Chapter 5

Quantum entropy and source coding

The *von Neumann entropy* of a quantum state is an information-theoretic measure of the amount of randomness or uncertainty that is inherent to that state, and the *quantum relative entropy* of one quantum state with respect to another is a related measure of the degree to which the first state differs from the second. This chapter defines these function, establishes some of their fundamental properties, and explains their connections to the task of *source coding*.

5.1 Classical entropy

The von Neumann entropy and quantum relative entropy functions are quantum analogues of classical information-theoretic notions: the Shannon entropy and (classical) relative entropy functions. It is appropriate to begin the chapter with a discussion of these classical notions, as an investigation of the mathematical properties of the von Neumann entropy and quantum relative entropy functions builds naturally on their classical counterparts.

5.1.1 Definitions of classical entropic functions

With respect to the definition that follows, the Shannon entropy is specified for every vector with nonnegative entries, over any real Euclidean space. Although it is most common that this function is considered in the case that

its argument is a probability vector, it is convenient nevertheless to extend its domain in this way.

Definition 5.1. Let Σ be an alphabet and let $u \in [0,\infty)^{\Sigma}$ be a vector of non-negative real numbers indexed by Σ. One defines the *Shannon entropy* of the vector u as

$$\mathrm{H}(u) = -\sum_{\substack{a\in\Sigma \\ u(a)>0}} u(a)\log(u(a)). \tag{5.1}$$

The Shannon entropy $\mathrm{H}(p)$ of a probability vector $p \in \mathcal{P}(\Sigma)$ is sometimes described as the amount of randomness inherent to the distribution described by p, measured in bits. Alternatively, $\mathrm{H}(p)$ may be described as the number of bits of uncertainty one has regarding the outcome of a random process described by p before the outcome is learned, or as the number of bits of information one gains as a result of learning which element $a \in \Sigma$ has been produced by such a process.

In the simple case that $\Sigma = \{0,1\}$ and $p(0) = p(1) = 1/2$, for instance, it holds that $\mathrm{H}(p) = 1$. This is natural, as one would expect that the amount of uncertainty of a uniformly generated random bit, measured in bits, would be 1 bit of uncertainty. In contrast, for a completely deterministic process, meaning one in which p is an elementary unit vector, there is no randomness or uncertainty, and no information gain when the selection is learned. Correspondingly, one has that the entropy $\mathrm{H}(p)$ is zero in this case.

It is important to recognize, however, that intuitive descriptions of the Shannon entropy, as a measure of randomness, uncertainty, or information gain, must be viewed as representing *expectations* rather than absolute or definitive measures. The following example illustrates this point.

Example 5.2. Let m be a positive integer, let

$$\Sigma = \left\{0, 1, \ldots, 2^{m^2}\right\}, \tag{5.2}$$

and define a probability vector $p \in \mathcal{P}(\Sigma)$ as follows:

$$p(a) = \begin{cases} 1 - \frac{1}{m} & \text{if } a = 0 \\ \frac{1}{m} 2^{-m^2} & \text{if } 1 \leq a \leq 2^{m^2}. \end{cases} \tag{5.3}$$

A calculation reveals that $\mathrm{H}(p) > m$, and yet the outcome 0 appears with probability $1 - 1/m$ in a random selection described by p. So, as m grows,

one becomes more and more "certain" that the outcome will be 0, and yet the "uncertainty" (as measured by the entropy) increases.

This example does not represent a paradox or suggest that the Shannon entropy is not reasonably viewed as a measure of uncertainty. If one considers an experiment in which a very large number of elements of Σ are selected independently, each according to the probability vector p, then the value $\mathrm{H}(p)$ indeed does correspond more intuitively to the average or expected amount of uncertainty of each random selection.

Sometimes one speaks of the Shannon entropy of a classical register X, with the notation $\mathrm{H}(\mathsf{X})$ being used for this purpose. This is a convenient shorthand to be interpreted as meaning $\mathrm{H}(p)$, for the probability vector p describing the probabilistic state of X at the moment under consideration. Notations such as $\mathrm{H}(\mathsf{X},\mathsf{Y})$ and $\mathrm{H}(\mathsf{X}_1,\ldots,\mathsf{X}_n)$ are used in place of $\mathrm{H}((\mathsf{X},\mathsf{Y}))$ and $\mathrm{H}((\mathsf{X}_1,\ldots,\mathsf{X}_n))$ when referring to the Shannon entropy of compound registers. Along similar lines, the notation $\mathrm{H}(\alpha_1,\ldots,\alpha_n)$ will be used in place of $\mathrm{H}((\alpha_1,\ldots,\alpha_n))$ when it is convenient to refer to the entropy of a vector written as $(\alpha_1,\ldots,\alpha_n)$.

The *relative entropy* function, which is also known as the *Kullback–Leibler divergence*, is closely related to the Shannon entropy. For the purposes of this book, the primary motivation for its introduction is that it serves as a useful analytic tool for reasoning about the Shannon entropy.

Definition 5.3. Let Σ be an alphabet and let $u, v \in [0,\infty)^{\Sigma}$ be vectors of nonnegative real numbers indexed by Σ. The *relative entropy* $\mathrm{D}(u\|v)$ of u with respect to v is defined as follows. If it is the case that $\operatorname{supp}(u) \subseteq \operatorname{supp}(v)$ (i.e., $u(a) > 0$ implies $v(a) > 0$ for all $a \in \Sigma$), then $\mathrm{D}(u\|v)$ is defined as

$$\mathrm{D}(u\|v) = \sum_{\substack{a\in\Sigma \\ u(a)>0}} u(a)\log\left(\frac{u(a)}{v(a)}\right). \tag{5.4}$$

For all other choices of u and v, one defines $\mathrm{D}(u\|v) = \infty$.

Like the Shannon entropy function, the relative entropy is most typically considered in cases where its arguments are probability vectors, but again it is convenient to extend its domain to arbitrary nonnegative real vectors.

For a given pair of probability vectors $p, q \in \mathcal{P}(\Sigma)$, the relative entropy $\mathrm{D}(p\|q)$ may be viewed as a measure of how much p differs from q in a

certain information-theoretic sense. Analytically speaking, it fails to satisfy the requirements of being a true metric: it is not symmetric, it takes infinite values for some pairs of inputs, and it does not satisfy the triangle inequality. When extended to arbitrary vectors of the form $u, v \in [0, \infty)^{\Sigma}$, it may also take negative values. Despite these apparent shortcomings, the relative entropy is an indispensable information-theoretic tool.

Two additional functions derived from the Shannon entropy function are the *conditional Shannon entropy* and the *mutual information*. Both concern correlations between two classical registers X and Y, and are functions of the joint probabilistic state of the pair (X, Y). The conditional Shannon entropy of X given Y is defined as

$$\mathrm{H}(\mathsf{X} \,|\, \mathsf{Y}) = \mathrm{H}(\mathsf{X}, \mathsf{Y}) - \mathrm{H}(\mathsf{Y}). \tag{5.5}$$

Intuitively speaking, this quantity represents the expected amount of uncertainty regarding the classical state of X one would have upon learning the classical state of Y. The *mutual information* between X and Y is defined as

$$\mathrm{I}(\mathsf{X} : \mathsf{Y}) = \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X}, \mathsf{Y}). \tag{5.6}$$

This quantity can alternately be expressed as

$$\mathrm{I}(\mathsf{X} : \mathsf{Y}) = \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{Y}|\mathsf{X}) = \mathrm{H}(\mathsf{X}) - \mathrm{H}(\mathsf{X}|\mathsf{Y}). \tag{5.7}$$

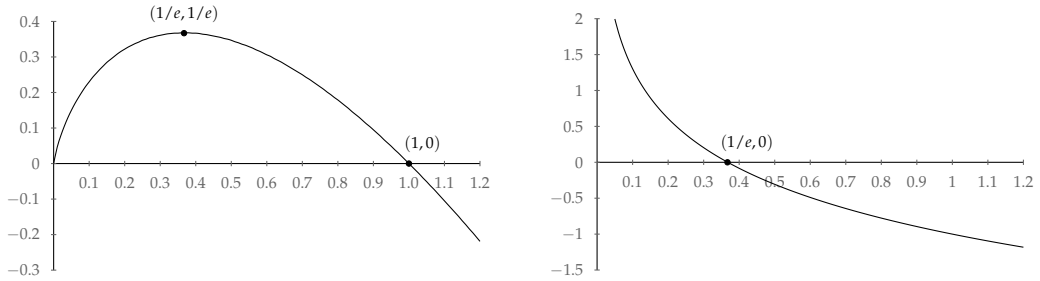
One typically views this quantity as representing the expected amount of information about X that one gains by learning the classical state of Y, or (equivalently) that one gains about Y by learning the classical state of X.

5.1.2 Properties of classical entropic functions

The Shannon and relative entropy functions possess a variety of useful and interesting properties. This section establishes several basic properties of these functions.

Scalar analogues of Shannon entropy and relative entropy

For the purposes of establishing basic analytic properties of the Shannon and relative entropy functions, it is helpful to define functions representing scalar analogues of these functions. These scalar functions are to be defined with respect to the natural logarithm rather than the base-2 logarithm, as



(a) $\eta(\alpha) = -\alpha \ln(\alpha)$ (b) $\eta'(\alpha) = -(1 + \ln(\alpha))$

Figure 5.1: Plots of the functions η and η'.

this will simplify some of the calculations to follow, particularly when they make use of differential calculus.

The first function $\eta : [0, \infty) \to \mathbb{R}$, which represents a scalar analogue of the Shannon entropy, is defined as follows:

$$\eta(\alpha) = \begin{cases} -\alpha \ln(\alpha) & \alpha > 0 \\ 0 & \alpha = 0. \end{cases} \tag{5.8}$$

The function η is continuous everywhere on its domain, and derivatives of η of all orders exist for all positive real numbers. In particular,

$$\eta'(\alpha) = -(1 + \ln(\alpha)) \tag{5.9}$$

and

$$\eta^{(n+1)}(\alpha) = \frac{(-1)^n}{\alpha^n} \tag{5.10}$$

for $n \geq 1$, for all $\alpha > 0$. Plots of the function η its first derivative η' are shown in Figure 5.1. As the second derivative of η is negative for all $\alpha > 0$, one has that η is a concave function:

$$\eta(\lambda\alpha + (1 - \lambda)\beta) \geq \lambda\eta(\alpha) + (1 - \lambda)\eta(\beta) \tag{5.11}$$

for all $\alpha, \beta \geq 0$ and $\lambda \in [0, 1]$.

The second function $\theta : [0, \infty)^2 \to (-\infty, \infty]$, which represents a scalar analogue of the relative entropy, is defined as follows:

$$\theta(\alpha, \beta) = \begin{cases} 0 & \text{if } \alpha = 0 \\ \infty & \text{if } \alpha > 0 \text{ and } \beta = 0 \\ \alpha \ln(\alpha) - \alpha \ln(\beta) & \text{if } \alpha > 0 \text{ and } \beta > 0. \end{cases} \tag{5.12}$$

It is evident from this definition that, when restricted to positive real number arguments $\alpha, \beta > 0$, the value $\theta(\alpha, \beta)$ is negative when $\alpha < \beta$, zero when $\alpha = \beta$, and positive when $\alpha > \beta$.

It is useful to note that the functions θ and η are related by the identity

$$\theta(\alpha, \beta) = -\beta \, \eta\left(\frac{\alpha}{\beta}\right), \tag{5.13}$$

which holds for all $\alpha \in [0, \infty)$ and $\beta \in (0, \infty)$. The function θ is continuous at every point (α, β) for which $\beta > 0$. It is not continuous at any point $(\alpha, 0)$, however, as every neighborhood of such a point contains both finite and infinite values.

The following useful lemma regarding the function θ is equivalent to a fact commonly known as the *log-sum inequality*.

Lemma 5.4. *Let $\alpha_0, \alpha_1, \beta_0, \beta_1 \in [0, \infty)$ be nonnegative real numbers. It holds that*

$$\theta(\alpha_0 + \alpha_1, \beta_0 + \beta_1) \leq \theta(\alpha_0, \beta_0) + \theta(\alpha_1, \beta_1). \tag{5.14}$$

Proof. If either of β_0 or β_1 is zero, the inequality is straightforward. More specifically, if $\beta_0 = 0$ and $\alpha_0 = 0$, the inequality is equivalent to

$$\theta(\alpha_1, \beta_1) \leq \theta(\alpha_1, \beta_1), \tag{5.15}$$

which is trivial, while if $\beta_0 = 0$ and $\alpha_0 > 0$, the right-hand side of (5.14) is infinite. A similar argument holds when $\beta_1 = 0$ by symmetry.

In the case that both β_0 and β_1 are positive, the inequality may be proved by combining the identity (5.13) with the concavity of η:

$$\begin{aligned}
&\theta(\alpha_0, \beta_0) + \theta(\alpha_1, \beta_1) \\
&\quad = -(\beta_0 + \beta_1) \left[\frac{\beta_0}{\beta_0 + \beta_1} \, \eta\left(\frac{\alpha_0}{\beta_0}\right) + \frac{\beta_1}{\beta_0 + \beta_1} \, \eta\left(\frac{\alpha_1}{\beta_1}\right) \right] \\
&\quad \geq -(\beta_0 + \beta_1) \, \eta\left(\frac{\alpha_0 + \alpha_1}{\beta_0 + \beta_1}\right) \\
&\quad = \theta(\alpha_0 + \alpha_1, \beta_0 + \beta_1),
\end{aligned} \tag{5.16}$$

as claimed. □

Elementary properties of Shannon entropy and relative entropy

The Shannon entropy function may be expressed in terms of the η-function as follows:

$$\mathrm{H}(u) = \frac{1}{\ln(2)} \sum_{a \in \Sigma} \eta(u(a)), \tag{5.17}$$

for every choice of an alphabet Σ and a vector $u \in [0, \infty)^{\Sigma}$. As the function η is continuous everywhere on its domain, the Shannon entropy function is continuous everywhere on its domain as well. The concavity of η implies the concavity of the Shannon entropy, as the following proposition states.

Proposition 5.5 (Concavity of Shannon entropy)**.** *Let Σ be an alphabet, let $u, v \in [0, \infty)^{\Sigma}$ be vectors, and let $\lambda \in [0, 1]$. It holds that*

$$\mathrm{H}(\lambda u + (1 - \lambda)v) \geq \lambda \, \mathrm{H}(u) + (1 - \lambda) \, \mathrm{H}(v). \tag{5.18}$$

Proof. By the concavity of the function η, one has

$$\begin{aligned} \mathrm{H}(\lambda u + (1 - \lambda)v) &= \frac{1}{\ln(2)} \sum_{a \in \Sigma} \eta(\lambda u(a) + (1 - \lambda)v(a)) \\ &\geq \frac{\lambda}{\ln(2)} \sum_{a \in \Sigma} \eta(u(a)) + \frac{1 - \lambda}{\ln(2)} \sum_{a \in \Sigma} \eta(v(a)) \\ &= \lambda \, \mathrm{H}(u) + (1 - \lambda) \, \mathrm{H}(v), \end{aligned} \tag{5.19}$$

as required. □

The next proposition states two identities that involve the Shannon entropy of direct sums and tensor products of vectors. Both identities may be verified through direct calculations.

Proposition 5.6. *Let Σ and Γ be alphabets and let $u \in [0, \infty)^{\Sigma}$ and $v \in [0, \infty)^{\Gamma}$ be vectors. It holds that*

$$\mathrm{H}(u \oplus v) = \mathrm{H}(u) + \mathrm{H}(v) \tag{5.20}$$

and

$$\mathrm{H}(u \otimes v) = \mathrm{H}(u) \sum_{b \in \Gamma} v(b) + \mathrm{H}(v) \sum_{a \in \Sigma} u(a). \tag{5.21}$$

One may observe that, for any choice of probability vectors $p \in \mathcal{P}(\Sigma)$ and $q \in \mathcal{P}(\Gamma)$, the identity (5.21) implies that

$$\mathrm{H}(p \otimes q) = \mathrm{H}(p) + \mathrm{H}(q). \tag{5.22}$$

As a special case of the same identity, one finds that

$$\mathrm{H}(\alpha p) = \alpha \, \mathrm{H}(p) - \alpha \log(\alpha) \tag{5.23}$$

for every scalar $\alpha > 0$ and every probability vector $p \in \mathcal{P}(\Sigma)$.

The relative entropy function may be expressed using the θ-function as follows:

$$\mathrm{D}(u\|v) = \frac{1}{\ln(2)} \sum_{a \in \Sigma} \theta(u(a), v(a)), \tag{5.24}$$

for every choice of an alphabet Σ and two vectors $u, v \in [0,\infty)^{\Sigma}$. It therefore holds that the relative entropy function is continuous when its domain is restricted to choices of v for which $\mathrm{supp}(v) = \Sigma$, but is not continuous at any point (u, v) for which $\mathrm{supp}(v) \neq \Sigma$.

The next proposition, which implies that the relative entropy between any two probability vectors is nonnegative, represents one application of Lemma 5.4.

Proposition 5.7. *Let Σ be an alphabet and let $u, v \in [0,\infty)^{\Sigma}$ be vectors. If it holds that*

$$\sum_{a \in \Sigma} u(a) \geq \sum_{a \in \Sigma} v(a), \tag{5.25}$$

then $\mathrm{D}(u\|v) \geq 0$. In particular, $\mathrm{D}(p\|q) \geq 0$ for all choices of probability vectors $p, q \in \mathcal{P}(\Sigma)$.

Proof. By Lemma 5.4, it holds that

$$\mathrm{D}(u\|v) = \frac{1}{\ln(2)} \sum_{a \in \Sigma} \theta(u(a), v(a)) \geq \frac{1}{\ln(2)} \, \theta\left(\sum_{a \in \Sigma} u(a), \sum_{a \in \Sigma} v(a) \right). \tag{5.26}$$

The proposition follows from the fact that $\theta(\alpha, \beta) \geq 0$ for every choice of nonnegative real numbers $\alpha, \beta \in [0,\infty)$ satisfying $\alpha \geq \beta$. □

Remark 5.8. Theorem 5.17 establishes a quantitative lower-bound on the relative entropy $\mathrm{D}(p\|q)$ in terms of the 1-norm distance $\|p - q\|_1$ between any two probability vectors p and q.

Proposition 5.7 may be used to prove upper and lower bounds on the Shannon entropy, as in the proof of the following proposition.

Proposition 5.9. *Let Σ be an alphabet, let $u \in [0,\infty)^{\Sigma}$ be a vector, and let*

$$\alpha = \sum_{a\in\Sigma} u(a). \tag{5.27}$$

It holds that

$$0 \leq \mathrm{H}(u) + \alpha \log(\alpha) \leq \alpha \log(|\Sigma|). \tag{5.28}$$

In particular, it holds that $0 \leq \mathrm{H}(p) \leq \log(|\Sigma|)$ for every probability vector $p \in \mathcal{P}(\Sigma)$.

Proof. First, suppose $p \in \mathcal{P}(\Sigma)$ is a probability vector. The Shannon entropy $\mathrm{H}(p)$ may be written as

$$\mathrm{H}(p) = \sum_{\substack{a\in\Sigma \\ p(a)>0}} p(a) \log\biggl(\frac{1}{p(a)}\biggr), \tag{5.29}$$

which is a convex combination of nonnegative real numbers, by virtue of the fact that $p(a) \leq 1$ for each $a \in \Sigma$. It follows that $\mathrm{H}(p) \geq 0$.

Next, let $q \in \mathcal{P}(\Sigma)$ be the probability vector defined by $q(a) = 1/|\Sigma|$ for each $a \in \Sigma$. One may evaluate the relative entropy $\mathrm{D}(p\|q)$ directly from its definition, obtaining

$$\begin{aligned} \mathrm{D}(p\|q) &= \sum_{a\in\Sigma} p(a)\log(p(a)) - \sum_{a\in\Sigma} p(a) \log\left(\frac{1}{|\Sigma|}\right) \\ &= -\mathrm{H}(p) + \log(|\Sigma|). \end{aligned} \tag{5.30}$$

As p and q are probability vectors, Proposition 5.7 implies that the relative entropy $\mathrm{D}(p\|q)$ is nonnegative, and therefore $\mathrm{H}(p) \leq \log(|\Sigma|)$.

