Inequalities for quantum entropy: A review with conditions for equality

Cite as: J. Math. Phys. 43, 4358 (2002); https://doi.org/10.1063/1.1497701
Submitted: 28 April 2002 • Accepted: 16 May 2002 • Published Online: 20 August 2002

Mary Beth Ruskai

Inequalities for quantum entropy: A review with conditions for equality

Mary Beth Ruskai[a)]
Department of Mathematics, University of Massachusetts Lowell, Lowell, Massachusetts 01854

(Received 28 April 2002; accepted for publication 16 May 2002)

This article presents self-contained proofs of the strong subadditivity inequality for von Neumann's quantum entropy, $S(\rho)$, and some related inequalities for the quantum relative entropy, most notably its convexity and its monotonicity under stochastic maps. Moreover, the approach presented here, which is based on Klein's inequality and Lieb's theorem that the function $A \rightarrow \mathrm{Tr}\, e^{K+\log A}$ is concave, allows one to obtain conditions for equality. In the case of strong subadditivity, which states that $S(\rho_{123})+S(\rho_2) \leqslant S(\rho_{12})+S(\rho_{23})$ where the subscripts denote subsystems of a composite system, equality holds if and only if $\log \rho_{123} = \log \rho_{12} - \log \rho_2 + \log \rho_{23}$. Using the fact that the Holevo bound on the accessible information in a quantum ensemble can be obtained as a consequence of the monotonicity of relative entropy, we show that equality can be attained for that bound only when the states in the ensemble commute. The article concludes with an Appendix giving a short description of Epstein's elegant proof of Lieb's theorem. [DOI: 10.1063/1.1497701]

TABLE OF CONTENTS

[a)]Electronic mail: MaryBeth_Ruskai@uml.edu

I. INTRODUCTION

A. Quantum entropy

Quantum information science[32] is the study of the information carrying and processing properties of quantum mechanical systems. Recent work in this area has generated renewed interest in the properties of the quantum mechanical entropy. It is interesting to note that von Neumann[45,46] introduced the notion of mixed state, represented by a density matrix ρ (a positive semi-definite operator with $\mathrm{Tr}\rho=1$), into quantum theory and defined its entropy as $S(\rho) \equiv -\mathrm{Tr}(\rho \log \rho)$ in 1927, well before the corresponding classical quantity was introduced in Shannon's seminal work[41] on ''The Mathematical Theory of Communication'' in 1948. (Admittedly, von Neumann's motivation was the extension of the classical theory of statistical mechanics, developed by Gibbs *et al.* to the quantum domain rather than the development of a theory of quantum communication.) Many fundamental properties of the quantum entropy were proved in a remarkable, but little-known, 1936 paper of Delbrück and Molèiere.[9] For further discussion of the history of quantum entropy, see Refs. 33, 38 and 47 and the introductory remarks in Ref. 40.

One important class of inequalities relates the entropy of subsystems to that of a composite system, whose Hilbert space is a tensor product $\mathcal{H}_{12}=\mathcal{H}_1 \otimes \mathcal{H}_2$ of the Hilbert spaces for the subsystems. When the state of the composite system is described by the density matrix ρ_{12}, the states of the subsystems are given by the reduced density matrices, e.g., $\rho_1 = T_2(\rho_{12})$, obtained by taking the partial trace. The subadditivity inequality

$$S(\rho_{12}) \leqslant S(\rho_1) + S(\rho_2) \tag{1}$$

was proved in Refs. 9 and 24. [It should not be confused with the concavity

$$S(x\rho' + (1-x)\rho'') \geqslant xS(\rho') + (1-x)S(\rho''), \tag{2}$$

which can actually be obtained *from* subadditivity by considering block matrices.[27,28,47]] In the more complex situation in which the composite system is composed of three subsystems, the following stronger inequality, known as strong subadditivity (SSA), holds:

$$S(\rho_{123}) + S(\rho_2) \leqslant S(\rho_{12}) + S(\rho_{23}). \tag{3}$$

This inequality was conjectured by Lanford and Robinson in Ref. 24 and proved in Refs. 28 and 29. In this article, we review its proof in a form that easily yields the following condition for equality.

Theorem 1: *Equality holds in strong subadditivity (3) if and only if*

$$\log \rho_{123} - \log \rho_{12} = \log \rho_{23} - \log \rho_2 . \tag{4}$$

We have suppressed implicit tensor products with the identity so that, e.g., $\log \rho_{12}$ means $(\log \rho_{12}) \otimes I_3$. Rewriting (4) as $\log \rho_{123} + \log \rho_2 = \log \rho_{12} + \log \rho_{23}$, multiplying by ρ_{123} and taking the trace immediately establishes the sufficiency of this equality condition. In Sec. IV, we will show that it is also necessary.

B. Relative entropy

The SSA inequality can be restated as a property of the *quantum relative entropy* which is defined as

$$H(\rho,\gamma)\equiv \mathrm{Tr}\rho(\log\rho-\log\gamma). \tag{5}$$

It is usually assumed that ρ, γ are density matrices, although (5) is well-defined for any pair of positive semi-definite matrices for which $\ker(\gamma)\subset\ker(\rho)$. Strong subadditivity can now be restated as

$$H(\rho_{12},\rho_2)\leqslant H(\rho_{123},\rho_{23}), \tag{6}$$

where we again write, e.g., ρ_{23} for $I_1\otimes\rho_{23}$. More generally, the relative entropy is monotone under completely positive, trace-preserving maps (also known as "quantum operations" [32] and "stochastic maps" [1,20] and discussed in more detail in Sec. III D), i.e.,

$$H[\Phi(\rho),\Phi(\gamma)]\leqslant H(\rho,\gamma). \tag{7}$$

This monotonicity implies (6) when $\Phi=T_3$ is the partial trace operation; perhaps surprisingly, the converse is also true.[31] This, and other connections between strong subadditivity and relative entropy, are discussed in Sec. V C.

The approach to SSA presented here can also be used to obtain conditions for equality in properties of relative entropy, including its joint convexity and monotonicity. The explicit statements are postponed to later sections. Since the monotonicity can be used to give a simple proof of the celebrated Holevo bound[14,32] on accessible information, we show how our results can be used to recover the equality conditions in that bound. As discussed in Sec. II C, Petz[33,36] has also obtained several equality conditions in different, but equivalent, forms. However, Theorem 8, which applies to the most general form of monotonicity, appears to be new.

C. Lieb's convex trace functions

One of the most frequently cited approaches to strong subadditivity is to present it as a consequence of the concavity of a quantity known as the Wigner–Yanase–Dyson entropy.[49] This property, conjectured by Bauman,[6] is equivalent to the joint concavity in A and B of the map

$$(A,B)\rightarrow \mathrm{Tr}A^sK^\dagger B^{(1-s)}K \quad \text{for} \quad A,B>0,\ \ 0<s<1 \tag{8}$$

(where † is used to denote the adjoint). Lieb's proof[26] of the concavity of the WYD function (8) and his realization of a connection between SSA and Bauman's concavity conjecture was a crucial breakthrough. However, concavity of the WYD function was only one of several concave trace functions studied in Ref. 26; the following result was also established by Lieb.

Theorem 2: *For any fixed self-adjoint matrix K, the function* $A\mapsto F(A)=\mathrm{Tr}\ e^{K+\log A}$ *is concave in* $A>0$.

This result played a fundamental role in the original proof[28,29] of SSA and the closely related property of joint concavity of the relative entropy.[28–30] Although SSA is a deep theorem, a complete proof is not as forbidding as is sometimes implied. Therefore, for completeness, we include Epstein's elegant proof[11] of Theorem 2 in the Appendix, and then follow the original strategies of Lieb and Ruskai[29] to show how it implies SSA.

D. Overview

Although this article grew out of questions about the conditions for equality in strong subadditivity and related inequalities, it seems useful to present these conditions within a more comprehensive exposition. For simplicity, we confine our discussion to finite dimensions, and assume that, unless otherwise stated, the density matrices under consideration are strictly positive.

The remainder of the article is structured as follows. In Sec. II we discuss some consequences and interpretations of the SSA equality condition. In Sec. III we summarize some mathematical results needed for the proofs in the sections that follow. Section IV, which might be regarded as the heart of the paper, presents the proof of strong subadditivity in a form which easily yields the equality conditions. (A reader primarily interested in this proof can proceed directly to Sec. IV with a willingness to accept the results of Sec. III.) Section V presents proofs with equality conditions for the monotonicity of the relative entropy under partial traces, the joint convexity of the relative entropy, and the general monotonicity under stochastic maps. This section also contains a discussion of the connection between these properties, SSA and their proofs. Section VI contains the proof of the equality conditions for monotonicity of relative entropy. Section VII considers bounds, most notably the Holevo bound, on the accessible information that can be extracted from an ensemble of quantum states, and the conditions under which they can be attained. The article concludes with some additional historical comments in Sec. VIII.

