Quantum Subentropy: Subadditivity and Sharp Bounds on Discrimination Error

Pratik Patil

Department of Statistics and Data Sciences, University of Texas at Austin, Austin, USA; pratikpatil@utexas.edu

Abstract

Subentropy gives a tight spectral lower bound on the accessible information of ensembles of pure states. We prove its subadditivity for arbitrary bipartite states in finite dimensions and derive a sharp relation between conditional subentropy and the minimum discrimination error. This relation strengthens a conjectured comparison with conditional min-entropy. Its inverse gives the strongest universal lower bound on the error determined only by conditional subentropy and the number of labels. Both main results follow from a comparison of conditional subentropy with the average subentropy of measurement posteriors, valid for mixed signal states and every finite measurement. We characterize all equality cases in these two results and bound posterior deviations near equality. We also determine the optimal Lipschitz constant in trace distance for conditional subentropy on classical–quantum states, including the exact constant when the quantum marginal is fixed. A determinant inequality strengthens subadditivity and proves strong subadditivity for classical distributions, while an exact counterexample shows that strong subadditivity fails for quantum states.

Lean 4 proofs of all theoretical results are available at https://pratikpatil.io/lean/subentropy/.

Keywords: subentropy; quantum state discrimination; guessing probability; Fano inequalities; subadditivity; strong subadditivity; continuity bounds; quantitative stability

MSC: Primary 81P17; Secondary 81P18, 94A17

---

1. Introduction

1.1. Subentropy and quantum state discrimination

Suppose a source prepares a quantum signal ρ_x with prior probability p_x. In minimum-error discrimination, a receiver uses a measurement outcome to guess x, with the aim of minimizing the probability of error. We study bounds on this optimum that use only the spectra of the joint state of label and signal and of the average signal state. These bounds give guarantees when the relative eigenvectors of the signal states are not known. We ask how subentropy provides such a bound and how it behaves when quantum systems are combined.

Subentropy was introduced by Jozsa, Robb, and Wootters [1] to bound accessible information, the largest Shannon mutual information between the label and a measurement outcome. For ensembles of pure states with a prescribed average state, the subentropy of that state is the greatest universal lower bound on accessible information. The Scrooge ensemble attains this bound. Representations through random measurements and Gaussian vectors provide useful tools for studying subentropy [2–5].

This bound on accessible information does not directly give a lower bound on discrimination error. Subentropy also appears in bounds on guesswork, the minimum expected number of successive guesses [6]. The present question concerns the error probability of a single guess.

Datta, Dorlas, Jozsa, and Benatti [2] posed two conjectures connecting these structural and operational questions. Conjecture 1 asserts subadditivity for arbitrary bipartite states; they proved it for product states. Conjecture 2 bounds conditional subentropy by conditional min-entropy, the negative logarithm of the optimal guessing probability [7]. They proved this comparison for binary ensembles of pure signals. Zhang, Singh, and Pati [8] subsequently highlighted subadditivity and

strong subadditivity (SSA) as open questions. We prove subadditivity and strengthen the conditional min-entropy comparison to a sharp bound on the optimal discrimination error for each fixed number of labels. We characterize equality in both main results.

While this manuscript was under journal review, we became aware of two related developments. The Omega Institute's public Lean source repository [9] contains formalizations of subadditivity and the linear guessing bound $Q(X|B) \leq 1 - p_{\text{guess}}$. Zhang, Zhang, and Ma [10] establish subadditivity and its equality cases for the Nichols–Wootters family, including subentropy, and prove failure of quantum strong subadditivity. They also obtain a sharp two-qubit subadditivity deficit bound that strengthens the elementary estimate given here. Our results and proofs were developed independently of both works. The present paper additionally establishes the sharp discrimination bound and its equality cases, quantitative posterior stability, and optimal continuity bounds for conditional subentropy on classical–quantum states. Further results include pointwise determinant inequalities, classical strong subadditivity with equality cases, and bounds that account for a specified ancillary state.

1.2. Definitions and the main results

We work with finite-dimensional Hilbert spaces and finite label alphabets. Logarithms are natural, and entropy is measured in nats. We write $Q(p)$ for the subentropy of a probability vector and $Q(\rho)$ for the subentropy of the spectrum of a density operator. If its eigenvalues are $\lambda_1, \ldots, \lambda_d$, then

$$Q(\rho) = -\sum_{i=1}^{d} \frac{\lambda_i^d \log \lambda_i}{\prod_{j \neq i}(\lambda_i - \lambda_j)}. \tag{1}$$

The formula is interpreted by continuity when eigenvalues coincide or vanish. Appending zeros does not change Q, and pure states have subentropy zero.

For a bipartite state, $\rho_A = \operatorname{Tr}_B \rho_{AB}$ and $\rho_B = \operatorname{Tr}_A \rho_{AB}$ are its marginal states. We also write $Q(AB)$ for $Q(\rho_{AB})$, and similarly for other registers. Following [2], we define conditional subentropy by $Q(A|B) = Q(\rho_{AB}) - Q(\rho_B)$.

For matrices, positive means positive semidefinite; positive definiteness is stated explicitly. Matrix inequalities use the positive semidefinite order.

An ensemble of signal states with n labels is encoded by the classical–quantum (CQ) state $\rho_{XB} = \sum_x |x\rangle\langle x| \otimes A_x$, where $A_x = p_x \rho_x \geq 0$ and $\sum_x \operatorname{Tr} A_x = 1$. The classical register X records the label, and B holds the quantum side information. For this state, the conditional subentropy, average signal state, and optimal guessing probability are

$$Q(X|B) = Q(\rho_{XB}) - Q(\rho_B), \qquad \rho_B = \sum_x A_x, \qquad p_{\text{guess}} = \max_{\{E_x\}} \sum_x \operatorname{Tr}(E_x A_x). \tag{2}$$

The maximum is over positive operator-valued measures (POVMs) with outcomes labeled by x: $E_x \geq 0$ and $\sum_x E_x = I$. Each outcome is interpreted as the receiver's guess of the label.

The notation $Q(X|B)$ always denotes the spectral difference in (2). Measurement posteriors are the conditional label distributions given the observed outcomes. Their average subentropy is generally different from conditional subentropy. Nonnegativity of conditional subentropy for CQ states was established by Datta et al. [2, Proposition 5]. Proposition 6 recovers this fact and gives an upper bound in terms of the measurement posteriors.

Signals with orthogonal supports can be distinguished perfectly, and their conditional subentropy is zero. If all signals are the same pure state, no measurement reveals the label: the optimal rule is to guess a label with largest prior probability, and conditional subentropy is the subentropy of the prior. At a fixed largest prior probability, subentropy is maximized when all remaining probabilities are equal. This extremal prior determines the curve in the next theorem.

For $n \geq 2$, define

$$\Phi_n(u) = Q\Big(1 - u, \frac{u}{n-1}, \ldots, \frac{u}{n-1}\Big), \qquad 0 \leq u \leq 1. \tag{3}$$

The range $0 \le u \le 1 - 1/n$ contains all optimal discrimination errors. Defining the curve on the full interval simplifies the concavity argument.

Theorem 1 (Sharp guessing bound)**.** *For every classical–quantum ensemble with $n \ge 2$ labels, its optimal error $e = 1 - p_{\text{guess}}$ lies in $[0, 1 - 1/n]$ and satisfies*

$$0 \le Q(X|B) \le \Phi_n(e) \le e \le -\log p_{\text{guess}}. \tag{4}$$

The bound $Q(X|B) \le \Phi_n(e)$ is pointwise optimal at fixed n: equality is achievable at every admissible error.

For $e > 0$, equality holds exactly when, up to label permutation, all signals are the same pure state and $p = (1 - e, e/(n-1), \ldots, e/(n-1))$. Moreover, $Q(X|B) = 0$ if and only if $e = 0$, equivalently when the signals with positive prior probability have pairwise orthogonal supports.

For $e > 0$, $\Phi_n(e) < e$; the coefficient 1 in the linear bound $Q(X|B) \le e$ is nevertheless optimal uniformly over ensembles. For $n = 1$, both the error and conditional subentropy are zero.

Since $H_{\min}(X|B) = -\log p_{\text{guess}}$ [7], Theorem 1 proves the conditional min-entropy comparison proposed in [2]. The sharper bound can be inverted to give a lower bound on discrimination error. On $[0, 1 - 1/n]$, Φ_n is strictly increasing, so

$$e \ge \Phi_n^{-1}\big(Q(X|B)\big). \tag{5}$$

The right-hand side depends only on the joint and average spectra and requires no optimization over measurements. At the upper endpoint, $\Phi_n(1 - 1/n) = \log n - H_n + 1$, where $H_n = \sum_{k=1}^{n} k^{-1}$.

The equality cases in Theorem 1 show that (5) is the strongest universal lower bound on the error determined by $Q(X|B)$ and n alone. For a particular ensemble, the inequality may be strict.

When the full operators A_x are known, the optimization in (2) can be solved numerically as a semidefinite program [11, Section I]. Spectral data can also be obtained without reconstructing those operators: unitarily invariant collective measurements on repeated preparations can estimate the signal spectra and the spectrum of the average state without estimating eigenvectors [12]. The known priors and signal spectra determine the joint spectrum because ρ_{XB} is block diagonal. Thus (5) gives a guarantee from these data even when the relative eigenvectors required to specify the semidefinite program are unavailable. Section 6.1 gives two commuting ensembles with the same priors, individual signal spectra, and average state, but different optimal errors.

Our second main result establishes subadditivity and characterizes equality for arbitrary bipartite states.

Theorem 2 (Subadditivity)**.** *Every finite-dimensional bipartite state satisfies*

$$Q(\rho_{AB}) \le Q(\rho_A) + Q(\rho_B). \tag{6}$$

Equality holds exactly when at least one marginal is pure.

The equality cases of the guessing bound raise a stability question: for an optimal guessing measurement, does a small deficit in the sharp bound force the posterior distributions to be close to the maximizing vector? Proposition 12 shows that their average squared Euclidean distance is at most six times the deficit. Each posterior is compared with the vector in (3) at the optimal error, with the largest coordinate placed at the guessed label. The estimate concerns measurement posteriors. As Section 3.5 explains, a vanishing absolute deficit alone need not imply that the signal ensemble is close to an ensemble attaining equality at positive error.

To control the effect of state perturbations on the discrimination bound, we use the trace distance $D(\rho, \sigma) = \|\rho - \sigma\|_1/2$, where $\|\cdot\|_1$ denotes the trace norm. For states diagonal in a common basis, this is the total variation distance of their diagonal distributions. When there are at least two labels,

$Q(X|B)$ is Lipschitz continuous on CQ states with optimal constant one (Theorem 14). Corollary 15 identifies the exact constant when the marginal state of B is fixed. These estimates control the change in conditional subentropy before the inverse bound is applied.

1.3. Proof strategy and relation to earlier results

A measurement replaces quantum side information B by a classical outcome. The average subentropy of the resulting posteriors generally differs from conditional subentropy, which is a difference of two spectral quantities. Proposition 6 relates them: for every finite POVM with outcome probabilities q_y and posteriors b^y,

$$Q(X|B) \leq \sum_y q_y Q(b^y).$$

The guessing bound then follows from a classical maximization. At a fixed largest probability, subentropy is maximized when the remaining probabilities are equal. Applying this fact to each posterior and using concavity gives a bound in terms of the average guessing error. Choosing an optimal guessing measurement gives Theorem 1. The same comparison also proves subadditivity. Removing off-diagonal blocks in an eigenbasis of one marginal gives a CQ state with the same marginals. For a measurement with a single outcome, the posterior is the spectrum of the classical marginal.

