The Inequalities of Quantum Information Theory

Nicholas Pippenger*
(nicholas@cs.ubc.ca)

Department of Computer Science
The University of British Columbia
Vancouver, British Columbia V6T 1Z4
CANADA

Abstract: Let ϱ denote the density matrix of a quantum state having n parts $1, \ldots, n$. For $I \subseteq N = \{1, \ldots, n\}$, let $\varrho_I = \mathrm{Tr}_{N \setminus I}(\varrho)$ denote the density matrix of the state comprising those parts i such that $i \in I$, and let $S(\varrho_I)$ denote the von Neumann entropy of the state ϱ_I. The collection of $\nu = 2^n$ numbers $\{S(\varrho_I)\}_{I \subseteq N}$ may be regarded as a point, called the *allocation of entropy* for ϱ, in the vector space $\mathbf{R}^\nu$. Let $\mathcal{A}_n$ denote the set of points in $\mathbf{R}^\nu$ that are allocations of entropy for n-part quantum states. We show that $\overline{\mathcal{A}_n}$ (the topological closure of $\mathcal{A}_n$) is a closed convex cone in $\mathbf{R}^\nu$. Lieb and Ruskai have established a number of inequalities for multipartite quantum states (strong subadditivity and weak monotonicity). We give a set of independent linear inequalities from which all of these can be deduced by taking positive linear combinations. Let $\mathcal{B}_n$ denote the polyhedral cone in $\mathbf{R}^\nu$ determined by these inequalities. We show that $\overline{\mathcal{A}_n} = \mathcal{B}_n$ for $n \leq 3$. The status of this equality is open for $n \geq 4$. We also consider the case of *weakly symmetric* quantum states ϱ, for which $S(\varrho_I)$ depends on I only through the number $i = \#I$ of indices in I, and thus may be denoted $S(\varrho_i)$. The collection of $n+1$ numbers $\{S(\varrho_i)\}_{0 \leq i \leq n}$ may be regarded as a point, called the *symmetric* allocation of entropy for ϱ, in $\mathbf{R}^{n+1}$. We define the set $\mathcal{C}_n$ of points in $\mathbf{R}^{n+1}$ that are symmetric allocations of entropy in the obvious way. We give a set of independent linear inequalities determining a polyhedral cone $\mathcal{D}_n$ in $\mathbf{R}^{n+1}$, and show that $\overline{\mathcal{C}_n} = \mathcal{D}_n$ for all n. Thus if there are additional inequalities beyond the known inequalities of Lieb and Ruskai for states with $n \geq 4$, their symmetrizations must be positive linear combinations of these known inequalities.

Keywords: quantum information, von Neumann entropy, linear inequalities

* The work reported here was supported by an NSERC Research Grant and a Canada Research Chair.

1. Introduction

Central to information theory is the notion of the entropy $H(X)$ of a random variable X, introduced by Shannon [S1] in 1948. Let X be a random variable. That is, let Ω be a finite sample space, and let

$$p_x = \Pr[X = x],$$

where $p_x \geq 0$ and $\sum_{x\in\Omega} p_x = 1$, be the probability distribution for X on Ω. Shannon's entropy is defined by

$$H(X) = -\sum_{x\in\Omega} p_x \log p_x.$$

We agree as usual that $0 \log 0 = 0$. In the body of this paper, we shall take logarithms to base 2, which corresponds to measuring entropy in bits.

There is a notion of the entropy $S(\varrho)$ of a quantum state ϱ, introduced by von Neumann [N] in 1927, which actually contains Shannon's notion as a special case. Let ϱ be the density matrix of an quantum state. That is, let ϱ be a self-adjoint matrix whose rows and columns are indexed by Ω, and whose eigenvalues $\{\lambda_x\}_{x\in\Omega}$ satisfy $\lambda_x \geq 0$ and $\sum_{x\in\Omega} \lambda_x = 1$. Von Neumann's entropy is defined by

$$S(\varrho) = -\sum_{x\in\Omega} \lambda_x \log \lambda_x.$$

If f is a numerical function, and $\varrho = \sum_{x\in\Omega} \lambda_x |\psi_x\rangle\langle\psi_x|$, where $\{|\psi_x\rangle\}_{x\in\Omega}$ are normalized eigenvectors of ϱ, we shall agree that $f(\varrho) = \sum_{x\in\Omega} f(\lambda_x)|\psi_x\rangle\langle\psi_x|$. With this convention, von Neumann's entropy can be written

$$S(\varrho) = -\mathrm{Tr}(\varrho \log \varrho),$$

where $\mathrm{Tr}(\sigma)$ denotes the trace of the matrix σ. Clearly, diagonal density matrices (called *classical* states) correspond to probability distributions, and von Neumann's entropy reduces to Shannon's entropy in this case.

We shall be dealing in this paper with *multipartite* systems. Let $X = (X_1, \ldots, X_n)$ be an n-component random variable. That is, let $\Omega = \Omega_1 \times \cdots \times \Omega_n$, and let $N = \{1, \ldots, n\}$. For $I \subseteq N$, let $X_I = (X_i)_{i\in I}$ denote the $(\#I)$-component random variable obtained by projecting X onto $\Omega_I = \prod_{i\in I} \Omega_i$. The probability distribution function p_I of X_I is given by

$$p_{I,y} = \sum_{x\in\Omega, \pi(x)=y} p_x$$

for $y \in \Omega_I$, where $\pi : \Omega \to \Omega_I$ is the canonical projection. With each set $I \subseteq N$ of components, we may associate the entropy $H(X_I)$. The collection $\{H(X_I)\}_{I \subseteq N}$ will be called the *allocation of entropy* for the random variable X. The entropy $H(X_\emptyset)$ of the empty set of components is always 0, but we shall include it in the allocation of entropy for technical convenience, as it makes the discussion of inequalities more systematic. The allocation of entropy is a collection of $\nu = 2^n$ numbers, and thus it may be regarded as a vector in the space $\mathbf{R}^\nu$. Let $\mathcal{A}_n^{\text{class}} \subseteq \mathbf{R}^\nu$ denote the set of allocations of entropy of n-component probability distributions.

Now let ϱ be an n-part quantum state. That is, let $\Omega = \Omega_1 \times \cdots \times \Omega_n$. For $I \subseteq N$, let ϱ_I denote the $(\#I)$-part quantum state obtained by tracing over all parts of ϱ other than those in I. The density matrix ϱ_I is given by

$$\langle x|\varrho_I|y\rangle = \sum_{\substack{w,z\in\Omega,\pi(w)=x \\ \pi(z)=y,\tau(w)=\tau(z)}} \langle w|\varrho|z\rangle$$

for $x, y \in \Omega_I$, where $\tau : \Omega \to \Omega_{N\setminus I}$ is the canonical projection. Since each entry of ϱ_I is obtained from ϱ by taking the trace of a block of ϱ, we shall also use the "partial trace" notation, $\varrho_I = \mathrm{Tr}_{N\setminus I}(\varrho)$. With each set $I \subseteq N$ of parts, we may associate the entropy $S(\varrho_I)$. The collection $\{S(\varrho_I)\}_{I\subseteq N}$ will be called the *allocation of entropy* for the quantum state ϱ. Again we include the entropy $S(\varrho_\emptyset) = 0$ of the empty set of parts, and again we regard the allocation of entropy as a vector in $\mathbf{R}^\nu$. Let $\mathcal{A}_n \subseteq \mathbf{R}^\nu$ denote the set of allocations of entropy of n-part quantum states.

We shall say that an n-component random variable X is *weakly symmetric* if $H(X_I)$ depends on I only through the number $\#I$ of elements in I. (This may come about because X is actually symmetric; that is, because $\Omega_1 = \cdots = \Omega_n$ and

$$\Pr[X_1 = \omega_1, \ldots, X_n = \omega_n] = \Pr[X_1 = \omega_{\sigma(1)}, \ldots, X_n = \omega_{\sigma(n)}]$$

for any permutation σ of N. But weak symmetry requires only that the allocation of entropy be symmetric, and this is clearly a much weaker condition.) For $0 \le i \le n$, the allocation of entropy for a weakly symmetric random variable X has $\binom{n}{i}$ equal entries $H(X_I)$ for $\#I = i$. This allocation of entropy can thus be abbreviated $\{H(X_i)\}_{0\le i\le n}$, where $X_i = X_{\{1,\ldots,i\}}$. We shall refer to this abbreviation as the *symmetric* allocation of entropy for the weakly symmetric random variable X. The symmetric allocation of entropy is a collection of $n+1$ numbers, and thus it may be regarded as a vector in the space $\mathbf{R}^{n+1}$. Let $\mathcal{C}_n^{\text{class}} \subseteq \mathbf{R}^{n+1}$ denote the set of symmetric allocations of entropy of n-component weakly symmetric probability distributions.

We shall say that an n-part quantum state ϱ is *weakly symmetric* if $S(\varrho_I)$ depends on I only through the number $\#I$ of elements in I. For $0 \le i \le n$, the allocation of entropy for a weakly symmetric quantum state ϱ has $\binom{n}{i}$ equal entries $S(\varrho_I)$ for $\#I = i$. This allocation of entropy can thus be abbreviated $\{S(\varrho_i)\}_{0\le i\le n}$, where $\varrho_i = \varrho_{\{1,\ldots,i\}}$. We shall refer to this abbreviation as the *symmetric* allocation of entropy for the weakly symmetric quantum state ϱ. Again we regard the symmetric allocation of entropy as a vector in $\mathbf{R}^{n+1}$. Let $\mathcal{C}_n \subseteq \mathbf{R}^{n+1}$ denote the set of symmetric allocations of entropy of n-component weakly symmetric quantum states.

We shall be concerned in this paper with properties of the sets $\mathcal{A}_n$ and $\mathcal{C}_n$, and our discussion will yield as by-products some known properties of their classical counterparts $\mathcal{A}_n^{\text{class}}$ and $\mathcal{C}_n^{\text{class}}$. In Section 2, we shall show that the topological closures $\overline{\mathcal{A}_n}$, $\overline{\mathcal{C}_n}$, $\overline{\mathcal{A}_n^{\text{class}}}$ and $\overline{\mathcal{C}_n^{\text{class}}}$ of $\mathcal{A}_n$, $\mathcal{C}_n$, $\mathcal{A}_n^{\text{class}}$ and $\mathcal{C}_n^{\text{class}}$ are convex cones. The result for $\overline{\mathcal{A}_n^{\text{class}}}$ is due to Zhang and Yeung [Z1]. In Section 3, we shall present the known inequalities (due to Lieb and Ruskai [L2]) governing the allocation of entropy for quantum states. These inequalities determine a convex cone $\mathcal{B}_n$ in $\mathbf{R}^\nu$ and, upon symmetrization, a convex cone $\mathcal{D}_n$ in $\mathbf{R}^{n+1}$. We shall distinguish among these inequalities a set of *basic* inequalities, which have the property that all others can be obtained from them by taking positive linear combinations, and we prove that these basic inequalities are *independent*, in the sense that none can be obtained from the others by taking positive linear combinations. We also give corresponding results for a set of inequalities, identified by Fujishige [F] as *polymatroid* inequalities, governing the allocation of entropy for random variables. These polymatroid inequalities are simple consequences of the following properties of the logarithm: $\log x$ is a concave and non-decreasing function of x, and $\log 1 = 0$. The polymatroid inequalities determine a convex cone $\mathcal{B}_n^{\text{class}}$ in $\mathbf{R}^\nu$ and, upon symmetrization, a convex cone $\mathcal{D}_n^{\text{class}}$ in $\mathbf{R}^{n+1}$. In Section 4, we show that $\overline{\mathcal{A}_n} = \mathcal{B}_n$ and $\overline{\mathcal{A}_n^{\text{class}}} = \mathcal{B}_n^{\text{class}}$ for $n \le 3$. The classical result is implicit in the work of Han [H2], and was obtained explicitly by Zhang and Yeung [Z1]. It is an open question whether $\overline{\mathcal{A}_n} = \mathcal{B}_n$ for $n \ge 4$, but Zhang and Yeung [Z2] have shown that $\overline{\mathcal{A}_n^{\text{class}}} \ne \mathcal{B}_n^{\text{class}}$ for $n \ge 4$ by giving an explicit example of an inequality not implied by the polymatroid inequalities for 4-component random variables. In Section 5, we shall consider the symmetric case and show that $\overline{\mathcal{C}_n} = \mathcal{D}_n$ and $\overline{\mathcal{C}_n^{\text{class}}} = \mathcal{D}_n^{\text{class}}$ for all n. The classical result is again implicit in the work of Han [H1]. This implies that, if there are any further inequalities governing the allocation of entropy for quantum states, their symmetrizations must be positive linear combinations of known inequalities.

We shall assume throughout this paper that Ω is finite, so that random variable assume a finite number of values and quantum states are in finite dimensional spaces. This is done for technical convenience, and all the result presented can be extended to countably infinite Ω. In the quantum case, this extension is not entirely trivial, but the methods required do not affect the issues discussed in this paper.

2. Convex Cones

Let $\mathcal{E}$ be a set of points in $\mathbf{R}^k$. We shall say that $\mathcal{E}$ is a *convex cone* if

(CC_1) for every $X \in \mathcal{E}$ and every real $\lambda \geq 0$, we have $\lambda X \in \mathcal{E}$; and

(CC_2) for every $X \in \mathcal{E}$ and $Y \in \mathcal{E}$, and every real $0 \leq \lambda \leq 1$, we have $\lambda X + (1-\lambda) Y \in \mathcal{E}$.

Our goal in this section is to show that $\mathcal{A}_n$, $\mathcal{C}_n$, $\mathcal{A}_n^{\text{class}}$ and $\mathcal{C}_n^{\text{class}}$ are convex cones.

Let $\|X\|_\infty$ denote the norm $\|X\|_\infty = \max_{1 \leq L \leq M} |X_L|$. Since $\mathbf{R}^k$ is a finite dimensional vector space, this norm gives rise to the same topology (and in particular, to the same notion of closed sets) as the usual Euclidean norm, $\|X\|_2 = \left(\sum_{1 \leq L \leq M} X_L^2\right)^{1/2}$.

We shall say that a set $\mathcal{E}$ of points in $\mathbf{R}^k$ is *additive* if, for every $X \in \mathcal{E}$ and $Y \in \mathcal{E}$, we have $X + Y \in \mathcal{E}$. We shall say that $\mathcal{E}$ is *approximately diluable* if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every $X \in \mathcal{E}$ and real $0 \leq \lambda \leq \delta$, there exists a point $Y \in \mathcal{E}$ such that $\|\lambda X - Y\|_\infty \leq \varepsilon$.

Proposition 2.1: If $\mathcal{E}$ is a convex cone, then $\mathcal{E}$ is additive and approximately diluable.

Proof: Suppose that $\mathcal{E}$ is a convex cone. If $X \in \mathcal{E}$ and $Y \in \mathcal{E}$, then $\frac{1}{2}X + \frac{1}{2}Y \in \mathcal{E}$ by (CC_2) with $\lambda = \frac{1}{2}$, and thus $X + Y \in \mathcal{E}$ by (CC_1) with $\lambda = 2$. This shows that $\mathcal{E}$ is additive. Furthermore, if $X \in \mathcal{E}$ and $\lambda \geq 0$, we may take $Y = \lambda X \in \mathcal{E}$ by (CC_1), and have $\|\lambda X - Y\|_\infty \leq \varepsilon$ for any $\varepsilon \geq 0$. This shows that $\mathcal{E}$ is approximately diluable. $\triangle$

The converse to Proposition 2.1 is false: with $k = 1$, the set of non-negative rational numbers is clearly additive and is approximately diluable with $\delta = \varepsilon$, since the non-negative rationals are dense in the non-negative reals; but the non-negative rationals do not satisfy either condition for a convex cone. We do, however, have the following approximate converse.

Let $\overline{\mathcal{E}}$ denote the topological closure of $\mathcal{E}$.

Theorem 2.2: If $\mathcal{E}$ is additive and approximately diluable, then $\overline{\mathcal{E}}$ is a convex cone.

Proof: Suppose that $\mathcal{E}$ is additive and approximately diluable. Suppose that we are given $A \in \overline{\mathcal{E}}$ and real $\mu \geq 0$. We shall show that, for every $\eta > 0$, there is a point $Y \in \mathcal{E}$ such that $\|\mu A - Y\|_\infty \leq \eta$. This will imply that $\overline{\mathcal{E}}$ satisfies (CC_1). Take $\varepsilon = \eta/(\mu+1)$, and take $\delta > 0$ as in the definition of approximate diluability. Let $U \in \mathcal{E}$ be such that $\|A - U\|_\infty \leq \varepsilon$. By additivity of $\mathcal{E}$ and a simple induction on m, we have $mU \in \mathcal{E}$ for every positive integer m.