II. IMPLICATIONS OF THE EQUALITY CONDITIONS FOR SSA

A. Classical conditions

To describe the corresponding classical inequalities, let the subsystems A, B and C correspond to classical random variables. One can recover the classical Shannon entropy $-\Sigma_a p(a)\log p(a)$ from the von Neumann entropy by taking ρ to be a diagonal matrix with elements $p(a)$ on the diagonal. Employing a slight abuse of notation, we write $S[p(a)]$ for this quantity. Then the classical strong subadditivity inequality can be stated as

$$S[p(a,b,c)]+S[p(b)]\leqslant S[p(a,b)]+S[p(b,c)]. \tag{9}$$

The classical relative entropy of the distribution $q(a)$ with respect to $p(a)$ is $H[p(a),q(a)]=\Sigma_a p(a)\log p(a)/q(a)$. It is well-known (see, e.g., Ref. 23) that the convexity of the function $f(x)=x\log x$ implies that $H[p(a),q(a)]\geqslant 0$ and its strict convexity implies that equality holds if and only if $p(a)=q(a)\forall a$. (The generalization of this result to quantum situations is discussed in Sec. III A.)

The classical form (9) of SSA is equivalent to $H[p(a,b,c),q(a,b,c)]\geqslant 0$ when the second distribution is $q(a,b,c)=p(a,b)[p(b)]^{-1}p(b,c)$. Thus, equality holds in (9) if and only if

$$p(a,b,c)=p(a,b)[p(b)]^{-1}p(b,c)\quad\forall a,b,c, \tag{10}$$

which can be rewritten as

$$\log p(a,b,c)-\log p(a,b)=\log p(b,c)-\log p(b)\quad\forall a,b,c, \tag{11}$$

which is identical to what one would obtain from Theorem 1. Using $p(c|b)$ to denote the classical conditional probability distribution, (11) can be rewritten as

$$p(c|a,b)=p(c|b), \tag{12}$$

which is precisely the condition that the sequence $A\rightarrow B\rightarrow C$ forms a Markov chain.

B. Special cases of SSA equality

Some insight into equality condition (4) may be obtained by looking at special cases in which it is satisfied. The most obvious is when ρ_{123} is a tensor product of its three reduced density matrices. However, it is readily verified that (4) also holds when either $\rho_{123}=\rho_1\otimes\rho_{23}$ *or* $\rho_{123}=\rho_{12}\otimes\rho_3$. One can generalize this slightly further. If the subsystem 2 can be partitioned further into two subsystems $2'$ and $2''$, then one can verify equality holds if $\rho_{123}=\rho_{12'}\otimes\rho_{2''3}$, where $\rho_{12'}$ and $\rho_{2''3}$ are states of the composite systems 1, $2'$ and $2''$, 3 respectively.

However, such a decomposition into tensor products is not necessary; indeed, we have already seen that equality also holds for the case of classical Markov processes. Moreover, by comparison to (12) it is natural to regard (4) as a kind of quantum Markov condition. Thus, the conditions in Theorem 1 can also be viewed as a natural noncommutative analog of the conditions for equality in classical SSA. Another way of regarding (4) is as a concise statement of a subtle intertwining condition discussed below. Unfortunately, we have not found explicit examples which satisfy it other than the two classes discussed above, that is, a partial decomposition into tensor products or a classical Markov chain.

C. Petz's conditions

Using a completely different approach, Petz[33,36] gave conditions for equality in (7) when Φ can be identified with a mapping of an algebra onto a subalgebra, a situation which includes (6). In that case Petz's conditions become

$$\rho_{12}^{it}\rho_2^{-it} = \rho_{123}^{it}\rho_{23}^{-it}. \tag{13}$$

Taking the derivative of both sides of (13) at $t=0$ yields (4). Although (13) appears stronger than (4), it is not since, as noted above, (4) is sufficient for equality in (6). Moreover, since (4) implies

$$e^{it\log(\rho_{123})} = e^{it[\log\rho_{12} - \log\rho_{12} + \log\rho_{23}]}, \tag{14}$$

our results can be combined with those of Petz to see that equality holds in SSA$\Leftrightarrow$(4)$\Leftrightarrow$(13) and that any of these conditions suffices to imply

$$e^{it[\log\rho_{12} - \log\rho_2 + \log\rho_{23}]} = e^{it\log(\rho_{12})}e^{-it\log(\rho_2)}e^{it\log(\rho_{23})}. \tag{15}$$

Note that one can also relate Petz's conditions to those for equality in classical SSA by rewriting (10) as $p(a,b,c)[p(b,c)]^{-1} = p(a,b)[p(b)]^{-1}$ and then raising to the *it* power.

III. FUNDAMENTAL MATHEMATICAL TOOLS

A. Klein's inequality

The fact that the relative entropy is positive, i.e., $H(\rho,\gamma)\geqslant 0$ when $\mathrm{Tr}\rho=\mathrm{Tr}\gamma$, is an immediate consequence of the following fundamental convexity result due to Klein.[22,32,47]

Theorem 3 (Klein's inequality): *For A, B>0*

$$\mathrm{Tr}\ A(\log A - \log B) \geqslant \mathrm{Tr}(A-B), \tag{16}$$

with equality if and only if A=B.

The closely related Peierls–Bogoliubov inequality[33,47] is sometimes used instead of Klein's inequality. However, the equality conditions in Theorem 3 play a critical role in the sections that follow.

B. Lieb's golden corollary

The proofs in Sec. IV do not use Theorem 2 directly, but a related result generalizing the following inequality, which we will also need.

Theorem 4 (Golden–Thompson–Symanzik): *For self-adjoint matrices A and B,* $\mathrm{Tr}\ e^{A+B} \leqslant \mathrm{Tr}\ e^A e^B$ *with equality if and only if A and B commute.*

Although this inequality is extremely well-known, the conditions for equality do not appear explicitly in such standard references as Refs. 16, 42 and 47. However, one method of proof is based on the observation that $\mathrm{Tr}[e^{A/2^k}e^{B/2^k}]^{2^k}$ is monotone decreasing in k, yielding e^{A+B} in the limit as $k\to\infty$. The equality conditions then follow easily from those for the Schwarz inequality for the Hilbert–Schmidt inner product $\mathrm{Tr}C^\dagger D$. Indeed, $k=1$ yields

$$\mathrm{Tr}(e^{A/2}e^{B/2})(e^{A/2}e^{B/2}) \leqslant [\mathrm{Tr}\ e^{B/2}e^{A}e^{B/2}]^{1/2}[\mathrm{Tr}\ e^{A/2}e^{B}e^{A/2}]^{1/2} = \mathrm{Tr}\ e^A e^B$$

with $C=e^{B/2}e^{A/2}$ and $D=e^{A/2}e^{B/2}$. The equality condition that C is a multiple of D implies $e^{B/2}e^{A/2}=e^{A/2}e^{B/2}$ which holds if and only if A and B commute. One reference[33] that does discuss equality does so by making the interesting observation that (as shown in Ref. 37) Theorem 4 and its equality conditions can be derived as a consequence of the monotonicity of relative entropy, Theorem 7.

The natural extension to three matrices $\mathrm{Tr}\, e^{A+B+C} \leqslant |\mathrm{Tr}\, e^A e^B e^C|$, fails; see, for example, Problem 20 on pp. 512–513 of Ref. 16. Therefore, the following result of Lieb[26] is particularly noteworthy.

Theorem 5 (Lieb): *For any R,S,T>0*

$$\mathrm{Tr}\;\; e^{\log R-\log S+\log T} \geqslant \mathrm{Tr}\int_0^\infty R\,\frac{1}{S+uI}\,T\,\frac{1}{S+uI}\,du. \tag{17}$$

One might expect that equality holds if and only if R,S,T commute. Although this is sufficient, it is not necessary. One easily checks that both sides of (17) equal $\mathrm{Tr}\rho_1\otimes\rho_{23}$ when $R=\rho_1\otimes\rho_2\otimes I_3$, $S=I_1\otimes\rho_2\otimes I_3$, and $T=I_1\otimes\rho_{23}$, even when T does not commute with R or S.

Proof: Lieb's proof of (17) begins with the easily established fact[39] that if $F(A)$ is concave and homogeneous in the sense $F(xA)=xF(A)$, then

$$\lim_{x\to 0}\frac{F(A+xB)-F(A)}{x} \geqslant F(B). \tag{18}$$

Applying this to the functions in Theorem 2 with $A=S$, $B=T$, $K=\log R-\log S$ yields

$$\mathrm{Tr}\;\; e^{\log R-\log S+\log T} \leqslant \lim_{x\to 0}\frac{\mathrm{Tr}\;\; e^{\log R-\log S+\log(S+xT)}-\mathrm{Tr}R}{x}. \tag{19}$$

To complete the proof, we need the well-known integral representation

$$\log(S+xT)-\log S=\int_0^\infty \frac{1}{S+uI}\,xT\,\frac{1}{S+xT+uI}\,du. \tag{20}$$

Substituting (20) into (19) and noting that

$$\mathrm{Tr}\; e^{\log R+x\int_0^\infty[1/(S+uI)]T[1/(S+xT+uI)]du}=\mathrm{Tr}R+x\mathrm{Tr}R\int_0^\infty \frac{1}{S+uI}\,T\,\frac{1}{S+uI}\,du+O(x^2)$$

yields the desired result. Q.E.D.