To prove the comparison, we construct a Gram matrix with the same nonzero spectrum as the average signal state. Its block diagonal has the same nonzero spectrum as the joint state. The construction for mixed signals and the refinement of measurements follow Montanaro [11, Section III]. A related refinement of pretty good measurements for mixed states appears in Barnum and Knill [13, Section V]. The key property is convexity of the increase in subentropy under pinching, the operation that removes off-diagonal blocks. A Gaussian representation expresses this increase as a classical relative entropy. Joint convexity of relative entropy then bounds the increase for the Gram matrix by its average over the pure components induced by the refined measurement. Concavity combines the refined outcomes to give the displayed comparison.

The maximizing spectrum in (3) was identified for subentropy in [2, Proposition 4]. The bound on posterior averages is a specialization of Rioul's Fano framework [14, Theorems 4 and 5]; see also [15]. These classical results concern posterior averages; the measurement comparison relates them to conditional subentropy. A spectrum with all but one eigenvalue equal also arises in optimization at fixed purity [16]. Here the constraint is discrimination error.

Strong concavity makes the posterior argument quantitative and gives the stability estimate. The optimal modulus of continuity for subentropy itself, including its spectral equality cases, follows from the general theorem of Hanson and Datta [17]. Our proof uses maximal coupling to relate this known modulus to the guessing bound. For conditional subentropy, the shared CQ block structure yields the continuity bounds.

A separate proof of subadditivity uses the reciprocal determinant representation of subentropy. Merging rows and columns of a classical probability table gives a determinant inequality at each value of the integration parameter. Hadamard's determinant inequality transfers the resulting bipartite comparison to quantum states. The merging inequalities also yield classical strong subadditivity. An exact counterexample shows that strong subadditivity fails for general quantum states.

The theoretical results established in this paper have been formalized in Lean 4. Parts of the Lean formalization adapt code from the Omega Institute's development [9]. The companion proof index at https://pratikpatil.io/lean/subentropy/ links the statements and proofs and includes certificates for the numerical bounds in Tables 1 and A1.

The remainder of the paper is organized as follows. Section 2 gives the representations used in the proofs. Sections 3–5 develop the measurement comparison, stability and continuity estimates, and determinant inequalities. Section 6 compares spectral lower bounds on the error and studies independent ancillary states. The appendices give polynomial and asymptotic calculations, details of the continuity equality cases, and an example with three labels.

2. Representations of subentropy

The spectral definition is convenient for evaluation. Our proofs use two equivalent representations: a Gaussian expectation and an integral of reciprocal determinants. The Gaussian representation relates concavity of subentropy to scalar convexity and leads to the measurement comparison. The determinant representation reduces composition inequalities to determinant inequalities at each value of the integration parameter.

Throughout the paper, the norms $\|\cdot\|_1$, $\|\cdot\|_2 = \|\cdot\|_{\mathrm{HS}}$, and $\|\cdot\|_\infty$ are the trace, Hilbert–Schmidt, and operator norms, respectively. On vectors, the subscript two denotes the Euclidean norm.

Let $H_d = \sum_{k=1}^{d} k^{-1}$ and let γ be Euler's constant. The random measurement representation [2] gives

$$Q(\lambda) = -d\mathbb{E}[(\lambda \cdot X)\log(\lambda \cdot X)] - (H_d - 1), \tag{7}$$

where X is uniform on the simplex of probability vectors with the same number of entries as the spectrum.

Gaussian constructions for the Scrooge ensemble are discussed by Wootters [4]; McGinley and Schuster [5] state the Gaussian subentropy identity explicitly. The determinant formula is equivalent, under a reciprocal change of variable, to the integral representation of Jozsa and Mitchison [3]. We include a derivation that also covers singular states.

Lemma 3 (Gaussian and integral forms)**.** *Let g be a standard circular complex Gaussian vector, with independent coordinates satisfying $\mathbb{E}|g_i|^2 = 1$. Then*

$$Q(\rho) = 1 - \gamma - F(\rho), \qquad F(\rho) = \mathbb{E}[(g^*\rho g)\log(g^*\rho g)], \tag{8}$$

$$Q(\rho) = \int_0^\infty \Big[\frac{1}{1+t} - \frac{1}{\det(I+t\rho)}\Big]\frac{\mathrm{d}t}{t^2}. \tag{9}$$

Both formulas remain valid for singular states.

Proof. Unitary invariance of the Gaussian law allows us to work in an eigenbasis of the state. The squared moduli of the Gaussian coordinates are independent exponential variables with mean one. Normalizing by their sum gives a vector that is uniform on the probability simplex and independent of that sum. Thus we may write $|g_i|^2 = RX_i$, where the sum R has a Gamma$(d,1)$ distribution and is independent of X. Since $\mathbb{E}R = d$ and $\mathbb{E}[R\log R] = d(H_d - \gamma)$, comparison with (7) gives (8).

For the determinant formula, we express the logarithm as an integral of exponentials. Set $T = g^*\rho g$. Independence of the Gaussian coordinates in an eigenbasis of the state gives its Laplace transform, $L_\rho(t) = \mathbb{E}e^{-tT} = \det(I+t\rho)^{-1}$. For $y > 0$, the scalar identity

$$y\log y = \int_0^\infty (ye^{-t} - ye^{-ty})\frac{\mathrm{d}t}{t}$$

has an integrand whose absolute value integrates to $y|\log y|$. Since $0 \le T \le \|g\|^2$ and Gaussian moments are finite, the positive part of $T\log T$ is integrable. The negative part is bounded by $1/\exp(1)$. Thus the integral of the absolute value has finite expectation, and Fubini's theorem allows us to interchange expectation and integration. For the exponential distribution with mean one, the Laplace transform is $L_0(t) = (1+t)^{-1}$ and the corresponding value of $\mathbb{E}[T\log T]$ is $1-\gamma$. Subtracting the two expectation formulas cancels the terms involving their common first moment and gives $F(\rho) - (1-\gamma) = \int_0^\infty [L'_\rho(t) - L'_0(t)]\,\mathrm{d}t/t$. Integrating by parts on a truncated interval and then taking limits gives

$$F(\rho) - (1-\gamma) = \int_0^\infty [L_\rho(t) - L_0(t)]\frac{\mathrm{d}t}{t^2}.$$

The boundary terms vanish at both endpoints. The two distributions have the same mean and finite second moments, so the difference of their Laplace transforms is $O(t^2)$ at zero. At infinity, both

transforms are bounded, so their difference divided by t vanishes. This proves (9) for every density operator, including singular states.

Finally, the Gaussian formula recovers concavity of Q [2] and shows that it is strict. Indeed, $y \log y$ is strictly convex on $[0, \infty)$, and for distinct Hermitian matrices R, S, the quadratic form $g^*(R - S)g$ is nonzero on a set of positive Gaussian measure. Thus F is strictly convex on density operators, and subentropy is strictly concave. □

3. Measurement comparison, guessing, and subadditivity

A classical Fano argument bounds the subentropy of each posterior distribution in terms of its largest probability, then uses concavity to average over measurement outcomes. This bounds the average posterior subentropy. Conditional subentropy in (2) is a difference of spectral quantities and need not equal that average. We first bound conditional subentropy by the posterior average for every finite measurement. Choosing an optimal guessing measurement then gives the sharp error bound.

The joint and average states act on different spaces. We use a Gram matrix with the same nonzero spectrum as the average state. The block diagonal of this matrix has the spectrum of the joint state. Conditional subentropy is therefore the increase under block pinching. Convexity bounds this increase by its average over a decomposition into pure states. This property does not follow from concavity of subentropy alone; a Gaussian relative entropy identity establishes the required convexity.

We use Montanaro's construction for mixed signals [11, Section III], which factors the signals weighted by their priors. Refining a measurement then gives a pure decomposition of the Gram matrix, extending the Gram method for pure signals. Barnum and Knill describe related refinement for mixed signals and pretty good measurements [13, Section V, Equations (35)–(37)]. Gram constructions also connect coherence and discrimination [18].

Related coherence results concern averages over bases or properties of dephasing channels. The Haar average of the relative entropy of coherence can be expressed in terms of $S - Q$, where S is the von Neumann entropy [19]. Subentropies of stochastic columns also enter expressions for the coherence-generating power of dephasing channels [20]. The functional defined next measures the increase in subentropy of a fixed state under a specified pinching.

3.1. A convex functional for block pinching

Fix mutually orthogonal projections $P_1, \ldots, P_n$ on a finite-dimensional space, with $\sum_x P_x = I$. The associated pinching map removes the off-diagonal blocks. Define this map and the resulting increase in subentropy by

$$\Delta(R) = \sum_x P_x R P_x, \qquad C_Q(R) = Q(\Delta(R)) - Q(R). \tag{10}$$

Pinching is an average over unitaries that change the relative phases of the blocks. Together with the Gaussian representation, this averaging identifies the increase in (10) with classical relative entropy between two probability measures. Their densities are Gaussian quadratic forms and depend linearly on the state, so joint convexity of relative entropy gives convexity of the increase.

Lemma 4 (Convexity of the increase in subentropy under pinching)**.** *The functional C_Q is nonnegative and convex on density operators. For a normalized vector ψ, put $b_x = \|P_x\psi\|^2$. Then*

$$C_Q(|\psi\rangle\langle\psi|) = Q(b_1, \ldots, b_n). \tag{11}$$

Proof. In the Gaussian representation, the quadratic form multiplying the logarithm defines a probability density relative to Gaussian measure. Let Γ be standard complex Gaussian probability measure and define

$$a_R(g) = g^* R g, \qquad b_R(g) = g^* \Delta(R) g, \qquad \mathrm{d}\mu_R = a_R \, \mathrm{d}\Gamma, \quad \mathrm{d}\mu_{\Delta R} = b_R \, \mathrm{d}\Gamma.$$

Both measures have mass one because $\operatorname{Tr} R = \operatorname{Tr}\Delta(R) = 1$. To identify the cross term in relative entropy, average over the relative phases of the blocks. Put $\omega = \exp(2\pi i/n)$ and $U_k = \sum_x \omega^{kx} P_x$, $0 \le k < n$. Then $n^{-1}\sum_k U_k^* R U_k = \Delta(R)$, and $b_R(U_k g) = b_R(g)$. Gaussian invariance and this finite average yield

$$\int a_R \log b_R \, \mathrm{d}\Gamma = \frac{1}{n}\sum_k \int a_R(U_k g) \log b_R(g) \, \mathrm{d}\Gamma = \int b_R \log b_R \, \mathrm{d}\Gamma.$$

The identities remain valid for singular states. To justify the integrals, Cauchy–Schwarz gives

$$a_R(g) = \Big\| \sum_x R^{1/2} P_x g \Big\|^2 \le n \sum_x \|R^{1/2} P_x g\|^2 = n b_R(g).$$

Thus $b_R = 0$ implies $a_R = 0$, and $a_R|\log b_R| \le n b_R |\log b_R|$ off this zero set. The right-hand side is integrable by the Gaussian moment bounds used in Lemma 3. The positive and negative parts of $a_R \log(a_R/b_R)$ are bounded by $a_R \log n$ and $b_R/\exp(1)$, respectively. Integrands on the common zero set are defined to be zero.

With D denoting classical relative entropy, the phase average and (8) give

$$D(\mu_R \| \mu_{\Delta R}) = \int a_R \log(a_R/b_R) \, \mathrm{d}\Gamma = F(R) - F(\Delta(R)) = C_Q(R). \tag{12}$$

Relative entropy is nonnegative and jointly convex. Since both densities depend linearly on R, (12) proves nonnegativity and convexity of C_Q.