Now consider $u \in [0,\infty)^{\Sigma}$ and α, as in the statement of the proposition. If it is the case that $\alpha = 0$, then it must hold that u is the zero vector, in which case the proposition may be verified directly. Otherwise, let $p \in \mathcal{P}(\Sigma)$ be the probability vector defined by the equation $\alpha p = u$. By (5.23), one has

$$\mathrm{H}(u) = \mathrm{H}(\alpha p) = \alpha\,\mathrm{H}(p) - \alpha\log(\alpha). \tag{5.31}$$

Given that $0 \leq \mathrm{H}(p) \leq \log(|\Sigma|)$, it follows that

$$-\alpha\log(\alpha) \leq \mathrm{H}(u) \leq \alpha\log(|\Sigma|) - \alpha\log(\alpha), \tag{5.32}$$

which completes the proof. □

Proposition 5.7 also leads to a proof that the Shannon entropy is subadditive, in the sense described by the proposition that follows. Intuitively speaking, this property reflects the idea that the amount of uncertainty one has about a compound register cannot be greater than the total uncertainty one has about its individual registers.

Proposition 5.10 (Subadditivity of Shannon entropy)**.** *Let* X *and* Y *be classical registers. With respect to an arbitrary probabilistic state of these registers, it holds that*

$$\mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}). \tag{5.33}$$

Proof. Let $p \in \mathcal{P}(\Sigma \times \Gamma)$ denote an arbitrary probabilistic state of the pair (X,Y), for Σ and Γ being the classical state sets of X and Y, respectively. A calculation based on the definition of the relative entropy and elementary properties of logarithms reveals the equality

$$\mathrm{D}\bigl(p \bigl\| p[\mathsf{X}] \otimes p[\mathsf{Y}]\bigr) = \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}). \tag{5.34}$$

As the relative entropy of one probability vector with respect to another is nonnegative by Proposition 5.7, the required inequality follows. □

One may observe that Proposition 5.10 is equivalent to the statement that the mutual information $\mathrm{I}(\mathsf{X}:\mathsf{Y})$ between two registers is necessarily nonnegative, or equivalently that the conditional Shannon entropy $\mathrm{H}(\mathsf{Y}|\mathsf{X})$ of one register Y given another register X is no larger than the (unconditional) Shannon entropy $\mathrm{H}(\mathsf{Y})$ of the register Y alone: $\mathrm{H}(\mathsf{Y}|\mathsf{X}) \leq \mathrm{H}(\mathsf{Y})$.

The next proposition establishes a related fact: the Shannon entropy of a pair of classical registers (X,Y) cannot be less than the Shannon entropy of either of the registers viewed in isolation. Equivalently, the conditional Shannon entropy $\mathrm{H}(\mathsf{X}|\mathsf{Y})$ is nonnegative for all possible probabilistic states of the pair (X,Y).

Proposition 5.11. *Let* X *and* Y *be classical registers. With respect to an arbitrary probabilistic state of these registers, it holds that*

$$\mathrm{H}(\mathsf{X}) \leq \mathrm{H}(\mathsf{X},\mathsf{Y}). \tag{5.35}$$

Proof. Let Σ and Γ denote the classical state sets of X and Y, respectively, and let $p \in \mathcal{P}(\Sigma \times \Gamma)$ be an arbitrary probabilistic state of (X,Y). The logarithm

is an increasing function, and therefore

$$\log(p(a,b)) \leq \log\left(\sum_{c\in\Gamma} p(a,c)\right) \tag{5.36}$$

for every pair $(a,b) \in \Sigma \times \Gamma$. It follows that

$$\begin{aligned} \mathrm{H}(\mathsf{X},\mathsf{Y}) &= -\sum_{a\in\Sigma}\sum_{b\in\Gamma} p(a,b)\log(p(a,b)) \\ &\geq -\sum_{a\in\Sigma}\left(\sum_{b\in\Gamma} p(a,b)\right)\log\left(\sum_{c\in\Gamma} p(a,c)\right) = \mathrm{H}(\mathsf{X}), \end{aligned} \tag{5.37}$$

as required. □

Remark 5.12. It should be noted that Proposition 5.11 does not carry over to the von Neumann entropy of quantum states (cf. Theorem 5.27).

The next theorem represents a direct and straightforward application of Lemma 5.4. A quantum analogue of this theorem, which is stated and proved in Section 5.2.3, is not known to have nearly so straightforward a proof.

Theorem 5.13. *Let Σ be an alphabet and let $u_0, u_1, v_0, v_1 \in [0,\infty)^\Sigma$ be vectors of nonnegative real numbers indexed by Σ. It holds that*

$$\mathrm{D}(u_0 + u_1 \| v_0 + v_1) \leq \mathrm{D}(u_0\|v_0) + \mathrm{D}(u_1\|v_1). \tag{5.38}$$

Proof. By Lemma 5.4 it holds that

$$\begin{aligned} &\mathrm{D}(u_0 + u_1 \| v_0 + v_1) \\ &\quad= \frac{1}{\ln(2)}\sum_{a\in\Sigma} \theta(u_0(a) + u_1(a), v_0(a) + v_1(a)) \\ &\quad\leq \frac{1}{\ln(2)}\sum_{a\in\Sigma} \big(\theta(u_0(a), v_0(a)) + \theta(u_1(a), v_1(a))\big) \\ &\quad= \mathrm{D}(u_0\|v_0) + \mathrm{D}(u_1\|v_1), \end{aligned} \tag{5.39}$$

as claimed. □

For all vectors $u, v \in [0,\infty)^\Sigma$ and scalars $\alpha, \beta \in [0,\infty)$ it holds that

$$\mathrm{D}(\alpha u \| \beta v) = \alpha\,\mathrm{D}(u\|v) + \frac{1}{\ln(2)}\theta(\alpha,\beta)\sum_{a\in\Sigma} u(a), \tag{5.40}$$

provided one makes the interpretation $0 \cdot \infty = 0$ in the case that $\alpha = 0$ and $\mathrm{D}(u\|v) = \infty$, or in the case that $\theta(\alpha, \beta) = \infty$ and $u = 0$. This can be verified through a direct calculation. As $\theta(\alpha, \alpha) = 0$ for all $\alpha \in [0, \infty)$, one obtains the identity

$$\mathrm{D}(\alpha u\|\alpha v) = \alpha\,\mathrm{D}(u\|v), \tag{5.41}$$

where again it is to be interpreted that $0 \cdot \infty = 0$. Alternately, one may verify that this identity holds by observing

$$\theta(\alpha\beta, \alpha\gamma) = \alpha\,\theta(\beta, \gamma) \tag{5.42}$$

for all nonnegative real numbers $\alpha, \beta, \gamma \in [0, \infty)$. Through this identity, one obtains the following corollary to Theorem 5.13.

Corollary 5.14 (Joint convexity of the relative entropy)**.** *Let Σ be an alphabet, let $u_0, u_1, v_0, v_1 \in [0, \infty)^\Sigma$ be vectors of nonnegative real numbers indexed by Σ, and let $\lambda \in [0, 1]$. It holds that*

$$\begin{aligned} &\mathrm{D}(\lambda u_0 + (1-\lambda)u_1\|\lambda v_0 + (1-\lambda)v_1) \\ &\qquad \leq \lambda\,\mathrm{D}(u_0\|v_0) + (1-\lambda)\,\mathrm{D}(u_1\|v_1). \end{aligned} \tag{5.43}$$

Through a similar argument, one may prove that the relative entropy of one vector with respect to another cannot increase under the action of any stochastic operation performed simultaneously on the two vectors.

Theorem 5.15. *Let Σ and Γ be alphabets, let $u, v \in [0, \infty)^\Sigma$ be vectors, and let $A \in \mathrm{L}\big(\mathbb{R}^\Sigma, \mathbb{R}^\Gamma\big)$ be a stochastic operator. It holds that*

$$\mathrm{D}(Au\|Av) \leq \mathrm{D}(u\|v). \tag{5.44}$$

Proof. By Lemma 5.4 along with the identity (5.42), it holds that

$$\begin{aligned} \mathrm{D}(Au\|Av) &= \frac{1}{\ln(2)} \sum_{a\in\Gamma} \theta\left(\sum_{b\in\Sigma} A(a,b)u(b), \sum_{b\in\Sigma} A(a,b)v(b)\right) \\ &\leq \frac{1}{\ln(2)} \sum_{a\in\Gamma}\sum_{b\in\Sigma} A(a,b)\,\theta(u(b), v(b)) \\ &= \frac{1}{\ln(2)} \sum_{b\in\Sigma} \theta(u(b), v(b)) \\ &= \mathrm{D}(u\|v), \end{aligned} \tag{5.45}$$

as required. □

Quantitative bounds on Shannon entropy and relative entropy

Two bounds, one concerning the Shannon entropy and one concerning the relative entropy, will now be proved. The first bound is a quantitative form of the statement that the Shannon entropy function is continuous on the set of all probability vectors.

Theorem 5.16 (Audenaert)**.** *Let $p_0, p_1 \in \mathcal{P}(\Sigma)$ be probability vectors, for Σ being an alphabet with $|\Sigma| \geq 2$. It holds that*

$$|\mathrm{H}(p_0) - \mathrm{H}(p_1)| \leq \lambda \log(|\Sigma| - 1) + \mathrm{H}(\lambda, 1 - \lambda) \tag{5.46}$$

for $\lambda = \frac{1}{2} \| p_0 - p_1 \|_1$.

Proof. The theorem holds trivially when $p_0 = p_1$, so it will be assumed that this is not the case. Let $\Sigma_0, \Sigma_1 \subseteq \Sigma$ be disjoint sets defined as

$$\begin{aligned} \Sigma_0 &= \big\{a \in \Sigma \,:\, p_0(a) > p_1(a)\big\}, \\ \Sigma_1 &= \big\{a \in \Sigma \,:\, p_0(a) < p_1(a)\big\}, \end{aligned} \tag{5.47}$$

and let vectors $u_0, u_1 \in [0,1]^{\Sigma}$ be defined as

$$u_0(a) = \begin{cases} p_0(a) - p_1(a) & \text{if } a \in \Sigma_0 \\ 0 & \text{otherwise,} \end{cases} \tag{5.48}$$

$$u_1(a) = \begin{cases} p_1(a) - p_0(a) & \text{if } a \in \Sigma_1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.49}$$

for every $a \in \Sigma$. It holds that $p_0 - p_1 = u_0 - u_1$ and $u_0(a)u_1(a) = 0$ for all $a \in \Sigma$, and moreover

$$\sum_{a \in \Sigma} u_0(a) = \lambda = \sum_{a \in \Sigma} u_1(a). \tag{5.50}$$

Taking $w \in [0,1]^{\Sigma}$ to be defined as

$$w(a) = \min\{p_0(a), p_1(a)\} \tag{5.51}$$

for every $a \in \Sigma$, one finds that $p_0 = u_0 + w$, $p_1 = u_1 + w$, and

$$\sum_{a \in \Sigma} w(a) = 1 - \lambda. \tag{5.52}$$

Next, observe that the identity

$$\begin{aligned}(\alpha+\beta)\log(\alpha+\beta) - \alpha\log(\alpha) - \beta\log(\beta)& \\ = (\alpha+\beta)\,\mathrm{H}\biggl(\frac{\alpha}{\alpha+\beta}, \frac{\beta}{\alpha+\beta}\biggr)&\end{aligned} \tag{5.53}$$

holds for every choice of nonnegative real numbers α and β, assuming at least one of them is positive (and, as is to be expected, interpreting $0\log(0)$ as 0 if either α or β is 0). Through this identity, the following two expressions are obtained:

$$\mathrm{H}(u_0) + \mathrm{H}(w) - \mathrm{H}(p_0) = \sum_{a\in\Sigma_0} p_0(a)\;\mathrm{H}\biggl(\frac{u_0(a)}{p_0(a)}, \frac{w(a)}{p_0(a)}\biggr), \tag{5.54}$$

$$\mathrm{H}(u_1) + \mathrm{H}(w) - \mathrm{H}(p_1) = \sum_{a\in\Sigma_1} p_1(a)\;\mathrm{H}\biggl(\frac{u_1(a)}{p_1(a)}, \frac{w(a)}{p_1(a)}\biggr). \tag{5.55}$$

In both cases, the restriction of the sums to the sets Σ_0 and Σ_1 reflects the exclusion of 0 summands. Both sums include only nonnegative summands, and therefore

$$\mathrm{H}(p_0) \leq \mathrm{H}(u_0) + \mathrm{H}(w) \quad \text{and} \quad \mathrm{H}(p_1) \leq \mathrm{H}(u_1) + \mathrm{H}(w). \tag{5.56}$$

Furthermore, by setting

$$\alpha_0 = \sum_{a\in\Sigma_0} p_0(a) \quad \text{and} \quad \alpha_1 = \sum_{a\in\Sigma_1} p_1(a), \tag{5.57}$$

one has that $\alpha_0, \alpha_1 \in [\lambda, 1]$, and the following two inequalities are obtained from the concavity of the Shannon entropy (Proposition 5.5):

$$\mathrm{H}(u_0) + \mathrm{H}(w) - \mathrm{H}(p_0) \leq \alpha_0\;\mathrm{H}\biggl(\frac{\lambda}{\alpha_0}, 1 - \frac{\lambda}{\alpha_0}\biggr), \tag{5.58}$$

$$\mathrm{H}(u_1) + \mathrm{H}(w) - \mathrm{H}(p_1) \leq \alpha_1\;\mathrm{H}\biggl(\frac{\lambda}{\alpha_1}, 1 - \frac{\lambda}{\alpha_1}\biggr). \tag{5.59}$$

Given that the function

$$f_\lambda(\alpha) = \alpha\;\mathrm{H}\biggl(\frac{\lambda}{\alpha}, 1 - \frac{\lambda}{\alpha}\biggr) \tag{5.60}$$

is strictly increasing on the interval $[\lambda, 1]$, it follows that

$$\begin{aligned}0 \leq \mathrm{H}(u_0) + \mathrm{H}(w) - \mathrm{H}(p_0) \leq \mathrm{H}(\lambda, 1-\lambda),& \\ 0 \leq \mathrm{H}(u_1) + \mathrm{H}(w) - \mathrm{H}(p_1) \leq \mathrm{H}(\lambda, 1-\lambda).&\end{aligned} \tag{5.61}$$

By the triangle inequality together with (5.61), one may therefore conclude that

$$\begin{aligned} &|\mathrm{H}(p_0) - \mathrm{H}(p_1)| - |\mathrm{H}(u_0) - \mathrm{H}(u_1)| \\ &\quad \leq |(\mathrm{H}(p_0) - \mathrm{H}(u_0) - \mathrm{H}(w)) - (\mathrm{H}(p_1) - \mathrm{H}(u_1) - \mathrm{H}(w))| \\ &\quad \leq \mathrm{H}(\lambda, 1-\lambda). \end{aligned} \tag{5.62}$$

To complete the proof, it suffices to prove

$$|\mathrm{H}(u_0) - \mathrm{H}(u_1)| \leq \lambda \log(|\Sigma| - 1). \tag{5.63}$$

For any alphabet Γ and any vector $v \in [0,\infty)^\Gamma$ with

$$\sum_{b \in \Gamma} v(b) = \lambda, \tag{5.64}$$

it holds that

$$-\lambda \log(\lambda) \leq \mathrm{H}(v) \leq \lambda \log(|\Gamma|) - \lambda \log(\lambda), \tag{5.65}$$

as was demonstrated in Proposition 5.9. Given that u_0 and u_1 are supported on disjoint subsets of Σ and have entries summing to the same value λ, it follows that

$$|\mathrm{H}(u_0) - \mathrm{H}(u_1)| \leq \lambda \log(|\Gamma|) - \lambda \log(\lambda) + \lambda \log(\lambda) = \lambda \log(|\Gamma|), \tag{5.66}$$

for Γ being a proper subset of Σ. The largest value obtained for the upper bound occurs when Γ has one fewer element than Σ, yielding the required inequality (5.63), which completes the proof. □

The second bound, which concerns the relative entropy function, is a quantitative form of Proposition 5.7. It lower-bounds the relative entropy $\mathrm{D}(p_0 \| p_1)$, for probability vectors p_0 and p_1, by a quantity determined by their 1-norm distance $\| p_0 - p_1 \|_1$.

Theorem 5.17 (Pinsker's inequality)**.** *Let Σ be an alphabet and $p_0, p_1 \in \mathcal{P}(\Sigma)$ be probability vectors indexed by Σ. It holds that*

$$\mathrm{D}(p_0 \| p_1) \geq \frac{1}{2\ln(2)} \| p_0 - p_1 \|_1^2. \tag{5.67}$$

The proof of Theorem 5.17 will make use of the following lemma, which is equivalent to a special case of the theorem in which $|\Sigma| = 2$.

Lemma 5.18. *For all choices of real numbers $\alpha, \beta \in [0,1]$ it holds that*

$$\theta(\alpha,\beta) + \theta(1-\alpha,1-\beta) \geq 2(\alpha-\beta)^2. \tag{5.68}$$

Proof. The inequality in the statement of the lemma is immediate in the case that $\beta \in \{0,1\}$. In the case that $\alpha \in \{0,1\}$ and $\beta \in (0,1)$, the inequality in the statement of the lemma is equivalent to

$$-\ln(\beta) \geq 2(1-\beta)^2, \tag{5.69}$$

which can be verified using elementary calculus. It remains to consider the case where $\alpha, \beta \in (0,1)$. Under this assumption it may be verified that

$$\begin{aligned} &\theta(\alpha,\beta) + \theta(1-\alpha,1-\beta) \\ &\qquad = (\eta(\beta) + \eta(1-\beta)) - (\eta(\alpha) + \eta(1-\alpha)) \\ &\qquad\quad + (\alpha-\beta)(\eta'(\beta) - \eta'(1-\beta)) \\ &\qquad = f(\beta) - f(\alpha) + (\alpha-\beta)f'(\beta) \end{aligned} \tag{5.70}$$

for $f : [0,1] \to \mathbb{R}$ defined as $f(\gamma) = \eta(\gamma) + \eta(1-\gamma)$ for all $\gamma \in [0,1]$. By Taylor's theorem it holds that

$$f(\alpha) = f(\beta) + (\alpha-\beta)f'(\beta) + \frac{1}{2}(\alpha-\beta)^2 f''(\gamma) \tag{5.71}$$

for some choice of γ being a convex combination of α and β. Equation (5.71) therefore holds for some choice of $\gamma \in (0,1)$. Evaluating the second derivative of f yields

$$f''(\gamma) = -\left(\frac{1}{\gamma} + \frac{1}{1-\gamma}\right), \tag{5.72}$$

whereby it follows that $f''(\gamma) \leq -4$ for all $\gamma \in (0,1)$. This implies the inequality (5.68), which completes the proof. □

Proof of Theorem 5.17. Define disjoint sets $\Sigma_0, \Sigma_1, \Gamma \subseteq \Sigma$ as

$$\Sigma_0 = \{a \in \Sigma : p_0(a) > p_1(a)\}, \tag{5.73}$$

$$\Sigma_1 = \{a \in \Sigma : p_0(a) < p_1(a)\}, \tag{5.74}$$

$$\Gamma = \{a \in \Sigma : p_0(a) = p_1(a)\}, \tag{5.75}$$

and define a stochastic operator $A \in \mathrm{L}\big(\mathbb{R}^{\{0,1\}}, \mathbb{R}^{\Sigma}\big)$ as

$$A = \sum_{a\in\Sigma_0} E_{0,a} + \sum_{a\in\Sigma_1} E_{1,a} + \frac{1}{2}\sum_{a\in\Gamma}(E_{0,a} + E_{1,a}). \tag{5.76}$$

Let $\alpha = (Ap_0)(0)$ and $\beta = (Ap_1)(0)$, and note that $(Ap_0)(1) = 1 - \alpha$ and $(Ap_1)(1) = 1 - \beta$, as p_0 and p_1 are probability vectors and A is stochastic. It holds that

$$\alpha - \beta = \sum_{a \in \Sigma_0} (p_0(a) - p_1(a)) = \sum_{a \in \Sigma_1} (p_1(a) - p_0(a)) = \frac{1}{2} \| p_0 - p_1 \|_1. \quad (5.77)$$

By Theorem 5.15 and Lemma 5.18, one finds that

$$\begin{aligned} \mathrm{D}(p_0 \| p_1) \geq \mathrm{D}(Ap_0 \| Ap_1) &= \frac{1}{\ln(2)} \left(\theta(\alpha, \beta) + \theta(1 - \alpha, 1 - \beta) \right) \\ &\geq \frac{2}{\ln(2)} (\alpha - \beta)^2 = \frac{1}{2 \ln(2)} \| p_0 - p_1 \|_1^2, \end{aligned} \quad (5.78)$$

as required. □

5.2 Quantum entropy

The von Neumann entropy and quantum relative entropy functions, which extend the Shannon entropy and relative entropy functions from nonnegative vectors to positive semidefinite operators, are defined in this section. Fundamental properties of these functions are established, including the key properties of joint convexity of the quantum relative entropy and strong subadditivity of the von Neumann entropy.