Take $m = \lceil \mu/\delta \rceil$, and take $\lambda = \mu/m$. Then $X = mU \in \mathcal{E}$, and by approximate diluability there exists $Y \in \mathcal{E}$ such that $\|\lambda X - Y\|_\infty = \|\mu U - Y\|_\infty \le \varepsilon$. By the triangle inequality, we have

$$\|\mu A - Y\|_\infty \le \|\mu A - \mu U\|_\infty + \|\mu U - Y\|_\infty \le \mu\varepsilon + \varepsilon = \eta,$$

which completes the proof of (CC_1).

Now suppose that we are given $X \in \overline{\mathcal{E}}$, $Y \in \overline{\mathcal{E}}$ and real $0 < \lambda < 1$. We shall show that for every $\zeta > 0$, there is a point $Z \in \mathcal{E}$ such that $\|\lambda X + (1-\lambda)Y - Z\|_\infty \le \zeta$. This will imply that $\overline{\mathcal{E}}$ satisfies (CC_2). Applying the result of the previous paragraph with $A = X$, $\mu = \lambda$ and $\eta = \lambda\zeta$, we obtain a point $V \in \mathcal{E}$ such that $\|\lambda X - V\|_\infty \le \lambda\zeta$. Applying the same result with $A = Y$, $\mu = 1 - \lambda$ and $\eta = (1-\lambda)\zeta$, we obtain a point $W \in \mathcal{E}$ such that $\|(1-\lambda)Y - W\|_\infty \le (1-\lambda)\zeta$. By additivity, we have $Z = V + W \in \mathcal{E}$, and by the triangle inequality, we have

$$\|\lambda X + (1-\lambda)Y - Z\|_\infty \le \|\lambda X - V\|_\infty + \|(1-\lambda)Y - W\|_\infty \le \lambda\zeta + (1-\lambda)\zeta = \zeta,$$

which completes the proof of (CC_2). $\triangle$

Recall that $\mathcal{A}_n \subseteq \mathbf{R}^\nu$, where $\nu = 2^n$, denotes the set of allocations of entropy of n-part quantum states.

Theorem 2.3: The set $\overline{\mathcal{A}_n}$ is a convex cone.

Proof: By Proposition 2.2, it will suffice to show that $\mathcal{A}_n$ is additive and approximately diluable.

Suppose that the n-part quantum state ϱ_X has rows and columns indexed by $\Omega_X = \prod_{1\le i\le n} \Omega_{X,i}$, and allocation of entropy $\{S(\varrho_{X,I})\}_{I\subseteq N}$, and suppose that ϱ_Y has rows and columns indexed by $\Omega_Y = \prod_{1\le i\le n} \Omega_{Y,i}$, and allocation of entropy $\{S(\varrho_{Y,I})\}_{I\subseteq N}$. We shall construct an n-part quantum state ϱ_Z, with rows and columns indexed by $\Omega_Z = \prod_{1\le i\le n} \Omega_{Z,i}$, and allocation of entropy $\{S(\varrho_{Z,I})\}_{I\subseteq N}$, where

$$S(\varrho_{Z,I}) = S(\varrho_{X,I}) + S(\varrho_{Y,I}) \qquad (2.1)$$

for all $I \subseteq N$. This will prove additivity.

To do this we take $\Omega_{Z,i} = \Omega_{X,i} \times \Omega_{Y,I}$ for all $1 \le i \le n$, and take $\varrho_Z = \varrho_X \otimes \varrho_Y$. If the eigenvalues of $\varrho_{X,I}$ are $\{\lambda_{X,I,x}\}_{x\in\Omega_{X,I}}$ and the eigenvalues of $\varrho_{Y,I}$ are $\{\lambda_{Y,I,y}\}_{y\in\Omega_{Y,I}}$, then the eigenvalues of $\varrho_{X,I} \otimes \varrho_{Y,I}$ are $\{\lambda_{X,I,x}\,\lambda_{Y,I,y}\}_{x\in\Omega_X, y\in\Omega_Y}$. This implies (2.1) by the definition of von Neumann's entropy, and completes the proof of additivity.

Now suppose that $\varepsilon > 0$ and take $\delta > 0$ sufficiently small that $\delta \leq 1/2$ and $h(\delta) \leq \varepsilon$, where

$$h(p) = -p \log p - (1-p) \log(1-p).$$

Suppose that $\lambda \leq \delta$ and that the n-part quantum state ϱ_X has rows and columns indexed by $\Omega_X = \prod_{1 \leq i \leq n} \Omega_{X,i}$, and allocation of entropy $\{S(\varrho_{X,I})\}_{I \subseteq N}$. We shall construct an n-part quantum state ϱ_Y, with rows and columns indexed by $\Omega_Y = \prod_{1 \leq i \leq n} \Omega_{Y,i}$, and allocation of entropy $\{S(\varrho_{Y,I})\}_{I \subseteq N}$, where

$$\|\lambda S(\varrho_{X,I}) - S(\varrho_{Y,I})\|_\infty \leq \varepsilon \tag{2.2}$$

for all $I \subseteq N$. This will prove approximate diluability.

To do this, we take ω to be a new element not in $\bigcup_{1 \leq i \leq n} \Omega_{X,i}$, and take $\Omega_{Y,i} = \Omega_{X,i} \cup \{\omega\}$ for all $1 \leq i \leq n$. Let z denote the element $(\omega, \ldots, \omega)$ in Ω_Y. Define the density matrix ϱ_Y by

$$\langle \psi_x | \varrho_Y | \psi_y \rangle = \begin{cases} \lambda \langle \psi_x | \varrho_X | \psi_y \rangle, & \text{if } x, y \in \Omega_X; \\ (1-\lambda), & \text{if } x = y = \omega; \\ 0, & \text{otherwise.} \end{cases}$$

If the eigenvalues of $\varrho_{X,I}$ are $\Lambda = \{\lambda_{X,I,x}\}_{x \in \Omega_{X,I}}$, then the eigenvalues of $\varrho_{Y,I}$ are $\Lambda \cup K$, where K contains the eigenvalue $1 - \lambda$ with multiplicity 1 and the eigenvalue 0 with multiplicity $\#\Omega_{X,I} - 1$. From the definition of von Neumann's entropy, we have

$$S(\varrho_{Y,I}) = \lambda S(\varrho_{X,I}) + h(\lambda)$$

for all $I \subseteq N$. Since h is an increasing function on the interval $[0, 1/2]$, this implies (2.2). This completes the proof of approximate diluability, and thus of the theorem. $\triangle$

Let $\mathcal{C}_n \subseteq \mathbf{R}^{n+1}$ denote the set of symmetric allocations of entropy of n-part weakly symmetric quantum states.

Corollary 2.4: The set $\overline{\mathcal{C}_n}$ is a convex cone.

Proof: We observe that the constructions showing that $\mathcal{A}_n$ is additive and approximately diluable take weakly symmetric states into weakly symmetric states. Thus they also establish that $\mathcal{C}_n$ is additive and approximately diluable, from which the corollary follows by Theorem 2.2. $\triangle$

Recall that $\mathcal{A}_n^{\text{class}} \subseteq \mathbf{R}^\nu$, where $\nu = 2^n$, denotes the set of allocations of entropy of n-component probability distributions.

Corollary 2.5: The set $\overline{\mathcal{A}_n^{\text{class}}}$ is a convex cone.

Proof: We observe that the constructions showing that $\mathcal{A}_n$ is additive and approximately diluable take classical states (represented by diagonal density matrices) into classical states. Thus they also establish that $\mathcal{A}_n^{\text{class}}$ is additive and approximately diluable, from which the corollary follows by Theorem 2.2. $\triangle$

Let $\mathcal{C}_n^{\text{class}} \subseteq \mathbf{R}^{n+1}$ denote the set of symmetric allocations of entropy of n-component weakly symmetric probability distiributions.

Corollary 2.5: The set $\overline{\mathcal{C}_n^{\text{class}}}$ is a convex cone.

Proof: We observe that the constructions showing that $\mathcal{A}_n$ is additive and approximately diluable take weakly symmetrical classical states into weakly symmetrical classical states. Thus they also establish that $\mathcal{C}_n^{\text{class}}$ is additive and approximately diluable, from which the corollary follows by Theorem 2.2. $\triangle$

3. Basic Inequalities

In this section we shall describe the known inequalities governing allocations of quantum entropy. These inequalities were established by Lieb and Ruskai [L2], though some special cases were prove earlier by Araki and Lieb [A2]. Our goal is to distinguish among them a set of *basic* inequalities, which have the property that all others can be obtained from them by taking positive linear combinations, and to prove that these basic inequalities are *independent*, in the sense that none of them can be expressed as a positive linear combination of the others. Simple modifications of our arguments will give corresponding results for the analogous set of inequalities (the *polymatroid* inequalities) governing allocations of classical entropy.

In 1973, Lieb and Ruskai [L2] established the inequalities

$$S(\varrho_{123}) + S(\varrho_3) \leq S(\varrho_{13}) + S(\varrho_{23}) \tag{3.1}$$

and

$$S(\varrho_1) + S(\varrho_2) \leq S(\varrho_{13}) + S(\varrho_{23}). \tag{3.2}$$

To simplify notation, we shall often omit braces and commas in subscripts, writing ϱ_{123} for $\varrho_{\{1,2,3\}}$, for example. The inequality (3.1) is referred to as *strong subadditivity* (or *submodularity*), and we shall refer to (3.2) as *weak monotonicity.* The special case of (3.1) with $\varrho_3 = 1$, which reduces to

$$S(\varrho_{12}) \leq S(\varrho_1) + S(\varrho_2) \tag{3.3}$$

and is called *weak subadditivity* (or simply *subadditivity*), was proved earlier by Araki and Lieb [A2], as was the special case of (3.2) with $\varrho_2 = 1$, which after renumbering of parts can be written

$$S(\varrho_1) \le S(\varrho_{12}) + S(\varrho_2) \tag{3.4}$$

or

$$S(\varrho_2) \le S(\varrho_{12}) + S(\varrho_1). \tag{3.5}$$

The inequalities (3.3), (3.4) and (3.5) taken together show that the entropies $S(\varrho_1)$, $S(\varrho_2)$ and $S(\varrho_{12})$ associated with a bipartite state can be interpreted as the lengths of the sides of a triangle, and thus are sometimes referred to as *triangle inequalities.*

Inequalities (3.1) and (3.2) are equivalent, in the sense that each can be deduced from the other using the principles of quantum mechanics. To due this, we use two lemmas proved by Araki and Lieb [A2] (which, however, they describe as "well known").

Lemma 3.1: Given any quantum state ϱ_1, there exists a pure quantum state ϱ_{12} (that is, a state with a density matrix of rank 1, or equivalently a state with entropy 0) such that $\varrho_1 = \mathrm{Tr}_2(\varrho_{12})$.

Proof: Let $\varrho_1 = \sum_{1\le i\le m} \lambda_i \, |\psi_{1,i}\rangle\langle\psi_{1,i}|$, where $\lambda_i > 0$ and $\{|\psi_{1,i}\rangle\}_{1\le i\le m}$ are orthonormal. Let $\{|\psi_{2,i}\rangle\}_{1\le i\le m}$ be an arbitrary orthonormal basis for a space of dimension m, and take $\varrho_{12} = |\phi\rangle\langle\phi|$, where $|\phi\rangle = \sum_{1\le i\le m} \lambda_i^{1/2} \, |\psi_{1,i}\rangle \otimes |\psi_{2,i}\rangle$. Then ϱ_{12} is a pure state and $\varrho_1 = \mathrm{Tr}_2(\varrho_{12})$. $\triangle$

Lemma 3.2: If ϱ_{12} is a pure quantum state, then $S(\varrho_1) = S(\varrho_2)$.

Proof: Let $\varrho_{12} = |\phi\rangle\langle\phi|$, where $|\phi\rangle = \sum_{1\le i\le m} \lambda_i \, |\psi_{1,i}\rangle \otimes |\psi_{2,i}\rangle$, where $\{|\psi_{1,i}\rangle\}_{1\le i\le m}$ and $\{|\psi_{2,i}\rangle\}_{1\le i\le m}$ are orthonormal, and where the phases have been chosen so that $\lambda_i > 0$. Then $\varrho_1 = \sum_{1\le i\le m} \lambda_i^2 \, |\psi_{1,i}\rangle\langle\psi_{1,i}|$ and $\varrho_2 = \sum_{1\le i\le m} \lambda_i^2 \, |\psi_{2,i}\rangle\langle\psi_{2,i}|$. Thus ϱ_1 and ϱ_2 have the same eigenvalues, with multiplicities, except possibly for the eigenvalue 0, and so they have the same entropy. $\triangle$

Let us show that (3.1) implies (3.2). Given ϱ_{123}, apply Lemma 3.1 to obtain the pure state ϱ_{1234}. We then have

$$\begin{aligned} S(\varrho_1) + S(\varrho_2) &= S(\varrho_{234}) + S(\varrho_2) \\ &\le S(\varrho_{23}) + S(\varrho_{24}) \\ &= S(\varrho_{23}) + S(\varrho_{13}), \end{aligned}$$

where the equalities follow from Lemma 3.2 and the inequality follows from (3.1). The result is (3.2).

Next let us show that (3.2) implies (3.1). Given ϱ_{123}, apply Lemma 3.1 to obtain the pure state ϱ_{1234}. We then have

$$\begin{aligned} S(\varrho_{123}) + S(\varrho_3) &= S(\varrho_4) + S(\varrho_3) \\ &\le S(\varrho_{14}) + S(\varrho_{13}) \\ &= S(\varrho_{23}) + S(\varrho_{13}), \end{aligned}$$

where the equalities follow from Lemma 3.2 and the inequality follows from (3.2). The result is (3.1).

In the Appendix, we give the simplest elementary proof we know of (3.2), and thus also of (3.1).

Inequalities (3.1) and (3.2) can be applied to an n-part quantum state in many ways. From (3.1) we obtain

$$S(\varrho_{I\cup J}) + S(\varrho_{I\cap J}) \le S(\varrho_I) + S(\varrho_J) \tag{3.3}$$

for all $I, J \subseteq N$, and from (3.2) we obtain

$$S(\varrho_{I\setminus J}) + S(\varrho_{J\setminus I}) \le S(\varrho_I) + S(\varrho_J) \tag{3.4}$$

for all $I, J \subseteq N$. To these we add the trivial

$$S(\varrho_\emptyset) = 0. \tag{3.5}$$

Let $\mathcal{B}_n$ denote the convex cone in $\mathbf{R}^\nu$, where $\nu = 2^n$, defined by (3.3), (3.4) and (3.5).

Our next goal is to distinguish a subset of the inequalities (3.3) and (3.4) having the property that all others can be deduced from them by taking positive linear combinations. Let us define

$$\Delta(I, J) = S(\varrho_I) + S(\varrho_J) - S(\varrho_{I\cup J}) - S(\varrho_{I\cap J}),$$

so that (3.3) becomes

$$\Delta(I, J) \ge 0, \tag{3.6}$$

and define

$$E(I, J) = S(\varrho_I) + S(\varrho_J) - S(\varrho_{I\setminus J}) - S(\varrho_{J\setminus I}),$$

so that (3.4) becomes

$$E(I, J) \ge 0. \tag{3.7}$$

We shall distinguish those instances of (3.6) for which $I \setminus J = \{i\}$ and $J \setminus I = \{j\}$ are disjoint singletons (so that $i \ne j$). This gives us each distinguished inequality twice (since we get the same inequality if we exchange I and J). We can eliminate this redundancy by imposing the additional condition that $i < j$. Thus our conditions for (3.6) become

$$I \setminus J = \{i\},\ J \setminus I = \{j\} \text{ and } i < j. \tag{3.8}$$

We observe that the number of ways of choosing I and J satisfying (3.8) is $n(n-1)2^{n-3}$, since we may choose $i < j$ in $\binom{n}{2}$ ways, and then choose $I \cap J$ to be a subset of $N \setminus \{i, j\}$ in 2^{n-2} ways. These choices are in an obvious one-to-one correspondence with the faces (2-dimensional subcubes) of the n-dimensional Boolean cube. We shall distinguish those instances of (3.7) for which $I \cap J = \{k\}$ is a singleton and $I \cup J = N$. If $n \geq 2$, this also gives us each distinguished inequality twice (since we again get the same inequality if we exchange I and J). We can eliminate this redundancy by imposing the additional condition that $k + 1 \in I$ (where we take $k + 1$ to be 1 if $k = n$). We observe that this additional condition is implied by the other two when $n = 1$, so we may impose it for all $n \geq 1$. Thus our conditions for (3.7) become

$$I \cap J = \{k\},\ I \cup J = N \text{ and } k + 1 \in I. \tag{3.9}$$

We observe that the number of ways of choosing I and J satisfying (3.9) is 1 if $n = 1$ and $n2^{n-2}$ if $n \geq 2$, since in the latter case we may choose k in n ways, and then choose J to be a subset of $N \setminus \{k, k+1\}$ in 2^{n-2} ways. These choices are in an obvious one-to-one correspondence with the antipodal pairs of edges (1-dimensional subcubes) of the n-dimensional Boolean cube (except when $n = 1$, when there is just one edge in the cube).