For a pure state, each nonzero block $P_x|\psi\rangle\langle\psi|P_x$ has the single eigenvalue b_x. The blocks have orthogonal supports, so the nonzero eigenvalues of the pinched state are these block probabilities. The original pure state has subentropy zero, and adding zero eigenvalues does not change Q. This proves (11). □

Gaussian density tilting is also used in the subentropy analysis of [5, Lemma 3]. The phase average above supplies the cross term for the specified pinching map.

Equality in the measurement comparison depends on equality in the convexity statement above. For pure decompositions, the relative entropy identity makes this condition explicit.

Lemma 5 (Equality for pure decompositions)**.** *Let* $R = \sum_{\ell=1}^N w_\ell |v_\ell\rangle\langle v_\ell|$, *where* $w_\ell > 0$, $\sum_\ell w_\ell = 1$, *and* $\|v_\ell\| = 1$. *Then*

$$C_Q(R) = \sum_\ell w_\ell C_Q(|v_\ell\rangle\langle v_\ell|)$$

if and only if either each v_ℓ *lies in a single pinching block, or all* v_ℓ *are collinear. In particular, equality with* $C_Q(R) > 0$ *forces* R *to have rank one.*

Proof. Write $a_v(g) = |v^* g|^2$ and $b_v(g) = \sum_x |(P_x v)^* g|^2$. For each nonzero v, $b_v > 0$ almost everywhere under the Gaussian measure. At such a g, the ratio for the mixture is a weighted average of the component ratios. The relative entropy inequality for mixtures is therefore Jensen's inequality for the strictly convex function $u \log u$, with weights $w_\ell b_{v_\ell}/b_R$. The Jensen gap is nonnegative, and the weights are strictly positive. Integrating the gap against the mixture density shows that equality holds exactly when all ratios a_{v_ℓ}/b_{v_ℓ} agree almost everywhere. The integrability established in Lemma 4 justifies this equality argument. Clearing denominators gives, for every pair v, w in the decomposition,

$$a_v(g) b_w(g) = a_w(g) b_v(g). \tag{13}$$

The identity initially holds almost everywhere. Both sides are polynomials in the real and imaginary parts of g, so continuity and full support of the Gaussian measure make it valid everywhere. In

particular, we may evaluate it on a hyperplane, even though that hyperplane has Gaussian measure zero.

Suppose one vector v has nonzero components in at least two blocks. The complex hyperplane $H = \{g : v^* g = 0\}$ has codimension one. The zero set $\ker b_v$ has codimension at least two, since the nonzero vectors $P_x v$ are orthogonal. Hence $b_v > 0$ on a nonempty relatively open subset of H. Equation (13) implies $w^* g = 0$ on that subset for every vector w in the decomposition. A linear form vanishing on a relatively open subset of a linear space vanishes throughout that space. Thus $H \subseteq \ker w^*$, which makes w collinear with v. If no such v exists, every vector in the decomposition lies in a single block.

Conversely, collinear vectors represent the same pure state, so the convexity inequality is an equality. If each vector lies in a single block, $R = \Delta(R)$ and all subentropy increases vanish. Thus both alternatives give equality. Equality with a positive increase forces the collinear case, and hence a state of rank one. □

3.2. Comparison with measurement posteriors

Each outcome of a measurement on the signal determines a posterior distribution of the label. Conditional subentropy is bounded above by the average subentropy of these posteriors, as the following proposition shows. The measurement outcomes need not be guesses of the signal label.

For a finite POVM $\{F_y\}$ on system B, write

$$q_y = \operatorname{Tr}(F_y \rho_B), \qquad b_x^y = \frac{\operatorname{Tr}(F_y A_x)}{q_y} \quad (q_y > 0), \qquad \rho_B = \sum_x A_x. \tag{14}$$

Thus b^y is the posterior distribution of X given outcome y.

Proposition 6 (Comparison with measurement posteriors)**.** *For every finite POVM,*

$$0 \le Q(X|B) \le \sum_{y:q_y>0} q_y Q(b^y). \tag{15}$$

If $Q(X|B) > 0$, upper equality holds if and only if all signals with positive prior probability are identical and pure; in that case it holds for every finite POVM. If $Q(X|B) = 0$, upper equality holds if and only if every posterior associated with an outcome of positive probability is a point mass.

Proof. The joint state and the average state act on different spaces. The Gram construction expresses their subentropy difference in terms of a single state with diagonal blocks A_x. On $\mathcal{H} = \bigoplus_x \mathcal{H}_B$, form

$$M = [\sqrt{A_1} \ \cdots \ \sqrt{A_n}] : \mathcal{H} \to \mathcal{H}_B, \qquad G = M^* M,$$

and use the projections onto these n summands. Then G is a density operator, $\operatorname{Tr} G = 1$, $MM^* = \rho_B$, and $\Delta(G) = \bigoplus_x A_x$. The nonzero spectra of M^*M and MM^* coincide, and Q is unchanged by adding zero eigenvalues. Consequently $Q(X|B) = C_Q(G)$, which is nonnegative by Lemma 4.

To use the formula for pure states, refine each measurement effect into rank-one terms, $F_y = \sum_k |f_{yk}\rangle\langle f_{yk}|$, using eigenvectors scaled by the square roots of the nonzero eigenvalues. Set $v_{yk} = M^* f_{yk}$ and $w_{yk} = \|v_{yk}\|^2$. Completeness of the POVM gives

$$G = \sum_{y,k} |v_{yk}\rangle\langle v_{yk}|, \qquad \sum_{y,k} w_{yk} = 1, \qquad \sum_k w_{yk} = q_y.$$

The weights are the probabilities of the refined outcomes. Discard terms with $w_{yk} = 0$. The normalized vector $v_{yk}/\sqrt{w_{yk}}$ has block probabilities

$$b_x^{yk} = \frac{\|P_x v_{yk}\|^2}{w_{yk}} = \frac{\langle f_{yk}, A_x f_{yk}\rangle}{w_{yk}}.$$

Thus the block probabilities are the posterior probabilities for the refined outcome. Convexity of the increase under pinching and (11) therefore yield

$$Q(X|B) \leq \sum_{y,k:w_{yk}>0} w_{yk} Q(b^{yk}). \tag{16}$$

For $q_y > 0$, the posterior for the original outcome is the mixture $b^y = \sum_{k:w_{yk}>0} (w_{yk}/q_y) b^{yk}$. Concavity of Q, established in Lemma 3, then gives

$$\sum_{k:w_{yk}>0} w_{yk} Q(b^{yk}) \leq q_y Q(b^y). \tag{17}$$

If $q_y = 0$, every w_{yk} is zero and that outcome contributes nothing. Combining (16) and (17), and summing over the original outcomes, proves the bound. No assumption about ranks or optimality is needed.

For equality with positive conditional subentropy, both the refinement and the coarsening inequalities must be equalities. Lemma 5 applied to the finite decomposition into pure states above implies that G has rank one. Its nonzero spectrum is that of ρ_B, so ρ_B is pure. Positivity of A_x and $\sum_x A_x = \rho_B$ imply $A_x = p_x \rho_B$: all signals with positive prior probability are that same pure state. Conversely, for identical pure signals every posterior equals p, and $Q(X|B) = Q(p)$, giving equality for every POVM.

When $Q(X|B) = 0$, equality means that the posterior average vanishes. Each term is nonnegative, so this happens exactly when every posterior associated with an outcome of positive probability has zero subentropy, that is, is a point mass. □

Zero conditional subentropy has a direct operational interpretation: it is equivalent to perfect discrimination.

Corollary 7 (Zero conditional subentropy)**.** *The following are equivalent:* $Q(X|B) = 0$; $p_{\text{guess}} = 1$; *and the supports of the signals with positive priors are pairwise orthogonal.*

Proof. The Gram construction identifies $Q(X|B)$ with $D(\mu_G \| \mu_{\Delta G})$. This vanishes if and only if $a_G = b_G$ almost everywhere. The quadratic forms are continuous and Gaussian measure has full support, so they agree everywhere. This implies equality of the operators: $G = \Delta(G)$. Its off-diagonal blocks are $\sqrt{A_x}\sqrt{A_z}$. These vanish exactly when the supports of the corresponding signals with positive priors are orthogonal. In that case their support projections distinguish the signals perfectly. Conversely, a perfectly distinguishing POVM has posteriors that are point masses; Proposition 6 then gives $Q(X|B) = 0$. Signals with zero priors impose no constraint. □

3.3. The classical extremal spectrum and the sharp bound

At a fixed largest probability, subentropy is maximized by equalizing the remaining probabilities. The spectrum in (3) follows by Schur concavity and was used in [2, Proposition 4 and Equation (12)]. The following lemma specializes Rioul's classical Fano bounds [14, Theorems 4 and 5] to subentropy. Rioul's conditional quantity is the posterior average; Proposition 6 connects it to spectral conditional subentropy.

Lemma 8 (Classical error curve)**.** *For a probability vector b with $n \geq 2$ entries, set $\varepsilon(b) = 1 - \max_x b_x$. Then $Q(b) \leq \Phi_n(\varepsilon(b))$, and Φ_n is concave on $[0,1]$. For posteriors b^y with probabilities q_y, their maximum a posteriori (MAP) error $\varepsilon_{\mathrm{MAP}} = \sum_y q_y \varepsilon(b^y)$ satisfies*

$$\sum_y q_y Q(b^y) \leq \Phi_n(\varepsilon_{\mathrm{MAP}}). \tag{18}$$

Zero-probability outcomes are omitted.

Proof. Choose a largest coordinate $b_{x_*} = 1 - \varepsilon(b)$ and average b over permutations of its other $n-1$ coordinates. The averaged vector is $(1-\varepsilon(b), \varepsilon(b)/(n-1), \ldots)$, up to permutation. Concavity and permutation invariance of Q give the first bound. The vector in (3) depends affinely on its argument, so its subentropy is concave on $[0,1]$. Applying Jensen's inequality to the posterior errors gives (18). □

Concavity bounds subentropy by its tangent at a point mass [2, Proposition 4]:

Lemma 9 (Comparison with the tangent at a pure state)**.** *For a probability vector b,*

$$Q(b) \leq 1 - \max_x b_x, \tag{19}$$

with strict inequality unless b is a point mass. For $n \geq 2$, $\Phi_n'(0) = 1$.