5.2.1 Definitions of quantum entropic functions

The von Neumann entropy function represents a natural extension of the Shannon entropy function from nonnegative vectors to positive semidefinite operators; as the following definition states, the von Neumann entropy is defined as the Shannon entropy of a given positive semidefinite operator's vector of eigenvalues.

Definition 5.19. Let $\mathcal{X}$ be a complex Euclidean space and let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator. The *von Neumann entropy* of P is defined as

$$\mathrm{H}(P) = \mathrm{H}(\lambda(P)), \quad (5.79)$$

for $\lambda(P)$ being the vector of eigenvalues of P.

The von Neumann entropy may also be expressed as

$$\mathrm{H}(P) = -\operatorname{Tr}\left(P \log(P)\right). \tag{5.80}$$

Formally speaking, this expression assumes that the operator $P\log(P)$ is defined for all positive semidefinite operators $P \in \operatorname{Pos}(\mathcal{X})$, despite the fact that $\log(P)$ is only defined for positive definite operators P. The natural interpretation is that $P\log(P)$ refers to the operator obtained by extending the scalar function

$$\alpha \mapsto \begin{cases} \alpha \log(\alpha) & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha = 0 \end{cases} \tag{5.81}$$

to positive semidefinite operators in the usual way (q.v. Section 1.1.3).

Similar to the Shannon entropy usually being considered for probability vectors, it is most common that one considers the von Neumann entropy function on density operator inputs. Also similar to the Shannon entropy, it is convenient to speak of the von Neumann entropy $\mathrm{H}(\mathsf{X})$ of a register X, which means the quantity $\mathrm{H}(\rho)$ for $\rho \in \mathrm{D}(\mathcal{X})$ representing the state of X at the moment being considered. Once again, the notation $\mathrm{H}(\mathsf{X},\mathsf{Y})$ is taken to mean $\mathrm{H}((\mathsf{X},\mathsf{Y}))$, and likewise for other forms of compound registers.

The study of the von Neumann entropy is aided by the consideration of the *quantum relative entropy*, which is an extension of the ordinary relative entropy from vectors to positive semidefinite operators.

Definition 5.20. Let $P, Q \in \operatorname{Pos}(\mathcal{X})$ be positive semidefinite operators, for a complex Euclidean space $\mathcal{X}$. The *quantum relative entropy* of P with respect to Q is defined as

$$\mathrm{D}(P\|Q) = \begin{cases} \operatorname{Tr}(P\log(P)) - \operatorname{Tr}(P\log(Q)) & \text{if } \operatorname{im}(P) \subseteq \operatorname{im}(Q) \\ \infty & \text{otherwise.} \end{cases} \tag{5.82}$$

This definition is deserving of a short explanation because, as before, the logarithm is really only defined for positive definite operators. However, the operator $P\log(Q)$ has a natural interpretation for positive semidefinite operators P and Q that satisfy $\operatorname{im}(P) \subseteq \operatorname{im}(Q)$. The action of this operator on the subspace $\operatorname{im}(Q)$ is well-defined, as Q is a positive definite operator when restricted to this subspace, while its action on the subspace $\ker(Q)$ is taken to be the zero operator. This interpretation is equivalent to identifying $0\log(0)$ with 0, as the condition $\operatorname{im}(P) \subseteq \operatorname{im}(Q)$ implies that P acts as the

zero mapping on $\ker(Q)$. The operator $P\log(P)$ is defined for all positive semidefinite operators P, as was discussed previously.

It will be convenient to make note of a concrete expression for the value $\mathrm{D}(P\|Q)$, assuming $\mathrm{im}(P) \subseteq \mathrm{im}(Q)$. Let $n = \dim(\mathcal{X})$ and suppose that

$$P = \sum_{j=1}^{n} \lambda_j(P)\, x_j x_j^* \quad \text{and} \quad Q = \sum_{k=1}^{n} \lambda_k(Q)\, y_k y_k^* \tag{5.83}$$

are spectral decompositions of P and Q. Let $r = \mathrm{rank}(P)$ and $s = \mathrm{rank}(Q)$, and observe that the expressions of P and Q in (5.83) may be truncated to r and s terms, respectively. It then holds that

$$\mathrm{D}(P\|Q) = \sum_{j=1}^{r} \sum_{k=1}^{s} |\langle x_j, y_k\rangle|^2\, \lambda_j(P) \big(\log(\lambda_j(P)) - \log(\lambda_k(Q))\big). \tag{5.84}$$

The omission of the indices $j \in \{r+1, \ldots, n\}$ and $k \in \{s+1, \ldots, n\}$ in the sums is consistent with the identification $0\log(0) = 0$ suggested above. In particular, if k is such that $\lambda_k(Q) = 0$, then it must hold that

$$|\langle x_j, y_k\rangle|^2 \lambda_j(P) = 0 \tag{5.85}$$

for all $j \in \{1, \ldots, n\}$ by the assumption $\mathrm{im}(P) \subseteq \mathrm{im}(Q)$. An alternative expression for the quantum relative entropy $\mathrm{D}(P\|Q)$, for P and Q having spectral decompositions (5.83), which is valid for all choices of P and Q, is given by

$$\mathrm{D}(P\|Q) = \frac{1}{\ln(2)} \sum_{j=1}^{n} \sum_{k=1}^{n} \theta\Big(|\langle x_j, y_k\rangle|^2 \lambda_j(P), |\langle x_j, y_k\rangle|^2 \lambda_k(Q)\Big). \tag{5.86}$$

The *conditional von Neumann entropy* and *quantum mutual information* are defined in an analogous manner to the conditional Shannon entropy and mutual information. More precisely, for two registers X and Y in a given state of interest, one defines the conditional von Neumann entropy of X given Y as

$$\mathrm{H}(\mathsf{X}|\mathsf{Y}) = \mathrm{H}(\mathsf{X},\mathsf{Y}) - \mathrm{H}(\mathsf{Y}), \tag{5.87}$$

and one defines the quantum mutual information between X and Y as

$$\mathrm{I}(\mathsf{X}:\mathsf{Y}) = \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}). \tag{5.88}$$

5.2.2 Elementary properties of quantum entropic functions

This section discusses elementary properties of the von Neumann entropy and quantum relative entropy functions. Specifically, these are properties that may be established without making essential use of the joint convexity of the quantum relative entropy, which is proved in the section following this one, or other equivalent statements.

Continuity of the von Neumann entropy

The von Neumann entropy function is continuous, owing to the fact that it is a composition of continuous functions: the Shannon entropy function is continuous at every point in its domain, as is the function

$$\lambda : \mathrm{Herm}(\mathcal{X}) \to \mathbb{R}^n, \tag{5.89}$$

for $n = \dim(\mathcal{X})$.

Simple identities concerning quantum entropy

The three propositions that follow are stated as propositions for the sake of convenience. They may be verified directly through the definitions of the von Neumann entropy and quantum relative entropy functions.

Proposition 5.21. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces for which it holds that $\dim(\mathcal{X}) \leq \dim(\mathcal{Y})$, let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, and let $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$ be an isometry. It holds that*

$$\mathrm{H}\big(VPV^*\big) = \mathrm{H}(P) \quad \textit{and} \quad \mathrm{D}\big(VPV^* \big\| VQV^*\big) = \mathrm{D}(P\|Q). \tag{5.90}$$

Proposition 5.22. *Let $P \in \mathrm{Pos}(\mathcal{X})$ and $Q \in \mathrm{Pos}(\mathcal{Y})$ be positive semidefinite operators, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$. It holds that*

$$\mathrm{H}\left(\begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}\right) = \mathrm{H}(P) + \mathrm{H}(Q) \tag{5.91}$$

and

$$\mathrm{H}(P \otimes Q) = \mathrm{Tr}(Q)\,\mathrm{H}(P) + \mathrm{Tr}(P)\,\mathrm{H}(Q). \tag{5.92}$$

In particular, it holds that

$$\mathrm{H}(\rho \otimes \sigma) = \mathrm{H}(\rho) + \mathrm{H}(\sigma) \tag{5.93}$$

for all choices of density operators $\rho \in \mathrm{D}(\mathcal{X})$ and $\sigma \in \mathrm{D}(\mathcal{Y})$.

Proposition 5.23. *Let $P_0, Q_0 \in \mathrm{Pos}(\mathcal{X})$ and $P_1, Q_1 \in \mathrm{Pos}(\mathcal{Y})$ be positive semidefinite operators, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and assume that P_0 and P_1 are nonzero. It holds that*

$$\mathrm{D}(P_0 \otimes P_1 \| Q_0 \otimes Q_1) = \mathrm{Tr}(P_1)\ \mathrm{D}(P_0 \| Q_0) + \mathrm{Tr}(P_0)\ \mathrm{D}(P_1 \| Q_1). \tag{5.94}$$

As a consequence of the tensor product identities in the second and third of these propositions, one finds that the following two identities hold for all choices of a complex Euclidean space $\mathcal{X}$, positive semidefinite operators $P, Q \in \mathrm{Pos}(\mathcal{X})$, and scalars $\alpha, \beta \in (0, \infty)$:

$$\mathrm{H}(\alpha P) = \alpha\, \mathrm{H}(P) - \alpha \log(\alpha)\, \mathrm{Tr}(P), \tag{5.95}$$

$$\mathrm{D}(\alpha P \| \beta Q) = \alpha\, \mathrm{D}(P \| Q) + \alpha \log(\alpha / \beta)\, \mathrm{Tr}(P). \tag{5.96}$$

Klein's inequality

An analogous statement to Proposition 5.7 in the quantum setting is known as *Klein's inequality*. It implies that the quantum relative entropy function is nonnegative for density operator inputs.

Proposition 5.24 (Klein's inequality)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators satisfying $\mathrm{Tr}(P) \geq \mathrm{Tr}(Q)$. It holds that $\mathrm{D}(P \| Q) \geq 0$. In particular, it holds that $\mathrm{D}(\rho \| \sigma) \geq 0$ for every choice of density operators $\rho, \sigma \in \mathrm{D}(\mathcal{X})$.*

Proof. Let $n = \dim(\mathcal{X})$ and let

$$P = \sum_{j=1}^{n} \lambda_j(P)\, x_j x_j^* \quad \text{and} \quad Q = \sum_{k=1}^{n} \lambda_k(Q)\, y_k y_k^* \tag{5.97}$$

be spectral decompositions of P and Q. By Lemma 5.4, it holds that

$$\begin{aligned} \mathrm{D}(P \| Q) &= \frac{1}{\ln(2)} \sum_{j,k} \theta\Big(|\langle x_j, y_k \rangle|^2 \lambda_j(P),\ |\langle x_j, y_k \rangle|^2 \lambda_k(Q) \Big) \\ &\geq \frac{1}{\ln(2)}\, \theta\Bigg(\sum_{j,k} |\langle x_j, y_k \rangle|^2\, \lambda_j(P), \sum_{j,k} |\langle x_j, y_k \rangle|^2\, \lambda_k(Q) \Bigg) \\ &= \frac{1}{\ln(2)}\, \theta(\mathrm{Tr}(P), \mathrm{Tr}(Q)), \end{aligned} \tag{5.98}$$

where the sums are over all $j, k \in \{1, \ldots, n\}$. If it holds that $\mathrm{Tr}(P) \geq \mathrm{Tr}(Q)$ then $\theta(\mathrm{Tr}(P), \mathrm{Tr}(Q)) \geq 0$, which completes the proof. □

Concavity and subadditivity of von Neumann entropy

Similar to the Shannon entropy, the von Neumann entropy is concave and subadditive, as the following two theorems establish.

Theorem 5.25 (Concavity of von Neumann entropy)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let $P, Q \in \operatorname{Pos}(\mathcal{X})$ be positive semidefinite operators, and let $\lambda \in [0,1]$. It holds that*

$$\mathrm{H}(\lambda P + (1-\lambda)Q) \geq \lambda \,\mathrm{H}(P) + (1-\lambda)\,\mathrm{H}(Q). \tag{5.99}$$

Proof. A straightforward computation reveals that

$$\mathrm{D}\left(\begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \middle\| \begin{pmatrix} \frac{P+Q}{2} & 0 \\ 0 & \frac{P+Q}{2} \end{pmatrix}\right) = 2\,\mathrm{H}\left(\frac{P+Q}{2}\right) - \mathrm{H}(P) - \mathrm{H}(Q). \tag{5.100}$$

As the operators

$$\begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{P+Q}{2} & 0 \\ 0 & \frac{P+Q}{2} \end{pmatrix} \tag{5.101}$$

have the same trace, the quantity represented by (5.100) is nonnegative by Klein's inequality (Proposition 5.24). It therefore holds that

$$\mathrm{H}\left(\frac{P+Q}{2}\right) \geq \frac{1}{2}\,\mathrm{H}(P) + \frac{1}{2}\,\mathrm{H}(Q) \tag{5.102}$$

which implies that the von Neumann entropy is midpoint concave on the domain $\operatorname{Pos}(\mathcal{X})$. As the von Neumann entropy function is continuous on all of $\operatorname{Pos}(\mathcal{X})$, it follows that it is in fact a concave function on this domain, which completes the proof. □

Theorem 5.26 (Subadditivity of von Neumann entropy)**.** *Let X and Y be registers. For every state of the register (X,Y), it holds that*

$$\mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}). \tag{5.103}$$

Proof. The inequality in the statement of the proposition may equivalently be written

$$\mathrm{H}(\rho) \leq \mathrm{H}(\rho[\mathsf{X}]) + \mathrm{H}(\rho[\mathsf{Y}]) \tag{5.104}$$

for $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ denoting an arbitrary state of the pair (X,Y). Using the formula

$$\log(P \otimes Q) = \log(P) \otimes \mathbb{1} + \mathbb{1} \otimes \log(Q), \tag{5.105}$$

together with the fact that $\operatorname{im}(\rho) \subseteq \operatorname{im}(\rho[\mathsf{X}] \otimes \rho[\mathsf{Y}])$, it may be observed that

$$\mathrm{D}\big(\rho \big\| \rho[\mathsf{X}] \otimes \rho[\mathsf{Y}]\big) = -\mathrm{H}(\rho) + \mathrm{H}(\rho[\mathsf{X}]) + \mathrm{H}(\rho[\mathsf{Y}]). \tag{5.106}$$

It holds that (5.106) is nonnegative by Klein's inequality (Proposition 5.24), and therefore the inequality (5.104) follows. □

Von Neumann entropy and purifications

Let X and Y be registers, and assume the compound register (X, Y) is in a pure state uu^*, for $u \in \mathcal{X} \otimes \mathcal{Y}$ being a unit vector. By means of the Schmidt decomposition, one may write

$$u = \sum_{a \in \Sigma} \sqrt{p(a)}\, x_a \otimes y_a \tag{5.107}$$

for some choice of an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and orthonormal sets

$$\{x_a : a \in \Sigma\} \subset \mathcal{X} \quad \text{and} \quad \{y_a : a \in \Sigma\} \subset \mathcal{Y}. \tag{5.108}$$

It holds that

$$(uu^*)[\mathsf{X}] = \sum_{a \in \Sigma} p(a)\, x_a x_a^* \quad \text{and} \quad (uu^*)[\mathsf{Y}] = \sum_{a \in \Sigma} p(a)\, y_a y_a^*, \tag{5.109}$$

and therefore

$$\mathrm{H}(\mathsf{X}) = \mathrm{H}(p) = \mathrm{H}(\mathsf{Y}). \tag{5.110}$$

This simple observation, when combined with the notion of purifications of states, provides a useful tool for reasoning about the von Neumann entropy of collections of registers. The proof of the following theorem offers one example along these lines.

Theorem 5.27. *Let* X *and* Y *be registers. For every state of the register* (X, Y)*, it holds that*

$$\mathrm{H}(\mathsf{X}) \leq \mathrm{H}(\mathsf{Y}) + \mathrm{H}(\mathsf{X}, \mathsf{Y}). \tag{5.111}$$

Proof. Let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be a state of the pair (X, Y), and introduce a new register Z whose associated complex Euclidean space $\mathcal{Z}$ has dimension at least $\operatorname{rank}(\rho)$. By Theorem 2.9, there must exist a unit vector $u \in \mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}$ such that

$$\rho = \mathrm{Tr}_{\mathcal{Z}}(uu^*). \tag{5.112}$$

Now, consider the situation in which the compound register $(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$ is in the pure state uu^*, which is consistent with the state of (X, Y) being ρ by the requirement (5.112). By the argument suggested above, one finds that

$$\mathrm{H}(\mathsf{X}) = \mathrm{H}(\mathsf{Y}, \mathsf{Z}) \quad \text{and} \quad \mathrm{H}(\mathsf{X}, \mathsf{Y}) = \mathrm{H}(\mathsf{Z}). \tag{5.113}$$

By the subadditivity of the von Neumann entropy (Theorem 5.26), one has

$$\mathrm{H}(\mathsf{Y}, \mathsf{Z}) \leq \mathrm{H}(\mathsf{Y}) + \mathrm{H}(\mathsf{Z}), \tag{5.114}$$

and therefore

$$\mathrm{H}(\mathsf{X}) \leq \mathrm{H}(\mathsf{Y}) + \mathrm{H}(\mathsf{X}, \mathsf{Y}). \tag{5.115}$$

The required inequality has therefore been established for all choices of the state ρ, which completes the proof. □

The Fannes–Audenaert inequality

The next theorem establishes an upper bound on the difference between the values of the von Neumann entropy function of two density operators. It may be seen as a quantitative form of the statement that the von Neumann entropy is continuous, restricted to density operator inputs. It is essentially a quantum generalization of Theorem 5.16, and its proof is based on that theorem.

Theorem 5.28 (Fannes–Audenaert inequality)**.** *Let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be density operators, for $\mathcal{X}$ being a complex Euclidean space of dimension $n \geq 2$, and let*

$$\delta = \frac{1}{2} \left\| \rho_0 - \rho_1 \right\|_1. \tag{5.116}$$

It holds that

$$|\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| \leq \delta \log(n-1) + \mathrm{H}(\delta, 1-\delta). \tag{5.117}$$

The following lemma relating the trace distance between two Hermitian operators to the 1-norm distance between vectors of their eigenvalues is used to reduce Theorem 5.28 to Theorem 5.16.