Proposition 3.3: Every instance of (3.6) can be obtained as a sum of distinguished instances of (3.6).

Proof: Consider the instance $\Delta(I, J) \geq 0$ of (3.6). If $I \subseteq J$ or $J \subseteq I$, this instance is trivial. Take $I \setminus J = \{i_1, \ldots, i_a\}$, $J \setminus I = \{j_1, \ldots, j_b\}$ and $K = I \cap J$. For $0 \leq r \leq a$, take $I_r = \{i_1, \ldots, i_r\}$ and for $0 \leq s \leq b$, take $J_s = \{j_1, \ldots, j_s\}$. Then we have

$$\Delta(I, J) = \sum_{1 \leq r \leq a} \sum_{1 \leq s \leq b} \Delta(I_r \cup J_{s-1} \cup K, I_{r-1} \cup J_s \cup K),$$

since each term $S(\cdots)$ on the left-hand side appears once with the correct sign on the right-hand side, while every other $S(\cdots)$ term appears on the right-hand side just as often with a positive sign as with a negative sign. Since every term $\Delta(\cdots)$ on the right-hand side satisfies (3.8) (either as written or in the equivalent form with its arguments exchanged), this completes the proof. $\triangle$

Proposition 3.4: Every instance of (3.7) can be obtained as a sum of distinguished instances of (3.6) and (3.7).

Proof: Consider the instance $E(I,J) \geq 0$ of (3.7). Using the identities

$$I \cup (N \setminus J) = N \setminus (J \setminus I),$$
$$I \cap (N \setminus J) = I \setminus J,$$
$$J \cup (N \setminus I) = N \setminus (I \setminus J)$$

and

$$J \cap (N \setminus I) = J \setminus I,$$

we have

$$E(I,J) = \Delta(I, N \setminus J) + \Delta(J, N \setminus I) + E\big(N \setminus (I \setminus J), N \setminus (J \setminus I)\big),$$

since each term $S(\cdots)$ on the left-hand side appears once with the correct sign on the right-hand side, while every other term $S(\cdots)$ appears on the right-hand side just as often with a positive sign as with a negative sign. Thus we have expressed the given instance of (3.7) as a sum of two instances of (3.6) and the instance $E((N \setminus (I \setminus J), N \setminus (J \setminus I)\big) \geq 0$ of (3.7). By Proposition 3.3, both instances of (3.6) can be expressed as sums of distinguished instances of (3.6). Writing $I' = N \setminus (I \setminus J)$ and $J' = N \setminus (J \setminus I)$, we have $I' \cup J' = N$. Thus it will suffice to show that every instance of (3.7) satisfying $I \cup J = N$ can be expressed as a sum of distinguished instances of (3.7).

Let $E(I,J) \geq 0$ be an instance of (3.7) satisfying $I \cup J = N$. If $I \cap J = \emptyset$, this instance is trival. Otherwise, take $I \cap J = \{k_1, \ldots, k_c\}$ and $L = J \setminus I$. For $0 \leq t \leq c$, let $K_t = \{k_1, \ldots, k_t\}$. Then we have

$$E(I,J) = \sum_{1 \leq t \leq c} E\big(K_t \cup L, N \setminus (K_{t-1} \cup L)\big),$$

since each term $S(\cdots)$ on the left-hand side appears once with the correct sign on the right-hand side, while every other term $S(\cdots)$ appears on the right-hand side just as often with a positive sign as with a negative sign. Since every term $E(\cdots)$ on the right-hand side satisfies (3.9) (either as written or in the equivalent form with its arguments exchanged), this completes the proof. $\triangle$

Proposition 3.5: None of the distinguished instances of (3.6) or (3.7) can be deduced from the other distinguished instances of (3.6) and (3.7).

Proof: To prove an instance of (3.6) or (3.7) cannot be deduced from the others, we shall find a collection $\{S_K\}_{K\subseteq N}$ of numbers, with $S_\emptyset = 0$, that violate the given instance (when substituted into the instance in the obvious way), but satisfy all the other distinguished instances of (3.6) and (3.7). (This collection is not, of course, the allocation of entropy of any quantum state, since it does not satisfy all the instances of (3.6) and (3.7).)

We begin by taking

$$R_K = (n+1)k - k^2,$$

where $k = \#K$. The collection $\{R_K\}_{K\subseteq N}$ satisfies $R_\emptyset = 0$, and satisfies every distinguished instance of (3.6) and (3.7) by a margin of 2.

If $n = 1$, there is just one distinguished instance, so we may assume that $n \geq 2$. If we are given a distinguished instance $\Delta(I, J) \geq 0$ of (3.6) we shall take

$$S_K = \begin{cases} R_K - 1, & \text{if } K = I \text{ or } K = J; \\ R_K + 1, & \text{if } K = I \cup J; \\ R_K, & \text{otherwise.} \end{cases}$$

Then we have $S_\emptyset = 0$, since none of I, J and $I \cup J$ can be empty. Three of the terms appearing in the given instance are modified in $\{S_K\}_{K\subseteq N}$ in such way as to reduce the margin of the given instance by 3 to -1, thus violating this inequality. If on the other hand we are given a distinguished instance $E(I, J) \geq 0$ of (3.7), we shall take

$$S_K = \begin{cases} R_K - 1, & \text{if } K = I \text{ or } K = J; \\ R_K + 1, & \text{if } K = I \setminus J; \\ R_K, & \text{otherwise.} \end{cases}$$

Then we again have $S_\emptyset = 0$, since none of I, J and $I \setminus J$ can be empty. And again, three of the terms appearing in the given instance are modified in $\{S_K\}_{K\subseteq N}$ in such way as to reduce the margin of the given instance by 3 to -1, thus violating this inequality.

It remains to show that no other distinguished instance of (3.6) or (3.7) is violated by this modification. No other distinguished instance can be violated unless it involves all three modified terms, since if it involves at most two modified terms its margin can be reduced by at most 2 to 0, and thus it will remain satisfied. No two distinguished instances of (3.6) have the same values of I and J, and no two distinguished instances of (3.7) have the same values of I and J, so we only need to consider the case in which a distinguished instance of (3.6) has the same values of I and J as a distinguished instance of (3.7). It is easy to check that if (3.8) and (3.9) are satisfied, $I \cup J$ cannot equal either $I \setminus J$ or $J \setminus I$, and $I \setminus J$ cannot equal either $I \cup J$ or $I \cap J$. This completes the proof. $\triangle$

Corollary 3.6: The cone $\mathcal{B}_n$ lies in the hyperplane (3.5) and is bounded by the facets (3.6) satisfying (3.8) and (3.7) satisfying (3.9).

Proof: This follows immediately from Propositions 3.3, 3.4 and 3.5. $\triangle$

We shall now turn to the classical counterparts of the results we have just derived. The classical version of strong subadditivity is

$$H(X_{123}) + H(X_3) \leq H(X_{13}) + H(X_{23}). \tag{3.10}$$

In the classical case, we have *strong monotonicity*,

$$H(X_1) \leq H(X_{12}). \tag{3.11}$$

Inequalities (3.10) and (3.11) are consequences of the concavity and monotonicity, respectively, of the logarithm. Inequality (3.11) can be written $H(X_{12} \mid X_1) \geq 0$, where $H(X_{12} \mid X_1) = H(X_{12}) - H(X_1)$ is the *conditional entropy* of X_{12} relative to X_1, and inequality (3.10) can be written $I(X_{13}; X_{23} \mid X_3) \geq 0$, where $I(X_{13}; X_{23} \mid X_3) = H(X_{13}) + H(X_{23}) - H(X_{123}) - H(X_3)$ is the *conditional mutual information* between X_{13} and X_{23} relative to X_3.

For an n-component random variable, we have from (3.10),

$$H(X_{I \cup J}) + H(X_{I \cap J}) \leq H(X_I) + H(X_J), \tag{3.12}$$

and from (3.11),

$$H(X_{I \setminus J}) \leq H(X_I). \tag{3.13}$$

To these we add the trivial

$$H(X_\emptyset) = 0. \tag{3.14}$$

Let $\mathcal{B}_n^{\text{class}}$ denote the convex cone in $\mathbf{R}^\nu$, where $\nu = 2^n$, defined by (3.12), (3.13) and (3.14).

The inequalities (3.12) (in the form $I(X_I; X_J \mid X_{I \cap J}) \geq 0$) and (3.13) (in the form $H(X_I \mid X_{I \setminus J}) \geq 0$) were know to Shannon [S1], and thus have been called *Shannon-type* inequalities by Zhang and Yeung [Z1]. Fujishige [F] has pointed out that (3.12), (3.13) and (3.14) are equivalent to saying that the map $K \mapsto H(X_K)$ is a polymatroid (see Welsh [W]), and thus we shall refer to them as *polymatroid* inequalities.

Our next goal is to distinguish a subset of the inequalities (3.12) and (3.13) having the property that all others can be deduced from them by taking positive linear combinations. This was first done by Yeung [Y]. Let us define

$$F(I, J) = H(X_I) + H(X_J) - H(X_{I \cup J}) - H(X_{I \cap J}),$$

so that (3.12) becomes

$$F(I, J) \geq 0, \tag{3.15}$$

and define

$$G(I, J) = H(X_I) - H(X_{I \setminus J}),$$

so that (3.13) becomes

$$G(I, J) \geq 0. \tag{3.16}$$

As in the quantum case, we shall distinguish those instances of (3.15) for which

$$I \setminus J = \{i\},\ J \setminus I = \{j\} \text{ and } i < j. \tag{3.17}$$

We shall distinguish those instances of (3.16) for which

$$I = N \text{ and } J = \{j\}. \tag{3.18}$$

We observe that the number of ways of choosing I and J satisfying (3.18) is n, since we may choose I in 1 way and J in n ways.

Proposition 3.7: Every instance of (3.15) can be obtained as a sum of distinguished instances of (3.15).

Proof: The proof is the same as that for Proposition 3.3. $\triangle$

Proposition 3.8: Every instance of (3.16) can be obtained as a sum of distinguished instances of (3.15) and (3.16).

Proof: Consider the instance $G(I, J) \geq 0$ of (3.16). We may assume that $J \subseteq I$. If $J = \emptyset$, this instance is trivial. Otherwise, take $J = \{j_1, \ldots, j_b\}$ and, for $1 \leq s \leq b$, take $I_s = I \setminus \{j_1, \ldots, j_{s-1}\}$ and $J_s = \{j_s\}$. Then we have

$$G(I, J) = \sum_{1 \leq s \leq b} G(I_s, J_s),$$

since each term $H(\cdots)$ on the left-hand side appears once with the correct sign on the right-hand side, while every other term $H(\cdots)$ appears on the right-hand side just as often with a positive sign as with a negative sign. Each instance $G(I', J') \geq 0$ of (3.16) on the right-hand side has $J' = \{j'\}$ a singleton. Thus it will suffice to show how to express an instance $G(I, J) \geq 0$ of (3.16) with J a singleton as a sum of distinguished instances of (3.15) and (3.16).

Consider an instance $G(I, J) \geq 0$ of (3.16) with $J = \{j\}$. Using the identities

$$I \cup (N \setminus J) = N$$

and

$$I \cap (N \setminus J) = I \setminus J,$$

we have

$$G(I, J) = F(I, N \setminus J) + G(N, N \setminus J),$$

since each term $H(\cdots)$ on the left-hand side appears once with the correct sign on the right-hand side, while every other term $H(\cdots)$ appears on the right-hand side just as often with a positive sign as with a negative sign. Thus we have expressed the given instance of (3.16) as the sum of an instance of (3.15) and a distinguished instance of (3.16). By Proposition 3.7, the instance of (3.15) can be expressed as a sum of distinguished instances of (3.15). This completes the proof. △

Proposition 3.9: None of the distinguished instances of (3.15) or (3.16) can be deduced from the other distinguished instances of (3.15) and (3.16).

Proof: To prove an instance of (3.15) or (3.16) cannot be deduced from the others, we shall find a collection $\{H_K\}_{K \subseteq N}$ of numbers, with $H_\emptyset = 0$, that violate the given instance (when substituted into the instance in the obvious way), but satisfy all the other distinguished instances of (3.15) and (3.16). (This collection is not, of course, the allocation of entropy of any random variable, since it does not satisfy all the instances of (3.15) and (3.16).)

We begin by taking

$$R_K = 2nk - k^2,$$

where $k = \#K$. The collection $\{R_K\}_{K \subseteq N}$ satisfies $R_\emptyset = 0$, and satisfies every distinguished instance of (3.15) by a margin of 2, and every distinguished instance of (3.16) by a margin of 1.

If $n = 1$, there is just one distinguished instance, so we may assume that $n \geq 2$. If we are given a distinguished instance $F(I, J) \geq 0$ of (3.15) we shall take

$$H_K = \begin{cases} R_K - 2, & \text{if } K = I \text{ or } K = J; \\ R_K, & \text{otherwise.} \end{cases}$$

Then we have $H_\emptyset = 0$, since neither I nor J can be empty. Two of the terms appearing in the given instance are modified in $\{H_K\}_{K \subseteq N}$ in such way as to reduce the margin of the given instance by 4 to -2, thus violating this inequality. If on the other hand we are given a distinguished instance $G(N, J) \geq 0$ of (3.7) with $J = \{j\}$, we shall take

$$H_K = \begin{cases} R_K - 1, & \text{if } K = N; \\ R_K + 1, & \text{if } K = N \setminus J; \\ R_K, & \text{otherwise.} \end{cases}$$

Then we again have $H_\emptyset = 0$, since neither N nor $N \setminus J$ can be empty. And again, two of the terms appearing in the given instance are modified in $\{H_K\}_{K \subseteq N}$ in such way as to reduce the margin of the given instance by 2 to -1, thus violating this inequality.

It remains to show that no other distinguished instance of (3.15) or (3.16) is violated by this modification. If the given distinguished instance of (3.15), then another distinguished instance of (3.15) cannot have the same values of both I and J, and thus cannot involve more that one of the modified terms in $\{H_K\}_{K \subseteq N}$, and thus cannot have its margin reduced by more than 2, and thus will remain satisfied. And a distinguished instance of (3.16) can only have its margin increased by this modification, and thus will remain satisfied. If on the other hand the given distinguished instance is of (3.16), then another distinguished instance of (3.16) cannot have the same value of J, and thus cannot have its margin reduced by more than 1, and thus will remain satisfied. And a distinguished instance of (3.15) cannot have its margin reduced by more than 1 by this modification, and thus will remain satisfied. This completes the proof. △

Corollary 3.10: The cone $\mathcal{B}_n^{\text{class}}$ lies in the hyperplane (3.14) and is bounded by the facets (3.15) satisfying (3.17) and (3.16) satisfying (3.18).

Proof: This follows immediately from Propositions 3.7, 3.8 and 3.9. △

4. Bipartite and Tripartite Allocations

Our goal in this section is to show that $\overline{\mathcal{A}_n} = \mathcal{B}_n$ for $n \leq 3$. We shall also prove the classical counterpart of this result, $\overline{\mathcal{A}_n^{\text{class}}} = \mathcal{B}_n^{\text{class}}$ for $n \leq 3$, which is implicit in the work of Han [H2], and was obtained explicitly by Zhang and Yeung [Z2].