Proof. Relabel a largest coordinate as b_1 and write $h = b - (1, 0, \ldots, 0)$. Let $F(p) = \mathbb{E}[(p \cdot Z) \log(p \cdot Z)]$, where the independent exponential random variables Z_i have mean one. To justify boundary differentiation, put $S = \sum_i Z_i$. Along any affine segment in the simplex from the pure spectrum, $Z_1/2 \leq p \cdot Z \leq S$ sufficiently near that endpoint. The derivative integrand has an integrable majorant $CS(1 + S + |\log Z_1|)$, so dominated convergence applies. The derivatives at the pure spectrum are

$$\partial_1 F = \mathbb{E}[Z_1(1 + \log Z_1)] = 2 - \gamma, \qquad \partial_j F = 1 - \gamma \quad (j > 1).$$

Since $\sum_j h_j = 0$ and $Q = 1 - \gamma - F$, the directional derivative of Q is $-h_1 = 1 - b_1$. Concavity gives (19); strict concavity from Lemma 3 makes the bound strict unless b is a point mass. Along the affine curve defining Φ_n, its direction has first coordinate -1, giving $\Phi_n'(0) = 1$. □

The binary expansion shows that the linear bound is asymptotically attained:

$$Q(1-\varepsilon, \varepsilon) = \varepsilon + \varepsilon^2(\log \varepsilon + \tfrac{1}{2}) + O(\varepsilon^3 |\log \varepsilon|). \tag{20}$$

Proof of Theorem 1. Choose an optimal guessing POVM $\{E_x\}_{x=1}^n$; the optimum is attained because the set of such POVMs is compact in finite dimension. For every outcome x with positive probability, the reported label x must maximize its posterior probability. Otherwise some label z has $\mathrm{Tr}(E_x A_z) > \mathrm{Tr}(E_x A_x)$; replacing E_z by $E_z + E_x$ and E_x by zero gives a valid POVM with a larger success probability. Ties may be resolved arbitrarily. Thus the classical MAP rule has exactly the optimal quantum error:

$$\varepsilon_{\mathrm{MAP}} = 1 - \sum_x \mathrm{Tr}(E_x A_x) = 1 - p_{\mathrm{guess}} = e.$$

Applying the measurement comparison in Proposition 6 and the classical bound in Lemma 8 gives

$$0 \leq Q(X|B) \leq \sum_{x : q_x > 0} q_x Q(b^x) \leq \Phi_n(e).$$

Always guessing a label with largest prior probability shows $p_{\mathrm{guess}} \geq 1/n$, hence $e \leq 1 - 1/n$. On this interval the largest entry of the spectrum in (3) is $1 - e$. Equation (19) gives $\Phi_n(e) \leq e$, strictly for $e > 0$. The elementary inequality $1 - u \leq -\log u$ gives the final comparison in (4).

For $n = 1$, the joint and average states have the same nonzero spectrum and the label is always guessed correctly.

For exact attainment, take $\rho_x = |\varphi\rangle\langle\varphi|$ for every label, with priors

$$p_1 = 1 - e, \qquad p_x = \frac{e}{n-1} \quad (2 \le x \le n).$$

The average state is pure and the joint spectrum is the prior vector padded with zeros. Measurement outcomes are independent of the label, so the optimal success probability is $\max_x p_x = 1 - e$. Therefore $Q(X|B) = \Phi_n(e)$ for every admissible e, including both endpoints.

For necessity at $e > 0$, $\Phi_n(e) > 0$. Equality in the guessing bound forces equality in Proposition 6 for an optimal guessing POVM. Thus all signals with positive prior probability are identical and pure, so $e = 1 - \max_x p_x$ and $Q(X|B) = Q(p)$. Strict concavity in the classical permutation average forces every coordinate other than a selected largest one to equal $e/(n-1)$. Thus all n priors are positive and have the stated form. At $e = 0$, Corollary 7 gives the complete equality condition, including orthogonal mixed signals.

Finally, two identical pure signals with priors $(1 - \varepsilon, \varepsilon)$ have error ε and conditional subentropy $Q(1 - \varepsilon, \varepsilon)$. Equation (20) proves that the coefficient 1 in the linear relaxation cannot be reduced uniformly. □

Strict concavity of Q from Lemma 3 makes Φ_n strictly concave. Its unique maximum occurs at $u = 1 - 1/n$, where the spectrum is uniform; permutation averaging proves this maximality. Strict concavity therefore makes Φ_n strictly increasing on $[0, 1 - 1/n]$, which justifies the inverse in (5). Removing labels with zero prior probability allows the theorem to use the number of positive priors.

3.4. Subadditivity as a consequence of the comparison

The main proof of Theorem 2 uses Proposition 6. Pinching in a marginal eigenbasis preserves both marginals and permits the posterior comparison.

Proof of Theorem 2. Pinch A in any eigenbasis of ρ_A, and set $\sigma_{AB} = \Delta_A(\rho_{AB})$. As pinching averages unitary conjugations, concavity and unitary invariance of Q give $Q(\rho_{AB}) \le Q(\sigma_{AB})$. The pinched state is CQ, with B marginal ρ_B. For the POVM with one outcome $\{I_B\}$, the posterior is the spectrum of ρ_A. Proposition 6 therefore gives

$$Q(\rho_{AB}) \le Q(\sigma_{AB}) \le Q(\rho_B) + Q(\rho_A).$$

If ρ_A is mixed, $Q(\rho_A) > 0$. Equality in the second inequality would make the pinched state's conditional subentropy positive. The proposition then forces the signals on B with positive priors to be identical and pure, so ρ_B is pure. Subadditivity is therefore strict when both marginals are mixed. Conversely, a pure marginal forces a product state with the other marginal's nonzero spectrum, giving equality. Zero eigenvalues and degeneracy of the marginal eigenbasis do not affect the argument. □

Corollary 10 (Triangle inequality)**.** *Every bipartite state satisfies* $|Q(\rho_A) - Q(\rho_B)| \le Q(\rho_{AB})$*. Equality holds exactly when at least one of the states* ρ_A*,* ρ_B*, and* ρ_{AB} *is pure.*

Proof. Purify ρ_{AB} on ABC. Complementary reduced states have identical nonzero spectra, so

$$Q(BC) = Q(A), \qquad Q(C) = Q(AB).$$

Apply Theorem 2 to BC and use these identities; interchange A, B for the other direction.

Equality in $Q(A) - Q(B) \le Q(AB)$ holds, by Theorem 2, exactly when ρ_B or ρ_C is pure. Interchanging A, B gives the condition that ρ_A or ρ_C is pure for the other bound. Equality in the triangle inequality holds if and only if one of these two bounds is an equality. Since ρ_C and ρ_{AB} have the same nonzero spectrum, this proves the stated condition. □

3.5. Quantitative posterior stability

The sharp guessing curve is attained by a specific prior and identical pure signals. Near equality, quantitative concavity controls the average squared deviation of measurement posteriors from the maximizing distribution. The relevant Jensen gap is the subentropy of the average minus the average subentropy. The Gaussian Hessian bounds this gap below by a multiple of the variance, with a coefficient bounded away from zero in every dimension.

Lemma 11 (An estimate from strong concavity)**.** *For density operators ρ_j on a common d-dimensional space, $d \geq 2$, weights $w_j \geq 0$ summing to one, and $\bar{\rho} = \sum_j w_j \rho_j$,*

$$Q(\bar{\rho}) - \sum_j w_j Q(\rho_j) \geq \frac{\mu_d}{2} \sum_j w_j \|\rho_j - \bar{\rho}\|_2^2, \qquad \mu_d = \frac{d}{3d-2}. \tag{21}$$

Here $\|H\|_2^2 = \operatorname{Tr}(H^*H)$ is the squared Hilbert–Schmidt norm. The strong concavity parameter satisfies $\mu_d \geq 1/3$. The coefficient of the variance in (21) is $\mu_d/2$, which is at least $1/6$; its optimality is not asserted. In dimension one the variance vanishes.

Proof. At a positive definite state ρ, take a nonzero traceless Hermitian direction H and put $X = g^*Hg$, $T = g^*\rho g$, and $a = \operatorname{Tr} H^2$. To justify differentiation under the expectation, write $\lambda = \lambda_{\min}(\rho) > 0$ and $R = \|g\|^2$. For $|s| \leq \lambda/(2\|H\|_\infty)$, the state $\rho + sH$ has trace one and is bounded below by $(\lambda/2)I$. Thus $T_s = g^*(\rho + sH)g$ satisfies $(\lambda/2)R \leq T_s \leq R$. The first two derivatives of $T_s \log T_s$ obey, almost surely,

$$|X(1 + \log T_s)| \leq \|H\|_\infty R\big(1 + |\log R| + \log(2/\lambda)\big),$$
$$\frac{X^2}{T_s} \leq \frac{2\|H\|_\infty^2}{\lambda} R.$$

Both bounds are integrable under the Gaussian law, since R and $R|\log R|$ have finite expectations. Dominated convergence therefore permits both differentiations. Differentiation of (8) and weighted Cauchy–Schwarz give

$$D^2 F_\rho[H,H] = \mathbb{E}\frac{X^2}{T} \geq \frac{(\mathbb{E}X^2)^2}{\mathbb{E}[X^2 T]} = \frac{a^2}{a + 2\operatorname{Tr}(\rho H^2)}.$$

The last identity uses the complex Gaussian second and third moments. In the third moment, the six pairings in Wick's formula leave $\operatorname{Tr} H^2 \operatorname{Tr}\rho + 2\operatorname{Tr}(\rho H^2)$ because $\operatorname{Tr} H = 0$. Replacing ρ by I would remove the state dependence from the denominator. If h is an eigenvalue of H of largest absolute value, the remaining eigenvalues sum to $-h$. Their squared sum is at least $h^2/(d-1)$, so $\|H\|_\infty^2 \leq (d-1)a/d$. Consequently $\operatorname{Tr}(\rho H^2) \leq (d-1)a/d$ and the Hessian is at least $\mu_d a$. Thus $F(\rho) - \mu_d\|\rho\|_2^2/2$ is convex on positive definite states. Its Jensen inequality gives (21). Mixing every state with the same I/d and letting the mixing weight tend to zero extends the bound to boundary states by continuity. □

An earlier dimension-independent estimate follows from the information bound for uniform measurements of Matthews, Wehner, and Winter [21, Section 5]. Combining that bound with the identity for Q using random measurements gives a Jensen gap of at least $1/18$ times the Hilbert–Schmidt variance, in nats. Equation (21) gives a larger explicit coefficient by the short argument using third moments above.

The permutation average relates a small gap in the guessing bound to posterior deviations. The variance term controls both the spread of posterior errors and the imbalance among labels other than the MAP guess.

Proposition 12 (Quantitative posterior stability)**.** *Let a finite POVM on a CQ ensemble with $n \geq 2$ labels have outcome probabilities q_y and posteriors b^y. Omit outcomes with $q_y = 0$, choose any MAP label $j(y)$,*

and put $e_M = \sum_y q_y(1 - b^y_{j(y)})$. *Let* $v_j(e_M)$ *have entry* $1 - e_M$ *at* j *and* $e_M/(n-1)$ *elsewhere. Then, with* $\Delta_M = \Phi_n(e_M) - Q(X|B)$,

$$\sum_y q_y \|b^y - v_{j(y)}(e_M)\|_2^2 \le \Big(6 - \frac{4}{n}\Big)\Delta_M \le 6\Delta_M. \tag{22}$$

In particular, for every optimal guessing POVM this holds with $e_M = e$ *and* $\Delta_M = \Phi_n(e) - Q(X|B)$.

Proof. Move a chosen MAP coordinate of each b^y to the first position and average uniformly over permutations of the other coordinates. The mean of the resulting random vector is $v_1(e_M)$. Subentropy is invariant under these permutations, and their squared distances to $v_1(e_M)$ are $\|b^y - v_{j(y)}(e_M)\|_2^2$. Apply Lemma 11 to these diagonal states in dimension n to obtain

$$\Phi_n(e_M) - \sum_y q_y Q(b^y) \ge \frac{n}{2(3n-2)} \sum_y q_y \|b^y - v_{j(y)}(e_M)\|_2^2.$$

Proposition 6 bounds the left side above by Δ_M. The proof of Theorem 1 above shows that reporting each outcome of an optimal guessing POVM is a MAP decision, so its posterior error is e. □

For $\varepsilon_y = 1 - b^y_{j(y)}$, the quantity controlled here is

$$\sum_y q_y \|b^y - v_{j(y)}(e_M)\|_2^2 = \sum_y q_y \sum_{x \ne j(y)} \Big(b^y_x - \frac{\varepsilon_y}{n-1}\Big)^2 + \frac{n}{n-1}\,\mathrm{Var}_q(\varepsilon_y). \tag{23}$$

The deviations among labels other than the guess sum to zero. They are therefore orthogonal to the change caused by varying the total error, which proves the identity. Thus a small deficit bounds both the variation in posterior errors and the departure from uniformity among labels other than the guess. The estimate applies to a specified measurement without computing the optimal discrimination error.