C. Purification

Araki and Lieb[5,27] observed that one could obtain useful new entropy inequalities by applying what is now known as the ''purification process'' to known inequalities. Any density ρ_1 can be extended to a pure state density matrix ρ_{12} on a tensor product space; moreover, $S(\rho_1)=S(\rho_2)$. Applying this to the subadditivity inequality (1), i.e., $S(\rho_{12})\leqslant S(\rho_1)+S(\rho_2)$, yields the equivalent result $S(\rho_3)\leqslant S(\rho_{23})+S(\rho_2)$ which can be combined with (1) to give the triangle inequality[5,27]

$$|S(\rho_1)-S(\rho_2)|\leqslant S(\rho_{12})\leqslant S(\rho_1)+S(\rho_2). \tag{21}$$

By purifying ρ_{123} to ρ_{1234} one can similarly show that SSA (3) is equivalent to

$$S(\rho_4)+S(\rho_2)\leqslant S(\rho_{12})+S(\rho_{14}). \tag{22}$$

D. Lindblad's representation of stochastic maps

Stochastic maps arise naturally in quantum information as a description of the effect on a subsystem A interacting with the environment in the pure state $\gamma_B = |\psi_B\rangle\langle\psi_B|$ via the unitary operation U_{AB},

$$\rho_A \to \mathrm{Tr}_B(U_{AB}\rho_A \otimes \gamma_B U_{AB}^\dagger). \tag{23}$$

Lindblad[31] used Stinespring's representation to show that any completely positive trace-preserving map Φ which maps an algebra into itself can be represented as if it arose in this way. That is, given such a map Φ one can always find an auxiliary system, $\mathcal{H}_B$, a density matrix γ_B on $\mathcal{H}_B$, and a unitary map U_{AB} on the combined system $\mathcal{H}_A \otimes \mathcal{H}_B$ (where A denotes the original system) such that

$$\Phi(\rho) = \mathrm{Tr}_B(U_{AB}\rho \otimes \gamma_B U_{AB}^\dagger) \tag{24}$$

where Tr_B denotes the partial trace over the auxiliary system.

Using the Kraus representation $\Phi(\rho) = \Sigma_k F_k \rho F_k^\dagger$ (and noting that the requirement that Φ be trace-preserving is equivalent to $\Sigma_k F_k^\dagger F_k = I$), one can give a construction equivalent to Lindblad's by initially defining U_{AB} as

$$U_{AB}|\psi\rangle \otimes |\beta\rangle \equiv \sum_k F_k|\psi\rangle \otimes |k\rangle, \tag{25}$$

where $|\beta\rangle$ is a fixed normalized state of the auxiliary system, and $\{|k\rangle\}$ is some orthonormal basis for the auxiliary system. Then U_{AB} is a partial isometry from $\mathcal{H}_A \otimes |\beta\rangle\langle\beta|$ to $\mathcal{H}_A \otimes \mathcal{H}_B$ which can be extended to a unitary operator on all of $\mathcal{H}_A \otimes \mathcal{H}_B$. This yields (24) with $\gamma_B = |\beta\rangle\langle\beta|$ a pure state.

However, U_{AB} can also be extended to $\mathcal{H}_A \otimes \mathcal{H}_B$ in other ways. In particular, it can be extended, instead, to the partial isometry for which $U_{AB}^\dagger U_{AB}$ is the projection onto $\mathcal{H}_A \otimes |\beta\rangle\langle\beta|$ so that $U_{AB} = 0$ on the orthogonal complement of $\mathcal{H}_A \otimes |\beta\rangle\langle\beta|$. We describe this in more detail when Φ requires at most m Kraus operators F_k, in which case one can choose the auxiliary system to be $\mathbf{C}^m$. One can also choose $|k\rangle = |e_k\rangle$, and $|\beta\rangle = |e_1\rangle$ with $|e_k\rangle$ the standard basis of column vectors with elements $c_j = \delta_{jk}$. Then (25) depends only on the first column of U_{AB} which we denote V and regard as a map from $\mathcal{H}$ to $\mathcal{H} \otimes \mathbf{C}^m$. In block form

$$V_\rho V^\dagger = U_{AB}\rho \otimes |e_1\rangle\langle e_1| U_{AB}^\dagger = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix} \rho (F_1^\dagger \;\; F_2^\dagger \;\; \cdots \;\; F_m^\dagger) = \begin{pmatrix} F_1\rho F_1^\dagger & F_1\rho F_2^\dagger & \cdots & F_1\rho F_m^\dagger \\ F_2\rho F_1^\dagger & F_2\rho F_2^\dagger & \cdots & F_2\rho F_m^\dagger \\ \vdots & \vdots & & \vdots \\ F_m\rho F_1^\dagger & & \cdots & F_m\rho F_m^\dagger \end{pmatrix} \tag{26}$$

from which it easily follows that $\mathrm{Tr}_B(V\rho V^\dagger) = \Sigma_k F_k \rho F_k^\dagger = \Phi(\rho)$. The requirement that Φ be trace-preserving gives $V^\dagger V = \Sigma_k F_k^\dagger F_k = I$ which again implies that V is a partial isometry. Moreover, $V\rho V^\dagger$ has the same nonzero eigenvalues as $(V\sqrt{\rho})^\dagger (V\sqrt{\rho}) = \rho$ so that $S[V\rho V^\dagger] = S(\rho)$.

This construction can be readily extended to situations in which Φ maps operators acting on one Hilbert space $\mathcal{H}_A$ to those acting on another space $\mathcal{H}_{A'}$, e.g., $\Phi : \mathcal{B}(\mathcal{H}_A) \mapsto \mathcal{B}(\mathcal{H}_{A'})$. In this case, the Kraus operators $F_k : \mathcal{H}_A \mapsto \mathcal{H}_{A'}$, and U_{AB} is a partial isometry from $\mathcal{H}_A \otimes |\beta\rangle\langle\beta|$ to a subspace of $\mathcal{H}_{A'} \otimes \mathcal{H}_B$. Alternatively, V can be defined as a partial isometry from $\mathcal{H}_A$ to $\mathcal{H}_{A'} \otimes \mathbf{C}^m$.

E. Measurements and their representations

A von Neumann or *projective measurement* is a partition of the identity $I=\Sigma_b E_b$ into mutually orthogonal projections, i.e., $E_b E_c=\delta_{bc}E_b$. A positive operator valued measurement (POVM) is a set of positive semi-definite operators E_b such that $\Sigma_b E_b=I$, i.e., the orthogonality condition is dropped. It is well-known that a general POVM can be represented as a projective measurement on a tensor product space.[32]

In fact, by noting that the map $\rho\mapsto\Sigma_b\sqrt{E_b}\rho\sqrt{E_b}$ is completely positive and trace-preserving with Kraus operators $F_b=\sqrt{E_b}$ one can use the construction above. Write $V=\Sigma_b\sqrt{E_b}\otimes|b\rangle$ where $|b\rangle$ is an orthonormal basis for $\mathbf{C}^M$ and M is the number of measurements in the POVM, i.e., $b=1,\ldots,M$. Then $V\rho V^\dagger=\Sigma_{b,c}\sqrt{E_b}\rho\sqrt{E_c}\otimes|b\rangle\langle c|$. Now, if $F_b=I\otimes|b\rangle\langle b|$, then $\{F_b\}$ is a projective measurement on $\mathcal{H}\otimes\mathbf{C}^M$ and $\mathrm{Tr}F_b(V\rho V^\dagger)=\mathrm{Tr}E_b\rho$.

F. Adjoint maps

It is sometimes useful to consider the adjoint, which we denote $\widehat{\Phi}$, of a stochastic map Φ with respect to the Hilbert–Schmidt inner product $\langle A,B\rangle=\mathrm{Tr}A^\dagger B$. When Φ acts on $n\times n$ matrices, this adjoint (or dual) is fully defined by the requirement

$$\mathrm{Tr}[\Phi(A)]^\dagger B=\mathrm{Tr}A^\dagger\widehat{\Phi}(B) \tag{27}$$

for all $n\times n$ matrices, A,B. Indeed, when $\Phi(\rho)=\Sigma_k F_k\rho F_k^\dagger$, the adjoint is given by $\widehat{\Phi}(\rho)=\Sigma_k F_k^\dagger\rho F_k$. Moreover, Φ is trace-preserving if and only if $\widehat{\Phi}$ is unital, i.e., $\widehat{\Phi}(I)=I$. When Φ is the partial trace, T_2, its adjoint takes $A\mapsto A\otimes I_2$.

IV. SUBADDITIVITY PROOFS

To understand the proof of strong subadditivity, it is instructive to first understand how Klein's inequality can be used to prove two weaker inequalities. First, we consider the subadditivity inequality (1). Substituting $A=\rho_{12}$ and $B=\rho_1\otimes\rho_2$ into Klein's inequality (16) yields

$$-S(\rho_{12})+S(\rho_1)+S(\rho_2)\geqslant\mathrm{Tr}(\rho_{12}-\rho_1\otimes\rho_2)=0, \tag{28}$$

which is equivalent to subadditivity. Furthermore, the well-known conditions for equality in subadditivity follow from the conditions for equality in Klein's inequality, namely that equality holds if and only if ρ_{12} is a tensor product, that is, $\rho_{12}=\rho_1\otimes\rho_2$.