Since $\mathcal{A}_n \subseteq \mathcal{B}_n$ and $\mathcal{B}_n$ is topologically closed, we have $\overline{\mathcal{A}_n} \subseteq \mathcal{B}_n$. Thus it will suffice to show that $\overline{\mathcal{A}_n} \supseteq \mathcal{B}_n$. To do this, we shall find a hyperplane $\mathcal{H}$ that meets every ray (that is, every set $\{\lambda X : \lambda \geq 0\}$) in $\mathcal{B}_n$. Since $\overline{\mathcal{A}_n}$ and $\mathcal{B}_n$ are both convex cones, it will then suffice to show that $\overline{\mathcal{A}_n \cap \mathcal{H}} \supseteq \mathcal{B}_n \cap \mathcal{H}$. To find such a hyperplane, we shall use two lemmas.

Lemma 4.1: The cone $\mathcal{B}_n$ lies in the non-negative orthant of $\mathbf{R}^\nu$.

Proof: The instance $I = J = K$ of (3.4) shows that

$$S_K \geq S_\emptyset$$

for all $K \subseteq N$. By (3.5), we have $S_\emptyset = 0$, and thus

$$S_K \geq 0$$

for all $K \subseteq N$. $\triangle$

Lemma 4.2: If $S_{\{k\}} = 0$ for all $k \in N$, then $S_K = 0$ for all $K \subseteq N$.

Proof: Assume the hypothesis. We shall prove the conclusion by induction on $\#K$. The case $\#K = 0$ is (3.5), and the case $\#K = 1$ is the hypothesis. If $\#K \geq 2$, let i and j be distinct elements of K, and take $I = K \setminus \{i\}$ and $J = K \setminus \{j\}$. Then the corresponding instance of (3.3) yields

$$S_K = S_{I \cup J} \leq S_I + S_J - S_{I \cap J}.$$

The three terms on the right hand side all vanish by inductive hypothesis, so we have $S_K \leq 0$. Thus by Lemma 4.1 we have $S_K = 0$. $\triangle$

Lemmas 4.1 and 4.2 show that for any $c > 0$, the hyperplane

$$\mathcal{H}_c = \{S : \sum_{k \in N} S_{\{k\}} = c\}$$

meets every ray of $\mathcal{B}_n$.

Our strategy for showing that $\overline{\mathcal{A}_n \cap \mathcal{H}} \supseteq \mathcal{B}_n \cap \mathcal{H}$ will be to find a set $\{\varrho^1, \ldots, \varrho^m\}$ of quantum states such that every point $\{S_K\}_{K \subseteq N}$ in $\mathcal{B}_n \cap \mathcal{H}$ can be expressed as a convex combination of the allocations of entropy of the states in $\{\varrho^1, \ldots, \varrho^m\}$; that is, such that for every point $\{S_K\}_{K \subseteq N}$ in $\mathcal{B}_n \cap \mathcal{H}$, we can find real numbers $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0$ such that $\sum_{1 \leq l \leq m} \lambda_l = 1$ and

$$S_K = \sum_{1 \leq l \leq m} \lambda_l \, S(\varrho^l_K)$$

for all $K \subseteq N$.

For $n = 1$, we take the hyperplane $\mathcal{H}_1$, which intersects the cone $\mathcal{B}_1$ in the point defined by $S_1 = 1$ and $S_\emptyset = 0$. Thus it will suffice to take $m = 1$ and $\varrho^1_1 = I_2/2$, where I_d denotes the $d \times d$ identity matrix, which corresponds to a classical random variable with 1 bit of entropy. This complete the proof that $\overline{\mathcal{A}_1} = \mathcal{B}_1$.

For $n=2$, we take the hyperplane $\mathcal{H}_2$, which intersects the cone in a triangular region defined by the equalities

$$S_1 + S_2 = 2 \tag{4.1}$$

and

$$S_\emptyset = 0,$$

and by the inequalities

$$S_{12} \leq S_1 + S_2, \tag{4.2}$$

$$S_1 \leq S_{12} + S_2 \tag{4.3}$$

and

$$S_2 \leq S_{12} + S_1. \tag{4.4}$$

We shall take $m=3$ and define ϱ^1, ϱ^2 and ϱ^3 as follows. Take $\varrho^1_1 = I_4/4$, $\varrho^1_2 = I_1$ and $\varrho^1_{12} = \varrho^1_1 \otimes \varrho^1_2$, which corresponds to a classical random variable with 2 bits of entropy in part 1. Thus $S(\varrho^1_{12}) = S(\varrho^1_1) = 2$ and $S(\varrho^1_2) = 0$. Similarly, take $\varrho^2_1 = I_1$, $\varrho^2_2 = I_4/4$ and $\varrho^2_{12} = \varrho^2_1 \otimes \varrho^2_2$, which corresponds to a classical random variable with 2 bits of entropy in part 2. Thus $S(\varrho^2_{12}) = S(\varrho^2_2) = 2$ and $S(\varrho^2_1) = 0$. Finally, take $\varrho^3_{12} = |\psi\rangle\langle\psi|$, where $|\psi\rangle = (|0,0\rangle + |1,1\rangle)/\sqrt{2}$ and $|x,y\rangle = |x\rangle \otimes |y\rangle$. Then ϱ^3_{12} corresponds to an Einstein-Podolsky-Rosen pair of entangled qubits shared between parts 1 and 2, so that $S(\varrho^3_1) = S(\varrho^3_2) = 1$ and $S(\varrho^3_{12}) = 0$.

The states ϱ^1, ϱ^2 and ϱ^3 satisfy

$$S(\varrho_1) + S(\varrho_2) = 2,$$

so their allocations of entropy lie in $\mathcal{H}_2$. Suppose that we are given $\{S_K\}_{K\subseteq\{1,2\}}$ in $\mathcal{B}_2 \cap \mathcal{H}_2$. Then if we take

$$\lambda_1 = (S_{12} + S_1 - S_2)/4,$$

$$\lambda_2 = (S_{12} + S_2 - S_1)/4,$$

and

$$\lambda_3 = (S_1 + S_2 - S_{12})/2,$$

we have $\lambda_l \geq 0$ for $1 \leq l \leq 3$ by virtue of (4.2), (4.3) and (4.4), and $\sum_{1\leq l\leq 3} \lambda_l = 1$ by virtue of (4.1). Furthermore, we have

$$S_K = \sum_{1\leq l\leq 3} \lambda_l \, S(\varrho^l_K)$$

for all $K \subseteq \{1,2\}$. This completes the proof that $\overline{\mathcal{A}_2} = \mathcal{B}_2$.

For $n = 3$, we take the hyperplane $\mathcal{H}_6$, which intersects the cone $\mathcal{B}_3$ in the polytope defined by the equalities

$$S_1 + S_2 + S_3 = 6$$

and

$$S_\emptyset = 0,$$

the six inequalities that are distinguished instances of (3.6) and the six inequalities that are distinguished instances of (3.7).

We shall take $m = 8$ and define the states $\varrho^1, \ldots, \varrho^8$ as follows. For $1 \le l \le 3$ and $1 \le k \le 3$, take

$$\varrho_k^l = \begin{cases} I_{64}/64, & \text{if } k = l; \\ I_1, & \text{otherwise;} \end{cases}$$

and take $\varrho_{123}^l = \varrho_1^l \otimes \varrho_2^l \otimes \varrho_3^l$. This corresponds to a classical random variable with 6 bits of entropy in part l, so for $1 \le l \le 3$ we have

$$S(\varrho_K^l) = \begin{cases} 6, & \text{if } l \in K; \\ 0, & \text{otherwise.} \end{cases}$$

For $4 \le l \le 6$, we take $\varrho_{123}^l = |\psi^l\rangle\langle\psi^l|$, where $|\psi^l\rangle = (\sum_{0\le w\le 7} |\psi_w^l\rangle)/\sqrt{8}$,

$$|\psi_w^l\rangle = \begin{cases} |0,w,w\rangle, & \text{if } l = 4; \\ |w,0,w\rangle, & \text{if } l = 5; \\ |w,w,0\rangle, & \text{if } l = 6; \end{cases}$$

and $|x,y,z\rangle = |x\rangle \otimes |y\rangle \otimes |z\rangle$. For $4 \le l \le 6$, ϱ_{123}^l corresponds to three Einstein-Podolsky-Rosen pairs of entangled qubits shared between the parts k such that $k \ne l-3$. Thus if we set $L_l = \{1,2,3\} \setminus \{l-3\}$, then for $4 \le l \le 6$ we have

$$S(\varrho_K^l) = \begin{cases} 3, & \text{if } K \setminus L_l \ne \emptyset \text{ and } L_l \setminus K \ne \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

We take $\varrho_{123}^7 = \left(\sum_{0\le w\le 3} |w,w,w\rangle\langle w,w,w|\right)/4$, which corresponds to a classical random variable with 2 bits of entropy common to parts 1, 2 and 3. Thus we have

$$S(\varrho_K^7) = \begin{cases} 2, & \text{if } K \ne \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

Finally, we take $\varrho_{123}^8 = \mathrm{Tr}_4(\varrho_{1234}^8)$, where $\varrho_{1234}^8 = |\psi^8\rangle\langle\psi^8|$,

$$|\psi^8\rangle = \frac{1}{4} \sum_{u,v \in GF(4)} |u, u+v, u+\zeta v, u+\zeta^2 v\rangle,$$

$|w,x,y,z\rangle = |w\rangle \otimes |x\rangle \otimes |y\rangle \otimes |z\rangle$ and $GF(4) = \{0, 1, \zeta, \zeta^2\}$ is the finite field with 4 elements. In the vector space $GF(4)^2$, any of the four elements $(1,0)$, $(1,1)$, $(1,\zeta)$ and $(1,\zeta^2)$ spans a subspace of dimension 1, and any two of these elements span a subspace of dimension 2. These facts, together with Lemma 3.2 and the fact that ϱ^8_{1234} is a pure state, imply that

$$S(\varrho^8_K) = \begin{cases} 2, & \text{if } \#K \in \{1,3\}; \\ 4, & \text{if } \#K = 2; \\ 0, & \text{otherwise.} \end{cases}$$

The states ϱ^l for $1 \leq l \leq 8$ all satisfy

$$S(\varrho^l_1) + S(\varrho^l_2) + S(\varrho^l_3) = 6,$$

so their allocations of entropy lie in $\mathcal{H}_6$. Suppose we are given $\{S_K\}_{K\subseteq\{1,2,3\}}$ in $\mathcal{B}_3 \cap \mathcal{H}_6$. We shall start by proving three lemmas.

Define

$$M = S_1 + S_2 + S_3 - S_{12} - S_{13} - S_{23} + S_{123},$$

and let $\mathcal{M}$ denote the hyperplane defined by $M = 0$.

Lemma 4.3: If $\{S_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}$, then $\{S_K\}_{K\subseteq\{1,2,3\}}$ is a convex combination of the allocations of entropy of $\varrho^1, \ldots, \varrho^6$.

Proof: Suppose we are given $\{S_K\}_{K\subseteq\{1,2,3\}}$ in $\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}$. Take

$$\begin{aligned} \lambda_1 &= E(12,13)/12, \\ \lambda_2 &= E(12,23)/12, \\ \lambda_3 &= E(13,23)/12, \\ \lambda_4 &= \Delta(2,3)/6, \\ \lambda_5 &= \Delta(1,3)/6 \end{aligned}$$

and

$$\lambda_6 = \Delta(1,2)/6.$$

Then $\lambda_l \geq 0$ for $1 \leq l \leq 6$, since $\{S_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3$, and $\sum_{1\leq l\leq 6} \lambda_l = \left(\sum_{1\leq k\leq 3} S_k\right)/6 = 1$ since $\{S_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{H}_6$. Furthermore

$$S_K = \sum_{1\leq l\leq 6} \lambda_l \, S(\varrho^l_K)$$

for $K \subseteq \{1,2,3\}$. $\triangle$

Let $\mathcal{M}^+$ denote the half-space defined by $M \geq 0$.

Lemma 4.4: If $\{S_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}^+$, then $\{S_K\}_{K\subseteq\{1,2,3\}}$ is a convex combination of the allocation of entropy of ϱ^7 and an allocation of entropy of lying in $\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}$.

Proof: Suppose we are given $\{S_K\}_{K\subseteq\{1,2,3\}}$ in $\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}^+$. Take $\lambda = M/2 \geq 0$. Define $S'_K = S_K - \lambda\, S(\varrho^7_K)$ for $K \subseteq \{1,2,3\}$. We shall show that $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $(1-\lambda)(\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}) = \mathcal{B}_3 \cap \mathcal{H}_{(1-\lambda)6} \cap \mathcal{M}$. This will complete the proof of the lemma.

Since

$$S(\varrho^7_1) + S(\varrho^7_2) + S(\varrho^7_3) - S(\varrho^7_{12}) - S(\varrho^7_{13}) - S(\varrho^7_{23}) + S(\varrho^7_{123}) = 2,$$

we have $M' = M - 2\lambda = 0$, and thus $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{M}$. Since

$$S(\varrho^7_1) + S(\varrho^7_2) + S(\varrho^7_3) = 6,$$

we have $S'_1 + S'_2 + S'_3 = (1-\lambda)6$, and thus $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{H}_{(1-\lambda)6}$. It remains to show that $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3$.

If we define

$$\Delta'(I,J) = S'_I + S'_J - S'_{I\cup J} - S'_{I\cap J}$$

then since

$$S(\varrho^7_{123}) + S(\varrho^7_1) = S(\varrho^7_{12}) + S(\varrho^7_{13}),$$

we have $\Delta'(12,13) = \Delta(12,13) \geq 0$, and we have $\Delta'(12,23) \geq 0$ and $\Delta'(13,23) \geq 0$ by permutation of the indices in this argument. Furthermore, since

$$S(\varrho^7_1) + S(\varrho^7_2) - S(\varrho^7_{12}) = 2,$$

we have $\Delta'(1,2) = \Delta(1,2) - 2\lambda = \Delta(1,2) - M$. But $M = \Delta(1,2) - \Delta(13,23)$, so $\Delta'(1,2) = \Delta(13,23) \geq 0$, and we have $\Delta'(1,3) \geq 0$ and $\Delta'(2,3) \geq 0$ by permutation of the indices in this argument. Thus $\{S'_K\}_{K\subseteq\{1,2,3\}}$ satisfies the distinguished instances of (3.6).

If we define

$$E'(I,J) = S'_I + S'_J - S'_{I\setminus J} - S'_{J\setminus I},$$

then since

$$S(\varrho^7_1) + S(\varrho^7_2) = S(\varrho^7_{13}) + S(\varrho^7_{23}),$$

we have $E'(13,23) = E(13,23) \geq 0$, and we have $E'(12,23) \geq 0$ and $E'(12,13) \geq 0$ by permutation of the indices in this argument. Furthermore, since

$$S(\varrho^7_{123}) + S(\varrho^7_3) - S(\varrho^7_{12}) = 2,$$

we have $E'(123,3) = E(123,3) - 2\lambda = E(123,3) - M$. But $M = E(123,3) - E(13,23)$, so $E'(123,3) = E(13,23) \geq 0$, and we have $E'(123,2) \geq 0$ and $E'(123,1) \geq 0$ by permutation of the indices in this argument. Thus $\{S'_K\}_{K\subseteq\{1,2,3\}}$ satisfies the distinguished instances of (3.7). This completes the proof that $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3$. $\triangle$

Let $\mathcal{M}^-$ denote the half-space defined by $M \leq 0$.

Lemma 4.5: If $\{S_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}^-$, then $\{S_K\}_{K\subseteq\{1,2,3\}}$ is a convex combination of the allocation of entropy of ϱ^8 and an allocation of entropy of lying in $\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}$.

Proof: Suppose we are given $\{S_K\}_{K\subseteq\{1,2,3\}}$ in $\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}^-$. Take $\lambda = -M/4 \geq 0$. Define $S'_K = S_K - \lambda\, S(\varrho^8_K)$ for $K \subseteq \{1,2,3\}$. We shall show that $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $(1-\lambda)(\mathcal{B}_3 \cap \mathcal{H}_6 \cap \mathcal{M}) = \mathcal{B}_3 \cap \mathcal{H}_{(1-\lambda)6} \cap \mathcal{M}$. This will complete the proof of the lemma.

Since

$$S(\varrho^8_1) + S(\varrho^8_2) + S(\varrho^8_3) - S(\varrho^8_{12}) - S(\varrho^8_{13}) - S(\varrho^8_{23}) + S(\varrho^8_{123}) = -4,$$

we have $M' = M + 4\lambda = 0$, and thus $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{M}$. Since

$$S(\varrho^8_1) + S(\varrho^8_2) + S(\varrho^8_3) = 6,$$

we have $S'_1 + S'_2 + S'_3 = (1-\lambda)6$, and thus $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{H}_{(1-\lambda)6}$. It remains to show that $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3$.