The stability bound concerns measurement posteriors. Near zero error, a small absolute deficit need not make the CQ state close to an ensemble attaining equality at positive error. To see this, take identical qubit signals $I_2/2$ with priors $(1-e, e)$, $0 < e < 1/2$. Their error is e and $0 < \Phi_2(e) - Q(X|B) \le e \to 0$. Every ensemble attaining the guessing bound with positive error has a pure marginal state P on B. Contraction under partial trace therefore puts the constructed ensemble at distance at least $D(I_2/2, P) = 1/2$ from any such equality case. Identical pure signals with the same priors attain this distance. For fixed dimensions and errors in a closed interval away from zero, compactness and the equality classification imply closeness to an equality case as the deficit tends to zero. This argument gives neither an explicit rate nor a bound uniform in the dimension.

4. Continuity of subentropy and conditional subentropy

Continuity estimates control perturbations of the joint and average subentropies in the discrimination bound. For subentropy, coupling the spectra connects continuity to the guessing bound. The shared CQ block structure gives the optimal Lipschitz bound for conditional subentropy.

4.1. Continuity of subentropy

Maximal coupling represents two spectra as the marginals of a joint classical distribution whose mismatch probability is their total variation distance. The guessing theorem then bounds the subentropy difference, which explains the shared maximizing spectrum. For $d \ge 2$, define

$$\omega_d(\delta) = \Phi_d(\min\{\delta, 1 - 1/d\}), \qquad 0 \le \delta \le 1; \qquad \omega_1(\delta) = 0.$$

The following theorem gives the optimal modulus of continuity: the largest subentropy difference under a bound on trace distance. The modulus and its spectral equality cases are consequences of Hanson and Datta's concave-type entropy theorem [17, arXiv version 3, Corollary 6.2, Equations

(22)–(23)]. Earlier subentropy estimates were of Fannes and Audenaert type [2, Equations (8)–(9)]. We give a new proof using maximal coupling and Theorem 1. The equality argument also determines the alignment of the eigenvectors.

Theorem 13 (Optimal modulus of continuity; after Hanson and Datta [17])**.** *For states on a common d-dimensional space,*

$$\sup_{D(\rho,\sigma)\le\delta} |Q(\rho) - Q(\sigma)| = \omega_d(\delta). \tag{24}$$

For $\delta \le 1 - 1/d$ in dimension at least two, a pure state and a commuting state with spectrum $(1-\delta, \delta/(d-1), \ldots)$ attain the supremum at distance δ when the first eigenvector represents the pure state. For larger allowed distances, the pair consisting of a pure state and the maximally mixed state attains the plateau $\log d - H_d + 1$ at distance $1 - 1/d$.

For $d \ge 2$ and $0 < \delta < 1 - 1/d$, the pairs attaining (24) are, up to exchange,

$$\rho = |v\rangle\langle v|, \qquad \sigma = (1-\delta)|v\rangle\langle v| + \frac{\delta}{d-1}(I - |v\rangle\langle v|), \quad \|v\| = 1. \tag{25}$$

For $\delta \ge 1 - 1/d$, they are exactly pairs consisting of a pure state and the maximally mixed state. For $\delta = 0$, they are identical pairs; dimension one is trivial.

On every common finite-dimensional space,

$$|Q(\rho) - Q(\sigma)| \le D(\rho, \sigma). \tag{26}$$

The inequality is strict for distinct states, and its dimension-independent coefficient 1 is optimal.

New proof by maximal coupling. The case $d = 1$ is immediate. For $d \ge 2$, let p, q be the ordered eigenvalue vectors of ρ, σ. The eigenvalue perturbation inequality [22] gives $\tau = \|p - q\|_1/2 \le D(\rho, \sigma)$. We adapt the proof of Shannon continuity using maximal coupling [23, Theorem 3]. If $\tau > 0$, put $s_i = \min(p_i, q_i)$, $p'_i = p_i - s_i$, $q'_i = q_i - s_i$ and define

$$P_{ij} = s_i \mathbf{1}_{i=j} + p'_i q'_j / \tau.$$

This has marginals p, q. The residual supports are disjoint and each has mass τ, so $\Pr[X \ne Y] = \tau$. For $\tau = 0$ use the diagonal coupling. Assume $Q(p) \ge Q(q)$. Nonnegativity of conditional subentropy for this classical coupling gives $Q(P) - Q(p) \ge 0$, hence

$$Q(p) - Q(q) = [Q(P) - Q(q)] - [Q(P) - Q(p)] \le Q(X|Y).$$

Apply Theorem 1 to this classical coupling with d labels for X. Guessing $X = Y$ has error τ, so the optimal error e satisfies $e \le \min\{\tau, 1 - 1/d\}$. Monotonicity of Φ_d gives

$$|Q(\rho) - Q(\sigma)| \le \Phi_d(e) \le \omega_d(\tau) \le \omega_d(\delta).$$

Choose the eigenvector corresponding to the first entry in (3) to represent the pure state. The resulting pair has trace distance δ and subentropy difference $\Phi_d(\delta)$ up to the uniform endpoint. Beyond that endpoint, a pure state and the maximally mixed state have distance at most δ and difference $\Phi_d(1 - 1/d)$. Both pairs attain the bound.

Appendix D proves the equality cases, including the alignment of the eigenvectors and the plateau. The strict linear bound follows from $\Phi_d(u) < u$ for $0 < u \le 1 - 1/d$; on the plateau the modulus stays below the uniform endpoint distance. The binary perturbation in (20) proves optimality of the coefficient. □

The determinant representation makes the connection to Hanson and Datta's theorem explicit. For $t > 0$ the kernel in (9) has the form

$$q_t(\rho) = h_t(\operatorname{Tr} \varphi_t(\rho)), \qquad \varphi_t(x) = \log(1 + tx), \qquad h_t(y) = \frac{1}{1+t} - e^{-y}.$$

The function φ_t is strictly concave and vanishes at zero; h_t is increasing and concave, with $h_t(\varphi_t(1)) = 0$. Their theorem bounds the change in each kernel by its value at the spectrum in (3), with the same cutoff at the uniform spectrum. The maximizing spectra are the same for every $t > 0$. Integration with the positive weight $\mathrm{d}t/t^2$ gives the modulus above and its spectral equality cases.

4.2. Optimal Lipschitz constant for conditional subentropy

Separate continuity bounds for the joint and marginal subentropies give a Lipschitz constant of two. The shared CQ block structure reduces it to one: the common scalar in the block gradients cancels, and the remaining terms share an upper bound in the positive semidefinite order. The matrix giving this bound is the population complex spatial sign covariance matrix [24, Theorem 1]. For a nonzero positive semidefinite matrix M, define

$$K(M) = \int_0^\infty \frac{M(I + tM)^{-1}}{\det(I + tM)}\,\mathrm{d}t = \mathbb{E}\frac{M^{1/2} g g^* M^{1/2}}{g^* M g}, \qquad \kappa(M) = \|K(M)\|_\infty. \tag{27}$$

The expectation uses a standard complex Gaussian vector of the same dimension as M. The normalized outer product is a rank-one state, so $\operatorname{Tr} K(M) = 1$ and $0 \le K(M) \le I$. Its expectation has the same support as M and is positive definite there. Both K and κ are continuous on the cone of nonzero positive semidefinite matrices; see Appendix B.

Theorem 14 (Optimal Lipschitz bound for conditional subentropy on CQ states)**.** *Let ρ and σ be CQ states on the same label alphabet and quantum system B, and set $\rho_s = (1-s)\rho + s\sigma$. Then*

$$|Q(X|B)_\rho - Q(X|B)_\sigma| \le D(\rho,\sigma) \int_0^1 \kappa((\rho_s)_B)\,\mathrm{d}s \le D(\rho,\sigma). \tag{28}$$

The bound with coefficient one is strict for distinct states. For every $n \ge 2$ and every dimension $m \ge 1$ of system B, its coefficient one is optimal. For $n = 1$, the functional is identically zero.

Proof. For positive definite CQ blocks $A_x(s)$, write their sum as $A(s)$ and their direct sum as $R(s)$. The Gaussian representation gives $Q(X|B)_{\rho_s} = F(A(s)) - F(R(s))$. Appendix B proves that the derivative with respect to block A_x is $cI + B_x$, where c is the same for every block and

$$0 \le B_x = K(A) - K_x(R) \le K(A) \le \kappa(A) I, \quad K_x(R) = \int_0^\infty \frac{A_x(I + tA_x)^{-1}}{\det(I + tR)}\,\mathrm{d}t. \tag{29}$$

For $H = \sigma - \rho$, the common scalar cancels because $\operatorname{Tr} H = 0$. The directional derivative is therefore $\operatorname{Tr}[(\bigoplus_x B_x)H]$. The factor one half follows from the trace constraint: if $0 \le B \le \kappa I$ and H is trace zero, its positive and negative parts each have trace $\|H\|_1/2$, giving $|\operatorname{Tr} BH| \le \kappa \|H\|_1/2$. Integrating this estimate along the segment proves (28). For arbitrary CQ states, mix both endpoints with the same $I_{nm}/(nm)$. Their distance scales by one minus the mixing weight. Continuity of Q and K, with $\kappa \le 1$, allows this weight to tend to zero. Dominated convergence justifies the limit under the path integral.

Unless both endpoint marginals are the same pure state, every interior marginal on B is mixed. The corresponding K has trace one and at least two positive eigenvalues, so $\kappa < 1$ and the path integral is strictly below one. Otherwise all blocks are scalar multiples of that pure state, so conditional subentropy is Q of the priors. Theorem 13 gives strictness. For sharpness, compare ensembles of identical pure signals with priors $(1, 0)$ and $(1-\varepsilon, \varepsilon)$, adding zero labels if needed. Their distance is ε and their difference in conditional subentropy is $\Phi_2(\varepsilon)$, whose ratio to ε tends to one by (20). For a

single label, the joint and marginal states have the same nonzero spectrum, so conditional subentropy vanishes. □

Corollary 15 (Optimal coefficient with a fixed marginal state). *If $\rho_B = \sigma_B = \tau$, then*

$$|Q(X|B)_\rho - Q(X|B)_\sigma| \le \kappa(\tau) D(\rho, \sigma). \tag{30}$$

For $n \ge 2$, $\kappa(\tau)$ is the optimal coefficient among CQ states with marginal τ. In particular, it is $1/m$ for $\tau = I_m/m$, one for pure τ, and strictly smaller than one for mixed τ.

Proof. The marginal is constant along the segment, giving the bound. For sharpness, restrict to the support of τ and choose a rank-one projection P onto a common eigenvector of τ and $K(\tau)$ for the largest eigenvalue of $K(\tau)$. Starting from blocks $(\tau, 0, \dots)$, replace the first two by $(\tau - \varepsilon P, \varepsilon P)$. For sufficiently small $\varepsilon > 0$ they are positive and their trace distance is ε. The boundary derivative of their conditional subentropy at zero is $\kappa(\tau)$, as computed in Appendix B. This proves optimality. Unitary covariance and trace one give $K(I_m/m) = I_m/m$. The pure and mixed cases follow from support and positivity. □

Corollary 16 (Continuity with an independent maximally mixed ancilla). *Let $u_k = I_k/k$ be the maximally mixed state of an independent ancillary register C. For any two CQ states on the original common spaces,*

$$|Q(X|BC)_{\rho\otimes u_k} - Q(X|BC)_{\sigma\otimes u_k}| \le \frac{D(\rho,\sigma)}{k}. \tag{31}$$

Proof. By unitary covariance, $K(\tau \otimes u_k)$ is invariant under every unitary $I \otimes U$. It has the form $T \otimes I_k$ with $T \ge 0$ and $k \operatorname{Tr} T = 1$, hence $\kappa(\tau \otimes u_k) \le 1/k$. Tensoring with u_k preserves trace distance, so (28) on the augmented segment gives the bound. □

The sharp guessing curve for a maximally mixed ancilla has slope $1/k$ at zero error (Section 6.3). Equation (56) combines this curve with Corollary 16 to bound the discrimination error under perturbations.