Lemma 5.29. *Let $X, Y \in \mathrm{Herm}(\mathcal{X})$ be Hermitian operators, for $\mathcal{X}$ being a complex Euclidean space of dimension n. It holds that*

$$\sum_{k=1}^{n} |\lambda_k(X) - \lambda_k(Y)| \leq \left\| X - Y \right\|_1 \leq \sum_{k=1}^{n} |\lambda_k(X) - \lambda_{n-k+1}(Y)|. \tag{5.118}$$

Proof. To prove the first inequality, let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be operators providing a Jordan–Hahn decomposition $X - Y = P - Q$, and let $Z = P + Y$ (which is equivalent to $Z = Q + X$). As $Z \geq X$, it follows from the Courant–Fischer theorem (Theorem 1.2) that $\lambda_k(Z) \geq \lambda_k(X)$ for all $k \in \{1, \ldots, n\}$. Thus,

$$\begin{aligned} \lambda_k(X) - \lambda_k(Y) &\leq \big(\lambda_k(X) - \lambda_k(Y)\big) + 2\big(\lambda_k(Z) - \lambda_k(X)\big) \\ &= 2\lambda_k(Z) - \big(\lambda_k(X) + \lambda_k(Y)\big). \end{aligned} \tag{5.119}$$

By similar reasoning it follows that

$$\lambda_k(Y) - \lambda_k(X) \leq 2\lambda_k(Z) - \big(\lambda_k(X) + \lambda_k(Y)\big)\,, \tag{5.120}$$

and therefore

$$|\lambda_k(X) - \lambda_k(Y)| \leq 2\lambda_k(Z) - \big(\lambda_k(X) + \lambda_k(Y)\big)\,. \tag{5.121}$$

Consequently, one has

$$\begin{aligned} &\sum_{k=1}^{n} |\lambda_k(X) - \lambda_k(Y)| \leq \sum_{k=1}^{n} \big(2\lambda_k(Z) - \big(\lambda_k(X) + \lambda_k(Y)\big)\big) \\ &\quad = 2\,\mathrm{Tr}(Z) - \mathrm{Tr}(X) - \mathrm{Tr}(Y) = \mathrm{Tr}(P) + \mathrm{Tr}(Q) = \big\|X - Y\big\|_1. \end{aligned} \tag{5.122}$$

To prove the second inequality, observe that

$$\|X - Y\|_1 = \langle 2\Pi - \mathbb{1}, X - Y\rangle \tag{5.123}$$

for some choice of a projection operator Π, owing to the fact that $X - Y$ is Hermitian. Let $r = \mathrm{rank}(\Pi)$, and note the following two inequalities:

$$\begin{aligned} \langle \Pi, X\rangle &\leq \lambda_1(X) + \cdots + \lambda_r(X), \\ \langle \Pi, Y\rangle &\geq \lambda_{n-r+1}(Y) + \cdots + \lambda_n(Y). \end{aligned} \tag{5.124}$$

It follows that

$$\begin{aligned} &\|X - Y\|_1 \\ &\quad \leq 2\big(\lambda_1(X) + \cdots + \lambda_r(X)\big) - 2\big(\lambda_{n-r+1}(Y) + \cdots + \lambda_n(Y)\big) \\ &\qquad\quad - \mathrm{Tr}(X) + \mathrm{Tr}(Y) \\ &\quad = \sum_{k=1}^{r} (\lambda_k(X) - \lambda_{n-k+1}(Y)) + \sum_{k=r+1}^{n} (\lambda_{n-k+1}(Y) - \lambda_k(X)) \\ &\quad \leq \sum_{k=1}^{n} |\lambda_k(X) - \lambda_{n-k+1}(Y)|, \end{aligned} \tag{5.125}$$

as required. □

Proof of Theorem 5.28. Define $\delta_0, \delta_1 \in [0,1]$ as

$$\delta_0 = \sum_{k=1}^{n} |\lambda_k(\rho_0) - \lambda_k(\rho_1)| \quad \text{and} \quad \delta_1 = \sum_{k=1}^{n} |\lambda_k(\rho_0) - \lambda_{n-k+1}(\rho_1)|. \tag{5.126}$$

By Lemma 5.29 it holds that $\delta_0 \leq \delta \leq \delta_1$, and therefore $\delta = \alpha\delta_0 + (1-\alpha)\delta_1$ for some choice of $\alpha \in [0,1]$. By Theorem 5.16 it holds that

$$\begin{aligned} &|\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| \\ &\qquad = |\mathrm{H}(\lambda_1(\rho_0), \ldots, \lambda_n(\rho_0)) - \mathrm{H}(\lambda_1(\rho_1), \ldots, \lambda_n(\rho_1))| \\ &\qquad \leq \delta_1 \log(n-1) + \mathrm{H}(\delta_1, 1-\delta_1) \end{aligned} \tag{5.127}$$

and

$$\begin{aligned} &|\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| \\ &\qquad = |\mathrm{H}(\lambda_1(\rho_0), \ldots, \lambda_n(\rho_0)) - \mathrm{H}(\lambda_n(\rho_1), \ldots, \lambda_1(\rho_1))| \\ &\qquad \leq \delta_0 \log(n-1) + \mathrm{H}(\delta_0, 1-\delta_0). \end{aligned} \tag{5.128}$$

Thus, by the concavity of the Shannon entropy function (Proposition 5.5), it follows that

$$\begin{aligned} |\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| &\leq (\alpha\delta_0 + (1-\alpha)\delta_1)\log(n-1) \\ &\qquad + \alpha\,\mathrm{H}(\delta_0, 1-\delta_0) + (1-\alpha)\,\mathrm{H}(\delta_1, 1-\delta_1) \\ &\leq \delta \log(n-1) + \mathrm{H}(\delta, 1-\delta), \end{aligned} \tag{5.129}$$

as required. □

The Fannes–Audenaert inequality is saturated for all values of $\delta \in [0,1]$ and $n \geq 2$. For instance, for any choice of $n \geq 2$ and $\Sigma = \{1, \ldots, n\}$, one may consider the density operators

$$\rho_0 = E_{1,1} \quad \text{and} \quad \rho_1 = (1-\delta)E_{1,1} + \frac{\delta}{n-1}\sum_{k=2}^{n} E_{k,k}. \tag{5.130}$$

It holds that

$$\delta = \frac{1}{2}\|\rho_0 - \rho_1\|_1 \tag{5.131}$$

and

$$|\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| = \mathrm{H}(\rho_1) = \mathrm{H}(\delta, 1-\delta) + \delta\log(n-1), \tag{5.132}$$

which saturates the Fannes–Audenaert inequality.

The quantum relative entropy as a limit of difference quotients

As the following proposition states, the quantum relative entropy can be expressed as the limit of a simple expression of its arguments. This fact will be useful in Section 5.2.3, for the task of proving that the quantum relative entropy is jointly convex.

Proposition 5.30. *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{D}(P\|Q) = \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P) - \langle P^{1-\varepsilon}, Q^{\varepsilon} \rangle}{\varepsilon}. \tag{5.133}$$

Proof. The proposition is immediate in the case that $\mathrm{im}(P) \not\subseteq \mathrm{im}(Q)$, for in this case

$$\lim_{\varepsilon \downarrow 0} \Big(\mathrm{Tr}(P) - \langle P^{1-\varepsilon}, Q^{\varepsilon} \rangle \Big) = \langle P, \mathbb{1} - \Pi_{\mathrm{im}(Q)} \rangle \tag{5.134}$$

is a positive real number. This implies that the limit in (5.133) evaluates to positive infinity, which is in agreement with the quantum relative entropy. The proposition is also immediate in the case that $P = 0$. It therefore remains to consider the case that P is a nonzero operator and $\mathrm{im}(P) \subseteq \mathrm{im}(Q)$, which is taken as an assumption for the remainder of the proof.

Let $r = \mathrm{rank}(P)$ and $s = \mathrm{rank}(Q)$. By the spectral theorem (as stated by Corollary 1.4), one may write

$$P = \sum_{j=1}^{r} \lambda_j(P)\, x_j x_j^* \quad \text{and} \quad Q = \sum_{k=1}^{s} \lambda_k(Q)\, y_k y_k^* \tag{5.135}$$

for orthonormal collections of vectors $\{x_1, \ldots, x_r\}$ and $\{y_1, \ldots, y_s\}$. Define a function $f : \mathbb{R} \to \mathbb{R}$ as

$$f(\alpha) = \sum_{j=1}^{r} \sum_{k=1}^{s} |\langle x_j, y_k \rangle|^2 \, \lambda_j(P)^{1-\alpha} \, \lambda_k(Q)^{\alpha} \tag{5.136}$$

for all $\alpha \in \mathbb{R}$. This function is differentiable at every point $\alpha \in \mathbb{R}$, with its derivative given by

$$f'(\alpha) = -\sum_{j=1}^{r} \sum_{k=1}^{s} |\langle x_j, y_k \rangle|^2 \, \lambda_j(P)^{1-\alpha} \, \lambda_k(Q)^{\alpha} \ln\!\left(\frac{\lambda_j(P)}{\lambda_k(Q)} \right). \tag{5.137}$$

Now, it holds that

$$f(\alpha) = \left\langle P^{1-\alpha}, Q^{\alpha} \right\rangle \tag{5.138}$$

for every $\alpha \in (0,1)$, while

$$f(0) = \left\langle P, \Pi_{\mathrm{im}(Q)} \right\rangle = \mathrm{Tr}(P). \tag{5.139}$$

Evaluating the derivative of f at 0 yields

$$f'(0) = -\ln(2)\, \mathrm{D}(P\|Q), \tag{5.140}$$

while the definition of the derivative, as the limit of difference quotients, yields

$$f'(0) = \lim_{\varepsilon \downarrow 0} \frac{f(\varepsilon) - f(0)}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{\langle P^{1-\varepsilon}, Q^{\varepsilon} \rangle - \mathrm{Tr}(P)}{\varepsilon}. \tag{5.141}$$

The proposition follows by combining equations (5.141) and (5.140). □

5.2.3 Joint convexity of quantum relative entropy

This section contains a proof of a fundamental fact concerning the quantum relative entropy, which is that it is a jointly convex function. By making use of this key fact, one may prove that several other important properties of the von Neumann entropy and quantum relative entropy functions hold.

Proof of the joint convexity of the quantum relative entropy

Multiple proofs of the joint convexity of the quantum relative entropy are known. The proof to be presented below will make use of the following technical lemma relating the diagonal and off-diagonal blocks of any 2-by-2 positive semidefinite block operator, under the assumption that the blocks are Hermitian and the diagonal blocks commute.

Lemma 5.31. *Let $\mathcal{X}$ be a complex Euclidean space, let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators such that $[P, Q] = 0$, and let $H \in \mathrm{Herm}(\mathcal{X})$ be a Hermitian operator for which*

$$\begin{pmatrix} P & H \\ H & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X}). \tag{5.142}$$

It holds that $H \leq \sqrt{P}\sqrt{Q}$.

Proof. The lemma will first be proved for P and Q being positive definite operators. By Lemma 3.18 it follows that

$$\left\| P^{-\frac{1}{2}} H Q^{-\frac{1}{2}} \right\| \leq 1, \tag{5.143}$$

which implies that every eigenvalue of the operator $P^{-\frac{1}{2}} H Q^{-\frac{1}{2}}$ is bounded by 1 in absolute value. As P and Q commute, it holds that the eigenvalues of $P^{-\frac{1}{4}} Q^{-\frac{1}{4}} H Q^{-\frac{1}{4}} P^{-\frac{1}{4}}$ agree with those of $P^{-\frac{1}{2}} H Q^{-\frac{1}{2}}$, and therefore

$$\lambda_1 \left(P^{-\frac{1}{4}} Q^{-\frac{1}{4}} H Q^{-\frac{1}{4}} P^{-\frac{1}{4}} \right) \leq 1. \tag{5.144}$$

The inequality (5.144) is equivalent to

$$P^{-\frac{1}{4}} Q^{-\frac{1}{4}} H Q^{-\frac{1}{4}} P^{-\frac{1}{4}} \leq \mathbb{1}, \tag{5.145}$$

which, again by the commutativity of P and Q, implies $H \leq \sqrt{P}\sqrt{Q}$.

In the general case where P and Q are not necessarily positive definite, the argument above may be applied to $P + \varepsilon \mathbb{1}$ and $Q + \varepsilon \mathbb{1}$ in place of P and Q, respectively, to obtain

$$H \leq \sqrt{P + \varepsilon \mathbb{1}} \sqrt{Q + \varepsilon \mathbb{1}} \tag{5.146}$$

for all $\varepsilon > 0$. The function $\varepsilon \mapsto \sqrt{P + \varepsilon \mathbb{1}} \sqrt{Q + \varepsilon \mathbb{1}} - H$ is continuous on the domain $[0, \infty)$, and so the preimage of the closed set $\mathrm{Pos}(\mathcal{X})$ under this function is closed. Given that every $\varepsilon > 0$ is contained in this preimage, it follows that 0 is contained in the preimage as well: $\sqrt{P}\sqrt{Q} - H$ is positive semidefinite, which proves the lemma. □

The next step toward the joint convexity of the quantum relative entropy is to prove the following theorem. It is one formulation of a fact known as *Lieb's concavity theorem*.

Theorem 5.32 (Lieb's concavity theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $A_0, A_1 \in \mathrm{Pos}(\mathcal{X})$ and $B_0, B_1 \in \mathrm{Pos}(\mathcal{Y})$ be positive semidefinite operators. For every choice of a real number $\alpha \in [0, 1]$ it holds that*

$$(A_0 + A_1)^{\alpha} \otimes (B_0 + B_1)^{1-\alpha} \geq A_0^{\alpha} \otimes B_0^{1-\alpha} + A_1^{\alpha} \otimes B_1^{1-\alpha}. \tag{5.147}$$

Remark 5.33. Within the context of this theorem and its proof, one should make the interpretation $P^0 = \Pi_{\mathrm{im}(P)}$ for every positive semidefinite operator P.

Proof of Theorem 5.32. For every real number $\alpha \in [0,1]$, define operators as follows:

$$\begin{aligned} X(\alpha) &= A_0^{\alpha} \otimes B_0^{1-\alpha}, \\ Y(\alpha) &= A_1^{\alpha} \otimes B_1^{1-\alpha}, \\ Z(\alpha) &= (A_0 + A_1)^{\alpha} \otimes (B_0 + B_1)^{1-\alpha}. \end{aligned} \tag{5.148}$$

The operators within these three individual collections commute, meaning

$$[X(\alpha), X(\beta)] = 0, \quad [Y(\alpha), Y(\beta)] = 0, \quad \text{and} \quad [Z(\alpha), Z(\beta)] = 0 \tag{5.149}$$

for every choice of $\alpha, \beta \in [0,1]$, and moreover it holds that

$$\sqrt{X(\alpha)}\sqrt{X(\beta)} = X\left(\frac{\alpha+\beta}{2}\right), \tag{5.150}$$

$$\sqrt{Y(\alpha)}\sqrt{Y(\beta)} = Y\left(\frac{\alpha+\beta}{2}\right), \tag{5.151}$$

$$\sqrt{Z(\alpha)}\sqrt{Z(\beta)} = Z\left(\frac{\alpha+\beta}{2}\right). \tag{5.152}$$

With respect to these operators, the statement of the theorem is equivalent to the claim that

$$Z(\alpha) \geq X(\alpha) + Y(\alpha) \tag{5.153}$$

for every $\alpha \in [0,1]$. The function

$$\alpha \mapsto Z(\alpha) - (X(\alpha) + Y(\alpha)) \tag{5.154}$$

defined on the interval $[0,1]$ is continuous, and therefore the preimage of the closed set $\mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ under this function is closed. It therefore suffices to prove that the set of all $\alpha \in [0,1]$ for which (5.153) holds is dense in $[0,1]$.

Now, suppose it has been proved that

$$Z(\alpha) \geq X(\alpha) + Y(\alpha) \quad \text{and} \quad Z(\beta) \geq X(\beta) + Y(\beta) \tag{5.155}$$

for some particular choice of real numbers $\alpha, \beta \in [0,1]$. It holds that

$$\begin{pmatrix} \sqrt{X(\alpha)} \\ \sqrt{X(\beta)} \end{pmatrix} \begin{pmatrix} \sqrt{X(\alpha)} & \sqrt{X(\beta)} \end{pmatrix} = \begin{pmatrix} X(\alpha) & X\left(\frac{\alpha+\beta}{2}\right) \\ X\left(\frac{\alpha+\beta}{2}\right) & X(\beta) \end{pmatrix} \tag{5.156}$$

is positive semidefinite, and likewise

$$\begin{pmatrix} \sqrt{Y(\alpha)} \\ \sqrt{Y(\beta)} \end{pmatrix} \begin{pmatrix} \sqrt{Y(\alpha)} & \sqrt{Y(\beta)} \end{pmatrix} = \begin{pmatrix} Y(\alpha) & Y\left(\frac{\alpha+\beta}{2}\right) \\ Y\left(\frac{\alpha+\beta}{2}\right) & Y(\beta) \end{pmatrix} \tag{5.157}$$

is positive semidefinite. The sum of these two matrices is therefore positive semidefinite, and given the inequalities (5.155) it therefore follows that

$$\begin{pmatrix} Z(\alpha) & X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) \\ X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) & Z(\beta) \end{pmatrix} \tag{5.158}$$

is positive semidefinite. Invoking Lemma 5.31, one finds that

$$X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) \leq \sqrt{Z(\alpha)}\sqrt{Z(\beta)} = Z\left(\frac{\alpha+\beta}{2}\right). \tag{5.159}$$

It trivially holds that $Z(0) \geq X(0) + Y(0)$ and $Z(1) \geq X(1) + Y(1)$. For any choice of $\alpha, \beta \in [0,1]$, one has that the inequalities (5.155) together imply that

$$Z\left(\frac{\alpha+\beta}{2}\right) \geq X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right). \tag{5.160}$$

The inequality (5.153) must therefore hold for every $\alpha \in [0,1]$ taking the form $\alpha = k/2^n$ for nonnegative integers k and n with $k \leq 2^n$. The set of all such α is dense in $[0,1]$, so the theorem is proved. □

Corollary 5.34. *Let $P_0, P_1, Q_0, Q_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, for $\mathcal{X}$ being any complex Euclidean space. It holds that*

$$\left\langle (P_0 + P_1)^\alpha, (Q_0 + Q_1)^{1-\alpha} \right\rangle \geq \left\langle P_0^\alpha, Q_0^{1-\alpha} \right\rangle + \left\langle P_1^\alpha, Q_1^{1-\alpha} \right\rangle \tag{5.161}$$

for every $\alpha \in [0,1]$.

Proof. By making the substituting $A_0 = P_0$, $A_1 = P_1$, $B_0 = Q_0^{\mathsf{T}}$, and $B_1 = Q_1^{\mathsf{T}}$ in Theorem 5.32, one finds that

$$(P_0 + P_1)^\alpha \otimes (Q_0^{\mathsf{T}} + Q_1^{\mathsf{T}})^{1-\alpha} \geq P_0^\alpha \otimes (Q_0^{\mathsf{T}})^{1-\alpha} + P_1^\alpha \otimes (Q_1^{\mathsf{T}})^{1-\alpha}, \tag{5.162}$$

and therefore

$$\begin{aligned} &\operatorname{vec}(\mathbb{1}_{\mathcal{X}})^* \left((P_0 + P_1)^\alpha \otimes (Q_0^{\mathsf{T}} + Q_1^{\mathsf{T}})^{1-\alpha} \right) \operatorname{vec}(\mathbb{1}_{\mathcal{X}}) \\ &\quad \geq \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^* \left(P_0^\alpha \otimes (Q_0^{\mathsf{T}})^{1-\alpha} + P_1^\alpha \otimes (Q_1^{\mathsf{T}})^{1-\alpha} \right) \operatorname{vec}(\mathbb{1}_{\mathcal{X}}). \end{aligned} \tag{5.163}$$

Simplifying the two sides of this inequality yields (5.161), as required. □

The joint convexity of the quantum relative entropy now follows from a combination of Corollary 5.34 with Proposition 5.30.