Since

$$S(\varrho^8_{12}) = S(\varrho^8_1) + S(\varrho^8_2),$$

we have $\Delta'(1,2) = \Delta(1,2) \geq 0$, and we have $\Delta'(1,3) \geq 0$ and $\Delta'(2,3) \geq 0$ by permutation of the indices in this argument. Furthermore, since

$$S(\varrho^8_{13}) + S(\varrho^8_{23}) - S(\varrho^8_{123}) - S(\varrho^8_3) = 4,$$

we have $\Delta'(13,23) = \Delta(13,23) - 4\lambda = \Delta(1,2) + M$. But $M = \Delta(1,2) - \Delta(13,23)$, so $\Delta'(13,23) = \Delta(1,2) \geq 0$, and we have $\Delta'(12,23) \geq 0$ and $\Delta'(12,13) \geq 0$ by permutation of the indices in this argument. Thus $\{S'_K\}_{K\subseteq\{1,2,3\}}$ satisfies the distinguished instances of (3.6).

Since

$$S(\varrho_{12}^8) = S(\varrho_{123}^8) + S(\varrho_3^8),$$

we have $E'(123,3) = E(123,3) \geq 0$, and we have $E'(123,2) \geq 0$ and $E'(123,1) \geq 0$ by permutation of the indices in this argument. Furthermore, since

$$S(\varrho_{13}^8) + S(\varrho_{23}^8) - S(\varrho_1^8) - S(\varrho_2^8) = 4,$$

we have $E'(13,23) = E(13,23) - 4\lambda = E(123,3) + M$. But $M = E(123,12) - E(13,23)$, so $E'(13,23) = E(123,12) \geq 0$, and we have $E'(123,13) \geq 0$ and $E'(123,23) \geq 0$ by permutation of the indices in this argument. Thus $\{S'_K\}_{K\subseteq\{1,2,3\}}$ satisfies the distinguished instances of (3.7). This completes the proof that $\{S'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3$. $\triangle$

We are now in a position to prove the main results of this section.

Theorem 4.6: If $\{S_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3 \cap \mathcal{H}_6$, then $\{S_K\}_{K\subseteq\{1,2,3\}}$ is a convex combination of the allocations of entropy of $\varrho^1, \ldots, \varrho^8$.

Proof: According as $M \geq 0$ or $M < 0$, we apply Lemma 4.4 or Lemma 4.6. Application of Lemma 4.3 to the resulting $\{S'_K\}_{K\subseteq\{1,2,3\}}$ then completes the proof. $\triangle$

Theorem 4.7: The allocations of entropy of the states $\varrho^1, \ldots, \varrho^8$ are extreme points of the polytope $\mathcal{B}_3 \cap \mathcal{H}_6$; that is, none of these allocations of entropy can be expressed as a convex combination of the other seven.

Proof: For each of the eight allocations, we shall find a linear inequality violated by that allocation, but satisfied by the other seven. Since the set of allocations satisfying a given linear inequality is closed under taking convex combinations, this will suffice to prove the proposition.

Take

$$P = S(\varrho_{12}) + S(\varrho_{13}) + S(\varrho_{23}).$$

Then $P - E(1,123) > 0$ for each of $\varrho^1, \ldots, \varrho^8$ except ϱ^1. The extremity of the allocations for ϱ^2 and ϱ^3 is established in the same way, with a permutation of the indices. Furthermore, $P - \Delta(12,13) > 0$ for each of $\varrho^1, \ldots, \varrho^8$ except ϱ^4. The extremity of the allocations for ϱ^5 and ϱ^6 is established in the same way, with a permutation of the indices. Finally, $M \leq 0$ for each of $\varrho^1, \ldots, \varrho^8$ except ϱ^7, and $M \geq 0$ for each of $\varrho^1, \ldots, \varrho^8$ except ϱ^8. $\triangle$

It is implicit in the proofs of Theorems 4.6 and 4.7 that the polytope $\mathcal{B}_3 \cap \mathcal{H}_6$ is the union of two 7-dimensional simplices, one with extreme points at the allocations of $\varrho^1, \ldots, \varrho^6, \varrho^7$ and the other with extreme points at the allocations of $\varrho^1, \ldots, \varrho^6, \varrho^8$, and whose intersection is the 6-dimensional simplex with extreme points at the allocations of $\varrho^1, \ldots, \varrho^6$.

We shall now turn to the classical counterparts of the results we have just derived. Specifically, we shall prove that $\overline{\mathcal{A}_n^{\text{class}}} = \mathcal{B}_n^{\text{class}}$ for $n \leq 3$.

Since $\mathcal{A}_n^{\text{class}} \subseteq \mathcal{B}_n^{\text{class}}$ and $\mathcal{B}_n^{\text{class}}$ is topologically closed, we have $\overline{\mathcal{A}_n^{\text{class}}} \subseteq \mathcal{B}_n^{\text{class}}$. Thus it will suffice to show that $\overline{\mathcal{A}_n^{\text{class}}} \supseteq \mathcal{B}_n^{\text{class}}$. Our strategy will again be find a hyperplane $\mathcal{H}$ that meets every ray in $\mathcal{B}_n^{\text{class}}$. Since $\overline{\mathcal{A}_n^{\text{class}}}$ and $\mathcal{B}_n^{\text{class}}$ are both convex cones, it will then suffice to show that $\overline{\mathcal{A}_n^{\text{class}} \cap \mathcal{H}} \supseteq \mathcal{B}_n^{\text{class}} \cap \mathcal{H}$.

Lemma 4.8: For all n, $\mathcal{B}_n^{\text{class}} \subseteq \mathcal{B}_n$.

Proof: This follows from the fact that strong monotonicity implies weak monotonicity. $\triangle$

Lemma 4.8 shows that for any $c > 0$, the hyperplane

$$\mathcal{H}_c = \{H : \sum_{k \in N} H_{\{k\}} = c\}$$

meets every ray of $\mathcal{B}_n^{\text{class}}$.

Our strategy for showing that $\overline{\mathcal{A}_n^{\text{class}} \cap \mathcal{H}} \supseteq \mathcal{B}_n^{\text{class}} \cap \mathcal{H}$ will be to find a set $\{X^1, \ldots, X^m\}$ of random variables such that every point $\{H_K\}_{K \subseteq N}$ in $\mathcal{B}_n^{\text{class}} \cap \mathcal{H}$ can be expressed as a convex combination of the allocations of entropy of the random variables in $\{X^1, \ldots, X^m\}$; that is, such that for every point $\{H_K\}_{K \subseteq N}$ in $\mathcal{B}_n^{\text{class}} \cap \mathcal{H}$, we can find real numbers $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0$ such that $\sum_{1 \leq l \leq m} \lambda_l = 1$ and

$$H_K = \sum_{1 \leq l \leq m} \lambda_l \, H(X_K^l)$$

for all $K \subseteq N$.

For $n = 1$, we take the hyperplane $\mathcal{H}_1$, which intersects the cone $\mathcal{B}_1^{\text{class}}$ in the point defined by $H_1 = 1$ and $H_\emptyset = 0$. Thus it will suffice to take $m = 1$ and X_1^1 to be a random variable with 1 bit of entropy. This complete the proof that $\overline{\mathcal{A}_1^{\text{class}}} = \mathcal{B}_1^{\text{class}}$. We observe that the case $n = 1$ is the only one for which the classical and quantum inequalities are identical.

For $n = 2$, we take the hyperplane $\mathcal{H}_2$, which intersects the cone in a triangular region defined by the equalities

$$H_1 + H_2 = 2 \tag{4.5}$$

and

$$H_\emptyset = 0,$$

and by the inequalities

$$H_{12} \leq H_1 + H_2, \tag{4.6}$$

$$H_1 \leq H_{12} \tag{4.7}$$

and

$$H_2 \leq H_{12}. \tag{4.8}$$

We shall take $m = 3$ and define X^1, X^2 and X^3 as follows. Take X^1, to be a random variable with 2 bits of entropy in part 1. Thus $H(X^1_{12}) = H(X^1_1) = 2$ and $H(X^1_2) = 0$. Similarly, take X^2 to be a random variable with 2 bits of entropy in part 2. Thus $H(X^2_{12}) = H(X^2_2) = 2$ and $H(X^2_1) = 0$. Finally, take X^3 to be a random variable with 1 bit of entropy common to parts 1 and 2. Then $H(X^3_1) = H(X^3_2) = H(X^3_{12}) = 1$. We observe that $\{H(X^3_K)\}_{K\subseteq\{1,2\}}$ does not lie in an extreme ray of the quantum cone $\mathcal{B}_2$, since $H(X^3_K) = S(\varrho^1_K)/4 + S(\varrho^2_K)/4 + S(\varrho^3_K)/2$, but it will turn out to lie in an extreme ray of $\mathcal{B}_2^{\text{class}}$.

The random variables X^1, X^2 and X^3 satisfy

$$H(X_1) + H(X_2) = 2,$$

so their allocations of entropy lie in $\mathcal{H}_2$. Suppose that we are given $\{H_K\}_{K\subseteq\{1,2\}}$ in $\mathcal{B}_2^{\text{class}} \cap \mathcal{H}_2$. Then if we take

$$\lambda_1 = (H_{12} - H_2)/2,$$

$$\lambda_2 = (H_{12} - H_1)/2,$$

and

$$\lambda_3 = (H_1 + H_2 - H_{12}),$$

we have $\lambda_l \geq 0$ for $1 \leq l \leq 3$ by virtue of (4.6), (4.7) and (4.8), and $\sum_{1\leq l\leq 3} \lambda_l = 1$ by virtue of (4.5). Furthermore, we have

$$H_K = \sum_{1\leq l\leq 3} \lambda_l H(X^l_K)$$

for all $K \subseteq \{1,2\}$. This completes the proof that $\overline{\mathcal{A}_2^{\text{class}}} = \mathcal{B}_2^{\text{class}}$.

For $n = 3$, we take the hyperplane $\mathcal{H}_6$, which intersects the cone $\mathcal{B}_3^{\text{class}}$ in the polytope defined by the equalities

$$H_1 + H_2 + H_3 = 6$$

and

$$H_\emptyset = 0,$$

the six inequalities that are distinguished instances of (3.15) and the three inequalities that are distinguished instances of (3.16).

We shall take $m = 8$ and define the random variables $X^1, \ldots, X^8$ as follows. For $1 \le l \le 3$, take X^l to be a random variable with 6 bits of entropy in part l. Thus for $1 \le l \le 3$ we have

$$H(X^l_K) = \begin{cases} 6, & \text{if } l \in K; \\ 0, & \text{otherwise.} \end{cases}$$

For $4 \le l \le 6$, we take X^l to be a random variable with 3 bits of entropy common to the parts k such that $k \ne l-3$. Thus if we set $L_l = \{1,2,3\} \setminus \{l-3\}$, then for $4 \le l \le 6$ we have

$$H(X^l_K) = \begin{cases} 3, & \text{if } K \cap L_l \ne \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

We observe that for $4 \le l \le 6$, $\{H(X^l_K)\}_{K \subseteq \{1,2,3\}}$ does not lie in an extreme ray of the quantum cone $\mathcal{B}_3$, since $H(X^l_K) = \left(\sum_{k \in L_l} S(\varrho^k_K)\right)/4 + S(\varrho^l_K)/2$, but it will turn out to lie in an extreme ray of $\mathcal{B}_3^{\text{class}}$. We take X^7 to be a random variable with 2 bits of entropy common to parts 1, 2 and 3. Thus we have

$$H(X^7_K) = \begin{cases} 2, & \text{if } K \ne \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

Finally, we take $X^8 = (X^8_1, X^8_2, X^8_3)$, where X^8_1 and X^8_2 are independent random variables uniformly distributed over $GF(4)$ (which therefore each have 2 bits of entropy) and $X^8_3 = X^8_1 + X^8_2$ is their sum in $GF(4)$. Then any two of components of X^8 are independent, and any two components determine the third. These facts imply

$$H(X^8_K) = \begin{cases} 2, & \text{if } \#K = 1; \\ 4, & \text{if } \#K \in \{2,3\}; \\ 0, & \text{otherwise.} \end{cases}$$

We observe that $\{H(X^8_K)\}_{K \subseteq \{1,2,3\}}$ does not lie in an extreme ray of the quantum cone $\mathcal{B}_3$, since $H(X^8_K) = \left(\sum_{1 \le l \le 3} S(\varrho^l_K)\right)/6 + S(\varrho^8_K)/2$, but it will turn out to lie in an extreme ray of $\mathcal{B}_3^{\text{class}}$.

The random variables X^l for $1 \le l \le 8$ all satisfy

$$H(X^l_1) + H(X^l_2) + H(X^l_3) = 6,$$

so their allocations of entropy lie in $\mathcal{H}_6$. Suppose we are given $\{H_K\}_{K\subseteq\{1,2,3\}}$ in $\mathcal{B}_3^{\text{class}}\cap\mathcal{H}_6$. We shall again prove three lemmas.

Define

$$M = H_1 + H_2 + H_3 - H_{12} - H_{13} - H_{23} + H_{123},$$

and let $\mathcal{M}$ denote the hyperplane defined by $M = 0$.

Lemma 4.9: If $\{H_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}$, then $\{H_K\}_{K\subseteq\{1,2,3\}}$ is a convex combination of the allocations of entropy of $X^1, \ldots, X^6$.

Proof: Suppose we are given $\{H_K\}_{K\subseteq\{1,2,3\}}$ in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}$. Take

$$\begin{aligned}
\lambda_1 &= G(123, 1)/6, \\
\lambda_2 &= G(123, 2)/6, \\
\lambda_3 &= G(133, 3)/6, \\
\lambda_4 &= F(2, 3)/3, \\
\lambda_5 &= F(1, 3)/3
\end{aligned}$$

and

$$\lambda_6 = F(1, 2)/3.$$

Then $\lambda_l \geq 0$ for $1 \leq l \leq 6$, since $\{H_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}}$, and $\sum_{1\leq l\leq 6} \lambda_l = \left(\sum_{1\leq k\leq 3} H_k\right)/6 = 1$ since $\{H_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{H}_6$. Furthermore

$$H_K = \sum_{1\leq l\leq 6} \lambda_l \, H(X_K^l)$$

for $K \subseteq \{1, 2, 3\}$. $\triangle$

Let $\mathcal{M}^+$ denote the half-space defined by $M \geq 0$.

Lemma 4.10: If $\{H_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}^+$, then $\{H_K\}_{K\subseteq\{1,2,3\}}$ is a convex combination of the allocation of entropy of X^7 and an allocation of entropy of lying in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}$.

Proof: Suppose we are given $\{H_K\}_{K\subseteq\{1,2,3\}}$ in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}^+$. Take $\lambda = M/2 \geq 0$. Define $H'_K = H_K - \lambda \, H(X_K^7)$ for $K \subseteq \{1, 2, 3\}$. We shall show that $\{H'_K\}_{K\subseteq\{1,2,3\}}$ lies in $(1-\lambda)(\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}) = \mathcal{B}_3^{\text{class}} \cap \mathcal{H}_{(1-\lambda)6} \cap \mathcal{M}$. This will complete the proof of the lemma.

Since

$$H(X_1^7) + H(X_2^7) + H(X_3^7) - H(X_{12}^7) - H(X_{13}^7) - H(X_{23}^7) + H(X_{123}^7) = 2,$$

we have $M' = M - 2\lambda = 0$, and thus $\{H'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{M}$. Since

$$H(X^7_1) + H(X^7_2) + H(X^7_3) = 6,$$

we have $H'_1 + H'_2 + H'_3 = (1-\lambda)6$, and thus $\{H'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{H}_{(1-\lambda)6}$. It remains to show that $\{H'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}}$.

If we define

$$F'(I,J) = H'_I + H'_J - H'_{I\cup J} - H'_{I\cap J}$$

then since

$$H(X^7_{123}) + H(X^7_1) = H(X^7_{12}) + H(X^7_{13}),$$

we have $F'(12,13) = F(12,13) \geq 0$, and we have $F'(12,23) \geq 0$ and $F'(13,23) \geq 0$ by permutation of the indices in this argument. Furthermore, since

$$H(X^7_1) + H(X^7_2) - H(X^7_{12}) = 2,$$

we have $F'(1,2) = F(1,2) - 2\lambda = F(1,2) - M$. But $M = F(1,2) - F(13,23)$, so $F'(1,2) = F(13,23) \geq 0$, and we have $F'(1,3) \geq 0$ and $F'(2,3) \geq 0$ by permutation of the indices in this argument. Thus $\{H'_K\}_{K\subseteq\{1,2,3\}}$ satisfies the distinguished instances of (3.15).