The restriction to CQ states in Theorem 14 is essential. For two qubits, let $|\psi\rangle = (2|00\rangle + |11\rangle)/\sqrt{5}$, $\rho_0 = |\psi\rangle\langle\psi|$, and $\rho_t = (1-t)\rho_0 + t|10\rangle\langle 10|$. The pure components are orthogonal, so $D(\rho_0, \rho_t) = t$. The marginal spectra give

$$Q(A|B)_{\rho_t} - Q(A|B)_{\rho_0} = \Phi_2(t) + \Phi_2(1/5) - \Phi_2((1-t)/5). \tag{32}$$

Dividing by t and letting $t \downarrow 0$ gives $1 + \Phi_2'(1/5)/5 > 1$. At $t = 1/100$, rational bounds on the logarithms place the change between 0.01045371079 and 0.01045371080, strictly above the distance 0.01. Both state matrices have rational entries at this parameter value. The example does not determine the optimal coefficient for general quantum states.

5. Determinant inequalities and strong subadditivity

The reciprocal determinant representation admits a pointwise form of subadditivity at each value of the integration parameter. For classical tables, merging rows or columns cannot increase the deficit, which vanishes for a single row or column. The same comparison applies to the unnormalized slices needed for classical strong subadditivity. Hadamard's inequality transfers the bipartite comparison to quantum states in marginal eigenbases.

Proposition 17 (Pointwise determinant comparison). *For every bipartite state and every $t \ge 0$,*

$$\frac{1}{\det(I + t\rho_{AB})} + \frac{1}{1+t} \ge \frac{1}{\det(I + t\rho_A)} + \frac{1}{\det(I + t\rho_B)}. \tag{33}$$

For each $t > 0$, equality holds exactly when at least one marginal is pure. In particular, both mixed marginals give strict inequality at every positive parameter.

5.1. Coarse graining of classical distributions

For a nonnegative vector v, which need not sum to one, write $L_t(v) = \prod_i (1 + tv_i)^{-1}$.

Lemma 18 (Four cells). *For $a, b, c, d \geq 0$ with total mass M and every $t \geq 0$,*

$$L_t(a, b, c, d) + (1 + tM)^{-1} \geq L_t(a + b, c + d) + L_t(a + c, b + d).$$

For $t > 0$ and $M > 0$, the inequality is strict exactly when both binary marginal sum vectors have two positive entries.

Proof. Zero mass or $t = 0$ gives equality. Otherwise divide the entries by M and replace t by tM to obtain a table of total mass one. Clearing the positive denominators of the four terms leaves a polynomial with nonnegative coefficients; the full certificate is given in Appendix A. Its coefficient at degree two is $ad + bc$. This quantity is positive exactly when both marginal sum vectors have two positive entries. If it vanishes, the table is supported in one row or one column and the two sides agree by cancellation. □

For $X, Y \geq 0$, set

$$G(X, Y) = \frac{1}{1 + X + Y} - \frac{1}{(1 + X)(1 + Y)} \geq 0. \tag{34}$$

This difference is the increase in the reciprocal product when two masses are merged. The next lemma bounds the corresponding increase when these masses are distributed among several rows.

Lemma 19 (Binary splitting). *For nonnegative x_i, y_i, $X = \sum_i x_i$, $Y = \sum_i y_i$,*

$$\prod_i \frac{1}{1 + x_i + y_i} - \prod_i \frac{1}{(1 + x_i)(1 + y_i)} \leq \frac{1}{1 + X + Y} - \frac{1}{(1 + X)(1 + Y)}. \tag{35}$$

Proof. With one index, the two sides are equal. For the induction step, let A, B be the first and second products over the preceding indices and X, Y their partial sums. For the last index put $a = (1 + x + y)^{-1}$, $b = [(1 + x)(1 + y)]^{-1}$. Using the induction hypothesis and the bounds $a - b \geq 0$ and $B \leq D = [(1 + X)(1 + Y)]^{-1}$ gives

$$aA - bB = a(A - B) + (a - b)B \leq aG(X, Y) + (a - b)D = \frac{a}{1 + X + Y} - \frac{b}{(1 + X)(1 + Y)}.$$

Apply Lemma 18 to the unnormalized table $\left(\begin{smallmatrix} X & Y \\ x & y \end{smallmatrix}\right)$ at $t = 1$. The final expression is the reciprocal product of the row sums minus that of the cells. Rearranging this inequality gives the bound $G(X + x, Y + y)$. □

Proposition 20 (Classical merging and subadditivity). *Let $p = (p_{ij})$ be a finite nonnegative table of total mass M, with row sums $r = (r_i)$ and column sums $c = (c_j)$. For $t \geq 0$, define*

$$D_t(p) = (1 + tM)^{-1} + L_t(p) - L_t(r) - L_t(c).$$

If p' is obtained by merging rows or columns, then $D_t(p) \geq D_t(p') \geq 0$. In particular,

$$L_t(p) + (1 + tM)^{-1} \geq L_t(r) + L_t(c). \tag{36}$$

For $t > 0$, the deficit is strictly positive exactly when both marginal sum vectors have at least two positive entries.

Proof. Merge two columns x_i, y_i to obtain p', with column sum vector c'. The total mass and row sums are unchanged, so compare the reciprocal products over cells and column sums. Write $\mathcal{J}$ for the indices of the other columns and put

$$U = \prod_{j\in\mathcal{J}}\prod_i (1+tp_{ij})^{-1}, \qquad V = \prod_{j\in\mathcal{J}}(1+tc_j)^{-1},$$

$$H = \prod_i (1+t(x_i+y_i))^{-1}, \qquad K = \prod_i [(1+tx_i)(1+ty_i)]^{-1}.$$

Empty products equal one. For each remaining column, $\prod_i(1+tp_{ij}) \geq 1+tc_j$, so $U \leq V$. Set $X = \sum_i x_i$ and $Y = \sum_i y_i$. Lemma 19, applied to tx_i, ty_i, gives

$$L_t(p') - L_t(p) = U(H-K) \leq UG(tX,tY) \leq VG(tX,tY) = L_t(c') - L_t(c).$$

Thus $D_t(p) \geq D_t(p')$. Repeated merging leaves one column, where $D_t = 0$. Hence $D_t \geq 0$ for every table. Transposition gives the same monotonicity for row merging. If both marginal sum vectors have at least two positive entries, partition the rows into two groups of positive total mass and do the same for the columns. The resulting table has two positive row sums and two positive column sums. Lemma 18 makes its deficit strictly positive for $t > 0$; merging monotonicity gives the same conclusion for p. If a marginal sum vector has at most one positive entry, the table is supported in a single row or column, or has zero mass. Direct cancellation in D_t then gives equality. □

The comparison (36) keeps the same parameter t throughout the merging argument. Related determinant inequalities for partial traces use positive powers and dimension factors [25]; applied directly to $I + t\rho_{AB}$, they introduce dimension multiples of the identity in the marginal terms. The classical comparison retains the identity matrices in (33).

Proof of Proposition 17. Choose eigenbases of ρ_A and ρ_B. Set $p_{ij} = \langle i,j|\rho_{AB}|i,j\rangle$ in their product basis. Its row and column sums are the marginal eigenvalues. Apply Proposition 20, then use Hadamard's determinant inequality:

$$\det(I+t\rho_{AB}) \leq \prod_{i,j}(1+tp_{ij}), \qquad \det(I+t\rho_{AB})^{-1} \geq L_t(p).$$

The quantum deficit in (33) is at least $D_t(p)$. If both marginals are mixed, this lower bound is positive for every $t > 0$ by Proposition 20. A pure marginal forces ρ_{AB} to be its tensor product with the other marginal; the nonzero joint spectrum then equals the other marginal spectrum, giving pointwise cancellation. □

Integrating (33) and using (9) gives a second proof of Theorem 2, including its equality condition. This determinant argument is independent of Proposition 6. It uses the integral representation, the classical merging comparison in Proposition 20, and Hadamard's inequality.

For two qubits, the polynomial estimate also gives an explicit lower bound in terms of the marginal determinants.

Corollary 21 (Subadditivity deficit for two qubits)**.** *Every state ρ_{AB} on $\mathbb{C}^2 \otimes \mathbb{C}^2$ satisfies*

$$Q(\rho_A) + Q(\rho_B) - Q(\rho_{AB}) \geq \frac{1}{3}\det(\rho_A)\det(\rho_B). \tag{37}$$

The right side is positive when both marginals are mixed. If either marginal is pure, both sides vanish. The coefficient $1/3$ is not claimed to be optimal.

Proof. Choose marginal eigenbases and let a, b, c, d be the four diagonal entries of ρ_{AB} in their product basis, arranged as a table $p = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then $a+b+c+d = 1$ and

$$u = (a+b)(c+d) = \det(\rho_A), \qquad v = (a+c)(b+d) = \det(\rho_B).$$

Let $R(t)$ be the left side minus the right side of (33). Hadamard's inequality gives $R(t) \geq D_t(p)$. For this normalized table, let P_A, P_B, P and $N(t)$ be as in Appendix A. Writing $E = ad + bc$ and $x = abc + abd + acd + bcd = ad(b+c) + bc(a+d)$ gives $x \leq E$. The coefficients of N are nonnegative, and its terms of degrees two and three therefore give

$$N(t) \geq Et^2 + (2E - x)t^3 \geq Et^2(1+t).$$

Consequently

$$D_t(p) = \frac{N(t)}{(1+t)P_A P_B P} \geq \frac{Et^2}{P_A P_B P} \geq Et^2 e^{-3t}.$$

For the last inequality, each of P_A, P_B, P is a product of factors $1 + t\lambda_j$ with $\lambda_j \geq 0$ and $\sum_j \lambda_j = 1$. The bound $1 + y \leq e^y$ for $y \geq 0$ thus bounds each product by e^t. Also, using $a+b+c+d = 1$,

$$E - uv = ad(a+d-ad) + bc(b+c-bc) + 2abcd \geq 0,$$

since all four entries lie in $[0,1]$. Integrating with (9) now yields

$$Q(\rho_A) + Q(\rho_B) - Q(\rho_{AB}) = \int_0^\infty \frac{R(t)}{t^2}\, \mathrm{d}t \geq E \int_0^\infty e^{-3t}\, \mathrm{d}t = \frac{E}{3} \geq \frac{uv}{3}. \quad \square$$

5.2. Classical strong subadditivity and its quantum violation

For three classical variables, the reciprocal determinant products factor over the middle variable. We combine the inequalities for unnormalized slices using an auxiliary probability construction.

5.2.1. Classical distributions

Theorem 22 (Classical strong subadditivity)**.** *For every finite classical joint distribution p_{ijk},*

$$Q(p_{AB}) + Q(p_{BC}) \geq Q(p_B) + Q(p_{ABC}). \tag{38}$$

More strongly, for every $t \geq 0$,

$$L_t(p_B) + L_t(p_{ABC}) \geq L_t(p_{AB}) + L_t(p_{BC}).$$

The pointwise statement holds also for unnormalized nonnegative tables. For a probability distribution, equality in (38) holds exactly when $A = f(B)$ almost surely or $C = g(B)$ almost surely. For each fixed $t > 0$, the pointwise equality condition is the same.