Theorem 5.35. *Let $P_0, P_1, Q_0, Q_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, for $\mathcal{X}$ being any complex Euclidean space. It holds that*

$$\mathrm{D}(P_0 + P_1 \| Q_0 + Q_1) \leq \mathrm{D}(P_0 \| Q_0) + \mathrm{D}(P_1 \| Q_1). \tag{5.164}$$

Proof. By Proposition 5.30 together with Corollary 5.34 it holds that

$$\begin{aligned}
&\mathrm{D}(P_0 + P_1 \| Q_0 + Q_1) \\
&\quad = \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P_0 + P_1) - \langle (P_0 + P_1)^{1-\varepsilon}, (Q_0 + Q_1)^{\varepsilon} \rangle}{\varepsilon} \\
&\quad \leq \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P_0 + P_1) - \langle P_0^{1-\varepsilon}, Q_0^{\varepsilon} \rangle - \langle P_1^{1-\varepsilon}, Q_1^{\varepsilon} \rangle}{\varepsilon} \\
&\quad = \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P_0) - \langle P_0^{1-\varepsilon}, Q_0^{\varepsilon} \rangle}{\varepsilon} + \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P_1) - \langle P_1^{1-\varepsilon}, Q_1^{\varepsilon} \rangle}{\varepsilon} \\
&\quad = \mathrm{D}(P_0 \| Q_0) + \mathrm{D}(P_1 \| Q_1),
\end{aligned} \tag{5.165}$$

which proves the theorem. □

Corollary 5.36 (Joint convexity of quantum relative entropy)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let P_0, P_1, Q_0, $Q_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, and let $\lambda \in [0, 1]$. It holds that*

$$\begin{aligned}
&\mathrm{D}(\lambda P_0 + (1-\lambda) P_1 \| \lambda Q_0 + (1-\lambda) Q_1) \\
&\qquad \leq \lambda \, \mathrm{D}(P_0 \| Q_0) + (1-\lambda) \, \mathrm{D}(P_1 \| Q_1).
\end{aligned} \tag{5.166}$$

Proof. Combining Theorem 5.35 with the identity (5.96) yields

$$\begin{aligned}
&\mathrm{D}(\lambda P_0 + (1-\lambda) P_1 \| \lambda Q_0 + (1-\lambda) Q_1) \\
&\qquad \leq \mathrm{D}(\lambda P_0 \| \lambda Q_0) + \mathrm{D}((1-\lambda) P_1 \| (1-\lambda) Q_1) \\
&\qquad = \lambda \, \mathrm{D}(P_0 \| Q_0) + (1-\lambda) \, \mathrm{D}(P_1 \| Q_1),
\end{aligned} \tag{5.167}$$

as required. □

Monotonicity of quantum relative entropy

As was suggested above, the fact that the quantum relative entropy function is jointly convex has several interesting implications. One such implication

is that the quantum relative entropy function is monotonically decreasing under the action of any channel. The next proposition establishes that this is so for mixed-unitary channels, and the theorem that follows establishes that the same is true for all channels.

Proposition 5.37. *Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{C}(\mathcal{X})$ be a mixed-unitary channel, and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{D}\big(\Phi(P)\big\|\Phi(Q)\big) \leq \mathrm{D}(P\|Q). \tag{5.168}$$

Proof. As Φ is a mixed-unitary channel, there must exist an alphabet Σ, a collection of unitary operators $\{U_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{X})$, and a probability vector $p \in \mathcal{P}(\Sigma)$, such that

$$\Phi(X) = \sum_{a\in\Sigma} p(a) U_a X U_a^* \tag{5.169}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Applying Corollary 5.36, along with Proposition 5.21, one has

$$\begin{aligned}
\mathrm{D}(\Phi(P)\|\Phi(Q)) &= \mathrm{D}\Bigg(\sum_{a\in\Sigma} p(a) U_a P U_a^* \,\Bigg\|\, \sum_{a\in\Sigma} p(a) U_a Q U_a^*\Bigg) \\
&\leq \sum_{a\in\Sigma} p(a)\, \mathrm{D}\big(U_a P U_a^* \big\| U_a Q U_a^*\big) \\
&= \sum_{a\in\Sigma} p(a)\, \mathrm{D}(P\|Q) \\
&= \mathrm{D}(P\|Q),
\end{aligned} \tag{5.170}$$

as required. □

Theorem 5.38 (Monotonicity of quantum relative entropy)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{D}(\Phi(P)\|\Phi(Q)) \leq \mathrm{D}(P\|Q). \tag{5.171}$$

Proof. By Corollary 2.27 there must exist a complex Euclidean space $\mathcal{Z}$ and a linear isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ for which

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(AXA^*\big) \tag{5.172}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Let $\Omega \in \mathrm{C}(\mathcal{Z})$ denote the completely depolarizing channel, defined by

$$\Omega(Z) = \mathrm{Tr}(Z)\omega \tag{5.173}$$

for all $Z \in \mathrm{L}(\mathcal{Z})$, where

$$\omega = \frac{\mathbb{1}_{\mathcal{Z}}}{\dim(\mathcal{Z})} \tag{5.174}$$

denotes the completely mixed state with respect to the space $\mathcal{Z}$. As was demonstrated in Section 4.1.1, the channel Ω is a mixed-unitary channel, from which it follows that $\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Omega$ is also a mixed-unitary channel. By Proposition 5.37, together with Proposition 5.21, it therefore holds that

$$\begin{aligned} \mathrm{D}\big((\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Omega)(APA^*) \big\| (\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Omega)(AQA^*)\big) \\ \leq \mathrm{D}(APA^* \| AQA^*) = \mathrm{D}(P\|Q). \end{aligned} \tag{5.175}$$

As

$$(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Omega)(AXA^*) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \otimes \omega = \Phi(X) \otimes \omega \tag{5.176}$$

for all $X \in \mathrm{L}(\mathcal{X})$, it follows by Proposition 5.23 that

$$\mathrm{D}(\Phi(P)\|\Phi(Q)) = \mathrm{D}\big(\Phi(P) \otimes \omega \,\big\|\, \Phi(Q) \otimes \omega\big) \leq \mathrm{D}(P\|Q), \tag{5.177}$$

which completes the proof. □

Strong subadditivity of von Neumann entropy

Another implication of the joint convexity of quantum relative entropy is the following theorem, stating that the von Neumann entropy possesses a property known as *strong subadditivity*.

Theorem 5.39 (Strong subadditivity of von Neumann entropy)**.** *Let* X, Y, *and* Z *be registers. For every state of the register* $(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$ *it holds that*

$$\mathrm{H}(\mathsf{X}, \mathsf{Y}, \mathsf{Z}) + \mathrm{H}(\mathsf{Z}) \leq \mathrm{H}(\mathsf{X}, \mathsf{Z}) + \mathrm{H}(\mathsf{Y}, \mathsf{Z}). \tag{5.178}$$

Proof. Let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ be chosen arbitrarily and let

$$\omega = \frac{\mathbb{1}_{\mathcal{X}}}{\dim(\mathcal{X})} \tag{5.179}$$

denote the completely mixed state with respect to the space $\mathcal{X}$. These two equalities hold:

$$\begin{aligned} &\mathrm{D}\big(\rho[\mathsf{X},\mathsf{Y},\mathsf{Z}] \,\big\|\, \omega \otimes \rho[\mathsf{Y},\mathsf{Z}]\big) \\ &\qquad = -\mathrm{H}\big(\rho[\mathsf{X},\mathsf{Y},\mathsf{Z}]\big) + \mathrm{H}\big(\rho[\mathsf{Y},\mathsf{Z}]\big) + \log(\dim(\mathcal{X})) \end{aligned} \tag{5.180}$$

and

$$\begin{aligned} &\mathrm{D}\big(\rho[\mathsf{X},\mathsf{Z}] \,\big\|\, \omega \otimes \rho[\mathsf{Z}]\big) \\ &\qquad = -\mathrm{H}\big(\rho[\mathsf{X},\mathsf{Z}]\big) + \mathrm{H}\big(\rho[\mathsf{Z}]\big) + \log(\dim(\mathcal{X})). \end{aligned} \tag{5.181}$$

Taking the channel $\Phi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}, \mathcal{X} \otimes \mathcal{Z})$ to be the partial trace over $\mathcal{Y}$ in Theorem 5.38, one finds that

$$\mathrm{D}\big(\rho[\mathsf{X},\mathsf{Z}] \,\big\|\, \omega \otimes \rho[\mathsf{Z}]\big) \leq \mathrm{D}\big(\rho[\mathsf{X},\mathsf{Y},\mathsf{Z}] \,\big\|\, \omega \otimes \rho[\mathsf{Y},\mathsf{Z}]\big), \tag{5.182}$$

and therefore

$$\mathrm{H}(\rho[\mathsf{X},\mathsf{Y},\mathsf{Z}]) + \mathrm{H}(\rho[\mathsf{Z}]) \leq \mathrm{H}(\rho[\mathsf{X},\mathsf{Z}]) + \mathrm{H}(\rho[\mathsf{Y},\mathsf{Z}]), \tag{5.183}$$

which proves the theorem. □

The corollary that follows gives an equivalent statement to the strong subadditivity of von Neumann entropy, stated in terms of the quantum mutual information.

Corollary 5.40. *Let* X, Y, *and* Z *be registers. For every state of the register* $(\mathsf{X},\mathsf{Y},\mathsf{Z})$ *it holds that*

$$\mathrm{I}(\mathsf{X}:\mathsf{Y}) \leq \mathrm{I}(\mathsf{X}:\mathsf{Y},\mathsf{Z}). \tag{5.184}$$

Proof. By Theorem 5.39 it holds that

$$\mathrm{H}(\mathsf{X},\mathsf{Y},\mathsf{Z}) + \mathrm{H}(\mathsf{Y}) \leq \mathrm{H}(\mathsf{X},\mathsf{Y}) + \mathrm{H}(\mathsf{Y},\mathsf{Z}), \tag{5.185}$$

which is equivalent to

$$\mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{Y},\mathsf{Z}) - \mathrm{H}(\mathsf{X},\mathsf{Y},\mathsf{Z}). \tag{5.186}$$

Adding $\mathrm{H}(\mathsf{X})$ to both sides gives

$$\mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y},\mathsf{Z}) - \mathrm{H}(\mathsf{X},\mathsf{Y},\mathsf{Z}). \tag{5.187}$$

This inequality is equivalent to (5.184), which completes the proof. □

The quantum Pinsker inequality

The final implication of the joint convexity of quantum relative entropy to be presented in this section is a quantum analogue of Theorem 5.17 that establishes a lower bound on the quantum relative entropy between two density operators in terms of their trace distance.

Theorem 5.41 (Quantum Pinsker inequality)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be density operators. It holds that*

$$\mathrm{D}(\rho_0 \| \rho_1) \geq \frac{1}{2\ln(2)} \left\| \rho_0 - \rho_1 \right\|_1^2. \tag{5.188}$$

Proof. Let $\Sigma = \{0, 1\}$, let $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators providing the Jordan–Hahn decomposition $\rho_0 - \rho_1 = P_0 - P_1$, and define a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ as

$$\mu(0) = \Pi_{\mathrm{im}(P_0)} \quad \text{and} \quad \mu(1) = \mathbb{1} - \Pi_{\mathrm{im}(P_0)}. \tag{5.189}$$

This measurement is optimal for discriminating between the states ρ_0 and ρ_1 given with equal probability, as discussed in Section 3.1.1. For probability vectors $p_0, p_1 \in \mathcal{P}(\Sigma)$ defined as

$$p_0(a) = \langle \mu(a), \rho_0 \rangle \quad \text{and} \quad p_1(a) = \langle \mu(a), \rho_1 \rangle \tag{5.190}$$

for each $a \in \Sigma$, one has

$$\left\| p_0 - p_1 \right\|_1 = \left\| \rho_0 - \rho_1 \right\|_1. \tag{5.191}$$

Now let $\Phi \in \mathrm{C}(\mathcal{X}, \mathbb{C}^\Sigma)$ be the quantum-to-classical channel associated with μ, which satisfies

$$\Phi(X) = \langle \mu(0), X \rangle E_{0,0} + \langle \mu(1), X \rangle E_{1,1} \tag{5.192}$$

for each $X \in \mathrm{L}(\mathcal{X})$. It holds that

$$\mathrm{D}(\rho_0 \| \rho_1) \geq \mathrm{D}(\Phi(\rho_0) \| \Phi(\rho_1)) \tag{5.193}$$

by Corollary 5.36. By Theorem 5.17 it holds that

$$\begin{aligned} \mathrm{D}(\Phi(\rho_0) \| \Phi(\rho_1)) &= \mathrm{D}(\mathrm{Diag}(p_0) \| \mathrm{Diag}(p_1)) = \mathrm{D}(p_0 \| p_1) \\ &\geq \frac{1}{2\ln(2)} \left\| p_0 - p_1 \right\|_1^2 = \frac{1}{2\ln(2)} \left\| \rho_0 - \rho_1 \right\|_1^2, \end{aligned} \tag{5.194}$$

which completes the proof. □

5.3 Source coding

This section discusses the notion of *source coding*, as it relates to quantum information, and to the von Neumann entropy function in particular. The term *source coding*, as it is interpreted here, refers to the process of encoding information produced by given source in such a way that it may later be decoded. One natural goal of such a process is to compress the information produced by the source, in order to reduce costs of storage or transmission. Three principal variants of source coding will be discussed.

The first is a purely classical variant in which information from a given classical source is encoded into a fixed-length binary string in such a way that the information produced by the source can be decoded with high probability. A theorem known as *Shannon's source coding theorem* establishes asymptotic bounds on compression rates that are achievable for this task, given a standard assumption on the source.

The second variant of source coding to be discussed is a quantum analogue to the first; a source produces quantum information that is to be encoded into a sequence of qubits and then decoded. A theorem due to Schumacher, representing a quantum analogue of Shannon's source coding theorem, establishes asymptotic bounds on the rates of compression that are achievable for this task.

The third variant of source coding to be considered is one in which a source produces classical information, which is encoded into the quantum state of a collection of registers, and then decoded through a measurement performed on these registers. Theorems due to Holevo and Nayak establish fundamental limitations on two specific formulations of this task.

5.3.1 Classical source coding

In the first variant of source coding to be considered in the present section, a classical source produces a sequence of symbols, chosen independently from a known probability distribution. This sequence is to be encoded into a binary string in such a way that it may later be decoded, revealing the original sequence produced by the source with high probability.

The main purpose of this discussion, as it pertains to this book, is to introduce basic concepts and techniques regarding classical source coding that will carry over to the analogous quantum variant of this task. With this

purpose in mind, the discussion is limited to *fixed-length* coding schemes. These are schemes in which the length of each encoding is determined only by the number of symbols produced by the source, and not by the symbols themselves. A typical goal when designing such a scheme is to minimize the length of the binary string encodings while allowing for a recovery of the original sequence with high probability.

Shannon's source coding theorem establishes a fundamental connection between the rates of compression that can be achieved by such schemes and the Shannon entropy of the probability vector describing the source. While Shannon's source coding theorem is often stated in terms of *variable-length* coding schemes, with which one aims for a perfect recovery of the symbols produced by the source while minimizing the expected length of the binary string encodings, the fixed-length variant presented below translates more directly to the quantum setting.

Coding schemes and the statement of Shannon's source coding theorem

Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let $\Gamma = \{0,1\}$ denote the binary alphabet. For any choice of a positive integer n and real numbers $\alpha > 0$ and $\delta \in (0,1)$, and for $m = \lfloor \alpha n \rfloor$, a pair of mappings

$$\begin{aligned} f &: \Sigma^n \to \Gamma^m \\ g &: \Gamma^m \to \Sigma^n \end{aligned} \tag{5.195}$$

is said to be an (n, α, δ)-*coding scheme* for p if and only if it holds that

$$\sum_{a_1 \cdots a_n \in G} p(a_1) \cdots p(a_n) > 1 - \delta, \tag{5.196}$$

for

$$G = \big\{ a_1 \cdots a_n \in \Sigma^n \,:\, g(f(a_1 \cdots a_n)) = a_1 \cdots a_n \big\}. \tag{5.197}$$

(Here, and throughout the remainder of this chapter, elements of sets of the form Σ^n are written as strings $a_1 \cdots a_n$ rather than n-tuples $(a_1, \ldots, a_n)$, and likewise for Cartesian products of other alphabets.)

The expression on the left-hand side of (5.196) represents the probability that a random choice of symbols $a_1, \ldots, a_n \in \Sigma$, with each symbol chosen independently according to the probability vector p, results in a sequence satisfying

$$g(f(a_1 \cdots a_n)) = a_1 \cdots a_n. \tag{5.198}$$

The following scenario describes an abstract setting in which such coding schemes may be considered.

Scenario 5.42. Alice has a device (the source) that sequentially generates symbols chosen at random from an alphabet Σ. Each randomly generated symbol is independently distributed according to a single probability vector p. Alice allows the device to produce a string of n symbols $a_1 \cdots a_n$, and aims to communicate this string to Bob using as few bits of communication as possible.

To do this, Alice and Bob will use a coding scheme taking the form (5.195), which is assumed to have been agreed upon before the random generation of the symbols $a_1 \cdots a_n$. Alice *encodes* $a_1 \cdots a_n$ into a string of $m = \lfloor \alpha n \rfloor$ bits by computing $f(a_1 \cdots a_n)$, and sends the resulting binary string $f(a_1 \cdots a_n)$ to Bob. Bob *decodes* the string by applying the function g, obtaining $g(f(a_1 \cdots a_n))$. The coding scheme is said to be *correct* in the event that (5.198) holds, which is equivalent to $a_1 \cdots a_n \in G$, for then Bob will have obtained the correct string $a_1 \cdots a_n$.

If it is the case that the pair (f, g) is an (n, α, δ)-coding scheme for p, then the number δ is an upper bound on the probability that the coding scheme fails to be correct, so that Bob does not recover the string Alice obtained from the source, while α represents the average number of bits (as the value of n increases) needed to encode each symbol.

For a given probability vector p, it is evident that an (n, α, δ)-coding scheme will exist for some choices of the parameters n, α, and δ, and not others. The range of values of α for which coding schemes exist is closely related to the Shannon entropy $\mathrm{H}(p)$, as the following theorem establishes.

Theorem 5.43 (Shannon's source coding theorem)**.** *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let $\alpha > 0$ and $\delta \in (0, 1)$ be real numbers. The following statements hold:*

1. *If $\alpha > \mathrm{H}(p)$, then there exists an (n, α, δ)-coding scheme for p for all but finitely many choices of $n \in \mathbb{N}$.*
2. *If $\alpha < \mathrm{H}(p)$, then there exists an (n, α, δ)-coding scheme for p for at most finitely many choices of $n \in \mathbb{N}$.*

A proof of this theorem is presented below, following a discussion of the notion of a *typical string*, which is central to the proof. The general notion of

typicality, which can be formalized in various specific ways, will also play a major role in Chapter 8, which is devoted to the topic of quantum channel capacities.

Typical strings

The notion of a typical string, for a given distribution of symbols, a string length, and an error parameter, is defined as follows.

Definition 5.44. Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\varepsilon > 0$ be a positive real number. A string $a_1 \cdots a_n \in \Sigma^n$ is said to be *ϵ-typical (with respect to p)* if

$$2^{-n(\mathrm{H}(p)+\varepsilon)} < p(a_1) \cdots p(a_n) < 2^{-n(\mathrm{H}(p)-\varepsilon)}. \tag{5.199}$$

The notation $T_{n,\varepsilon}(p)$ refers to the set of all strings $a_1 \cdots a_n \in \Sigma^n$ for which the inequalities (5.199) hold, and when the probability vector p can safely be taken as being implicit, one may write $T_{n,\varepsilon}$ rather than $T_{n,\varepsilon}(p)$.

A random selection of a string $a_1 \cdots a_n \in \Sigma^n$, with each symbol being independently distributed according to $p \in \mathcal{P}(\Sigma)$, is increasingly likely to be ε-typical as n grows, as the following proposition demonstrates.

Proposition 5.45. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let $\varepsilon > 0$. It holds that*

$$\lim_{n\to\infty} \sum_{a_1\cdots a_n \in T_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) = 1. \tag{5.200}$$

Proof. Let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables, defined as follows: one first chooses $a \in \Sigma$ randomly according to the probability vector p, and then sets the value of the random variable to be the real number $-\log(p(a))$ for whichever value of a was selected. It holds that the expected value of each Y_k is

$$\mathrm{E}(Y_k) = -\sum_{a\in\Sigma} p(a)\log(p(a)) = \mathrm{H}(p). \tag{5.201}$$

The conclusion of the proposition may now be written

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n} Y_k - \mathrm{H}(p)\right| \geq \varepsilon\right) = 0, \tag{5.202}$$

which is true by the weak law of large numbers (Theorem 1.15). □

The proposition that follows establishes an upper bound on the number of ε-typical strings of a given length.