A second, more powerful subadditivity inequality was obtained by Araki and Lieb,[5]

$$S(\rho_{123})\leqslant S(\rho_{12})+S(\rho_{23}) \tag{29}$$

under the constraint $\mathrm{Tr}\rho_{123}=1$. To prove this, choose $A=\rho_{123}$ and $B=e^{\log\rho_{12}+\log\rho_{23}}$ in Klein's inequality to obtain

$$-S(\rho_{123})+S(\rho_{12})+S(\rho_{23})\geqslant 1-\mathrm{Tr}\;e^{\log\rho_{12}+\log\rho_{23}}. \tag{30}$$

Applying Theorem 4 to the right-hand side gives

$$-S(\rho_{123})+S(\rho_{12})+S(\rho_{23})\geqslant 1-\mathrm{Tr}_{123}\rho_{12}\rho_{23}=1-\mathrm{Tr}_2(\rho_2)^2\geqslant 1-\mathrm{Tr}_2\rho_2=0,$$

where the last line follows from $(\rho_2)^2\leqslant\rho_2$ (which is the *only* place the normalization condition $\mathrm{Tr}\rho_{123}=1$ is needed). Q.E.D.

The strategy for proving SSA is similar to that above, but with Theorem 4 replaced by Theorem 5. Let $A=\rho_{123}$ and choose B so that $\log B=\log\rho_{12}-\log\rho_2+\log\rho_{23}$. Then Klein's inequality implies

$$-S(\rho_{123})+S(\rho_{12})-S(\rho_2)+S(\rho_{23})\geqslant\mathrm{Tr}(\rho_{123}-e^{\log\rho_{12}-\log\rho_2+\log\rho_{23}}). \tag{31}$$

Applying Lieb's result (17) to the right-hand side above, we obtain

$$-S(\rho_{123})+S(\rho_{12})-S(\rho_2)+S(\rho_{23})\geqslant \mathrm{Tr}\left(\rho_{123}-\int_0^\infty \rho_{12}\frac{1}{\rho_2+uI}\,\rho_{23}\frac{1}{\rho_2+uI}\,du\right)$$

$$=\mathrm{Tr}_{123}\rho_{123}-\mathrm{Tr}_2\int_0^\infty \rho_2\frac{1}{\rho_2+uI}\,\rho_2\frac{1}{\rho_2+uI}\,du$$

$$=(\mathrm{Tr}_{123}\rho_{123}-\mathrm{Tr}_2\rho_2)=0.$$

This proves SSA. Moreover, this approach allows us to easily determine the conditions for equality, and thus complete the proof of Theorem 1. The first inequality in the derivation above is satisfied with equality if and only if $A=B$, which is just the condition (4). Although the conditions for equality in (17) are more difficult to analyze, this is not necessary here. When $A=B$, it immediately follows that $\mathrm{Tr}A=\mathrm{Tr}B$ so that the second inequality in the above derivation automatically becomes an equality when (4) holds.

V. INEQUALITIES FOR RELATIVE ENTROPY

A. Monotonicity under partial trace

We now show how the same strategy can be applied to obtain a proof with equality conditions for the monotonicity of relative entropy under partial trace.

Theorem 6: *When* $\rho_{12},\gamma_{12}>0$ *and* $\mathrm{Tr}\rho_{12}=\mathrm{Tr}\gamma_{12}$,

$$H(\rho_2,\gamma_2)\leqslant H(\rho_{12},\gamma_{12}) \tag{32}$$

with equality if and only if $\log\rho_{12}-\log\gamma_{12}=\log\gamma_2+\log\rho_2$.

This condition should be interpreted as $\log\rho_{12}-\log\gamma_{12}=I_1\otimes[\log\gamma_2-\log\rho_2]$. Since, as noted in Sec. III F, when $\Phi=T_1$, the action $\hat{\Phi}$ is precisely $I_1\otimes$, the equality condition can be written as $\log\rho_{12}-\log\gamma_{12}=\hat{T}_1[\log T_1(\gamma_{12})-\log T_1(\rho_{12})]$ which is a special case of the more general form (40) developed later.

SSA can be regarded as a special case of this monotonicity result via the correspondence $\rho_{12}\to\rho_{123}$, $\gamma_{12}\to\rho_{12}$, and Petz's form of the equality condition becomes $\rho_2^{it}\gamma_2^{-it}=\rho_{12}^{it}\gamma_{12}^{-it}$. It is interesting to note that in Ref. 29, Lieb and Ruskai actually obtained Eq. (32) *from* SSA using the convexity of the conditional entropy $S(\rho_1)-S(\rho_{12})$ and the inequality (18).

Proof: Let $A=\rho_{12}$, $\log B=\log\gamma_{12}-\log\gamma_2+\log\rho_2$. Then Klein's inequality and (17) imply

$$H(\rho_{12},\gamma_{12})-H(\rho_2,\gamma_2)\geqslant \mathrm{Tr}_{12}(\rho_{12}-e^{\log\gamma_{12}-\log\gamma_2+\log\rho_2})$$

$$\geqslant \mathrm{Tr}_{12}\left(\rho_{12}-\int_0^\infty \gamma_{12}\frac{1}{\gamma_2+uI}\,\rho_2\frac{1}{\gamma_2+uI}\,du\right)$$

$$=\mathrm{Tr}_{12}\rho_{12}-\mathrm{Tr}_2\int_0^\infty \gamma_2\frac{1}{\gamma_2+uI}\,\rho_2\,\frac{1}{\gamma_2+uI}\,du=\mathrm{Tr}_{12}\rho_{12}-\mathrm{Tr}_2\rho_2=0.$$

The equality condition is again precisely the condition $A=B$. Q.E.D.

B. Joint convexity of the relative entropy

The joint convexity of relative entropy can be obtained directly from Theorem 6 by choosing ρ_{12} (and similarly γ_{12}) to be a block diagonal matrix with blocks $\lambda_k\rho^{(k)}$ (and $\lambda_k\gamma^{(k)}$). We can interpret the partial trace as a sum over blocks so that $\rho\equiv\rho_2=\Sigma_k\lambda_k\rho^{(k)}$. However, it is worth giving a direct proof of the joint convexity since it demonstrates the central role of Theorem 2.

Theorem 7: *The relative entropy is jointly convex in its arguments, i.e., if* $\rho=\Sigma_k\lambda_k\rho^{(k)}$ *and* $\gamma=\Sigma_k\lambda_k\gamma^{(k)}$, *then*

$$H(\rho,\gamma)\leqslant\sum_k \lambda_k H(\rho^{(k)},\gamma^{(k)}) \tag{33}$$

with equality if and only if $\log\rho-\log\gamma=\log\rho^{(k)}-\log\gamma^{(k)}$ *for all* k.

Proof: Let $A=\rho^{(k)}$ and $\log B=\log\rho-\log\gamma+\log\gamma^{(k)}$ with $\rho=\Sigma_k\lambda_k\rho^{(k)}$ and $\gamma=\Sigma_k\lambda_k\gamma^{(k)}$. Then Klein's inequality implies

$$H(\rho^{(k)},\gamma^{(k)})-\mathrm{Tr}\rho^{(k)}[\log\rho-\log\gamma]\geqslant\mathrm{Tr}(\rho-e^{\log\rho-\log\gamma+\log\gamma^{(k)}}). \tag{34}$$

Multiplying this by λ_k with $\lambda_k>0$ and $\Sigma_k\lambda_k=1$ yields, after summation,

$$\begin{aligned}\sum_k \lambda_k H(\rho^{(k)},\gamma^{(k)})-H(\rho,\gamma)&\geqslant\mathrm{Tr}\left(\rho-\sum_k \lambda_k e^{\log\rho-\log\gamma+\log\gamma^{(k)}}\right)\\&\geqslant\mathrm{Tr}\left(\rho-e^{\log\rho-\log\gamma+\log\sum_k \lambda_k\gamma^{(k)}}\right)\\&=\mathrm{Tr}(\rho-e^{\log\rho})=0,\end{aligned}$$

where the second inequality is precisely the concavity of $C\to F(C)=\mathrm{Tr}\, e^{K+\log C}$ with $K=\log\rho-\log\gamma$ and $C=\Sigma_k\lambda_k\gamma^{(k)}$. Q.E.D.

C. Relationships among inequalities

We make some additional remarks about connections between SSA and various properties of relative entropy. To facilitate the discussion, we will use MONO to denote the general monotonicity inequality (7), MPT to denote the special case of monotonicity under partial traces, i.e., Theorem 6, and JC to denote the joint convexity, Theorem 7. Using the restatement of SSA in the form (6), it is easy to see that MONO⇒MPT⇒SSA. Before Theorem 7, we showed that MPT⇒JC. Similarly, by choosing ρ_{123} to be block diagonal with blocks ρ_{123}^k one can show that SSA implies that the map $\rho_{12}\mapsto S(\rho_1)-S(\rho_{12})$ is convex. In Ref. 29 it was observed that applying the convexity inequality (18) to this map (with $A+xB=\rho_{12}+x\gamma_{12}$) yields (32). This shows that SSA⇒MPT, so that we have the chain of implications

$$\text{MONO}\Rightarrow\text{MPT}\Leftrightarrow\text{SSA}\Rightarrow\text{JC}. \tag{35}$$

One can show that JC⇒MPT by using Uhlmann's observation[43] that the partial trace can be written as a convex combination of unitary transformations.

One can also show directly that JC⇒SSA by using the purification process described in Sec. III C to show that SSA is equivalent to

$$\rho_4+\rho_2\leqslant\rho_{12}+\rho_{14}. \tag{36}$$

Moreover, if ρ_{124} is pure, then $\rho_4=\rho_{12}$ and $\rho_2=\rho_{14}$ so that equality holds in (36). Since the extreme points of the convex set of density matrices are pure states, the inequality (36) then follows from the joint convexity, Theorem 7. Thus we have

$$\text{MONO}\Rightarrow\text{MPT}\Leftrightarrow\text{SSA}\Leftrightarrow\text{JC}. \tag{37}$$

Lindblad[31] completed this circuit by showing that MPT⇒MONO.