Proof. Fix $t > 0$. For each middle label j put

$$d_j = (1 + t p_B(j))^{-1}, \quad a_j = \prod_{i,k} (1 + t p_{ijk})^{-1},$$

$$x_j = \prod_i (1 + t p_{AB}(i,j))^{-1}, \quad y_j = \prod_k (1 + t p_{BC}(j,k))^{-1}.$$

Refinement gives $0 < a_j \leq x_j, y_j \leq d_j$. Proposition 20, applied to the unnormalized slice at j, gives $a_j + d_j \geq x_j + y_j$. Set $A_j = a_j/d_j$, $X_j = x_j/d_j$, and $Y_j = y_j/d_j$. The four numbers

$$A_j, \quad X_j - A_j, \quad Y_j - A_j, \quad 1 - X_j - Y_j + A_j$$

are nonnegative and sum to one. Use them as probabilities that two auxiliary events E_j, F_j both occur, only the first occurs, only the second occurs, or neither occurs. The marginal probabilities are X_j, Y_j, and the intersection has probability A_j. Construct these event pairs independently across j; they need not belong to the original probability space. Inclusion–exclusion gives

$$\prod_j X_j + \prod_j Y_j - \prod_j A_j = \Pr\Big(\bigcap_j E_j \cup \bigcap_j F_j\Big) \le 1.$$

Multiplying by $\prod_j d_j$ proves the pointwise statement. Zero slices give $a_j = x_j = y_j = d_j = 1$; $t = 0$ is immediate.

For equality at $t > 0$, put $u_j = X_j - A_j$, $v_j = Y_j - A_j$ and $w_j = 1 - X_j - Y_j + A_j$. The global deficit divided by $\prod_j d_j$ is the probability that both global intersections fail. If $w_j > 0$ at any slice, choose neither event there and both events in every other slice. Because every $A_\ell > 0$, this gives a strictly positive global deficit. Equality thus forces $w_j = 0$ throughout; by Proposition 20, each positive slice has one positive row or one positive column.

In a slice with a single row, $X_j = 1$, $A_j = Y_j$, and $v_j = 0$; if C varies there, strict refinement gives $u_j > 0$. In a slice with a single column, $Y_j = 1$, $A_j = X_j$, and $u_j = 0$; if A varies there, $v_j > 0$. If neither outer variable is determined by B, there is a slice where A varies and a different slice where C varies. Choose only F in the first, only E in the second, and both elsewhere. This event has positive probability and makes both global intersections fail. Thus equality requires the same outer variable to be deterministic in every positive middle slice.

Conversely, if $A = f(B)$, the nonzero entries of p_{AB} are those of p_B, and those of p_{ABC} are those of p_{BC}; both kernel and subentropy deficits vanish. The case $C = g(B)$ is symmetric. Otherwise the pointwise deficit is strictly positive at every $t > 0$. Integrating using (9) therefore proves (38) and its complete equality condition. □

The same outer variable must be deterministic across all positive middle slices. For example, let $\Pr[B = 0] = \Pr[B = 1] = 1/2$ and take the conditional (A, C) tables

$$p_{AC|B=0} = \begin{pmatrix} 1/2 & 1/2 \\ 0 & 0 \end{pmatrix}, \qquad p_{AC|B=1} = \begin{pmatrix} 1/2 & 0 \\ 1/2 & 0 \end{pmatrix}.$$

Each slice has a deterministic outer variable, so $A \perp C \mid B$. Nevertheless, the full distribution has pointwise deficit

$$\big[(1 + t/2)^{-1} - (1 + t/4)^{-2}\big]^2 > 0 \quad (t > 0),$$

and its subentropy deficit is $25/12 - 3\log 2 > 0$. Conditional independence therefore does not characterize subentropy SSA equality as it does for Shannon entropy.

5.2.2. An exact violation for three qubits

A Bell pair on AB with an independent maximally mixed qubit C already violates strong subadditivity, as recorded in [9]. Its deficit is $Q(I_4/4) - 2Q(I_2/2) = -1/12$: the AB state is pure, the BC state is maximally mixed, and the nonzero joint spectrum is that of $I_2/2$. The following example has A classical and shows failure even when discarding a subsystem implements an optimal discrimination measurement. Its state matrix is rational, and its deficit has a strictly negative determinant integrand at every positive parameter.

Proposition 23 (Quantum strong subadditivity fails)**.** *Let A be a classical binary register and B, C qubits. Define*

$$|\phi_0\rangle = \frac{2|00\rangle + |11\rangle}{\sqrt{5}}, \qquad |\phi_1\rangle = \frac{|00\rangle + 2|11\rangle}{\sqrt{5}},$$

$$\rho_{ABC} = \tfrac{1}{2}|0\rangle\langle 0| \otimes |\phi_0\rangle\langle\phi_0| + \tfrac{1}{2}|1\rangle\langle 1| \otimes |\phi_1\rangle\langle\phi_1|.$$

Then

$$\begin{aligned} &Q(AB) + Q(BC) - Q(B) - Q(ABC) \\ &\quad = 2\log 5 - \frac{256}{135}\log 2 - \frac{81}{40}\log 3 + \frac{5}{18} \simeq -0.04244871319 < 0. \end{aligned} \tag{39}$$

Its determinant integrand is strictly negative for every $t > 0$.

Proof. All matrix entries are rational. The two nonzero blocks, on the basis pairs $(000, 011)$ and $(100, 111)$ respectively, are

$$\frac{1}{10}\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}, \qquad \frac{1}{10}\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

The block structure and partial traces give the nonzero spectra

$$\begin{aligned} \operatorname{spec}_+(\rho_{ABC}) &= \operatorname{spec}_+(\rho_B) = (1/2, 1/2), \\ \operatorname{spec}_+(\rho_{AB}) &= (2/5, 2/5, 1/10, 1/10), \\ \operatorname{spec}_+(\rho_{BC}) &= (9/10, 1/10). \end{aligned}$$

Put $D_0 = (1 + t/2)^2$, $D_1 = (1 + 2t/5)^2(1 + t/10)^2$, and $D_2 = (1 + 9t/10)(1 + t/10)$. Exact expansion gives

$$\frac{2}{D_0} - \frac{1}{D_1} - \frac{1}{D_2} = -\frac{t^2(5000 + 2500t + 250t^2 + 75t^3 + 7t^4)}{62500 D_0 D_1 D_2} < 0 \quad (t > 0). \tag{40}$$

This is the reciprocal determinant combination for the SSA deficit. The weight in (9) is positive, so the subentropy deficit is strictly negative. Its exact value follows from (1), which gives

$$\begin{aligned} Q(AB) &= \log 5 - \frac{121}{135}\log 2 - \frac{13}{18}, \\ Q(BC) &= \log 10 - \frac{81}{40}\log 3, \qquad Q(B) = Q(ABC) = \log 2 - \frac{1}{2}. \end{aligned}$$

For the repeated spectrum, with $a = 2/5$, $b = 1/10$, and $f(z) = z^4 \log z$, the confluent divided difference is

$$Q(a, a, b, b) = -\frac{f'(a) + f'(b)}{(a - b)^2} + \frac{2[f(a) - f(b)]}{(a - b)^3}.$$

Substituting these values yields the stated subentropies and (39). □

The same state shows that an optimal discrimination measurement can decrease conditional subentropy. For the equally likely input signals $(2|0\rangle + |1\rangle)/\sqrt{5}$ and $(|0\rangle + 2|1\rangle)/\sqrt{5}$, the isometry $|0\rangle \mapsto |00\rangle$, $|1\rangle \mapsto |11\rangle$ gives the counterexample above. Discarding C implements dephasing in the computational basis, with output signals $\operatorname{diag}(4/5, 1/5)$ and $\operatorname{diag}(1/5, 4/5)$. The difference of the input signal states weighted by their priors is $\operatorname{diag}(3/10, -3/10)$, so this measurement is optimal and preserves $p_{\text{guess}} = 4/5$. Nevertheless, the negative deficit in (39) means $Q(A|B) < Q(A|BC)$. Thus conditional subentropy does not satisfy the usual data processing inequality, even when the label is classical. The comparison with measurement posteriors remains valid: it bounds conditional subentropy by an average of posterior subentropies.

6. Spectral lower bounds on the error and independent ancillary states

6.1. Two lower bounds from the same spectra

Both conditional subentropy and a difference of purities are determined by the joint and average spectra. Purity is the trace of the squared density operator; the difference below gives a complementary lower bound on the discrimination error. For $n \geq 2$ put

$$c(\rho) = \sum_{x<z} \operatorname{Tr}(A_x A_z) = \frac{1}{2}\left(\operatorname{Tr}\rho_B^2 - \operatorname{Tr}\rho_{XB}^2\right), \qquad g_n(u) = u - \frac{n}{2(n-1)}u^2. \tag{41}$$

The curve g_n is half the classical Fano upper bound for the noncollision probability (the probability that two independent draws from the same distribution differ) [14, Example 16, Equations (75)–(76)]. The same Gram decomposition used for subentropy gives a quantum bound in terms of the difference of purities.

Proposition 24 (A complementary quadratic lower bound)**.** *For every n-label CQ ensemble with optimal error e,*

$$0 \leq c(\rho) \leq g_n(e), \qquad e \geq g_n^{-1}(c(\rho)) = \frac{n-1}{n}\left(1 - \sqrt{1 - \frac{2n}{n-1}c(\rho)}\right). \tag{42}$$

The curve bound is pointwise sharp at fixed n and c.

Proof. For the Gram matrix $G = M^*M$ of Proposition 6, define $J(G) = \|G - \Delta G\|_{\mathrm{HS}}^2/2$. Pinching is a Hilbert–Schmidt orthogonal projection, so

$$J(G) = \frac{1}{2}\left(\operatorname{Tr}G^2 - \operatorname{Tr}(\Delta G)^2\right) = c(\rho).$$

The functional is convex because it is the squared norm of a linear map. On a pure state with block weights b, it equals $j(b) = (1 - \sum_x b_x^2)/2$. Apply convexity to the pure decomposition obtained by refining the measurement in (16), then use concavity of j to combine the refined outcomes belonging to each original outcome. This gives $c(\rho) \leq \sum_y q_y j(b^y)$ for every finite POVM. If $u = 1 - \max_x b_x$, Cauchy–Schwarz on the other $n-1$ coordinates gives $\sum_x b_x^2 \geq (1-u)^2 + u^2/(n-1)$, hence $j(b) \leq g_n(u)$. Concavity of the quadratic curve and Jensen's inequality applied to an optimal guessing POVM give $c \leq g_n(e)$. Since the curve is strictly increasing on $[0, 1 - 1/n]$, the quadratic formula gives the inverse bound. Identical pure signals with priors $(1-e, e/(n-1), \ldots)$ have $c = j(p) = g_n(e)$ and attain every point. □

Montanaro's lower bound [11, Theorem 1] is

$$e \geq \sum_{x<z} p_x p_z F(\rho_x, \rho_z), \qquad F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2.$$

We use squared fidelity. Since $F(\rho,\sigma) \geq \operatorname{Tr}(\rho\sigma)$, this already implies $e \geq c(\rho)$. Proposition 24 improves this scalar bound to $g_n^{-1}(c) > c$ for $c > 0$, since $g_n(u) < u$ for $u > 0$. The quadratic bound and Montanaro's full fidelity bound do not dominate each other. For two equiprobable identical signals, the quadratic and Montanaro bounds are $1/2$ and $1/4$ when the common state is pure, and $(1 - 1/\sqrt{2})/2$ and $1/4$ when it is $I_2/2$.