Proposition 5.46. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let $\varepsilon > 0$ be a positive real number, and let n be a positive integer. It holds that*

$$\bigl|T_{n,\varepsilon}(p)\bigr| < 2^{n(\mathrm{H}(p)+\varepsilon)}. \tag{5.203}$$

Proof. By the definition of ε-typicality, one has

$$1 \geq \sum_{a_1\cdots a_n \in T_{n,\varepsilon}(p)} p(a_1)\cdots p(a_n) > 2^{-n(\mathrm{H}(p)+\varepsilon)}\,|T_{n,\varepsilon}(p)|, \tag{5.204}$$

and therefore $|T_{n,\varepsilon}(p)| < 2^{n(\mathrm{H}(p)+\varepsilon)}$. □

Proof of Shannon's source coding theorem

Shannon's source coding theorem (Theorem 5.43) can be proved through a conceptually simple argument—a suitable coding scheme may be obtained for sufficiently large values of n by assigning a unique binary string to each typical string, with every other string being encoded arbitrarily. Conversely, any coding scheme that fails to account for a large fraction of the typical strings can be shown to fail with high probability.

Proof of Theorem 5.43. Assume first that $\alpha > \mathrm{H}(p)$, and choose $\varepsilon > 0$ so that $\alpha > \mathrm{H}(p) + 2\varepsilon$. A coding scheme of the form

$$\begin{aligned} f_n &: \Sigma^n \to \Gamma^m \\ g_n &: \Gamma^m \to \Sigma^n, \end{aligned} \tag{5.205}$$

for $m = \lfloor \alpha n \rfloor$, will be defined for every $n \in \mathbb{N}$ satisfying $n > 1/\varepsilon$. Observe, for each $n > 1/\varepsilon$, that the assumption $\alpha > \mathrm{H}(p) + 2\varepsilon$ implies that

$$m = \lfloor \alpha n \rfloor > n(\mathrm{H}(p) + \varepsilon). \tag{5.206}$$

By Proposition 5.46 it holds that

$$|T_{n,\varepsilon}| < 2^{n(\mathrm{H}(p)+\varepsilon)} < 2^m, \tag{5.207}$$

and one may therefore define a function $f_n : \Sigma^n \to \Gamma^m$ that is injective when restricted to $T_{n,\varepsilon}$, together with a function $g_n : \Gamma^m \to \Sigma^n$ that is chosen so that $g_n(f_n(a_1\cdots a_n)) = a_1\cdots a_n$ for every $a_1\cdots a_n \in T_{n,\varepsilon}$. Thus, for

$$G_n = \{a_1\cdots a_n \in \Sigma^n \,:\, g_n(f_n(a_1\cdots a_n)) = a_1\cdots a_n\}, \tag{5.208}$$

it holds that $T_{n,\varepsilon} \subseteq G_n$, and therefore

$$\sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \geq \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n). \tag{5.209}$$

It follows by Proposition 5.45 that the quantity on the right-hand side of (5.209) is greater than $1 - \delta$ for sufficiently large values of n. Therefore, for sufficiently large values of n it holds that the coding scheme (f_n, g_n) is an (n, α, δ)-coding scheme, which proves the first statement of the theorem.

Now assume that $\alpha < \mathrm{H}(p)$, let a coding scheme of the form (5.205) be fixed for each n, and let $G_n \subseteq \Sigma^n$ be as defined in (5.208). It must hold that

$$|G_n| \leq 2^m = 2^{\lfloor \alpha n \rfloor} \tag{5.210}$$

for each n, as the coding scheme cannot be correct for two or more distinct strings that map to the same encoding. To complete the proof, it suffices to prove that

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) = 0. \tag{5.211}$$

Toward this goal, observe that for every $n \in \mathbb{N}$ and $\varepsilon > 0$ it holds that

$$G_n \subseteq (\Sigma^n \setminus T_{n,\varepsilon}) \cup (G_n \cap T_{n,\varepsilon}), \tag{5.212}$$

and therefore, by the union bound, one has

$$\begin{aligned} &\sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \\ &\quad \leq \left(1 - \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n)\right) + 2^{-n(\mathrm{H}(p) - \varepsilon)} \, |G_n|. \end{aligned} \tag{5.213}$$

Choosing $\varepsilon > 0$ so that $\alpha < \mathrm{H}(p) - \varepsilon$, one has

$$\lim_{n \to \infty} 2^{-n(\mathrm{H}(p) - \varepsilon)} |G_n| = 0. \tag{5.214}$$

As Proposition 5.45 implies that

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n) = 1, \tag{5.215}$$

it follows that (5.211) holds, which completes the proof. □

5.3.2 Quantum source coding

There is a natural way to formulate a quantum analogue of classical source coding, which is as follows. It is assumed that a source produces a sequence of registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$, for some choice of a positive integer n, with all of these registers sharing a common classical state set Σ. Moreover, for some choice of a density operator $\rho \in \mathrm{D}(\mathbb{C}^{\Sigma})$, it is assumed that the state of the compound register $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ produced by the source is given by

$$\rho^{\otimes n} = \rho \otimes \cdots \otimes \rho \quad (n \text{ times}). \tag{5.216}$$

That is, the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are independent, and each is in a state that is described by ρ. The quantum information stored in these registers is to be encoded and decoded in a similar way to the classical setting, through the use of quantum channels rather than deterministic encoding and decoding functions.

Quantum coding schemes

A *quantum coding scheme* consists of a pair of channels (Φ, Ψ); the channel Φ represents the encoding process and Ψ represents the decoding process. The encoding channel Φ transforms $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ into $(\mathsf{Y}_1, \ldots, \mathsf{Y}_m)$, for some choice of an integer m, where $\mathsf{Y}_1, \ldots, \mathsf{Y}_m$ are registers having classical sets equal to the binary alphabet $\Gamma = \{0, 1\}$. In other words, each register Y_k represents a qubit. The decoding channel Ψ transforms $(\mathsf{Y}_1, \ldots, \mathsf{Y}_m)$ back into $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$.

The desired property of such a scheme is for the composition $\Psi\Phi$ to act trivially, or nearly trivially, on the compound register $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$, under the assumption that the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are independent and each in the state ρ as suggested above. It must be stressed that it is not sufficient to require that the state of $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ be close to $\rho^{\otimes n}$ after the decoding channel is applied—this would be a trivial requirement failing to recognize that there might initially be correlations among $\mathsf{X}_1, \ldots, \mathsf{X}_n$ and one or more other registers that must be respected by coding process. Indeed, for any complex Euclidean space $\mathcal{Z}$ and a state $\sigma \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \otimes \mathcal{Z})$ satisfying

$$\sigma[\mathsf{X}_1, \ldots, \mathsf{X}_n] = \rho^{\otimes n}, \tag{5.217}$$

it is required of a good coding scheme that the state $\big(\Psi\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\sigma)$ is approximately equal to σ.

The particular notion of approximate equality that will be considered is based on the fidelity function. This is a convenient choice, as it allows for the utilization of the closed-form expression of the channel fidelity given by Proposition 3.34. One could alternatively use the trace distance in place of the fidelity function, but this would not change the asymptotic behavior of the sorts of quantum coding schemes considered in this section, as the Fuchs–van de Graaf inequalities (Theorem 3.36) directly imply.

In accordance with the discussion above, quantum coding schemes are to be defined more precisely as follows. Let Σ be an alphabet, let $\rho \in \mathrm{D}(\mathbb{C}^{\Sigma})$ be a density operator, and let n be a positive integer. Also let $\Gamma = \{0,1\}$ denote the binary alphabet, let $\alpha > 0$ and $\delta \in (0,1)$ be real numbers, let $m = \lfloor \alpha n \rfloor$, and let $\mathcal{X}_1 = \mathbb{C}^{\Sigma}$, …, $\mathcal{X}_n = \mathbb{C}^{\Sigma}$ and $\mathcal{Y}_1 = \mathbb{C}^{\Gamma}$, …, $\mathcal{Y}_m = \mathbb{C}^{\Gamma}$ be complex Euclidean spaces. A pair of channels

$$\begin{aligned} \Phi &\in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) \\ \Psi &\in \mathrm{C}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n) \end{aligned} \tag{5.218}$$

is an (n, α, δ)-*quantum coding scheme* for ρ if and only if it holds that

$$\mathrm{F}\big(\Psi\Phi, \rho^{\otimes n}\big) > 1 - \delta, \tag{5.219}$$

for $\mathrm{F}\big(\Psi\Phi, \rho^{\otimes n}\big)$ denoting the channel fidelity of $\Psi\Phi$ with respect to $\rho^{\otimes n}$ (q.v. Section 3.2.3).

Schumacher's quantum source coding theorem

The following theorem is a quantum analogue to Shannon's source coding theorem (Theorem 5.43), establishing conditions on α for which quantum coding schemes exist.

Theorem 5.47 (Schumacher)**.** *Let Σ be an alphabet, let $\rho \in \mathrm{D}(\mathbb{C}^{\Sigma})$ be a density operator, and let $\alpha > 0$ and $\delta \in (0,1)$ be real numbers. The following statements hold:*

1. *If $\alpha > \mathrm{H}(\rho)$, then there exists an (n, α, δ)-quantum coding scheme for ρ for all but finitely many choices of $n \in \mathbb{N}$.*
2. *If $\alpha < \mathrm{H}(\rho)$, then there exists an (n, α, δ)-quantum coding scheme for ρ for at most finitely many choices of $n \in \mathbb{N}$.*

Proof. By the spectral theorem, one may write

$$\rho = \sum_{a \in \Sigma} p(a)\, u_a u_a^*, \tag{5.220}$$

for some choice of a probability vector $p \in \mathcal{P}(\Sigma)$ and an orthonormal basis $\{u_a : a \in \Sigma\}$ of $\mathbb{C}^\Sigma$. The association of the eigenvectors and eigenvalues of ρ with the elements of Σ may be chosen arbitrarily, and is assumed to be fixed for the remainder of the proof. By the definition of the von Neumann entropy, it holds that $\mathrm{H}(\rho) = \mathrm{H}(p)$.

Assume first that $\alpha > \mathrm{H}(\rho)$, and choose $\varepsilon > 0$ to be sufficiently small so that $\alpha > \mathrm{H}(\rho) + 2\varepsilon$. Along similar lines to the proof of Theorem 5.43, a quantum coding scheme (Φ_n, Ψ_n) of the form

$$\begin{aligned} \Phi_n &\in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) \\ \Psi_n &\in \mathrm{C}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n) \end{aligned} \tag{5.221}$$

will be defined for every $n > 1/\varepsilon$, where $m = \lfloor \alpha n \rfloor$. It will then be shown that (Φ_n, Ψ_n) is an (n, α, δ)-quantum coding scheme for sufficiently large values of n.

For a given choice of $n > 1/\varepsilon$, the quantum coding scheme (Φ_n, Ψ_n) is defined as follows. First, consider the set of ε-typical strings

$$T_{n,\varepsilon} = T_{n,\varepsilon}(p) \subseteq \Sigma^n \tag{5.222}$$

associated with the probability vector p, and define a projection operator $\Pi_{n,\varepsilon} \in \mathrm{Proj}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ as follows:

$$\Pi_{n,\varepsilon} = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} u_{a_1} u_{a_1}^* \otimes \cdots \otimes u_{a_n} u_{a_n}^*. \tag{5.223}$$

The subspace upon which this operator projects is the *ε-typical subspace* of $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ with respect to ρ. Notice that

$$\left\langle \Pi_{n,\varepsilon}, \rho^{\otimes n} \right\rangle = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n). \tag{5.224}$$

Now, by Shannon's source coding theorem (or, to be more precise, the proof of that theorem given in the previous subsection), there exists a classical coding scheme (f_n, g_n) for p that satisfies

$$g_n(f_n(a_1 \cdots a_n)) = a_1 \cdots a_n \tag{5.225}$$

for every ε-typical string $a_1 \cdots a_n \in T_{n,\varepsilon}$. Define a linear operator of the form

$$A_n \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) \tag{5.226}$$

as follows:

$$A_n = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} e_{f_n(a_1 \cdots a_n)} (u_{a_1} \otimes \cdots \otimes u_{a_n})^*. \tag{5.227}$$

Finally, define channels Φ_n and Ψ_n of the form (5.221) as

$$\Phi(X) = A_n X A_n^* + \langle \mathbb{1} - A_n^* A_n, X \rangle \, \sigma \tag{5.228}$$

$$\Psi(Y) = A_n^* Y A_n + \langle \mathbb{1} - A_n A_n^*, Y \rangle \, \xi \tag{5.229}$$

for all $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ and $Y \in \mathrm{L}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m)$, for density operators $\sigma \in \mathrm{D}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m)$ and $\xi \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ chosen arbitrarily.

It remains to prove that (Φ_n, Ψ_n) is an (n, α, δ)-quantum coding scheme for sufficiently large values of n. From the expressions (5.228) and (5.229) it follows that there must exist a Kraus representation of the channel $\Psi_n \Phi_n$ having the form

$$(\Psi_n \Phi_n)(X) = (A_n^* A_n) X (A_n^* A_n)^* + \sum_{k=1}^{N} C_{n,k} X C_{n,k}^* \tag{5.230}$$

for some choice of an integer N and a collection of operators $C_{n,1}, \ldots, C_{n,N}$, which will have no effect on the analysis that follows. By Proposition 3.34, it therefore holds that

$$\mathrm{F}\big(\Psi_n \Phi_n, \rho^{\otimes n}\big) \geq \big\langle \rho^{\otimes n}, A_n^* A_n \big\rangle = \big\langle \rho^{\otimes n}, \Pi_{n,\varepsilon} \big\rangle. \tag{5.231}$$

As

$$\lim_{n \to \infty} \big\langle \Pi_{n,\varepsilon}, \rho^{\otimes n} \big\rangle = 1, \tag{5.232}$$

it follows that (Φ_n, Ψ_n) is an (n, α, δ)-quantum coding scheme for all sufficiently large n, which proves the first statement in the theorem.

Now assume that $\alpha < \mathrm{H}(\rho)$, and suppose that Φ_n and Ψ_n are arbitrary channels of the form (5.221) for each $n \in \mathbb{N}$. It will be proved that, for any choice of $\delta \in (0, 1)$, the pair (Φ_n, Ψ_n) fails to be an (n, α, δ) quantum coding scheme for all sufficiently large values of n.

Fix any choice of $n \in \mathbb{N}$, and let

$$\Phi_n(X) = \sum_{k=1}^{N} A_k X A_k^* \quad \text{and} \quad \Psi_n(Y) = \sum_{k=1}^{N} B_k Y B_k^* \tag{5.233}$$

be Kraus representations of Φ_n and Ψ_n, where

$$\begin{aligned} A_1, \ldots, A_N &\in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m), \\ B_1, \ldots, B_N &\in \mathrm{L}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n). \end{aligned} \tag{5.234}$$

(The assumption that both representations have the same number of Kraus operators is made only for notational convenience. This assumption causes no loss of generality; one may include the zero operator as a Kraus operator for either channel any desired number of times.) It follows that

$$(\Psi_n \Phi_n)(X) = \sum_{1 \leq j,k \leq N} (B_k A_j) X (B_k A_j)^* \tag{5.235}$$

is a Kraus representation of the composition $\Psi_n \Phi_n$. For the purposes of this analysis, the key aspect of this Kraus representation is that

$$\operatorname{rank}(B_k A_j) \leq \dim(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) = 2^m \tag{5.236}$$

for all choices of $j, k \in \{1, \ldots, N\}$. Indeed, for each $k \in \{1, \ldots, N\}$, one may choose a projection operator $\Pi_k \in \operatorname{Proj}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ with $\operatorname{rank}(\Pi_k) \leq 2^m$ such that $\Pi_k B_k = B_k$. Therefore,

$$\begin{aligned} \mathrm{F}\big(\Psi_n \Phi_n, \rho^{\otimes n}\big)^2 &= \sum_{1 \leq j,k \leq N} \big|\big\langle B_k A_j, \rho^{\otimes n} \big\rangle\big|^2 \\ &= \sum_{1 \leq j,k \leq N} \big|\big\langle \Pi_k B_k A_j, \rho^{\otimes n} \big\rangle\big|^2 \\ &= \sum_{1 \leq j,k \leq N} \Big|\Big\langle B_k A_j \sqrt{\rho^{\otimes n}}, \Pi_k \sqrt{\rho^{\otimes n}} \Big\rangle\Big|^2 \\ &\leq \sum_{1 \leq j,k \leq N} \operatorname{Tr}\Big(B_k A_j \rho^{\otimes n} A_j^* B_k^*\Big) \big\langle \Pi_k, \rho^{\otimes n} \big\rangle, \end{aligned} \tag{5.237}$$

where the inequality follows from the Cauchy–Schwarz inequality. As each Π_k has rank bounded by 2^m, it follows that

$$\big\langle \Pi_k, \rho^{\otimes n} \big\rangle \leq \sum_{i=1}^{2^m} \lambda_i(\rho^{\otimes n}) = \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \tag{5.238}$$

for some subset $G_n \subseteq \Sigma^n$ having size at most 2^m. As the channel $\Psi_n \Phi_n$ is trace-preserving, it holds that

$$\sum_{1 \leq j,k \leq N} \operatorname{Tr}\Big(B_k A_j \rho^{\otimes n} A_j^* B_k^*\Big) = 1, \tag{5.239}$$

and, moreover, one has that each term in this sum is nonnegative. The final expression of (5.237) is therefore equal to a convex combination of values, each of which is bounded as in (5.238), which implies that

$$\mathrm{F}\bigl(\Psi_n \Phi_n, \rho^{\otimes n}\bigr)^2 \leq \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \tag{5.240}$$

for some subset $G_n \subseteq \Sigma^n$ having size at most 2^m.

Finally, reasoning precisely as in the proof of Theorem 5.43, one has that the assumption $\alpha < \mathrm{H}(\rho) = \mathrm{H}(p)$ implies that

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) = 0 \tag{5.241}$$

for any choice of sets $G_n \subseteq \Sigma^n$ having size bounded by 2^m. This implies that, for any fixed choice of $\delta \in (0,1)$, the pair (Φ_n, Ψ_n) fails to be a (n, α, δ) quantum coding scheme for all but finitely many values of n. □

5.3.3 Encoding classical information into quantum states

The final type of source coding to be discussed in this section is one in which classical information is encoded into a quantum state, and then decoded by means of a measurement. The following scenario represents one abstraction of this task.

Scenario 5.48. Let X and Z be classical registers having classical state sets Σ and Γ, respectively, and let Y be a quantum register. Also let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let $\{\rho_a : a \in \Sigma\} \subset \mathrm{D}(\mathcal{Y})$ be a collection of density operators, and let $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be a measurement.

Alice obtains an element $a \in \Sigma$, stored in the register X, that has been randomly generated by a source according to the probability vector p. She prepares Y in the state ρ_a and sends Y to Bob. Bob measures Y with respect to the measurement μ, and stores the outcome of this measurement in the classical register Z. This measurement outcome represents information that Bob has obtained regarding the classical state of X.

It is natural to consider the situation in which $\Gamma = \Sigma$ in this scenario, and to imagine that Bob aims to recover the symbol stored in Alice's register X; this is essentially the state discrimination problem discussed in Section 3.1.2. In

the discussion that follows, however, it will not be taken as an assumption that this is necessarily Bob's strategy.

Assuming that Alice and Bob operate as described in Scenario 5.48, the pair (X, Z) will be left in the probabilistic state $q \in \mathcal{P}(\Sigma \times \Gamma)$ defined by

$$q(a, b) = p(a)\langle \mu(b), \rho_a \rangle \tag{5.242}$$

for every pair $(a, b) \in \Sigma \times \Gamma$. For an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined as

$$\eta(a) = p(a)\,\rho_a \tag{5.243}$$

for each $a \in \Sigma$, the probability vector q may equivalently be expressed as

$$q(a, b) = \langle \mu(b), \eta(a) \rangle \tag{5.244}$$

for each $(a, b) \in \Sigma \times \Gamma$.