Using the representation described in Sec. III D, with V the partial isometry from $\mathcal{H}$ to $\mathcal{H}\otimes\mathbf{C}^m$ as in (26), one finds

$$H[\Phi(\rho),\Phi(\gamma)]=H[\mathrm{Tr}_B(V\rho V^\dagger),\ \ \mathrm{Tr}_B(V\gamma V^\dagger)]$$

$$\leq H[V\rho V^\dagger,V\gamma V^\dagger] \tag{38}$$

$$=H(\rho,\gamma). \tag{39}$$

since $\mathrm{Tr}V\rho V^\dagger \log(V\gamma V^\dagger)=\mathrm{Tr}\rho\log\gamma$ for a partial isometry V.

VI. EQUALITY IN MONOTONICITY UNDER STOCHASTIC MAPS

Conditions for equality in the general monotonicity inequality (7) may be more subtle since it is not always possible to achieve equality. Indeed, it was noted in Ref. 25 that $\sup_{\rho\neq\gamma}H[\Phi(\rho),\Phi(\gamma)]/H(\rho,\gamma)$ can be strictly less than 1. Using the reformulation (38) above, we prove the following result.

Theorem 8: *Equality holds in (7), $H[\Phi(\rho),\Phi(\gamma)]\leq H(\rho,\gamma)$, if and only if*

$$\log\rho-\log\gamma=\widehat{\Phi}[\log\Phi(\rho)-\log\Phi(\gamma)] \tag{40}$$

where $\widehat{\Phi}$ denotes the adjoint of Φ with respect to the Hilbert–Schmidt inner product as defined in (27).

To verify sufficiency, multiply (40) by ρ and take the trace to obtain

$$H(\rho,\gamma)=\mathrm{Tr}\rho\widehat{\Phi}[\log\Phi(\rho)-\log\Phi(\gamma)]=\mathrm{Tr}\Phi(\rho)[\log\Phi(\rho)-\log\Phi(\gamma)]=H[\Phi(\rho),\Phi(\gamma)].$$

It is tempting to follow our previous strategy and choose $A=\rho$, $\log B=\log\gamma+\widehat{\Phi}[\log\Phi(\rho)-\log\Phi(\gamma)]$. However, we have been unable to verify that $\mathrm{Tr}\,e^{\log\gamma+\widehat{\Phi}[\log\Phi(\rho)-\log\Phi(\gamma)]}\leq 1$ as required by this approach.

Instead, we use the representation (24) or (26). Rather than applying the equality conditions in Theorem 6 directly to (38), it is useful to repeat the argument for an appropriate choice of A and B.

Proof: Choose $A=V\rho V^\dagger$, $\log B=\log(V\gamma V^\dagger)+\log\mathrm{Tr}_2(V\rho V^\dagger)-\log\mathrm{Tr}_2(V\gamma V^\dagger)$ where V is again the partial isometry as in (26) of Sec. III D. B is defined so that the last two terms in $\log B$ are extended from $\mathcal{H}$ to $\mathcal{H}\otimes\mathbf{C}^m$ so that $\ker(B)\subset\ker(A)$. The condition for equality in (38) is then

$$\log(V\rho V^\dagger)-\log(V\gamma V^\dagger)=\log\mathrm{Tr}_2(V\rho V^\dagger)-\log\mathrm{Tr}_2(V\gamma V^\dagger)=\log\Phi(\rho)-\log\Phi(\gamma). \tag{41}$$

We can put this into a more useful form by noting that for a partial isometry V,

$$\log(V\rho V^\dagger)-\log(V\gamma V^\dagger)=V[\log\rho-\log\gamma]V^\dagger, \tag{42}$$

from which it follows that (41) is equivalent to

$$V[\log\rho-\log\gamma]V^\dagger=\log\Phi(\rho)-\log\Phi(\gamma). \tag{43}$$

Multiplying by $V^\dagger$ on the left and V on the right and using that $V^\dagger V=I$, one sees that (43) implies

$$\log\rho-\log\gamma=V^\dagger[\log\Phi(\rho)-\log\Phi(\gamma)]V. \tag{44}$$

Taking the partial trace Tr_2 over the auxiliary space in (44) yields (40) since $\widehat{\Phi}(P)=\sum_k F_k^\dagger P F_k=V^\dagger P V$ for all P in $\mathcal{H}$. Q.E.D.

Another useful necessary condition for equality in (7) can be obtained by multiplying both sides of (43) by the projection $VV^\dagger$. Since $V^\dagger V=I$, one finds

$$VV^\dagger[\log\Phi(\rho)-\log\Phi(\gamma)]=V[\log\rho-\log\gamma]V^\dagger=[\log\Phi(\rho)-\log\Phi(\gamma)]VV^\dagger, \tag{45}$$

i.e., the projection $VV^\dagger$ commutes with $[\log\Phi(\rho)-\log\Phi(\gamma)]$. Taking the partial trace and noting that $\Phi(I)=\mathrm{Tr}_2 VV^\dagger$ we can summarize this discussion in the following.

Corollary 9: If equality holds in (7), then

$$\Phi(\log\rho-\log\gamma)=\Phi(I)[\log\Phi(\rho)-\log\Phi(\gamma)]=[\log\Phi(\rho)-\log\Phi(\gamma)]\Phi(I). \tag{46}$$

Moreover, $\log\Phi(\rho)-\log\Phi(\gamma)$ *commutes with the projection* $VV^\dagger=\Sigma_{k,\ell}|k\rangle\langle\ell|F_kF_\ell^\dagger$ *where* $\{F_k\}$ *is a set of Kraus operators for* Φ, *i.e.,* $\Phi(\rho)=\Sigma_kF_k\rho F_k^\dagger$ *and* $|k\rangle$ *is an orthonormal basis for the auxiliary space* $\mathcal{H}_2$.

The results of this section also hold in the more general situation when $\Phi:\mathcal{B}(\mathcal{H}_A)\mapsto\mathcal{B}(\mathcal{H}_A')$ maps operators on one Hilbert space to those on another, in which case $F_k:\mathcal{H}_A)\mapsto\mathcal{H}_A'$.

VII. THE HOLEVO BOUND

A. Background

One reason for studying conditions for equality is that other results, such as Holevo's celebrated bound[14] on the accessible information, can be obtained rather easily from SSA or some form of the monotonicity of relative entropy. However, obtaining the corresponding conditions for equality is not as straightforward as one might hope because of the need to introduce an auxiliary system. Although Holevo's bound is quite general, it is often applied in situations where $\tilde{\rho}_j=\Phi(\rho_j)$ is the output of a noisy quantum channel Φ with input ρ_j. We use the tilde $\tilde{}$ as a reminder of this, as well as to ensure a distinction from other density matrices which arise.

For any fixed POVM and density matrix γ, $p(b)=\mathrm{Tr}(\gamma E_b)$ defines a classical probability distribution whose entropy we denote $S[\mathrm{Tr}\gamma E_b]$. The Holevo bound states that for any ensemble of density matrices $\mathcal{E}=\{\pi_j\tilde{\rho}_j\}$ with average density matrix $\tilde{\rho}=\Sigma_j\pi_j\tilde{\rho}_j$, the accessible information in the ensemble satisfies

$$I(\mathcal{E},\mathcal{M})\equiv S[\mathrm{Tr}\tilde{\rho}E_b]-\sum_j \pi_j S[\mathrm{Tr}\tilde{\rho}_jE_b] \tag{47}$$

$$\leq S(\tilde{\rho})-\sum_j \pi_j S(\tilde{\rho}_j) \tag{48}$$

for any POVM $\mathcal{M}=\{E_b\}$. If all of the $\tilde{\rho}_j$ commute, then it is easy to see that equality can be achieved by choosing the E_b to be the spectral projections which simultaneously diagonalize the density matrices $\tilde{\rho}_j$. We wish to show that this condition is also necessary, i.e., equality can only be achieved in (48) if all the $\tilde{\rho}_j$ commute.

