rev. Lett. **83**, 436 (1999).
[19] H.-K. Lo and S. Popescu, Phys. Rev. Lett. **83**, 1459 (1999).
[20] C.H. Bennett *et al.*, Phys. Rev. Lett. **70**, 1895 (1993).