One fundamental question regarding this scenario is the following: How much information can Bob's register Z contain about the state of Alice's register X? A theorem known as *Holevo's theorem* establishes an upper bound on this amount of information, as represented by the mutual information between Alice's register X and Bob's register Z. Holevo's theorem is phrased in terms of two functions of the ensemble η, the *accessible information* and the *Holevo information*, which are introduced below.

Accessible information

With Scenario 5.48 and the discussion above in mind, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble, let $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be a measurement, and let $q \in \mathcal{P}(\Sigma \times \Gamma)$ be the probability vector defined as in (5.244), representing a probabilistic state of the pair of classical registers (X, Z). The notation $\mathrm{I}_\mu(\eta)$ will denote the mutual information between X and Z, with respect to a probabilistic state defined in this way, so that

$$\mathrm{I}_\mu(\eta) = \mathrm{H}(q[\mathsf{X}]) + \mathrm{H}(q[\mathsf{Z}]) - \mathrm{H}(q) = \mathrm{D}\big(q \,\big\|\, q[\mathsf{X}] \otimes q[\mathsf{Z}]\big). \tag{5.245}$$

Now suppose that the ensemble η is fixed, while the measurement μ is unconstrained. The *accessible information* $\mathrm{I}_{\mathrm{acc}}(\eta)$ of the ensemble η is defined as the supremum value, ranging over all possible choices of a measurement μ, that may be obtained in this way. That is,

$$\mathrm{I}_{\mathrm{acc}}(\eta) = \sup_{\mu} \mathrm{I}_\mu(\eta), \tag{5.246}$$

where the supremum is over all choices of an alphabet Γ and a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$.

Although it is not necessarily apparent from its definition, the accessible information $\mathrm{I}_{\mathrm{acc}}(\eta)$ of an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ is indeed achieved by some choice of an alphabet Γ and a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$. The following lemma is useful for establishing this fact.

Lemma 5.49. *Let Σ and Γ be alphabets, let $\mathcal{Y}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble of states. Also let $\mu_0, \mu_1 : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be measurements and let $\lambda \in [0, 1]$ be a real number. It holds that*

$$\mathrm{I}_{\lambda\mu_0 + (1-\lambda)\mu_1}(\eta) \leq \lambda\, \mathrm{I}_{\mu_0}(\eta) + (1-\lambda)\, \mathrm{I}_{\mu_1}(\eta). \tag{5.247}$$

Proof. Let X and Z be classical registers having classical state sets Σ and Γ, respectively. Define a probability vector $p \in \mathcal{P}(\Sigma)$ as

$$p(a) = \mathrm{Tr}(\eta(a)) \tag{5.248}$$

for all $a \in \Sigma$. Also define probability vectors $q_0, q_1 \in \mathcal{P}(\Sigma \times \Gamma)$, representing probabilistic states of the pair (X, Z), as

$$q_0(a, b) = \langle \mu_0(b), \eta(a) \rangle \quad \text{and} \quad q_1(a, b) = \langle \mu_1(b), \eta(a) \rangle \tag{5.249}$$

for all $(a, b) \in \Sigma \times \Gamma$. By the joint convexity of the relative entropy function, it holds that

$$\begin{aligned}
&\mathrm{I}_{\lambda\mu_0 + (1-\lambda)\mu_1}(\eta) \\
&\quad = \mathrm{D}\big(\lambda q_0 + (1-\lambda) q_1 \,\big\|\, p \otimes (\lambda q_0[\mathsf{Z}] + (1-\lambda) q_1[\mathsf{Z}])\big) \\
&\quad \leq \lambda\, \mathrm{D}\big(q_0 \,\big\|\, p \otimes q_0[\mathsf{Z}]\big) + (1-\lambda)\, \mathrm{D}\big(q_1 \,\big\|\, p \otimes q_1[\mathsf{Z}]\big) \\
&\quad = \lambda\, \mathrm{I}_{\mu_0}(\eta) + (1-\lambda)\, \mathrm{I}_{\mu_1}(\eta)\,,
\end{aligned} \tag{5.250}$$

as required. □

Theorem 5.50. *Let Σ be an alphabet, let $\mathcal{Y}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble of states. There exists an alphabet Γ with $|\Gamma| \leq \dim(\mathcal{Y})^2$ and a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ such that $\mathrm{I}_{\mu}(\eta) = \mathrm{I}_{\mathrm{acc}}(\eta)$.*

Proof. Let $\nu : \Delta \to \mathrm{Pos}(\mathcal{Y})$ be a measurement, for an arbitrary choice of an alphabet Δ. By Lemma 5.49, the function $\mu \mapsto \mathrm{I}_{\mu}(\eta)$ is convex on the set of all measurements of the form $\mu : \Delta \to \mathrm{Pos}(\mathcal{Y})$. As every measurement of

this form can be written as a convex combination of extremal measurements of the same form, one has that there must exist an extremal measurement $\mu : \Delta \to \mathrm{Pos}(\mathcal{Y})$ satisfying $\mathrm{I}_\mu(\eta) \geq \mathrm{I}_\nu(\eta)$. By Corollary 2.48, the assumption that $\mu : \Delta \to \mathrm{Pos}(\mathcal{Y})$ is extremal implies that

$$\big|\big\{a \in \Delta \,:\, \mu(a) \neq 0\big\}\big| \leq \dim(\mathcal{Y})^2. \tag{5.251}$$

The value $\mathrm{I}_\mu(\eta)$ does not change if μ is restricted to the alphabet

$$\Gamma = \big\{a \in \Delta \,:\, \mu(a) \neq 0\big\}, \tag{5.252}$$

and therefore one has that there must exist a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$, for Γ satisfying $|\Gamma| \leq \dim(\mathcal{Y})^2$, such that $\mathrm{I}_\mu(\eta) \geq \mathrm{I}_\nu(\eta)$.

It follows that $\mathrm{I}_{\mathrm{acc}}(\eta)$ is equal to the supremum value of $\mathrm{I}_\mu(\eta)$, ranging over all measurements μ having at most $\dim(\mathcal{Y})^2$ measurement outcomes. The quantity $\mathrm{I}_\mu(\eta)$ is invariant under renaming the measurement outcomes of μ, so there is no loss of generality in restricting this supremum to the set of measurements having a single set Γ of measurement outcomes satisfying $|\Gamma| = \dim(\mathcal{Y})^2$. The supremum is therefore taken over a compact set, from which it follows that there exists a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ for which the supremum value is achieved, which completes the proof. □

The Holevo information

Again with Scenario 5.48 in mind, let X be a classical register, let Σ be the classical state set of X, let Y be a quantum register, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble. As described in Section 2.2.3, one associates the classical-quantum state

$$\sigma = \sum_{a \in \Sigma} E_{a,a} \otimes \eta(a) \tag{5.253}$$

of the pair (X, Y) with the ensemble η. The *Holevo information* (also called the *Holevo χ-quantity*) of the ensemble η, which is denoted $\chi(\eta)$, is defined as the quantum mutual information $\mathrm{I}(\mathsf{X} : \mathsf{Y})$ between the registers X and Y with respect to the state σ.

Under the assumption that the ensemble η is written as

$$\eta(a) = p(a)\,\rho_a \tag{5.254}$$

for each $a \in \Sigma$, for a probability vector $p \in \mathcal{P}(\Sigma)$ and a collection

$$\{\rho_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y}) \tag{5.255}$$

of states, the Holevo information of η may be calculated as follows:

$$\begin{aligned}
\chi(\eta) &= \mathrm{I}(\mathsf{X}:\mathsf{Y}) \\
&= \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}) \\
&= \mathrm{H}(p) + \mathrm{H}\biggl(\sum_{a\in\Sigma} p(a)\,\rho_a\biggr) - \mathrm{H}\biggl(\sum_{a\in\Sigma} p(a)\,E_{a,a}\otimes\rho_a\biggr) \\
&= \mathrm{H}\biggl(\sum_{a\in\Sigma} p(a)\,\rho_a\biggr) - \sum_{a\in\Sigma} p(a)\;\mathrm{H}\bigl(\rho_a\bigr),
\end{aligned} \tag{5.256}$$

where the last equality has made use of the identity (5.95). Alternatively, one may write

$$\chi(\eta) = \mathrm{H}\biggl(\sum_{a\in\Sigma}\eta(a)\biggr) - \sum_{\substack{a\in\Sigma \\ \eta(a)\neq 0}} \mathrm{Tr}(\eta(a))\;\mathrm{H}\biggl(\frac{\eta(a)}{\mathrm{Tr}(\eta(a))}\biggr), \tag{5.257}$$

or, equivalently,

$$\chi(\eta) = \mathrm{H}\biggl(\sum_{a\in\Sigma}\eta(a)\biggr) - \sum_{a\in\Sigma}\mathrm{H}(\eta(a)) + \mathrm{H}(p). \tag{5.258}$$

It follows from the concavity of the von Neumann entropy (Theorem 5.25), or by the subadditivity of von Neumann entropy (Proposition 5.10), that the Holevo information $\chi(\eta)$ is nonnegative for every ensemble η.

At an intuitive level, the Holevo information may be interpreted in the following way. When the pair of registers (X,Y) is in the classical-quantum state σ as described above, and the register Y is considered in isolation, its von Neumann entropy is given by

$$\mathrm{H}(\mathsf{Y}) = \mathrm{H}\biggl(\sum_{a\in\Sigma} p(a)\rho_a\biggr). \tag{5.259}$$

If one learns the classical state $a \in \Sigma$ of X, then from their perspective the von Neumann entropy of Y drops to $\mathrm{H}(\rho_a)$. The Holevo information $\chi(\eta)$ may therefore be viewed as representing the average decrease in the von Neumann entropy of Y that is expected when one learns the classical state of X.

It cannot be said that the Holevo information is convex in general, but the following proposition provides two conditions under which it is. The proof follows a similar argument to the proof of Lemma 5.49.

Proposition 5.51. *Let $\mathcal{Y}$ be a complex Euclidean space, let Σ be an alphabet, and let $\eta_0 : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ and $\eta_1 : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be ensembles of states. Suppose further that at least one of the following two conditions is satisfied:*

1. *The ensembles η_0 and η_1 have the same average state:*

$$\sum_{a \in \Sigma} \eta_0(a) = \rho = \sum_{a \in \Sigma} \eta_1(a), \tag{5.260}$$

   for some choice of $\rho \in \mathrm{D}(\mathcal{Y})$.

2. *The ensembles η_0 and η_1 yield equal probability distributions, over possibly different states:*

$$\mathrm{Tr}(\eta_0(a)) = p(a) = \mathrm{Tr}(\eta_1(a)) \tag{5.261}$$

   for each $a \in \Sigma$, for some choice of a probability vector $p \in \mathcal{P}(\Sigma)$.

For every real number $\lambda \in [0,1]$, it holds that

$$\chi(\lambda\eta_0 + (1-\lambda)\eta_1) \leq \lambda\chi(\eta_0) + (1-\lambda)\chi(\eta_1). \tag{5.262}$$

Proof. Let $\mathcal{X} = \mathbb{C}^{\Sigma}$, let X and Y be registers corresponding to the spaces $\mathcal{X}$ and $\mathcal{Y}$, and define classical-quantum states $\sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ as

$$\sigma_0 = \sum_{a \in \Sigma} E_{a,a} \otimes \eta_0(a) \quad \text{and} \quad \sigma_1 = \sum_{a \in \Sigma} E_{a,a} \otimes \eta_1(a). \tag{5.263}$$

For a given choice of $\lambda \in [0,1]$, define $\sigma = \lambda\sigma_0 + (1-\lambda)\sigma_1$. The Holevo information of the ensembles η_0, η_1, and $\lambda\eta_0 + (1-\lambda)\eta_1$ may be expressed as follows:

$$\begin{aligned} \chi(\eta_0) &= \mathrm{D}\big(\sigma_0 \big\| \sigma_0[\mathsf{X}] \otimes \sigma_0[\mathsf{Y}]\big), \\ \chi(\eta_1) &= \mathrm{D}\big(\sigma_1 \big\| \sigma_1[\mathsf{X}] \otimes \sigma_1[\mathsf{Y}]\big), \end{aligned} \tag{5.264}$$

and

$$\chi(\lambda\eta_0 + (1-\lambda)\eta_1) = \mathrm{D}\big(\sigma \big\| \sigma[\mathsf{X}] \otimes \sigma[\mathsf{Y}]\big). \tag{5.265}$$

Under the first condition in the statement of the proposition, it holds that $\sigma_0[\mathsf{Y}] = \sigma_1[\mathsf{Y}] = \sigma[\mathsf{Y}] = \rho$. In this case, the inequality (5.262) is equivalent to

$$\mathrm{D}\big(\sigma \big\| \sigma[\mathsf{X}] \otimes \rho\big) \leq \lambda\, \mathrm{D}\big(\sigma_0 \big\| \sigma_0[\mathsf{X}] \otimes \rho\big) + (1-\lambda)\, \mathrm{D}\big(\sigma_1 \big\| \sigma_1[\mathsf{X}] \otimes \rho\big), \tag{5.266}$$

which follows from the joint convexity of the quantum relative entropy function (Corollary 5.36). Under the second condition in the statement of the proposition, one has $\sigma_0[\mathsf{X}] = \sigma_1[\mathsf{X}] = \sigma[\mathsf{X}] = \mathrm{Diag}(p)$. Exchanging the roles of X and Y from the first condition, one has that the the proof follows by similar reasoning. □

Holevo's theorem

The next theorem, known as *Holevo's theorem*, establishes that the accessible information is upper-bounded by the Holevo information, for all ensembles of states.

Theorem 5.52 (Holevo's theorem)**.** *Let Σ be an alphabet, let $\mathcal{Y}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble of states. It holds that $I_{\mathrm{acc}}(\eta) \leq \chi(\eta)$.*

Proof. Let X be a classical register having classical state set Σ and let Y be a register whose associated complex Euclidean space is $\mathcal{Y}$. Define a state $\sigma \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ as

$$\sigma = \sum_{a \in \Sigma} E_{a,a} \otimes \eta(a), \tag{5.267}$$

and suppose that the pair (X, Y) is in the state σ. It holds that

$$\chi(\eta) = \mathrm{D}\big(\sigma \big\| \sigma[\mathsf{X}] \otimes \sigma[\mathsf{Y}]\big). \tag{5.268}$$

Next, let Γ be an alphabet, let Z be a classical register having classical state set Γ, and let $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be a measurement. Define a channel $\Phi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ as

$$\Phi(Y) = \sum_{b \in \Gamma} \langle \mu(b), Y \rangle E_{b,b} \tag{5.269}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$, which is the quantum-to-classical channel associated with the measurement μ, and consider the situation in which Y is transformed into Z by means of Φ. One has that

$$(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\sigma) = \sum_{a \in \Sigma} \sum_{b \in \Gamma} \big\langle \mu(b), \eta(a) \big\rangle E_{a,a} \otimes E_{b,b} = \mathrm{Diag}(q), \tag{5.270}$$

for $q \in \mathcal{P}(\Sigma \times \Gamma)$ being the probability vector defined as

$$q(a,b) = \big\langle \mu(b), \eta(a) \big\rangle \tag{5.271}$$

for all $a \in \Sigma$ and $b \in \Gamma$. It follows that

$$\begin{aligned} \mathrm{I}_{\mu}(\eta) &= \mathrm{D}(q \| q[\mathsf{X}] \otimes q[\mathsf{Z}]) \\ &= \mathrm{D}\big((\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\sigma) \big\| (\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\sigma[\mathsf{X}] \otimes \sigma[\mathsf{Y}])\big), \end{aligned} \tag{5.272}$$

and therefore $\mathrm{I}_{\mu}(\eta) \leq \chi(\eta)$, as the quantum relative entropy does not increase under the action of a channel (by Theorem 5.38). As this bound holds for all measurements μ, the theorem follows. □

For every collection of density operators $\{\rho_a : a \in \Sigma\} \subset \mathrm{D}(\mathcal{Y})$ and every probability vector $p \in \mathcal{P}(\Sigma)$, it holds that

$$\begin{aligned} \mathrm{H}\left(\sum_{a\in\Sigma} p(a)\rho_a\right) - \sum_{a\in\Sigma} p(a)\,\mathrm{H}(\rho_a) \\ \leq \mathrm{H}\left(\sum_{a\in\Sigma} p(a)\rho_a\right) \leq \log(\dim(\mathcal{Y})), \end{aligned} \tag{5.273}$$

and therefore the Holevo information of every ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ is upper-bounded by $\log(\dim(\mathcal{Y}))$. The following corollary of Theorem 5.52 is a consequence of this observation.

Corollary 5.53. *Let Σ be an alphabet, let $\mathcal{Y}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble of states. It holds that $\mathrm{I}_{\mathrm{acc}}(\eta) \leq \log(\dim(\mathcal{Y}))$.*

Although this is indeed a simple corollary to Theorem 5.52, it nevertheless establishes the following conceptually important fact: if two individuals share no prior correlations or shared resources, and one individual sends the other a quantum register of a given dimension n, then no more than $\log(n)$ bits of classical information will have been transmitted through this process.

Quantum random access codes

An interesting variation of source coding involves the notion of a *quantum random access code*. This is a coding scheme in which a sequence of classical symbols is encoded into a quantum state in such a way that one may obtain information about just one of the encoded symbols, chosen arbitrarily by the individual performing the decoding operation. The following scenario provides an abstraction of this type of scheme.

Scenario 5.54. Let Σ and Γ be alphabets, let n be a positive integer, let $\mathsf{X}_1, \ldots, \mathsf{X}_n$ be classical registers, each having classical state set Σ, let Z be a classical register having classical state set Γ, and let Y be a quantum register. Also let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let

$$\left\{\rho_{a_1\cdots a_n} : a_1 \cdots a_n \in \Sigma^n\right\} \subset \mathrm{D}(\mathcal{Y}) \tag{5.274}$$

be a collection of states indexed by Σ^n, and let $\mu_1, \ldots, \mu_n : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be measurements.

Alice obtains the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$, which have been independently prepared by a source, with p being the probabilistic state of each of these registers. She observes the classical state $a_1 \cdots a_n \in \Sigma^n$ of $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$, and prepares the register Y in the state $\rho_{a_1 \cdots a_n}$, which is then sent to Bob. Bob selects an index $k \in \{1, \ldots, n\}$ of his choice, measures Y with respect to the measurement μ_k, and stores the outcome in the classical register Z. The classical state of Z represents the information Bob has obtained regarding the classical state of X_k.

The following example describes an instance of this scenario in which Alice encodes two classical bits into one qubit in such a way that Bob can recover the encoded bit of his choice with a reasonably high probability of success.

Example 5.55. Let $\Sigma = \{0, 1\}$ denote the binary alphabet. For every real number θ, define a density operator $\sigma(\theta) \in \mathrm{D}(\mathbb{C}^\Sigma)$ as

$$\sigma(\theta) = \begin{pmatrix} \cos^2(\theta) & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \sin^2(\theta) \end{pmatrix}, \tag{5.275}$$

and observe that each of these operators is a rank one projection.

Alice obtains two classical registers X_1 and X_2, both having classical state set Σ. It is to be assumed that the probabilistic states of these registers are independent and uniformly distributed. She encodes the classical state $(a_1, a_2) \in \Sigma \times \Sigma$ of the pair $(\mathsf{X}_1, \mathsf{X}_2)$ into the quantum state $\rho_{a_1 a_2} \in \mathrm{D}(\mathbb{C}^\Sigma)$ defined as

$$\begin{aligned} \rho_{00} &= \sigma(\pi/8), \quad & \rho_{10} &= \sigma(3\pi/8), \\ \rho_{01} &= \sigma(7\pi/8), \quad & \rho_{11} &= \sigma(5\pi/8). \end{aligned} \tag{5.276}$$

Bob receives the qubit $\rho_{a_1 a_2}$ from Alice, and decides whether he wishes to learn the classical state a_1 of X_1 or the classical state a_2 of X_2. If he wishes to learn a_1, he measures the qubit with respect to the measurement μ_1 defined as

$$\mu_1(0) = \sigma(0) \quad \text{and} \quad \mu_1(1) = \sigma(\pi/2). \tag{5.277}$$

If he wishes to learn a_2, he measures the qubit with respect to the measurement μ_2 defined as

$$\mu_2(0) = \sigma(\pi/4) \quad \text{and} \quad \mu_2(1) = \sigma(3\pi/4). \tag{5.278}$$

Using the formula

$$\langle \sigma(\phi), \sigma(\theta) \rangle = \cos^2(\phi - \theta), \tag{5.279}$$

one concludes from a case analysis that, if Bob measures $\rho_{a_1 a_2}$ with respect to the measurement μ_k, he will obtain the measurement outcome a_k with probability

$$\cos^2(\pi/8) \approx 0.85 \tag{5.280}$$

in all cases.