It is known[21,50] that (48) can be obtained from (7). First, observe that

$$S(\tilde{\rho})-\sum_j \pi_j S(\tilde{\rho}_j)=\sum_j \pi_j H(\tilde{\rho}_j,\tilde{\rho}). \tag{49}$$

Now let $\Omega_{\mathcal{M}}$ be the map $\Omega_{\mathcal{M}}(A)=\Sigma_b|b\rangle\langle b|\mathrm{Tr}(AE_b)$ where $\mathcal{M}=\{E_b\}$. Then $\Omega_{\mathcal{M}}$ is a stochastic map of the special type known as a Q-C channel[15] and the Holevo bound (48) follows immediately from (49) and

$$H[\Omega_{\mathcal{M}}(\tilde{\rho}_j),\Omega_{\mathcal{M}}(\tilde{\rho})]\leq H(\tilde{\rho}_j,\tilde{\rho}). \tag{50}$$

B. Equality conditions

We will henceforth assume that $\{\pi_j,\tilde{\rho}_j\}$ is a fixed ensemble and seek conditions under which we can find a POVM satisfying the equality requirements. Since $\widehat{\Omega}_{\mathcal{M}}(D)=\Sigma_bE_b\langle b,Db\rangle$, applying Theorem 8 yields conditions for equality in (50). For equality in (48) these conditions must hold for every j and reduce to

$$\log \tilde{\rho}_j - \log \tilde{\rho} = \sum_b E_b \log \frac{\mathrm{Tr} E_b \tilde{\rho}_j}{\mathrm{Tr} E_b \tilde{\rho}} \; \forall j, \tag{51}$$

where this should be interpreted as a condition on $\ker(\tilde{\rho}_j)^\perp$ in which case all terms are well-defined. (Indeed, since the condition arises from the use of Klein's inequality and the requirement $A=B$, the operators in B must be defined to be zero on $\ker(A)$, which reduces to $\ker(\tilde{\rho}_j)$ in the situation considered here.) If the POVM $\{E_b\}$ consists of a set of mutually orthogonal projections, then it is immediate that the operators $Z_j \equiv \log \tilde{\rho}_j - \log \tilde{\rho}$ commute, since (51) can be regarded as the spectral decomposition of Z_j. To show that the $\tilde{\rho}_j$ themselves commute, observe that

$$\begin{aligned} 1 = \mathrm{Tr} \tilde{\rho}_j &= \mathrm{Tr}\; e^{\log \tilde{\rho} + [\log \tilde{\rho}_j - \log \tilde{\rho}]} \\ &\leq \mathrm{Tr} \tilde{\rho} e^{\log \tilde{\rho}_j - \log \tilde{\rho}} \\ &= \mathrm{Tr} \tilde{\rho} e^{\Sigma_b E_b \log (\mathrm{Tr} E_b \tilde{\rho}_j / \mathrm{Tr} E_b \rho)} \\ &= \mathrm{Tr} \tilde{\rho} \sum_b E_b \frac{\mathrm{Tr} E_b \tilde{\rho}_j}{\mathrm{Tr} E_b \tilde{\rho}} = \sum_b \mathrm{Tr} E_b \tilde{\rho}_j = 1, \end{aligned}$$

where we have used Theorem 4 with $A = \log \tilde{\rho}$, $B = \log \tilde{\rho}_j - \log \tilde{\rho}$, and the fact that for orthogonal projections $e^{\Sigma_b a_b E_b} = \Sigma_b e^{a_b} E_b$. The conditions for equality in Theorem 4 then imply that $\log \tilde{\rho}_j$ and $\log \tilde{\rho}$ commute for all j. Hence $\tilde{\rho}_j$ and $\tilde{\rho}_k$ also commute for all j,k when the POVM consists of mutually orthogonal projections.

Using King's observation in the next section, one can reduce the general case to that of projective measurements. However, we prefer to use the equality conditions to show directly that the elements of the POVM must be orthogonal. Moreover, the commutativity condition involving $VV^\dagger$ is reminiscent of the more sophisticated Connes cocyle approach used by Petz, and thus of some interest.

Since the Kraus operators for the Q-C map $\Omega_{\mathcal{M}}$ can be chosen as $F_{kb} = |b\rangle\langle k| \sqrt{E_b}$ where $|b\rangle$ and $|k\rangle$ are orthonormal bases, one finds

$$VV^\dagger = \sum_{b,c} \sum_{k,\ell} |b\rangle\langle c| \langle k \sqrt{E_b} \sqrt{E_c} \ell \rangle = \sum_{b,c} |b\rangle\langle c| \langle \phi \sqrt{E_b} \sqrt{E_c} \phi \rangle, \tag{52}$$

where $|\phi\rangle = \Sigma_k |k\rangle$. By (45), this must commute for all j with $\log \Omega_{\mathcal{M}}(\tilde{\rho}_j) - \log \Omega_{\mathcal{M}}(\tilde{\rho}_j)$ which can be written in the form $\Sigma_b z_{bj} |b\rangle\langle b|$ with $z_{bj} = \log(\mathrm{Tr} E_b \tilde{\rho}_j / \mathrm{Tr} E_b \rho)$. A diagonal operator of the form $\Sigma_b z_b |b\rangle\langle b|$ with all $z_b \neq 0$ will commute with the projection in (52) if and only if all off-diagonal terms are zero. This will hold if the POVM is a projective measurement, since then $\sqrt{E_b}\sqrt{E_c} = E_b E_c = E_b \delta_{bc}$. To see that this is necessary, note that the possibility that the vector ϕ is orthogonal to all E_b is precluded by the condition that $\Sigma_b E_b = I$. Moreover, since the orthonormal basis $|k\rangle$ is arbitrary, ϕ can be chosen to be arbitrary. The restriction that (51) hold only on $\ker(\tilde{\rho}_j)^\perp$ may permit some $z_{bj} = 0$; however, for each b there will always be at least one j for which $z_{bj} \neq 0$, and this suffices. Q.E.D.

One can obtain an alternate form of the equality conditions from Corollary 9. Since $\Phi(I) = \Sigma_b |b\rangle\langle b| \mathrm{Tr} E_b$, another necessary condition for equality in (48) is

$$\mathrm{Tr} E_b [\log \tilde{\rho}_j - \log \tilde{\rho}] = \mathrm{Tr} E_b (\log \mathrm{Tr} E_b \tilde{\rho}_j - \log \mathrm{Tr} E_b \tilde{\rho}) \quad \forall j, b. \tag{53}$$

Inserting this in (51) yields the requirement

$$\log \tilde{\rho}_j - \log \tilde{\rho} = \sum_b \frac{1}{\mathrm{Tr} E_b} E_b \mathrm{Tr} E_b [\log \tilde{\rho}_j - \log \tilde{\rho}], \tag{54}$$

which can be rewritten as

$$Z_j = \sum_b \frac{|E_b\rangle}{\mathrm{Tr}E_b}\langle E_b, Z_j\rangle \quad \forall j, \tag{55}$$

where $Z_j = \log\tilde{\rho}_j - \log\tilde{\rho}$ and the bra-ket now refer to the Hilbert–Schmidt inner product. This implies that $\Sigma_b|E_b\rangle\langle E_b|/\mathrm{Tr}E_b$ projects onto the span($\{Z_j\}$). However, this alone is not sufficient to imply that the E_b form a projective measurement.

C. Other approaches

Chris King has observed[19] that when the POVM is a projective measurement of the form $E_b = |b\rangle\langle b|$, one can obtain the Holevo bound from the joint convexity of relative entropy. Let $\beta(\tilde{\rho}) = \Sigma_b|b\rangle\langle b|\mathrm{Tr}E_b\tilde{\rho}$. Then applying Theorem 7 to $H[\tilde{\rho},\beta(\tilde{\rho})]$ yields

$$-S(\tilde{\rho}) + S(\mathrm{Tr}E_b\tilde{\rho}) \leqslant \sum_j \pi_j[-S(\tilde{\rho}_j) + S(\mathrm{Tr}E_b\tilde{\rho}_j)] \tag{56}$$

or

$$S(\mathrm{Tr}E_b\tilde{\rho}) - \sum_j \pi_j S(\mathrm{Tr}E_b\tilde{\rho}_j) \leqslant S(\tilde{\rho}) - \sum_j \pi_j S(\tilde{\rho}_j)$$

with equality if and only if

$$\log\tilde{\rho} - \sum_b |b\rangle\langle b|\log \mathrm{Tr}E_b\tilde{\rho} = \log\tilde{\rho}_j - \sum_b |b\rangle\langle b|\log \mathrm{Tr}E_b\tilde{\rho}_j \quad \forall j. \tag{57}$$

This is equivalent to (51) when $E_b = |b\rangle\langle b|$, and the argument can be extended to more general projective measurements.

King also pointed out that if $\{E_b\}$ is an arbitrary POVM, the construction in Sec. III E can be used to show that (48) and (51) are equivalent to the equalities obtained when $\tilde{\rho}_j$ is replaced by $V\tilde{\rho}_jV^\dagger$ and E_b by F_b. Since the $\{F_b\}$ form a projective measurement, we can conclude from the argument above that equality implies that all $V\tilde{\rho}_jV^\dagger$ commute, which implies that all $\tilde{\rho}_j$ also commute since $V^\dagger V = I$.

It should be noted that Petz was able to use his equality conditions to find the conditions for equality in the Holevo bound and this is sketched in Ref. 34. Indeed, Petz's analog of (57) is $\tilde{\rho}^{it}D^{-it} = \tilde{\rho}_j^{it}D_j^{-it}\forall j$ where D, D_j denotes the diagonal parts of $\tilde{\rho}$, $\tilde{\rho}_j$, respectively. Then

$$\tilde{\rho}_j^{it} = \tilde{\rho}^{it}D^{-it}D_j^{it}. \tag{58}$$

Since (58) holds for all real t, as well as all j, it also implies $\tilde{\rho}_j^{-it} = \tilde{\rho}^{-it}D^{it}D_j^{-it}$. However, taking the adjoint of (58) yields $\tilde{\rho}_j^{-it} = D_j^{-it}D^{it}\tilde{\rho}^{-it}$. Therefore, $\tilde{\rho}^{-it}$ commutes with the diagonal matrix $D^{it}D_j^{-it} = D_j^{-it}D^{it}$ and must also be diagonal. This gives a simultaneous diagonalization of all $\tilde{\rho}_j^{it}$ which means that all $\tilde{\rho}_j$ commute.

Holevo's original longer derivation[14] of the bound (48) also concluded that commutativity was necessary and sufficient for equality. Some simplifications of this argument were given by Fuchs[12] in his thesis.