If we define

$$G'(I,J) = H'_I - H'_{I\setminus J},$$

then since

$$H(X^7_{123}) = H(X^7_{12}),$$

we have $G'(123,3) = G(123,3) \geq 0$, and we have $G'(123,2) \geq 0$ and $G'(123,1) \geq 0$ by permutation of the indices in this argument. This completes the proof that $\{H'_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}}$. $\triangle$

Let $\mathcal{M}^-$ denote the half-space defined by $M \leq 0$.

Lemma 4.11: If $\{H_K\}_{K\subseteq\{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}^-$, then $\{H_K\}_{K\subseteq\{1,2,3\}}$ is a convex combination of the allocation of entropy of X^8 and an allocation of entropy of lying in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}$.

Proof: Suppose we are given $\{H_K\}_{K\subseteq\{1,2,3\}}$ in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}^-$. Take $\lambda = -M/2 \geq 0$. Define $H'_K = H_K - \lambda\, H(X^8_K)$ for $K \subseteq \{1,2,3\}$. We shall show that $\{H'_K\}_{K\subseteq\{1,2,3\}}$ lies in $(1-\lambda)(\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6 \cap \mathcal{M}) = \mathcal{B}_3^{\text{class}} \cap \mathcal{H}_{(1-\lambda)6} \cap \mathcal{M}$. This will complete the proof of the lemma.

Since

$$H(X_1^8) + H(X_2^8) + H(X_3^8) - H(X_{12}^8) - H(X_{13}^8) - H(X_{23}^8) + H(X_{123}^8) = -2,$$

we have $M' = M + 2\lambda = 0$, and thus $\{H'_K\}_{K \subseteq \{1,2,3\}}$ lies in $\mathcal{M}$. Since

$$H(X_1^8) + H(X_2^8) + H(X_3^8) = 6,$$

we have $H'_1 + H'_2 + H'_3 = (1-\lambda)6$, and thus $\{H'_K\}_{K \subseteq \{1,2,3\}}$ lies in $\mathcal{H}_{(1-\lambda)6}$. It remains to show that $\{H'_K\}_{K \subseteq \{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}}$.

Since

$$H(X_{12}^8) = H(X_1^8) + H(X_2^8),$$

we have $F'(1,2) = F(1,2) \geq 0$, and we have $F'(1,3) \geq 0$ and $F'(2,3) \geq 0$ by permutation of the indices in this argument. Furthermore, since

$$H(X_{13}^8) + H(X_{23}^8) - H(X_{123}^8) - H(X_3^8) = 2,$$

we have $F'(13,23) = F(13,23) - 2\lambda = F(1,2) + M$. But $M = F(1,2) - F(13,23)$, so $F'(13,23) = F(1,2) \geq 0$, and we have $F'(12,23) \geq 0$ and $F'(12,13) \geq 0$ by permutation of the indices in this argument. Thus $\{H'_K\}_{K \subseteq \{1,2,3\}}$ satisfies the distinguished instances of (3.15).

Since

$$H(X_{123}^8) = H(X_{12}^8),$$

we have $G'(123,3) = E(123,3) \geq 0$, and we have $G'(123,2) \geq 0$ and $G'(123,1) \geq 0$ by permutation of the indices in this argument. Thus $\{H'_K\}_{K \subseteq \{1,2,3\}}$ satisfies the distinguished instances of (3.16). This completes the proof that $\{H'_K\}_{K \subseteq \{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}}$. △

We are now in a position to prove the following theorems.

Theorem 4.12: If $\{H_K\}_{K \subseteq \{1,2,3\}}$ lies in $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6$, then $\{H_K\}_{K \subseteq \{1,2,3\}}$ is a convex combination of the allocations of entropy of $X^1, \ldots, X^8$.

Proof: According as $M \geq 0$ or $M < 0$, we apply Lemma 4.10 or Lemma 4.11. Application of Lemma 4.9 to the resulting $\{H'_K\}_{K \subseteq \{1,2,3\}}$ then completes the proof. △

Theorem 4.13: The allocations of entropy of the random variables $X^1, \ldots, X^8$ are extreme points of the polytope $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6$; that is, none of these allocations of entropy can be expressed as a convex combination of the other seven.

Proof: For each of the eight allocations, we shall find a linear inequality violated by that allocation, but satisfied by the other seven. Since the set of allocations satisfying a given linear inequality is closed under taking convex combinations, this will suffice to prove the proposition.

Take

$$Q = H(X_{123}).$$

Then $Q - G(123, 1) > 0$ for each of $X^1, \ldots, X^8$ except X^1. The extremity of the allocations for X^2 and X^3 is established in the same way, with a permutation of the indices. Furthermore, $Q - F(12, 13) > 0$ for each of $X^1, \ldots, X^8$ except X^4. The extremity of the allocations for X^5 and X^6 is established in the same way, with a permutation of the indices. Finally, $M \leq 0$ for each of $X^1, \ldots, X^8$ except X^7, and $M \geq 0$ for each of $X^1, \ldots, X^8$ except X^8. △

It is implicit in the proofs of Theorems 4.12 and 4.13 that the polytope $\mathcal{B}_3^{\text{class}} \cap \mathcal{H}_6$ is the union of two 7-dimensional simplices, one with extreme points at the allocations of $X^1, \ldots, X^6, X^7$ and the other with extreme points at the allocations of $X^1, \ldots, X^6, X^8$, and whose intersection is the 6-dimensional simplex with extreme points at the allocations of $X^1, \ldots, X^6$.

5. Symmetric Allocations

In this section, we shall consider the symmetric allocations of entropy $\{S(\varrho_i)\}_{0 \leq i \leq n}$. (In this section, ϱ_i denotes $\varrho_{\{1,\ldots,i\}}$, rather than $\mathrm{Tr}_{N \setminus \{i\}}(\varrho)$.) We begin by deriving a set of linear inequalities governing $\{S(\varrho_i)\}_{0 \leq i \leq n}$.

From (3.3) with (3.8), we obtain

$$S(\varrho_{i+1}) + S(\varrho_{i-1}) \leq 2S(\varrho_i) \tag{5.1}$$

for $1 \leq i \leq n-1$, a set of $n-1$ inequalities. From (3.4) with (3.9), we obtain

$$S(\varrho_{i-1}) + S(\varrho_{n-i}) \leq S(\varrho_i) + S(\varrho_{n-i+1}) \tag{5.2}$$

for $1 \leq i \leq n-1$. This inequality is unchanged if we replace i by $n-i+1$. Thus we obtain all the distinct instances of this inequality for $1 \leq i \leq \lceil n/2 \rceil$, a set of $\lceil n/2 \rceil$ distinct inequalities. To these we add the trivial

$$S(\varrho_0) = 0 \tag{5.3}$$

from (3.5). Let $\mathcal{D}_n$ denote the convex cone in $\mathbf{R}^{n+1}$ defined by (5.1), (5.2) and (5.3).

Proposition 5.1: None of the instances of (5.1) for $1 \le i \le n-1$ or (5.2) for $1 \le i \le \lceil n/2 \rceil$ can be deduced from the other instances of these inequalities.

Proof: To prove that a given instance cannot be deduced from the others, we shall find a collection $\{S_k\}_{0 \le k \le n}$ of numbers, with $S_0 = 0$, that violate the given instance (when substituted into the instance in the obvious way), but satisfy all the other instances.

Given an instance of (5.1) with $1 \le i \le n-1$, we take

$$S_k = \begin{cases} k, & \text{if } 0 \le k \le i-1; \\ k-1, & \text{if } i \le k \le n. \end{cases}$$

It is easy to check that $S_0 = 0$, that the given instance of (5.1) is violated, and that all other instances of (5.1) are satisfied. Furthermore, since S_k is a non-decreasing function of k, all instances of (5.2) are satisfied. This completes the proof for an instance of (5.1).

Given an instance of (5.2) with $1 \le i \le \lceil n/2 \rceil$, we take

$$S_k = \begin{cases} k, & \text{if } 0 \le k \le i-1; \\ i-1, & \text{if } i \le k \le n-i; \\ n-k-1, & \text{if } n-i+1 \le k \le n. \end{cases}$$

It is easy to check that $S_0 = 0$, that the given instance of (5.2) is violated, and that all other instances of (5.2) are satisfied. Furthermore, since S_k is a concave function of k, all instances of (5.1) are satisfied. This completes the proof for an instance of (5.1). $\triangle$

Corollary 5.2: The cone $\mathcal{D}_n$ lies in the hyperplane (5.3) and is bounded by the facets (5.1) with $1 \le i \le n-1$ and (5.2) with $1 \le i \le \lceil n/2 \rceil$.

Proof: This follows immediately from Proposition 5.1. $\triangle$

Our next goal is to show that $\overline{\mathcal{C}_n} = \mathcal{D}_n$ for all n. Since $\mathcal{C}_n \subseteq \mathcal{D}_n$ and $\mathcal{D}_n$ is topologically closed, we have $\overline{\mathcal{C}_n} \subseteq \mathcal{D}_n$. Thus it will sufice to show that $\overline{\mathcal{C}_n} \supseteq \mathcal{D}_n$. From Lemmas 4.1 and 4.2, we have that for any $c > 0$, the hyperplane

$$\mathcal{K}_c = \{S : S_1 = c\}$$

meets every ray of $\mathcal{D}_n$. Thus it will suffice to show that $\overline{\mathcal{C}_n \cap \mathcal{K}_m} \supseteq \mathcal{D}_n \cap \mathcal{K}_m$, where

$$m = \lceil \log_2(2n+1) \rceil.$$

To do this, we shall find a set of weakly symmetric quantum states such that every point in $\mathcal{D}_n \cap \mathcal{K}_m$ can be expressed as a convex combination of the symmetric allocations of entropy of the states in this set. It will be convenient to use double superscripts for this set of quantum states. Specifically, we shall find quantum states $\varrho^{a,b}$ with

$$1 \le a \le n \text{ and } \max\{a, n-a\} \le b \le n. \tag{5.4}$$

A simple calculation shows that this gives

$$\binom{\lfloor n/2 \rfloor + 2}{2} + \binom{\lceil n/2 \rceil + 1}{2} - 1$$

quantum states. We shall arrange that the symmetric allocation of entropy of $\varrho^{a,b}$ is given by

$$S(\varrho^{a,b}) = \begin{cases} im, & \text{if } 0 \le i \le a; \\ am, & \text{if } a+1 \le i \le b; \\ (a+b-i)m, & \text{if } b+1 \le i \le n. \end{cases} \tag{5.5}$$

(These allocations will not all be extreme points of $\mathcal{D}_n \cap \mathcal{K}_m$. We shall identify the extreme points later; for now it will be convenient to work with this larger set of allocations.)

Lemma 5.3: Suppose $1 \le a \le 2n \le 2^m - 1$. Then there exist $2n$ vectors $v_1, \ldots, v_{2n}$ in $GF(2^m)$ such that any a of these vectors are linearly independent.

Proof: Let $\sigma_1, \ldots, \sigma_{2n}$ be distinct non-zero elements of $GF(2^m)$, and let

$$v_i = (\sigma_i^0, \sigma_i^1, \ldots, \sigma_i^{a-1}).$$

Then any a of these vectors $v_{i_1}, \ldots, v_{i_a}$ are linearly independent, since the Vandermonde determinant

$$\det \begin{pmatrix} \sigma_{i_1}^0 & \ldots & \sigma_{i_1}^{a-1} \\ & \ldots & \\ \sigma_{i_a}^0 & \ldots & \sigma_{i_a}^{a-1} \end{pmatrix} = \prod_{1 \le j < k \le a} (\sigma_{i_k} - \sigma_{i_j})$$

does not vanish. △

We now take $\varrho^{a,b} = \mathrm{Tr}_{\{n+1,\ldots,a+b\}}(\varrho_{a+b}^{a,b})$, where $\varrho_{a+b}^{a,b} = |\psi^{a,b}\rangle \langle \psi^{a,b}|$,

$$|\psi^{a,b}\rangle = \frac{1}{2^{am/2}} \sum_{\beta \in GF(2^m)^a} |v_1 \cdot \beta, \ldots, v_{a+b} \cdot \beta\rangle,$$

$|\alpha_1, \ldots, \alpha_{a+b}\rangle = |\alpha_1\rangle \otimes \cdots \otimes |\alpha_{a+b}\rangle$, $\alpha \cdot \beta$ denotes the inner product in $GF(2^m)^a$, and $v_1, \ldots, v_{a+b}$ satisfy Lemma 5.3. For $0 \le i \le a$, any i of the $a+b$ vectors $v_1, \ldots, v_{a+b}$ span a space of dimension i. This fact, together with Lemma 3.2 and the fact that $\varrho_{a+b}^{a,b}$ is a pure state, imply that the quantum states $\varrho^{a,b}$ with (5.4) are weakly symmetric and the symmetric allocation of entropy of $\varrho^{a,b}$ is as given by (5.5).

The states $\varrho^{a,b}$ with (5.4) all satisfy $S(\varrho_1^{a,b}) = m$, so their symmetric allocations of entropy lie in $\mathcal{K}_m$. Suppose we are given $\{S_i\}_{0\le i\le n}$ in $\mathcal{D}_n \cap \mathcal{K}_m$. We shall show that $\{S_i\}_{0\le i\le n}$ is a convex combination of the allocations of $\varrho^{a,b}$ with (5.4).

For $0 \le i \le n$, define

$$T_i = \max_{0\le h\le i} S_h,$$

and set $T_{n+1} = T_n$. For $1 \le a \le n$, define

$$\mu_a = (2T_a - T_{a-1} - T_{a+1})/m.$$

By virtue of (5.1), we have $\mu_a \ge 0$, and since $T_1 = S_1 = m$, we have

$$\sum_{1\le a\le n} \mu_a = 1. \tag{5.6}$$

For $0 \le i \le n$, define

$$R_i = \max_{i\le h\le n} S_h,$$

and set $R_{n+1} = R_n - m$. For $1 \le b \le n$, define

$$\nu_a = (2R_b - R_{b-1} - R_{b+1})/m.$$

By virtue of (5.1), we have $\nu_a \ge 0$, and since $R_{n+1} = R_n - m$, we have

$$\sum_{1\le b\le n} \nu_b = 1. \tag{5.7}$$

We shall next find $\lambda_{a,b} \ge 0$ for $1 \le a \le n$ and $\max\{a, n-a\} \le b \le n$ such that

$$\sum_{\max\{a,n-a\}\le b\le n} \lambda_{a,b} = \mu_a \tag{5.8}$$

for $1 \le a \le n$ and

$$\sum_{n-b\le a\le b} \lambda_{a,b} = \nu_b \tag{5.9}$$

for $1 \le b \le n$. A straightforward calculation using (5.5) show that these conditions imply

$$S_i = \sum_{1\le a\le n} \; \sum_{\max\{a,n-a\}\le b\le n} \lambda_{a,b}\, S(\varrho_i^{a,b}), \tag{5.10}$$

which is the desired conclusion. (The sequence $\{\mu_a\}_{1\le a\le n}$ encodes the increases in the sequence $\{S_i\}_{0\le i\le n}$, and $\{\nu_b\}_{1\le b\le n}$ encodes the decreases. Condition (5.10) follows from (5.8) and (5.9) because $S(\varrho_i^{a,b})$ increases with i for $0\le i\le a$ and decreases for $b\le i\le n$.)

To fulfill (5.8) and (5.9), we shall define a double sequence $\{\kappa_{a,b}\}_{1\le a\le n, a\le b\le n+1}$. (The quantity $\kappa_{a,b}$ represents the amount of μ_a that remains to be allotted to $\lambda_{a,c}$ with $b\le c\le n$.) We shall define the $\kappa_{a,b}$ and $\lambda_{a,b}$ in order of increasing b. We take

$$\kappa_{b,b} = \mu_b, \tag{5.11}$$

$$\kappa_{a,b} = \kappa_{a,b-1} - \lambda_{a,b-1} \tag{5.12}$$

for $n-b+1\le a\le b-1$, and

$$\kappa_{a,b} = \kappa_{a,b-1}$$

for $1\le a\le n-b$. If $b\le n$, we then take

$$\lambda_{a,b} = \max\bigl\{0, \min\{\kappa_{a,b}, \nu_b - \sum_{a+1\le c\le b}\kappa_{c,b}\}\bigr\}$$

for $n-b\le a\le b$.