With Scenario 5.54 in mind, one may define a *quantum random access code* for a given choice of a positive integer n and a probability vector $p \in \mathcal{P}(\Sigma)$ as consisting of two objects: the first is the collection of density operators

$$\{\rho_{a_1 \cdots a_n} : a_1 \cdots a_n \in \Sigma^n\} \subseteq \mathrm{D}(\mathcal{Y}) \tag{5.281}$$

representing the encodings of the possible sequences $a_1 \cdots a_n \in \Sigma^n$, and the second is the sequence of measurements

$$\mu_1, \ldots, \mu_n : \Gamma \to \mathrm{Pos}(\mathcal{Y}) \tag{5.282}$$

that reveal information concerning one of the initial registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$.

The amount of information revealed by such a quantum random access code may be represented by a vector $(\alpha_1, \ldots, \alpha_n)$, where α_k represents the mutual information between X_k and Z, conditioned on the measurement μ_k having been performed and the outcome of that measurement stored in Z. The vector $(\alpha_1, \ldots, \alpha_n)$ may be defined in more precise terms as follows. First, one defines an ensemble $\eta : \Sigma^n \to \mathrm{Pos}(\mathcal{Y})$ as

$$\eta(a_1 \cdots a_n) = p(a_1) \cdots p(a_n)\, \rho_{a_1 \cdots a_n} \tag{5.283}$$

for each $a_1 \cdots a_n \in \Sigma^n$. Then, for each $k \in \{1, \ldots, n\}$, one defines

$$\alpha_k = \mathrm{I}(\mathsf{X}_k : \mathsf{Z}), \tag{5.284}$$

where the mutual information is defined with respect to the probabilistic state $q_k \in \mathcal{P}(\Sigma^n \times \Gamma)$ of the compound register $(\mathsf{X}_1, \ldots, \mathsf{X}_n, \mathsf{Z})$ given by

$$q_k(a_1 \cdots a_n, b) = \langle \mu_k(b), \eta(a_1 \cdots a_n) \rangle \tag{5.285}$$

for each $a_1 \cdots a_n \in \Sigma^n$ and $b \in \Gamma$.

Nayak's theorem

Although Example 5.55 suggests a potential for quantum random access codes to provide significant advantages over classical coding schemes, it is a false impression. The following theorem demonstrates that quantum random access codes are strongly limited in their capabilities.

Theorem 5.56 (Nayak's theorem)**.** *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let n be a positive integer. Also let $\mathcal{Y}$ be a complex Euclidean space, let Γ be an alphabet, and let*

$$\{\rho_{a_1\cdots a_n} : a_1\cdots a_n \in \Sigma^n\} \subset \mathrm{D}(\mathcal{Y}) \quad \text{and} \quad \mu_1,\ldots,\mu_n : \Gamma \to \mathrm{Pos}(\mathcal{Y}) \tag{5.286}$$

be a quantum random access code for p. If $(\alpha_1,\ldots,\alpha_n)$ is a vector representing the amount of information revealed by this code for the distribution p, in the manner defined above, then it must hold that

$$\sum_{k=1}^{n} \alpha_k \leq \chi(\eta) \tag{5.287}$$

for $\eta : \Sigma^n \to \mathrm{Pos}(\mathcal{Y})$ being the ensemble defined by

$$\eta(a_1\cdots a_n) = p(a_1)\cdots p(a_n)\rho_{a_1\cdots a_n} \tag{5.288}$$

for each $a_1\cdots a_n \in \Sigma^n$.

Proof. Let $\mathsf{X}_1,\ldots,\mathsf{X}_n$ be classical registers having classical state set Σ, and let Y be a register whose associated complex Euclidean space is $\mathcal{Y}$ (as in Scenario 5.54). Let

$$\sigma = \sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n) E_{a_1,a_1}\otimes\cdots\otimes E_{a_n,a_n}\otimes\rho_{a_1\cdots a_n} \tag{5.289}$$

be the classical-quantum state of the compound register $(\mathsf{X}_1,\ldots,\mathsf{X}_n,\mathsf{Y})$ corresponding to the ensemble η. With respect to the state σ, one has that

$$\mathrm{I}(\mathsf{X}_1,\ldots,\mathsf{X}_n:\mathsf{Y}) = \chi(\eta). \tag{5.290}$$

Now, it holds that

$$\begin{aligned}&\mathrm{I}(\mathsf{X}_1,\ldots,\mathsf{X}_n:\mathsf{Y})\\ &\quad = \mathrm{I}(\mathsf{X}_n:\mathsf{Y}) + \mathrm{I}(\mathsf{X}_1,\ldots,\mathsf{X}_{n-1}:\mathsf{X}_n,\mathsf{Y}) - \mathrm{I}(\mathsf{X}_1,\ldots,\mathsf{X}_{n-1}:\mathsf{X}_n).\end{aligned} \tag{5.291}$$

This identity (which is equivalent to an identity commonly known as the *chain rule* for quantum mutual information) holds independent of the state of these registers, and may be verified by expanding the definition of the quantum mutual information. In the particular case of the state σ, one has that

$$\mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_{n-1} : \mathsf{X}_n) = 0, \tag{5.292}$$

as the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are independent with respect to this state. Thus,

$$\begin{aligned} \mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_n : \mathsf{Y}) &= \mathrm{I}(\mathsf{X}_n : \mathsf{Y}) + \mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_{n-1} : \mathsf{X}_n, \mathsf{Y}) \\ &\geq \mathrm{I}(\mathsf{X}_n : \mathsf{Y}) + \mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_{n-1} : \mathsf{Y}), \end{aligned} \tag{5.293}$$

where the inequality holds by Corollary 5.53. By applying this inequality recursively, one finds that

$$\mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_n : \mathsf{Y}) \geq \sum_{k=1}^{n} \mathrm{I}(\mathsf{X}_k : \mathsf{Y}). \tag{5.294}$$

Finally, one may observe that $\alpha_k \leq \mathrm{I}(\mathsf{X}_k : \mathsf{Y})$ for each $k \in \{1, \ldots, n\}$, as a consequence of Holevo's theorem (Theorem 5.52). Thus,

$$\sum_{k=1}^{n} \alpha_k \leq \mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_n : \mathsf{Y}) = \chi(\eta), \tag{5.295}$$

as required. □

One interesting type of quantum random access code, which includes the code suggested by Example 5.55, is one in which Σ and Γ are equal to the binary alphabet, and one aims for the classical state of the register Z to agree with X_k for whichever index $k \in \{1, \ldots, n\}$ was measured. Theorem 5.56 implies a strong limitation on schemes of this sort. The following lemma, which is a special case of an inequality known as *Fano's inequality*, is useful for analyzing this special case.

Lemma 5.57. *Let* X *and* Y *be classical registers sharing the same classical state set* $\Sigma = \{0, 1\}$*, and assume the pair* (X, Y) *is in a probabilistic state* $q \in \mathcal{P}(\Sigma \times \Sigma)$ *for which* $q[\mathsf{X}](0) = q[\mathsf{X}](1) = 1/2$ *and*

$$q(0, 0) + q(1, 1) = \lambda \tag{5.296}$$

for $\lambda \in [0, 1]$*. (In words, the state of* X *is uniformly distributed and* Y *and* X *agree with probability* λ*.) It holds that* $\mathrm{I}(\mathsf{X} : \mathsf{Y}) \geq 1 - \mathrm{H}(\lambda, 1 - \lambda)$*.*

Proof. Define Z to be a classical register having classical state set Σ, and let $p \in \mathcal{P}(\Sigma \times \Sigma \times \Sigma)$ be the probability vector defined as

$$p(a,b,c) = \begin{cases} q(a,b) & \text{if } c = a \oplus b \\ 0 & \text{otherwise,} \end{cases} \tag{5.297}$$

where $a \oplus b$ denotes the exclusive-OR of the binary values a and b. In words, p describes the probabilistic state of $(\mathsf{X},\mathsf{Y},\mathsf{Z})$ for which (X,Y) is distributed according to q and Z is set to the exclusive-OR of X and Y. With respect to this state, one has

$$\mathrm{H}(\mathsf{Z}) = \mathrm{H}(\lambda, 1-\lambda). \tag{5.298}$$

Moreover, it holds that

$$\mathrm{H}(\mathsf{X}|\mathsf{Y}) = \mathrm{H}(\mathsf{Z}|\mathsf{Y}), \tag{5.299}$$

as the classical states of X and Z uniquely determine one another for each fixed classical state of Y. Finally, by the subadditivity of Shannon entropy (Proposition 5.10), one has that

$$\mathrm{H}(\mathsf{Z}|\mathsf{Y}) \leq \mathrm{H}(\mathsf{Z}). \tag{5.300}$$

Consequently,

$$\begin{aligned} \mathrm{I}(\mathsf{X}:\mathsf{Y}) &= \mathrm{H}(\mathsf{X}) - \mathrm{H}(\mathsf{X}|\mathsf{Y}) = 1 - \mathrm{H}(\mathsf{Z}|\mathsf{Y}) \\ &\geq 1 - \mathrm{H}(\mathsf{Z}) = 1 - \mathrm{H}(\lambda, 1-\lambda), \end{aligned} \tag{5.301}$$

as required. □

Corollary 5.58. *Let $\Sigma = \{0,1\}$ denote the binary alphabet, let n be a positive integer, let $\mathcal{Y}$ be a complex Euclidean space, and let $\lambda \in [1/2, 1]$ be a real number. Also let*

$$\{\rho_{a_1 \cdots a_n} \,:\, a_1 \cdots a_n \in \Sigma^n\} \subset \mathrm{D}(\mathcal{Y}) \tag{5.302}$$

be a collection of density operators, and let

$$\mu_1, \ldots, \mu_n : \Sigma \to \mathrm{Pos}(\mathcal{Y}) \tag{5.303}$$

be measurements. If it holds that

$$\langle \mu_k(a_k), \rho_{a_1 \cdots a_n} \rangle \geq \lambda \tag{5.304}$$

for every choice of $a_1 \ldots a_n \in \Sigma^n$, then

$$\log(\dim(\mathcal{Y})) \geq (1 - \mathrm{H}(\lambda, 1-\lambda))n. \tag{5.305}$$

Proof. Let $p \in \mathcal{P}(\Sigma)$ be the uniform distribution and define an ensemble $\eta : \Sigma^n \to \mathrm{Pos}(\mathcal{Y})$ as

$$\eta(a_1 \cdots a_n) = p(a_1) \cdots p(a_n) \rho_{a_1 \cdots a_n} = \frac{1}{2^n} \rho_{a_1 \cdots a_n} \tag{5.306}$$

for each string $a_1 \cdots a_n \in \Sigma^n$. Let $(\alpha_1, \ldots, \alpha_n)$ be the vector representing the amount of information revealed by the quantum random access code defined by the collection $\{\rho_{a_1 \cdots a_n} : a_1 \cdots a_n \in \Sigma^n\}$ and the measurements $\mu_1, \ldots, \mu_n$ for the distribution p. By combining Lemma 5.57 with the fact that $\mathrm{H}(\alpha, 1-\alpha)$ is a decreasing function of α on the interval $[1/2, 1]$, one finds that

$$\alpha_k \geq 1 - \mathrm{H}(\lambda, 1-\lambda) \tag{5.307}$$

for every $k \in \{1, \ldots, n\}$. Therefore, by Theorem 5.56, it holds that

$$\chi(\eta) \geq (1 - \mathrm{H}(\lambda, 1-\lambda))n. \tag{5.308}$$

As the Holevo information of η is upper-bounded by $\log(\dim(\mathcal{Y}))$, the proof is complete. □

Thus, for the special type of random access code under consideration, the number of qubits required to encode a binary string of length n is linear in n, with the constant of proportionality tending to 1 as the error tolerance decreases.

5.4 Exercises

5.1. Let X, Y and Z be registers. Prove that the following inequalities hold for all states $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ of these registers.

(a) $\mathrm{I}(\mathsf{X}, \mathsf{Y} : \mathsf{Z}) + \mathrm{I}(\mathsf{Y} : \mathsf{Z}) \geq \mathrm{I}(\mathsf{X} : \mathsf{Z})$.

(b) $\mathrm{H}(\mathsf{X}, \mathsf{Y}|\mathsf{Z}) + \mathrm{H}(\mathsf{Y}|\mathsf{Z}) \geq \mathrm{H}(\mathsf{X}|\mathsf{Z}) - 2\,\mathrm{H}(\mathsf{Z})$

5.2. Let X, Y, and Z be registers.

(a) Prove that, for every state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ of these registers, it holds that

$$\mathrm{I}(\mathsf{X}, \mathsf{Y} : \mathsf{Z}) \leq \mathrm{I}(\mathsf{Y} : \mathsf{X}, \mathsf{Z}) + 2\,\mathrm{H}(\mathsf{X}). \tag{5.309}$$

(b) Let Σ be the classical state set of X, let

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z}) \tag{5.310}$$

be a collection of density operators, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let

$$\rho = \sum_{a \in \Sigma} p(a)\, E_{a,a} \otimes \sigma_a \tag{5.311}$$

be a state of $(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$. Prove that, with respect to the state ρ, one has

$$\mathrm{I}(\mathsf{X}, \mathsf{Y} : \mathsf{Z}) \leq \mathrm{I}(\mathsf{Y} : \mathsf{X}, \mathsf{Z}) + \mathrm{H}(\mathsf{X}). \tag{5.312}$$

5.3. Let Σ be an alphabet, let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$ be a density operator, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let

$$\{\Phi_a \,:\, a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{X}, \mathcal{Y}) \tag{5.313}$$

be a collection of channels. Define an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$ as

$$\eta(a) = p(a)\big(\Phi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho) \tag{5.314}$$

for each $a \in \Sigma$. Prove that

$$\chi(\eta) \leq \mathrm{H}\left(\sum_{a \in \Sigma} p(a)\Phi_a\big(\mathrm{Tr}_{\mathcal{Z}}(\rho)\big)\right) + \sum_{a \in \Sigma} p(a)\,\mathrm{H}\big(\Phi_a\big(\mathrm{Tr}_{\mathcal{Z}}(\rho)\big)\big). \tag{5.315}$$

5.4. Let Σ be an alphabet and let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces. Also let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble, and define an ensemble $\Phi(\eta) : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ as

$$(\Phi(\eta))(a) = \Phi(\eta(a)) \tag{5.316}$$

for each $a \in \Sigma$. Prove that

$$\chi(\Phi(\eta)) \leq \chi(\eta). \tag{5.317}$$

5.5. Let X and Y be registers and let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be states of these registers. Prove that, for every choice of $\lambda \in [0, 1]$, it holds that

$$\begin{aligned} &\mathrm{H}\big(\lambda\rho_0 + (1-\lambda)\rho_1\big) - \mathrm{H}\big(\lambda\rho_0[\mathsf{Y}] + (1-\lambda)\rho_1[\mathsf{Y}]\big) \\ &\quad \geq \lambda\big(\mathrm{H}(\rho_0) - \mathrm{H}(\rho_0[\mathsf{Y}])\big) + (1-\lambda)\big(\mathrm{H}(\rho_1) - \mathrm{H}(\rho_1[\mathsf{Y}])\big). \end{aligned} \tag{5.318}$$

(Equivalently, prove that the conditional von Neumann entropy of X given Y is a concave function of the state of these registers.)

5.6. Let X and Y be registers and let $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$ be a state of these registers for which it holds that

$$\rho = \sum_{a \in \Sigma} p(a) \sigma_a \otimes \xi_a,$$

for some choice of an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and two collections of states $\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X})$ and $\{\xi_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y})$.

(a) Prove that, with respect to the state ρ, it holds that $\mathrm{I}(\mathsf{X} : \mathsf{Y}) \leq \mathrm{H}(p)$.

(b) Prove that

$$\mathrm{H}(\rho) \geq \sum_{a \in \Sigma} p(a) \, \mathrm{H}(\sigma_a) + \mathrm{H}\left(\sum_{a \in \Sigma} p(a) \xi_a \right). \tag{5.319}$$

5.5 Bibliographic remarks

The Shannon entropy function was defined in Shannon's 1948 paper [191], which is generally viewed as representing the birth of information theory. Several fundamental facts were proved in that paper, including Shannon's source coding theorem (of which Theorem 5.43 is a variant) and Shannon's channel coding theorem. Shannon also defined the conditional entropy in the same paper, considered the mutual information (although not under that name), and proved that the entropy function now bearing his name is the unique function of a probability vector, up to a normalization, satisfying a few simple axioms that a measure of information and uncertainty should naturally possess. Shannon observed the similarity in form of his entropy function to the notion of entropy in statistical mechanics in his 1948 paper, and was later quoted as saying that he used the name "entropy" on the advice of von Neumann [203]. More substantive connections between these different notions of entropy have been considered by several researchers. (See, for instance, Jaynes [176].)

The relative entropy function was defined by Kullback and Leibler in 1951 [139]. Theorem 5.16 is due to Audenaert [17]. A variant of Pinsker's inequality (Theorem 5.17, but with a smaller constant factor) was proved by Pinsker [173] and later refined by others, including Csiszár and Kullback. Further information on classical information theory can be found in books on the subject, including the books of Ash [15] and Cover and Thomas [53], among many others.

The von Neumann entropy was first defined by von Neumann in a 1927 paper [213] and then investigated in greater detail in his 1932 book [217], in both cases within the context of quantum statistical mechanics. Despite Shannon's reported discussion with von Neumann regarding the Shannon entropy, there is no evidence known to suggest that von Neumann ever considered the information-theoretic aspects of the von Neumann entropy function.

The quantum relative entropy was defined by Umegaki [211] in 1962. A fact from which Klein's inequality (as stated in Proposition 5.24) may be derived was proved many years earlier by Klein [131]. Theorem 5.27 was proved by Araki and Lieb [12], who also introduced the purification method through which it is proved in the same paper. A weaker version of the Fannes–Audenaert inequality (Theorem 5.28) was proved by Fannes [71], and was later strengthened by Audenaert [17] (through a reduction to the classical result stated in Theorem 5.16, which was proved in the same paper).

Lieb's concavity theorem was proved by Lieb [146] in 1973. The precise formulation of this theorem represented by Theorem 5.32 is due to Ando [10]. Multiple proofs of this theorem are known; the proof presented in this book is an adaptation of one appearing in the book of Simon [193] with simplifications inspired by Ando's methodology in [10]. Simon attributes the central idea of his proof to Uhlmann [210]. The strong subadditivity of von Neumann entropy was first conjectured by Lanford and Robinson [143] and proved by Lieb and Ruskai [147] using Lieb's concavity theorem. The joint convexity of quantum relative entropy was proved by Lindblad [148], also through the use of Lieb's concavity theorem. The quantum Pinsker inequality (Theorem 5.41) appears in a paper of Hiai, Ohya, and Tsukada [102], and may be obtained as a special case of a more general theorem due to Uhlmann [210].

Theorem 5.47 was proved by Schumacher [185] in 1995. Holevo [105] proved his eponymous theorem (Theorem 5.52) in 1973, through a different proof than the one presented in this chapter—Holevo's proof did not make use of the strong subadditivity of von Neumann entropy or Lieb's concavity theorem.

Quantum random access codes were proposed by Ambainis, Nayak, Ta-Shma, and Vazirani [8]; they proved a somewhat weaker limitation on quantum random access codes than what is established by Corollary 5.58, which

was proved by Nayak [162] a short time later. (The two previously referenced papers appeared in conference proceedings, and were consolidated as a journal paper [9].) Nayak's theorem, as stated in Theorem 5.56, follows from the proof of a closely related theorem that appears in Nayak's PhD thesis [161].