D. Another bound on accessible information

When ρ is a density matrix, the mapping $A \mapsto \rho^{-1/2}A\rho^{-1/2}$ and its inverse gives a duality between ensembles and POVMs. Hall[13] observed that this duality can be used to give another upper bound on the accessible information (47) in terms of the POVM and average density ρ, i.e.,

$$I(\mathcal{E},\mathcal{M}) \leqslant S(\rho) - \sum_b \tau_b S\left(\frac{1}{\tau_b}\sqrt{\rho}E_b\sqrt{\rho}\right) \tag{59}$$

$$= \sum_b \tau_b H\left(\frac{1}{\tau_b}\sqrt{\rho}E_b\sqrt{\rho},\rho\right), \tag{60}$$

where $\tau_b = \text{Tr } E_b\rho$. This inequality can be obtained from the monoticity of relative entropy under the Q-C map $\Omega_{\mathcal{E}}(A) = \Sigma_j |j\rangle\langle j| \pi_j \rho^{-1/2}\rho_j\rho^{-1/2}$ applied to $H\,((1/\tau_b)\sqrt{\rho}E_b\sqrt{\rho},\rho)$ as in (50); or as in Ref. 21 where an equivalent bound was given. The argument in Sec. VII B can then be used to show that equality can be achieved in (59) if and only if all $\sqrt{\rho}E_b\sqrt{\rho}$ commute. Hall[13] also found this condition and noted that it implies that ρ commutes with every E_b in the POVM.

One is often interested in (48) and (59) when one wants to optimize the accessible information after using a noisy quantum channel, Φ. It was observed in Ref. 21 that, since $\text{Tr}\Phi(\rho_j)E_b = \text{Tr}\rho_j \widehat{\Phi}(E_b)$, one can regard the noise as either acting to transform pure inputs ρ_j to mixed state outputs $\Phi(\rho_j)$ or as acting through the adjoint $\widehat{\Phi}$ on the POVM with uncorrupted outputs. In the first case, one can bound the right side of (59) by choosing the E_b to be the spectral projections of the average output state $\Phi(\rho)$ to yield $I[\Phi(\mathcal{E}),\mathcal{M}] \leqslant S[\Phi(\rho)]$ which is weaker than the corresponding Holevo bound. Moreover, since the optimal choice for $\Phi(\rho_j)$ need not be in the image of Φ, it is not necessarily achievable even though the commutativity condition holds. Hall[13] discussed other situations in which the bound can not be achieved despite the fact that all $\sqrt{\rho}E_b\sqrt{\rho}$ commute.

Viewing the noise as acting on the POVM, King and Ruskai[21] defined

$$U_{EP}(\Phi) = \sup_{\rho,\mathcal{M}}\left[S(\rho) - \sum_b \tau_b S\left(\frac{1}{\tau_b}\sqrt{\rho}\,\widehat{\Phi}(E_b)\sqrt{\rho}\right)\right] \tag{61}$$

with $\tau_b = \text{Tr}\rho\widehat{\Phi}(E_b) = \text{Tr}\Phi(\rho)E_b$. If the supremum in (61) is achieved with an average density and POVM for which $\sqrt{\rho}\,\widehat{\Phi}(E_b)\sqrt{\rho}$ do not commute, then $U_{EP}(\Phi)$ is strictly greater than the accessible information. The questions of whether or not (61) can actually exceed the optimal accessible information, and how it might then be interpreted, are under investigation.

VIII. CONCLUDING REMARKS

The proof presented here for each inequality, SSA, Theorem 6, Theorem 7 and the general monotonicity (7), is quite short—only half a page using results from Sec. III which require less than one additional page *and* Theorem 2. However, as shown in the Appendix, even this result does *not* require a long argument if one is permitted to use some powerful tools of complex analysis.

It is certainly not unusual to find that complex analysis can be extremely useful, even when the functions of interest are real-valued. Indeed, Lieb's original proof of the concavity of WYD entropy used a complex interpolation argument. In his influential book[42] on trace ideals, Simon (extracting ideas from Uhlmann[44]) gave a longer ''elementary'' proof using the Schwarz inequality, perhaps inadvertently reinforcing the notion that any complete proof of SSA is long and forbidding. Similar ideas are implicit in Ando,[3] who restates the result in terms of tensor product spaces and block matrices. Uhlmann[44] again demonstrated the power of complex interpolation by using it to prove the monotonicity of relative entropy under trace-preserving maps which satisfy the slightly weaker condition of two-positivity (rather than complete positivity). SSA then follows immediately as a special case. However, Uhlmann's approach, which has been extended by Petz,[35,33] was developed within the framework of the relative modular operator formalism introduced by Araki[4,7,33] for much more general situations. Recently, Lesniewski and Ruskai[25] observed that within this relative modular operator framework, monotonicity can be established directly using an argument based on the Schwarz inequality.

The approach of this review is similar to that of Wehrl[47] in that we view Theorem 2 as the "essential ingredient." Indeed, Uhlmann,[43,47] using a completely different approach, had independently recognized that Theorem 2 would imply SSA. However, Wehrl's otherwise excellent review stated (at the end of Sec. III B) that "Unfortunately, the proof of [this] is not easy at all." Later (in Sec. III C) Wehrl again states that "...the proof is surprisingly complicated. I want to indicate only that the concavity of Tr $e^{K+\log A}$ can be obtained from Lieb's theorem [on concavity of the WYD entropy] through a sequence of lemmas." Although aware that Epstein's approach,[11] which was developed shortly after Lieb announced his results, permitted a "direct" proof of Theorem 2, Wehrl does not seem to have fully appreciated it. The utility of Epstein's technique may have been underestimated, in part, because he presented his results in a form which applied to the full collection of convex trace functions studied in Ref. 26. Checking Epstein's hypotheses for the WYD function requires some nontrivial mapping theorems. This may have obscured the elegance of the argument in Appendix A.

It is worth noting that if the concavity of WYD entropy is regarded as the key result, it is not necessary to use the long sequence of lemmas Wehrl refers to in order prove SSA. Lindblad[30] gave a direct proof of the joint convexity, Theorem 7, directly by differentiating the WYD function. Once this is done, SSA follows via the purification argument sketched after Eq. (36) or, alternatively, the variant of Uhlmann's argument described in Refs. 42 and 47. Combining this with Lieb's original complex interpolation proof of the concavity of the WYD function yields another "short" proof of SSA, albeit one which does not appear to be well-suited to establishing conditions for equality.

Finally, we mention that Carlen and Lieb[8] obtained another proof of SSA by using Epstein's technique to prove some Minkowski type inequalities for L_p trace norms. Using a different approach, King[17,18] recently proved several additivity results for the minimal entropy and Holevo capacity of a noisy channel by using L_p inequalities in which Epstein's technique provided a critical estimate. This suggests that connections with L_p inequalities, as advocated by Amosov, Holevo and Werner,[2] may be a promising avenue for studying entropy and capacity in quantum information. Despite the results mentioned above, many open conjectures remain; see Refs. 2, 8, 17, 18, and 48 for further details.

ACKNOWLEDGMENTS

This work was partially supported by the National Security Agency (NSA) and Advanced Research and Development Activity (ARDA) under Army Research Office (ARO) Contract No. DAAG55-98-1-0374 and by the National Science Foundation under Grant Nos. DMS-9706981 and DMS-0074566. This article is the result of questions first posed to the author by M. Nielsen and it is a pleasure to acknowledge the value of extensive discussions with him. Some of these discussions took place, and parts of this article were written, when MBR and M. Nielsen were visiting the Institute for Theoretical Physics at the University of California, Santa Barbara, and thereby also partly supported by the National Science Foundation under Grant No. PHY-9907949. The author is extremely grateful to C. King for careful readings of earlier versions of the manuscript leading to many useful comments, and for communicating (and allowing to be included) several critical observations needed to obtain the equality conditions for the Holevo bound in Sec. VII. It is also a pleasure to thank C. Fuchs for explaining the history of the equality conditions in the Holevo bound, E. Lieb for clarifying remarks about Bauman's responsibility for the actual conjecture of the concavity of the WYD function, D. Petz for discussions about his approach, P. Shor for asking a question which stimulated the analysis in Sec. VII D, and M. J. W. Hall for bringing Ref. 13 to the author's attention.

APPENDIX: EPSTEIN'S PROOF OF CONCAVITY OF $A \rightarrow \mathrm{Tr}\, e^{K+\log A}$

Let $f(x) = \mathrm{Tr}\, e^{K+\log(A+xB)}$ with $A>0$ strictly positive and K, B self-adjoint. For sufficiently small x, the function $f(x)$ is well-defined and the concavity of $F(A)$ in Theorem 2 follows immediately if $f''(0)<0$ for all choices of $B=B^*$.