Since $\mu_b\ge 0$ and $\lambda_{a,b}\le\kappa_{a,b}$, we have

$$\kappa_{a,b}\ge 0 \tag{5.13}$$

for all $1\le a\le b\le n$. We shall prove (5.9) by induction on b. To establish (5.9), it will suffice to show that

$$\sum_{n-b\le a\le b}\kappa_{a,b}\ge\nu_b \tag{5.14}$$

for all $1\le b\le n$. Since S_i is an increasing and concave funtion of i for $0\le i\le\lceil n/2\rceil$, we have $\nu_b = 0$ for $1\le b\le\lceil n/2\rceil-1$, and this together with (5.13) implies (5.14) for $1\le b\le\lceil n/2\rceil-1$. For $\lceil n/2\rceil\le b\le n$, if $\nu_b>0$ we have

$$\nu_b = S_b - S_{b+1} - \sum_{c\le b-1}\nu_c \tag{5.15}$$

from the definition of ν_b. We have

$$\sum_{n-b\le a\le b}\mu_a = S_{n-b+1} - S_{n-b} \tag{5.16}$$

from the definition of μ_a. From (5.11) and (5.12), we obtain

$$\kappa_{a,b} = \mu_a - \sum_{\max\{a,n-a\} \le c \le b-1} \lambda_{a,c} \tag{5.17}$$

and

$$\sum_{n-b \le a \le b} \kappa_{a,b} = \sum_{n-b \le a \le b} \mu_a - \sum_{c \le b-1} \sum_{n-c \le a \le c} \lambda_{a,c}.$$

Using (5.16) and the inductive hypothesis, this becomes

$$\sum_{n-b \le a \le b} \kappa_{a,b} = S_{n-b+1} - S_{n-b} - \sum_{c \le b-1} \nu_c.$$

Combining this with (5.15), we obtain

$$\sum_{n-b \le a \le b} \kappa_{a,b} - \nu_b = (S_{n-b+1} - S_{n-b}) - (S_b - S_{b+1}),$$

so that (5.14) follows from (5.2).

It remains to establish (5.8). Since from (5.17) we have

$$\sum_{\max\{a,n-a\} \le b \le n} \lambda_{a,b} = \mu_a - \kappa_{a,n+1}, \tag{5.18}$$

it will suffice to show that $\kappa_{a,n+1} = 0$ for $1 \le a \le n$. Since $\kappa_{a,n+1} \ge 0$, it will suffice to show that $\sum_{1 \le a \le n} \kappa_{a,n+1} = 0$. This follows from our previous results:

$$\begin{aligned} 1 &= \sum_{1 \le b \le n} \nu_b \\ &= \sum_{1 \le b \le n} \sum_{n-b \le a \le b} \lambda_{a,b} \\ &= \sum_{1 \le a \le n} \sum_{\max\{a,n-a\} \le b \le n} \lambda_{a,b} \\ &= \sum_{1 \le a \le n} (\mu_a - \kappa_{a,n+1}) \\ &= 1 - \sum_{1 \le a \le n} \kappa_{a,n+1}, \end{aligned}$$

using (5.7), (5.9), exchanging the order of summation, (5.18) and (5.6).

This establishes the following theorem.

Theorem 5.4: If $\{S_i\}_{0 \le i \le n}$ lies in $\mathcal{D}_n \cap \mathcal{K}_m$, then $\{S_i\}_{0 \le i \le n}$ is a convex combination of the symmetric allocations of entropy of $\varrho^{a,b}$ for $1 \le a \le n$ and $\max\{a, n-a\} \le b \le n$.

It remains to identify the extreme points of the polytope $\mathcal{D}_n \cap \mathcal{K}_m$. We begin by identifying some non-extreme points.

If

$$a \ge \lceil n/2 \rceil \text{ and } a+1 \le b \le n, \tag{5.19}$$

then we have

$$S(\varrho_i^{a,b}) = \frac{S(\varrho_i^{a,a}) + S(\varrho_i^{b,b})}{2}$$

for all $0 \le i \le n$. Thus the allocations of $\varrho^{a,b}$ with (5.19) are not extreme. A simple calculation shows that there are

$$\binom{\lfloor n/2 \rfloor + 1}{2}$$

pairs (a, b) satisfying (5.19), and thus

$$\binom{\lceil n/2 \rceil + 1}{2} + \lfloor n/2 \rfloor$$

pairs satisfying the complementary conditions

$$(1 \le a \le \lceil n/2 \rceil - 1 \text{ and } n-a+1 \le b \le n) \text{ or } \lceil n/2 \rceil \le a = b \le n. \tag{5.20}$$

Theorem 5.5: The symmetric allocations of entropy of the states $\varrho^{a,b}$ satisfying (5.20) are extreme points of the polytope $\mathcal{D}_n \cap \mathcal{K}_m$; that is, none of these allocations can be expressed as a convex combination of the others.

Proof: For each of these allocations, we shall find a linear inequality violated by that allocation, but satisfied by the others.

For $1 \le i \le n-1$, let

$$\Delta(i) = 2S(\varrho_i) - S(\varrho_{i-1}) - S(\varrho_{i+1}),$$

so that (5.1) is equivalent to $\Delta(i) \ge 0$. For $1 \le i \le \lceil n/2 \rceil$, let

$$E(i) = S(\varrho_i) - S(\varrho_{i-1}) + S(\varrho_{n-i+1}) - S(\varrho_{n-i}),$$

so that (5.2) is equivalent to $E(i) \ge 0$. Let

$$\begin{aligned} W &= \sum_{1 \le i \le n-1} \Delta(i) \\ &= S(\varrho_1) + S(\varrho_{n-1}) - S(\varrho_n), \end{aligned}$$

so that $W > 0$ if and only if $\Delta(i) > 0$ for some $1 \leq i \leq n-1$.

If $1 \leq a \leq \lceil n/2 \rceil - 1$ and $n - a + 1 \leq b \leq n-1$, then $W - \Delta(a) - \Delta(b) > 0$ is violated by $\varrho^{a,b}$, but satisfied by all the others, and $W - \Delta(a) - E(1) > 0$ is violated by $\varrho^{a,n}$, but satisfied by all the others.

If n is even, then $W - E(n/2) > 0$ is violated by $\varrho^{n/2,n/2}$, but satisfied by all the others. If $\lfloor n/2 \rfloor + 1 \leq a \leq n-1$, then $W - \Delta(a) - E(\lceil n/2 \rceil) > 0$ is violated by $\varrho^{a,a}$, but satisfied by all the others and $W - E(\lceil n/2 \rceil) - E(1) > 0$ is violated by $\varrho^{n,n}$, but satisfied by all the others. $\triangle$

We shall now turn to the classical counterparts of the results we have just derived. We shall show that $\overline{\mathcal{C}_n^{\text{class}}} = \mathcal{D}_n^{\text{class}}$ for all n, a result implicit in the work of Han [H1]. (In this section, X_i denotes $X_{\{1,\ldots,i\}}$, rather than $\pi(X)$ with $\pi : \Omega \to \Omega_i$ the canonical projection.) We begin by deriving a set of linear inequalities governing $\{H(X_i)\}_{0 \leq i \leq n}$. From (3.12) with (3.17), we obtain

$$H(X_{i+1}) + H(X_{i-1}) \leq 2H(X_i) \tag{5.21}$$

for $1 \leq i \leq n-1$, and from (3.13) with (3.18), we obtain

$$H(X_{n-1}) \leq H(X_n). \tag{5.22}$$

To these we add the trivial

$$H(X_0) = 0. \tag{5.23}$$

Let $\mathcal{D}_n^{\text{class}}$ denote the convex cone in $\mathbf{R}^{n+1}$ defined by (5.21), (5.22) and (5.23).

Proposition 5.6: None of the instances of (5.21) for $1 \leq i \leq n-1$ or (5.22) can be deduced from other instances of these inequalities.

Proof: Given an instance of (5.21) with $1 \leq i \leq n-1$, we take

$$H_k = \begin{cases} k, & \text{if } 0 \leq k \leq n-1; \\ n-2, & \text{if } k = n. \end{cases}$$

The $H_0 = 0$, and the given instance of (5.21) is violated, but all other instances of (5.21) and (5.22) are satisfied.

Given the inequality (5.22), we take

$$H_k = \begin{cases} k, & \text{if } 0 \leq k \leq i-1; \\ k-1, & \text{if } i \leq k \leq n. \end{cases}$$

The $H_0 = 0$, and (5.22) is violated, but all instances of (5.21) are satisfied. $\triangle$

Corollary 5.7: The cone $\mathcal{D}_n^{\text{class}}$ lines in the hyperplane (5.23) and is bounded by the facets (5.21) with $1 \le i \le n-1$ and (5.22).

Proof: This follows immediately from Proposition 5.6. $\triangle$

Our next goal is to show that $\overline{\mathcal{C}_n^{\text{class}}} = \mathcal{D}_n^{\text{class}}$. Since $\mathcal{C}_n^{\text{class}} \subseteq \mathcal{D}_n^{\text{class}}$ and $\mathcal{D}_n^{\text{class}}$ is topologically closed, we have $\overline{\mathcal{C}_n^{\text{class}}} \subseteq \mathcal{D}_n^{\text{class}}$. Thus it will suffice to show that $\overline{\mathcal{C}_n^{\text{class}}} \supseteq \mathcal{D}_n^{\text{class}}$ for all n. Since $\mathcal{D}_n^{\text{class}} \subseteq \mathcal{D}_n$, we have that for any $c > 0$, the hyperplane

$$\mathcal{K}_c = \{H : H_1 = c\}$$

meets every ray of $\mathcal{D}_n^{\text{class}}$. Thus it wil suffice to show that $\overline{\mathcal{C}_n^{\text{class}}} \cap \mathcal{K}_m \supseteq \mathcal{D}_n^{\text{class}} \cap \mathcal{K}_m$, where

$$m = \lceil \log_2(2n+1) \rceil.$$

To do this we shall find a set $\{X^1, \ldots, X^n\}$ of weakly symmetric random variables such that every point $\{H_i\}_{0 \le i \le n}$ in $\overline{\mathcal{C}_n^{\text{class}}} \cap \mathcal{K}_m$ can be expressed as a convex combination of the symmetric allocations of entropy of these random variables.

For $1 \le a \le n$, we take

$$X^a = (v_1 \cdot \beta, \ldots, v_n \cdot \beta),$$

where $v_1, \ldots, v_n$ satisfy Lemma 5.3, $\alpha \cdot \beta$ denotes the inner product in $GF(2^m)^a$, and the random variable β is uniformly distributed over $GF(2^m)^a$. Then the random variables $X^1, \ldots, X^n$ are weakly symmetric and we have

$$H(X_i^a) = \begin{cases} im, & \text{if } 0 \le i \le a; \\ am, & \text{if } a \le i \le n. \end{cases} \tag{5.24}$$

Furthermore, the random variables $X^1, \ldots, X^n$ all satisfy $H(X_1^a) = m$, so their symmetric allocationsof entropy all lie in $\mathcal{K}_m$. Suppose we are given $\{H_i\}_{0 \le i \le n}$ in $\mathcal{D}_n^{\text{class}} \cap \mathcal{K}_m$. We take

$$\lambda_a = (2H_a - Ha + 1 - H_{a-1})/m$$

for $1 \le a \le n-1$, and

$$\lambda_n = (H_n - H_{n-1})/m.$$

By virtue of (5.21) and (5.22) we have $\lambda_a \ge 0$ for $1 \le a \le n$, and since $H_1 = m$ we have

$$\sum_{1 \le a \le n} \lambda_a = 1.$$

Finally, from (5.24) we have

$$H_i = \sum_{1\le a\le n} \lambda_a\, H(X_i^a)$$

for $0 \le i \le n$. This establishes the following theorem.

Theorem 5.8: If $\{H_i\}_{0\le i\le n}$ lies in $\mathcal{D}_n^{\text{class}} \cap \mathcal{K}_m$, then $\{H_i\}_{0\le i\le n}$ is a convex combination of the symmetric allocations of entropy of X^a for $1 \le a \le n$.

Finally, we have the following theorem.

Theorem 5.9: The symmetric allocations of entropy of $X^1, \ldots, X^n$ are extreme points of the polytope $\mathcal{D}_n^{\text{class}} \cap \mathcal{K}_m$; that is, none of these allocations of entropy can be expressed as a convex combination of the others.

Proof: For $1 \le i \le n-1$, let

$$F(i) = 2H(X_i) - H(X_{i+1}) - H(X_{i-1}),$$

so that (5.21) is equivalent to $F(i) \ge 0$. Let

$$G = H(X_n) - H(X_{n-1}),$$

so that (5.22) is equivalent to $G \ge 0$. Let

$$\begin{aligned} V &= \sum_{1\le i\le n-1} F(i) + G \\ &= H(X_1), \end{aligned}$$

so that $V > 0$ if and only if either $G > 0$ or $F(i) > 0$ for some $1 \le i \le n-1$. Then for $1 \le i \le n-1$, $V - F(i) > 0$ is violated by X^i, but satisfied by all of the other $n-1$ random variables $X^1, \ldots, X^n$. Furthermore, $V - G > 0$ is violated by X^n, but satisfied by the other $n-1$ random variables $X^1, \ldots, X^{n-1}$. $\triangle$

It is implicit in the proofs of Theorems 5.8 and 5.9 that the polytope $\mathcal{D}_n^{\text{class}} \cap \mathcal{K}_m$ is a simplex with the symmetric allocations of entropy of $X^1, \ldots, X^n$ as its extreme points.

6. Conclusion

The main question left open by the present work is of course whether $\overline{\mathcal{A}_n} = \mathcal{B}_n$ for $n \ge 4$. It should be noted in this connection that Zhang and Yeung [Z2] have shown that $\overline{\mathcal{A}_n^{\text{class}}} \subsetneq \mathcal{B}_n^{\text{class}}$ for $n \ge 4$, by giving explicit inequalities that are satisfied by all allocations of entropy of quadripartite random variables, but are not satisifed by all points of $\mathcal{B}_4^{\text{class}}$.

Another open question is whether $\mathcal{A}_n = \overline{\mathcal{A}_n}$ (that is, whether $\mathcal{A}_n$ is topologically closed). Here, Zhang and Yeung [Z1] have show that $\mathcal{A}_n^{\mathrm{class}} \underset{\neq}{\subset} \overline{\mathcal{A}_n^{\mathrm{class}}}$ for $n \geq 3$.

Another problem that remains is to obtain a more complete description of the geometry of the polytope $\mathcal{D}_n \cap \mathcal{K}_m$. We have enumerated its facets and its extreme points, but we have not expressed it as an essentially disjoint union of simplices, as we have for the other polytopes appearing in this paper. A well known theorem of Caratheodory [C] states that a point in d-dimensional space that lies in the convex hull of a set of other points in fact lies in the convex hull of a subset of at most $d+1$ of these other points. Thus a symmetric allocation of entropy is a convex combination of at most $n+1$ extreme allocations, and it would be interesting to replace Theorems 5.4 and 5.5 with a characterization of $\mathcal{D}_n \cap \mathcal{K}_m$ that exhibits this fact.

7. Acknowledgment

A variant of the state ϱ_{1234}^{8} used in Section 4 was shown to the author by John Smolin, who explained that it had been shown to him by Serge Massar, who constructed it after seeing a less symmetrical state with similar properties discovered by Smolin, Elitza Maneva and Ashish Thapliyal. This state is of course essential to the main result of Section 4, and its symmetry properties are crucial to the generalizations that are needed in Section 5.

8. References

[A1] T. Ando, "Concavity of Certain Maps on Positive Definite Matrices and Applications to Hadamard Products", *Lin. Alg. and Appl.*, 26 (1979) 203–241.

[A2] H. Araki and E. H. Lieb, "Entropy Inequalities", *Commun. Math. Phys.*, 18 (1970) 160–170.

[C] C. Carathéodory, "Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen", *Math. Ann.*, 64 (1907) 95–115.

[F] S. Fujishige, "Polymatroidal Dependence Structure of a Set of Random Variables", *Inform. and Control*, 39 (1978) 55–72.