Instead of dealing with f directly, Epstein considered the function $g(x)=xf(x^{-1})$ which is well-defined for $|x|>\mu^{-1}\equiv\|A^{-1}\|\|B\|$ and can be analytically continued to the upper half plane so that

$$g(z)=\mathrm{Tr}\ e^{K+\log(zA+B)}. \tag{A1}$$

There are a number of equivalent (when meaningful) ways of defining functions of matrices. For the purposes needed here it is natural to assume that the spectrum $\sigma(A)$ of the operator A is contained in the domain of an analytic function $F(z)$ and that

$$F(A)=\frac{1}{2\pi i}\oint\frac{F(z)}{zI-A}dz. \tag{A2}$$

One can then use the spectral mapping theorem $\sigma[F(A)]\subset F[\sigma(A)]$ for an appropriate sequence of functions to verify that

$$\begin{aligned}\Im z>0 \Rightarrow{}& \Im\ \omega(zA+B) > 0\\ \Rightarrow{}& \pi > \Im\ \omega[\log(zA+B)] > 0\\ \Rightarrow{}& \pi > \Im\ \omega[K+\log(zA+B)] > 0\\ \Rightarrow{}& \Im\ \omega[e^{K+\log(zA+B)}] > 0\\ \Rightarrow{}& \Im\ \mathrm{Tr}\ e^{K+\log(zA+B)} > 0,\end{aligned}$$

where $\Im$ denotes the imaginary part of a complex number and ω is used to denote an arbitrary element of the spectrum of the indicated operator. Thus, $g(z)$ maps the upper half plane into the upper half plane. Functions with this property have been studied extensively under various names, including, "operator monotone," "Herglotz" or "Pick." (See, for example, Refs. 3, 10 and 33.) It then follows that g has an integral representation of the form

$$g(z)=a+bz+\int_{-\mu}^{\mu}\frac{1}{t-z}dm(t) \tag{A3}$$

for some positive measure $\mu(t)$. This yields (via the change of variables $s=t^{-1}$)

$$f(x)=ax+b+\int_{-\mu}^{\mu}\frac{x^2}{tx-1}dm(t). \tag{A4}$$

Differentiation under the integral sign can then be used to establish that $f''(0)<0$ as desired by observing $x^2/(tx-1)=t^{-2}[(xt+1)+(xt-1)^{-1}]$. Q.E.D.

[1] Alberti, P. M. and Uhlmann, A., *Stochasticity and Partial Order* (Kluwer, Dordrecht, 1982).
[2] Amosov, G. G., Holevo, A. S., and Werner, R. F., "On some additivity problems in quantum information theory," Probl. Inf. Transm. **36**, 305–313 (2000).
[3] Ando,T., *Topics on Operator Inequalities*, Sapporo Lecture Notes (1978).
[4] Araki, H., "Relative entropy of state of von Neumann algebras," Publ. RIMS Kyoto Univ. **9**, 809–833 (1976).
[5] Araki, H. and Lieb, E., "Entropy inequalities," Commun. Math. Phys. **18**, 160–170 (1970).
[6] Bauman, F., "Bemerkungen úber quantenmechanische entropie ungleichungen," Helv. Phys. Acta **44**, 95–100 (1971).
[7] Bratteli, O. and Robinson, D., *Operator Algebras and Quantum Statistical Mechanics*, Vol. II (Springer-Verlag, New York, 1981; 2nd edition, 1997).
[8] Carlen, E. and Lieb, E., "A Minkowski type trace inequality and strong subadditivity of quantum entropy," in *Advances in Math. Sciences*, Am. Math. Soc. Trans. **189**, 59–62 (1999).
[9] Delbrürk, M. and Molèiere, G., "Statistische quantenmechanik und thermodynamik," Abh. Preuss. Akad. Wissenschaften **1**, 1–42 (1936).
[10] Donoghue, W., *Monotone Matrix Functions and Analytic Continuations* (Springer-Verlag, New York, 1974).

[11] Epstein, H., ''Remarks on two theorems of E. Lieb,'' Commun. Math. Phys. **31**, 317–325 (1973).
[12] Fuchs, C. A., ''Distinguishability and Accessible Information in Quantum Theory,'' Ph.D. thesis, University of New Mexico, Albuquerque, NM, 1996. See quant-ph/9601020.
[13] Hall, M. J. W., ''Quantum information and correlation bounds,'' Phys. Rev. A **55**, 100–112 (1997).
[14] Holevo, A. S., ''Information theoretical aspects of quantum measurement,'' Probl. Inf. Transm. USSR **9**, 31–42 (1973).
[15] Holevo, A. S., ''Quantum coding theorem,'' Russ. Math. Surveys **53**, 1295–1331 (1999); ''Coding Theorems for Quantum Channels,'' preprint, lanl:quant-ph/9809023.
[16] Horn, R. A. and Johnson, C. R., *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, 1991).
[17] King, C., ''Maximum of capacity and p-norms for some product channels,'' J. Math. Phys. **43**, 1247–1260 (2002); quant-ph/0103086.
[18] King, C., ''Additivity for a class of unital qubit channels,'' J. Math. Phys. (in press); quant-ph/0103156.
[19] King, C., private communication.
[20] King, C. and Ruskai, M. B., ''Minimal entropy of states emerging from noisy quantum channels,'' IEEE Trans. Inf. Theory **47**, 1–19 (2001).
[21] King, C. and Ruskai, M. B., ''Capacity of quantum channels using product measurements,'' J. Math. Phys. **42**, 87–98 (2001); quant-ph/0004062.
[22] Klein, O., ''Zur quantenmechanischen begründung des zweiten hauptsatzes der wärmelehre,'' Z. Phys. **72**, 767–775 (1931).
[23] Kullback, S., *Information Theory and Statistics* (Wiley, 1959; Dover, New York, 1968).
[24] Lanford, O. and Robinson, D., ''Mean entropy of states in quantum statistical mechanics,'' J. Math. Phys. **9**, 1120–1125 (1968).
[25] Lesniewski, A. and Ruskai, M. B., ''Relative entropy and monotone Riemannian metrics on noncommutative probability space,'' J. Math. Phys. **40**, 5702–5724 (1999).
[26] Lieb, E., ''Convex trace functions and the Wigner-Yanase-Dyson conjecture,'' Adv. Math. **11**, 267–288 (1973).
[27] Lieb, E., ''Some convexity and subadditivity properties of entropy,'' Bull. Am. Math. Soc. **81**, 1–13 (1975).
[28] Lieb, E. and Ruskai, M. B., ''A fundamental property of quantum mechanical entropy,'' Phys. Rev. Lett. **30**, 434–436 (1973).
[29] Lieb, E. and Ruskai, M. B., ''Proof of the strong subadditivity of quantum mechanical entropy,'' J. Math. Phys. **14**, 1938–1941 (1973).
[30] Lindblad, G., ''Expectations and entropy inequalities,'' Commun. Math. Phys. **39**, 111–119 (1974).
[31] Lindblad, G., ''Completely positive maps and entropy inequalities,'' Commun. Math. Phys. **40**, 147–151 (1975).
[32] Nielsen, M. A., and Chuang, I. L., *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[33] Ohya, M., and Petz, D., *Quantum Entropy and Its Use* (Springer-Verlag, New York, 1993).
[34] Ohya, M., Petz, D., and Watanabe, N., ''On capacities of quantum channels,'' Prob. Math. Stats. **17**, 170–196 (1997).
[35] Petz, D., ''Quasi-entropies for finite quantum systems,'' Rep. Math. Phys. **23**, 57–65 (1986).
[36] Petz, D., ''Sufficient subalgebras and the relative entropy of states of a von Neumann algebra,'' Commun. Math. Phys. **105**, 123–131 (1986).
[37] Petz, D., ''A variational expression for the relative entropy,'' Commun. Math. Phys. **114**, 345–349 (1988).
[38] Petz, D., ''Entropy, von Neumann and the von Neumann Entropy,'' in *John von Neumann and the Foundations of Quantum Physics*, edited by M. Rdei and M. Stltzner (Kluwer, Dordrecht, 2001).
[39] Rockafellar, R. T., *Convex Analysis* (Princeton University Press, Princeton, NJ, 1972).
[40] Ruskai, M. B. and Stillinger, F. H., ''Convexity inequalities for estimating free energy and relative entropy,'' J. Phys. A **23**, 2421–2437 (1990).
[41] Shannon, C., ''The mathematical theory of communication,'' Bell Syst. Tech. J. **27**, 379–423 (1948); **27**, 623–656 (1948).
[42] Simon, B., *Trace Ideals and Their Applications* (Cambridge University Press, Cambridge, 1979).
[43] Uhlmann, A., ''Endlich dimensionale dichtmatrizen, II,'' Wiss. Z.-Karl-Marx-Univ. Leipzig, Math.-Naturwiss. Reihe **22**(2), 139 (1973).
[44] Uhlmann, A., ''Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory,'' Commun. Math. Phys. **54**, 21–32 (1977).
[45] von Neumann, J., ''Thermodynamik quantenmechanischer gesamheiten,'' Gött. Nach. **1**, 273–291 (1927).
[46] von Neumann, J., *Matheatische Grundlagen der Quantenmechanik* (Berlin, 1932); English translation by R. T. Beyer, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1955).
[47] Wehrl, A., ''General properties of entropy,'' Rev. Mod. Phys. **50**, 221–260 (1978).
[48] Werner, R. F. and Holevo, A. S., ''Counterexample to an additivity conjecture for output purity of quantum channels,'' J. Math. Phys. (in press); preprint lanl:quant-ph/0203003.
[49] Wigner, E. P. and Yanase, M. M., ''Information content of distributions,'' Proc. Natl. Acad. Sci. U.S.A. **49**, 910–918 (1963).
[50] Yuen, H. P. and Ozawa, M., ''Ultimate information carrying limit of quantum systems,'' Phys. Rev. Lett. **70**, 363–366 (1993).