[E] H. Epstein, "Remarks on Two Theorems of E. Lieb", *Commun. Math. Phys.*, 31 (1973) 317–325.

[F] S. Fujishige, "Polymatroidal Dependence Structure of a Set of Random Variables", *Inform. and Control*, 39 (1978) 55–72.

[H1] T. S. Han, “Nonnegative Entropy Measures of Multivariate Symmetric Correlations”, *Inform. and Control*, 36 (1978) 133–156.

[H2] T. S. Han, “A Uniqueness of Shannon’s Information Distance and Related Nonnegativity Problems”, *J. Combinatorics, Information and System Sciences*, 6 (1981) 320–331.

[L1] E. H. Lieb, “Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture”, *Adv. in Math.*, 11 (1973) 267–288.

[L2] E. H. Lieb and M. B. Ruskai, “Proof of the Strong Subadditivity of Quantum-Mechanical Entropy”, *J. Math. Phys.*, 14 (1973) 1938–1941.

[N] J. von Neumann, “Thermodynamik quantenmechanischer Gesamtheiten”, *Gött. Nachr.*, (1927) 273–291.

[R] D. Ruelle, *Statistical Mechanics—Rigorous Results*, W. A.Benjamin, New York, 1969.

[S1] C. E. Shannon, “A Mathematical Theory of Communication”, *Bell Syst. Tech. J.*, 27 (1948) 379–423, 623–655.

[S2] B. Simon, *Trace Ideals and Their Applications*, Cambridge University Press, Cambridge, UK, 1979.

[U] A. Uhlmann, “Relative Entropy and the Wigner-Yanase-Dyson-Lieb Concavity in an Interpolation Theory”, *Comm. Math. Phys.*, 54 (1977) 21–32.

[W] D. J. A. Welsh, *Matroid Theory*, Academic Press, London, 1976.

[Y] R. W. Yeung, “A Framework for Linear Information Inequalities”, *IEEE Trans. Inform. Theory*, 43 (1997) 1924–1934.

[Z1] Z. Zhang and R. W. Yeung, “A Non-Shannon-Type Conditional Inequality of Information Quantities”, *IEEE Trans. Inform. Theory*, 43 (1997) 1982–1986.

[Z2] Z. Zhang and R. W. Yeung, “On Characterization of Entropy Function via Information Inequalities”, *IEEE Trans. Inform. Theory*, 44 (1998) 1440–1452.

Appendix: The Lieb-Ruskai Inequalities

The quantum information-theoretic inequalities (3.1) and (3.2) were established by Lieb and Ruskai [L2]. Their proof used two previous results. One of these is an elementary inequality

$$\mathrm{Tr}(A\log A - A\log B) \geq \mathrm{Tr}(A - B),$$

due to O. Klein, which is given a short proof by Ruelle [R, p. 26–27]. (In this appendix, we shall use natural logarithms. This has no effect on the validity of linear inequalities among entropies, such as (3.1) and (3.2).) The other is the much deeper result that

$$f_K(A) = \mathrm{Tr}\big(\exp(K + \log A)\big)$$

is a concave function of the positive martrix A, for every self-adjoint matrix K. This latter result was proved by Lieb [L1]; Lieb's proof is long, and relies the fact that

$$g_{K,t}(A, B) = \mathrm{Tr}(K^* A^{1-t} K B^t)$$

is jointly concave in the matrices A and B, for every matrix K and every real number $0 \leq t \leq 1$. Lieb's proof of this last result used complex-analytic properties of matrix functions. A much more direct proof of the concavity of $f_K(A)$, but one still relying on complex-analytic methods, was given by Epstein [E]. Another proof of the joint concavity of $g_{K,t}$, but one still relying on complex-analytic methods, was given by Ando [A1]. A short and completely elementary proof of (3.1) has been given by Uhlmann [U], however. We shall give below a version of his proof, mostly following the account of Simon [S2, pp. 102–105].

We have seen in Section 3 that (3.1) and (3.2) are equivalent. Therefore it will suffice to establish (3.2). We shall show below that the quantity

$$S(\varrho_{12} \mid \varrho_1) = S(\varrho_{12}) - S(\varrho_1)$$

is a concave function of ϱ_{12}. Given this, the quantity

$$\begin{aligned}\Delta(\varrho_{123}) &= S(\varrho_{13} \mid \varrho_1) + S(\varrho_{23} \mid \varrho_2) \\ &= S(\varrho_{13}) - S(\varrho_1) + S(\varrho_{23}) - S(\varrho_2),\end{aligned}$$

being a sum of two concave functions of the linear functions Tr_2 and Tr_1 of ϱ_{123}, is also a concave function of ϱ_{123}. Any state ϱ_{123} can be expressed as a convex combination of pure states:

$$\varrho_{123} = \sum_{1 \leq i \leq d} \vartheta_i \; \varrho_{123,i},$$

where d is the dimension of ϱ_{123}. Since Δ is concave, we have

$$\Delta(\varrho_{123}) \geq \sum_{1 \leq i \leq d} \vartheta_i \; \Delta(\varrho_{123,i}) \qquad (A.1)$$

For each pure state $\varrho_{123,i}$, Lemma 3.2 gives $S(\varrho_{13,i}) = S(\varrho_{2,i})$ and $S(\varrho_{23,i}) = S(\varrho_{1,i})$, so we have $\Delta(\varrho_{123,i}) = 0$. Thus by $(A.1)$ we have $\Delta(\varrho_{123}) \geq 0$. But this is equivalent to (3.2).

It remains to prove the concavity of $S(\varrho_{12} \mid \varrho_1)$. We shall show below that the quantity

$$f(A, B) = -\mathrm{Tr}(A \log A - A \log B)$$

is a jointly concave function of the positive matrices A and B. (A self-adjoint matrix X is *positive* if $\langle u|X|u\rangle \geq 0$ for every vector u.) Given this, we have

$$\begin{aligned} S(\varrho_{12} \mid \varrho_1) &= -\mathrm{Tr}_{12}(\varrho_{12} \log \varrho_{12}) + \mathrm{Tr}_1(\varrho_1 \log \varrho_1) \\ &= -\mathrm{Tr}_{12}\big(\varrho_{12} \log \varrho_{12} - \big(\varrho_1 \otimes (I_d/d)\big) \log\big(\varrho_1 \otimes (I_d/d)\big)\big) + \log d \\ &= -\mathrm{Tr}_{12}\big(\varrho_{12} \log \varrho_{12} - \varrho_{12} \log\big(\varrho_1 \otimes (I_d/d)\big)\big) + \log d, \end{aligned}$$

where d is the dimension of ϱ_2, and I_d is a $d \times d$ identity matrix. Thus the concavity of $S(\varrho_{12} \mid \varrho_1)$ follows from the joint concavity of $f(A, B)$ with the positive matrices $A = \varrho_{12}$ and $B = \varrho_1 \otimes (I_d/d)$.

It remains to prove the joint concavity of $f(A, B)$ in the positive matrices A and B. We shall show below that the quantity

$$g_t(A, B) = \mathrm{Tr}(A^{1-t}\, B^t)$$

is jointly concave in the positive matrices A and B, for every real $0 \leq t \leq 1$. Given this, we have $-g_0(A, B) = -\mathrm{Tr}(A)$, which is linear in A, and thus jointly concave in A and B. Thus

$$h_t(A, B) = \frac{g_t(A, B) - g_0(A, B)}{t},$$

being the sum of two concave functions, is also jointly concave in A and B. It follows that

$$\begin{aligned} \lim_{t \to 0} h_t(A, B) &= \left.\frac{dg_t(A, B)}{dt}\right|_{t=0} \\ &= -\mathrm{Tr}(A \log A - A \log B) \\ &= f(A, B) \end{aligned}$$

is also jointly concave in A and B.

It remains to prove the joint concavity of $g_t(A, B)$ in the positive matrices A and B. We shall show below that if

$$R_1 \geq S_1 + T_1 \tag{A.2}$$

and

$$R_2 \geq S_2 + T_2, \tag{A.3}$$

where R_1, R_2, S_1, S_2, T_1 and T_2 are positive matrices, and where R_1 commutes with R_2, S_1 commutes with S_2 and T_1 commutes with T_2, then

$$R_1^{1-t}\, R_2^t \geq S_1^{1-t}\, S_2^t + T_1^{1-t}\, T_2^t \tag{A.4}$$

for every $0 \leq t \leq 1$. (The inequality $X \geq Y$ between self-adjoint matrices means that $X - Y$ is positive.) Given this, we shall consider the $d \times d$ matrices A and B to be vectors in a vector space of dimension d^2, with the inner product

$$\langle X \mid Y \rangle = \mathrm{Tr}(X^* Y).$$

Let A_0, A_1, B_0 and B_1 be positive matrices, and let $0 \leq \vartheta \leq 1$. Define R_1, R_2, S_1, S_2, T_1 and T_2 by taking

$$\begin{aligned} R_1 X &= \big(\vartheta A_0 + (1-\vartheta) A_1\big) X, \\ R_2 X &= X \big(\vartheta B_0 + (1-\vartheta) B_1\big) X, \\ S_1 X &= \vartheta A_0 X, \\ S_2 X &= \vartheta X B_0, \\ T_1 X &= (1-\vartheta) A_1 X \end{aligned}$$

and

$$T_2 X = (1-\vartheta) X B_1,$$

for every vector X. The positivity of these matrices follow from the positivity of A_0, A_1, B_0 and B_1, and each of the three pairs of matrices commutes. We have $(A.2)$ and $(A.3)$, since in fact $R_1 = S_1 + T_1$ and $R_2 = S_2 + T_2$. Thus $(A.4)$ implies

$$\langle I | R_1^{1-t}\, R_2^t | I \rangle \geq \langle I | S_1^{1-t}\, S_2^t | I \rangle + \langle I | T_1^{1-t}\, T_2^t | I \rangle,$$

which is equivalent to

$$\mathrm{Tr}\Big(\big(\vartheta A_0 + (1-\vartheta) A_1\big)^{1-t} \big(\vartheta B_0 + (1-\vartheta) B_1\big)^t\Big) \geq \vartheta\, \mathrm{Tr}(A_0^{1-t}\, B_0^t) + (1-\vartheta)\, \mathrm{Tr}(A_1^{1-t}\, B_1^t),$$

and this verifies the joint concavity of $g_t(A, B)$.

It remains to prove that $(A.2)$ and $(A.3)$ imply $(A.4)$ for all $0 \leq t \leq 1$. We shall show below that $(A.2)$ and $(A.3)$ imply $(A.4)$ in the special case $t = 1/2$. Given this, let

$H \subseteq [0,1]$ be the set of values of t such that $(A.2)$ and $(A.3)$ imply $(A.4)$. If p and q belong to H, we may take

$$R_3 = R_1^{1-p}\, R_2^p,$$
$$R_4 = R_1^{1-q}\, R_2^q,$$
$$S_3 = S_1^{1-p}\, S_2^p,$$
$$S_4 = S_1^{1-q}\, S_2^q,$$
$$T_3 = T_1^{1-p}\, T_2^p$$

and

$$T_4 = T_1^{1-q}\, T_2^q.$$

Since p and q belong to H, $(A.2)$ and $(A.3)$ imply

$$R_3 \geq S_3 + T_3$$

and

$$R_4 \geq S_4 + T_4.$$

By the special case $t = 1/2$, we conclude that

$$R_3^{1/2}\, R_4^{1/2} \geq S_3^{1/2}\, S_4^{1/2} + T_3^{1/2}\, T_4^{1/2}.$$

But this is the conclusion $(A.4)$ with $t = (p+q)/2$. Since H contains 0 and 1 and is closed under taking midpoints $p, q \mapsto (p+q)/2$, H contains all dyadic rational t in the range $0 \leq t \leq 1$. Since the dyadic rationals are dense in the reals, by the continuity of exponentiation $t \mapsto X^t$, H contains all reals in the range $0 \leq t \leq 1$.

It remains to prove that $(A.2)$ and $(A.3)$ imply $(A.4)$ in the special case $t = 1/2$. In doing this, we may assume that R_1 and R_2 are invertible: since R_1 and R_2 are positive, $R_1 + \varepsilon I$ and $R_2 + \varepsilon I$ are invertible; we may prove the result with these values substituted for R_1 and R_2, then let ε tend to 0 and invoke the continuity of all the operations appearing in the conclusion. Let

$$R = R_1^{1/2}\, R_2^{1/2},$$
$$S = S_1^{1/2}\, S_2^{1/2}$$

and

$$T = T_1^{1/2}\, T_2^{1/2}.$$

Then, for any vectors v and w, we have

$$\big|\langle v|S+T|w\rangle\big| \leq \big|\langle v|S|w\rangle\big| + \big|\langle v|T|w\rangle\big|,$$

by the triangle inequality for the absolute value. Since $|\langle x \mid y\rangle| \leq \|x\| \, \|y\|$, where $\|z\| = |\langle z \mid z\rangle|^{1/2}$, we also have

$$\big|\langle v|S|w\rangle\big| + \big|\langle v|T|w\rangle\big| \leq \|S_1^{1/2}v\| \, \|S_2^{1/2}w\| + \|T_1^{1/2}v\| \, \|T_2^{1/2}w\|.$$

By the Cauchy-Schwarz inequality, $a_1b_1 + a_2b_2 \leq (a_1^2 + a_2^2)^{1/2} \, (b_1^2 + b_2^2)^{1/2}$, we have

$$\begin{aligned} &\|S_1^{1/2}v\| \, \|S_2^{1/2}w\| + \|T_1^{1/2}v\| \, \|T_2^{1/2}w\| \\ &\qquad \leq \big(\|S_1^{1/2}v\|^2 + \|T_1^{1/2}v\|^2\big)^{1/2} \, \big(\|S_2^{1/2}w\|^2 + \|T_2^{1/2}w\|^2\big)^{1/2}. \end{aligned}$$

Finally, we have

$$\begin{aligned} &\big(\|S_1^{1/2}v\|^2 + \|T_1^{1/2}v\|^2\big)^{1/2} \, \big(\|S_2^{1/2}w\|^2 + \|T_2^{1/2}w\|^2\big)^{1/2} \\ &\qquad = \langle v|S_1 + T_1|v\rangle^{1/2} \, \langle w|S_2 + T_2|w\rangle^{1/2}. \end{aligned}$$

Combining these observations yields

$$\big|\langle v|S + T|w\rangle\big| \leq \langle v|S_1 + T_1|v\rangle^{1/2} \, \langle w|S_2 + T_2|w\rangle^{1/2}.$$

Using $(A.2)$ and $(A.3)$, this implies

$$\big|\langle v|S + T|w\rangle\big| \leq \langle v|R_1|v\rangle^{1/2} \, \langle w|R_2|w\rangle^{1/2}.$$

This in turn implies that, for any unit vectors v and w, we have

$$\begin{aligned} \big|\langle v|R_1^{-1/2}(S + T)R_2^{-1/2}|w\rangle\big| &= \big|\langle R_1^{-1/2}v|(S + T)|R_2^{-1/2}w\rangle\big| \\ &\leq \langle R_1^{-1/2}v \mid R_1^{1/2}v\rangle^{1/2} \, \langle R_2^{-1/2}w \mid R_w^{1/2}w\rangle^{1/2} \\ &= 1. \end{aligned}$$

This conclusion can be written

$$\|R_1^{-1/2}(S + T)R_2^{-1/2}\| \leq 1,$$

where $\|X\| = \sup_{\|v\|=\|w\|=1} |\langle v|X|w\rangle|$ denotes the "operator norm" of X. Since all matrices here are self-adjoint, their operator norms are equal to their spectral radii, and thus we have $\|XY\| = \|YX\|$. The last inequality is thus equivalent to

$$\|R_2^{-1/4} \, R_1^{-1/4}(S + T)R_2^{-1/4} \, R_1^{-1/4}\| \leq 1.$$

For a positive matrix X, $\|X\| \leq 1$ if and only if $X \leq I$. Thus we have

$$R_2^{-1/4} \, R_1^{-1/4}(S + T)R_2^{-1/4} \, R_1^{-1/4} \leq I.$$

Since $X \leq Y$ if and only if $Z^*XZ \leq Z^*YZ$, the last inequality is equivalent to

$$\begin{aligned} S + T &\leq R_1^{1/4} \, R_2^{1/4} \, R_1^{1/4} \, R_2^{1/4} \\ &= R, \end{aligned}$$

which completes the proof that $(A.2)$ and $(A.3)$ imply $(A.4)$ in the special case $t = 1/2$.