Qiu's trace norm bound [26, Theorem 1] gives

$$e \geq \frac{1}{2}\left(1 - \frac{1}{n-1}\sum_{x<z}\|A_x - A_z\|_1\right).$$

Audenaert and Mosonyi [27, Equation (6)] restate this as a sum of weighted binary errors divided by $n-1$. For $n = 2$ it is the exact Helstrom error and hence is at least either spectral lower bound. In general it requires pairwise operator information beyond the two spectra.

Barnum and Knill [13, Sections IV–V] compare optimal discrimination with the pretty good measurement (PGM). Write

$$M_x^{\mathrm{PGM}} = \rho_B^{-1/2} A_x \rho_B^{-1/2}, \qquad e_{\mathrm{PGM}} = 1 - \sum_x \mathrm{Tr}(A_x M_x^{\mathrm{PGM}}).$$

The inverse is taken on the support of ρ_B, and the POVM is completed arbitrarily outside that support. Their guarantee $(1-e)^2 \le 1 - e_{\mathrm{PGM}}$, together with the fidelity upper bound of Audenaert and Mosonyi [27, Theorem 3.1], yields in our fidelity convention

$$1 - \sqrt{1 - e_{\mathrm{PGM}}} \le e \le e_{\mathrm{PGM}} \le \sum_{x<z} \sqrt{p_x p_z F(\rho_x, \rho_z)}.$$

The last inequality improves the original Barnum–Knill fidelity upper bound by a factor of two. These guarantees use the ensemble operators or pairwise fidelities; the lower bounds in (42) and Theorem 1 use the joint and average spectra.

The two lower bounds can be used together:

$$e \ge \max\{\Phi_n^{-1}(Q(X|B)),\, g_n^{-1}(c(\rho))\}.$$

Each bound is optimal when its statistic is fixed separately; optimality with both statistics fixed does not follow. Even fixing the priors, individual signal spectra, and average state does not in general determine the optimal error. Consider three commuting signals on $\mathbb{C}^2$, represented by the joint probability tables

$$T = \frac{1}{21}\begin{pmatrix} 6 & 2 \\ 1 & 5 \\ 7 & 0 \end{pmatrix}, \qquad \widetilde{T} = \frac{1}{21}\begin{pmatrix} 2 & 6 \\ 5 & 1 \\ 7 & 0 \end{pmatrix}. \tag{43}$$

Row x gives $A_x = \mathrm{diag}(T_{x1}, T_{x2})$, and similarly for $\widetilde{T}$. The row sums give the same priors $p = (8,6,7)/21$, and each row has the same pair of entries in both tables. Thus each signal has the same spectrum in the two ensembles. Both average states are $\rho_B = \mathrm{diag}(14,7)/21$, and both joint spectra are $(0,1,2,5,6,7)/21$. Dephasing a POVM in the common eigenbasis leaves its success probability unchanged. Optimizing its diagonal entries independently in each column therefore gives

$$p_{\mathrm{guess}}(T) = \sum_{y=1}^{2} \max_{1 \le x \le 3} T_{xy} = \frac{12}{21}, \qquad p_{\mathrm{guess}}(\widetilde{T}) = \frac{13}{21}.$$

The optimal errors are $3/7$ and $8/21$, respectively, although each spectral lower bound has the same value on both ensembles.

The following examples show that neither spectral bound dominates the other.

6.2. Comparison of the bounds

For two pure signals with priors $p, 1-p$ and overlap s,

$$c = p(1-p)s^2, \qquad g_2^{-1}(c) = \frac{1 - \sqrt{1 - 4p(1-p)s^2}}{2} = e. \tag{44}$$

Here $P_i = |\psi_i\rangle\langle\psi_i|$ are the signal states and $s = |\langle\psi_0|\psi_1\rangle|$ is their overlap. Indeed, $pP_0 - (1-p)P_1$ has trace $2p-1$ and determinant $-p(1-p)(1-s^2)$ on the two-dimensional signal span when $s < 1$. Its trace norm is $\sqrt{1 - 4p(1-p)s^2}$, giving the Helstrom minimum error. The cases of identical signals or zero priors follow directly or by continuity. Thus the quadratic bound equals the optimal error for binary pure signals with arbitrary priors.

Table 1. Binary pure signals with equal priors. Montanaro's bound is $s^2/4$; the subentropy bound is $\Phi_2^{-1}(Q(X|B))$. The quadratic bound equals the actual error. Actual errors are rounded to nearest; lower bounds are rounded downward to five places using rational enclosures. Endpoints are evaluated analytically.

Overlap s	Actual error	Montanaro bound	Subentropy bound
0.00	0.00000	0.00000	0.00000
0.25	0.01588	0.01562	0.01097
0.50	0.06699	0.06250	0.04957
0.80	0.20000	0.16000	0.16664
0.95	0.34388	0.22562	0.31592
1.00	0.50000	0.25000	0.50000

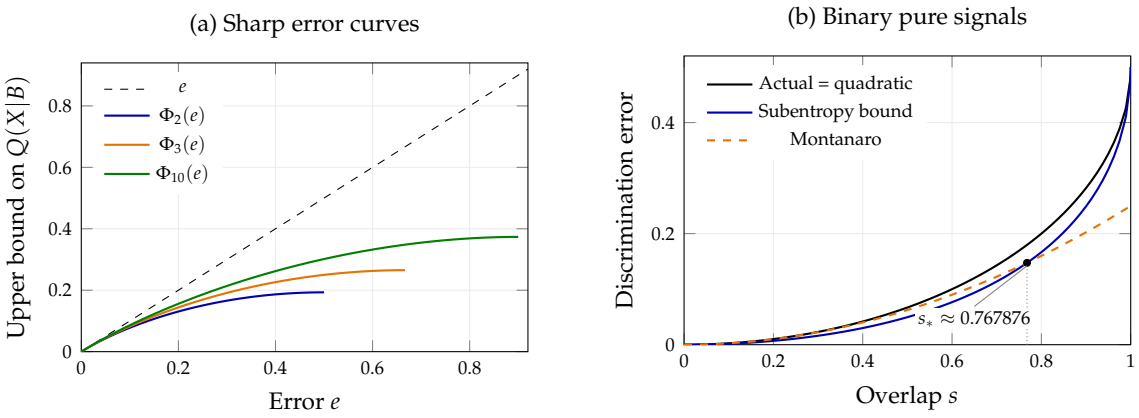


Figure 1. Sharp curves and their operational inversion. (a) Each curve stops at its largest admissible optimal error, $1 - 1/n$. Identical pure signals attain every point of the corresponding curve. (b) For the pure signals with equal priors in (45), the quadratic bound equals the actual error; the other curves are the subentropy bound and Montanaro's bound $s^2/4$. The marked crossover is $s_* \approx 0.767876$: Montanaro's bound is stronger for $0 < s < s_*$, and the subentropy bound is stronger for $s_* < s \leq 1$. The subentropy bound is exact at the orthogonal and identical endpoints. Plot coordinates are rounded to ten significant digits.

For equal priors, the overlap $s = |\langle\psi_0|\psi_1\rangle|$ also determines the two spectra: the joint nonzero spectrum is $(1/2, 1/2)$ and the average spectrum is $((1+s)/2, (1-s)/2)$. Hence

$$Q(X|B) = \log 2 - \frac{1}{2} - \Phi_2\Big(\frac{1-s}{2}\Big), \qquad e = \frac{1-\sqrt{1-s^2}}{2}. \tag{45}$$

The error formula is (44) specialized to equal priors. For $0 < u < 1/2$ the subentropy curve has the expression

$$\Phi_2(u) = -\frac{(1-u)^2\log(1-u) - u^2\log u}{1-2u},$$

with continuous values at the endpoints. For this binary pure family, Montanaro's bound is $s^2/4 = c$. Table 1 and Figure 1 compare it with the subentropy bound and the exact error. For small overlap, $Q(X|B) = s^2/6 + O(s^4)$ whereas $e = s^2/4 + O(s^4)$. Since $\Phi_2(u) \sim u$ at zero, its inverse captures two thirds of the leading error. Montanaro's bound is strictly stronger than the subentropy bound for $0 < s < s_*$, where $s_* \approx 0.767876$ is the unique crossover in $(0, 1)$, determined by

$$\Phi_2(s_*^2/4) = \log 2 - \frac{1}{2} - \Phi_2\Big(\frac{1-s_*}{2}\Big).$$

The subentropy bound is strictly stronger for $s_* < s \leq 1$; at $s = 0$ both bounds vanish. Appendix F.3 proves uniqueness and certifies the numerical value. The quadratic bound equals the optimal error throughout this family.

The subentropy bound can be stronger for three signals. To see this, take pure signals with priors $p = (2/3, 1/6, 1/6)$ and weighted Gram matrix

$$G_\varepsilon = \begin{pmatrix} 2/3 & 0 & 0 \\ 0 & 1/6 & \varepsilon \\ 0 & \varepsilon & 1/6 \end{pmatrix}, \qquad 0 < \varepsilon < 1/6. \tag{46}$$

One realization is $\psi_1 = e_1$, $\psi_2 = e_2$, $\psi_3 = 6\varepsilon e_2 + \sqrt{1 - 36\varepsilon^2}\, e_3$ in an orthonormal basis. The average spectrum is $(2/3, 1/6 + \varepsilon, 1/6 - \varepsilon)$ and the joint nonzero spectrum is p, giving $c = \varepsilon^2$. Writing $a = 2/3$ and $b = 1/6$, the integral representation gives the exact identity

$$Q(X|B) = \varepsilon^2 \int_0^\infty \frac{\mathrm{d}t}{(1+at)(1+bt)^2((1+bt)^2 - t^2\varepsilon^2)}.$$

Setting the perturbation to zero in the integrand gives the leading coefficient. To justify the expansion and control the remainder, suppose $|\varepsilon| \le b/2$. After removing the prefactor, the integrand is bounded by $4/[3(1+at)(1+bt)^4]$. Its difference from the $\varepsilon = 0$ integrand is bounded by $O(\varepsilon^2)$ times this integrable function, since $t/(1+bt) \le 1/b$. Therefore

$$Q(X|B) = \alpha\varepsilon^2 + O(\varepsilon^4), \qquad \alpha = \int_0^\infty \frac{\mathrm{d}t}{(1+2t/3)(1+t/6)^4} = \frac{256\log 2 - 150}{27} > 1. \tag{47}$$

For an explicit evaluation of the coefficient, substitute $t = 6x$ and use

$$\frac{1}{(1+4x)(1+x)^4} = \frac{256/81}{1+4x} - \frac{64/81}{1+x} - \frac{16/27}{(1+x)^2} - \frac{4/9}{(1+x)^3} - \frac{1/3}{(1+x)^4}.$$

Integrating the first two terms together gives the stated value. The first three positive terms in the power series for the inverse hyperbolic tangent give $\log 2 > 842/1215 > 177/256$, proving $\alpha > 1$. Both inverse curves have right derivative one at zero, by Lemma 9 and the quadratic formula. Consequently $\Phi_3^{-1}(Q(X|B)) = \alpha\varepsilon^2 + o(\varepsilon^2)$, while $g_3^{-1}(c) = \varepsilon^2 + O(\varepsilon^4)$. The subentropy bound is therefore stronger for every sufficiently small positive perturbation. Together with the binary example, this proves that neither bound dominates the other over all ensembles. Appendix E gives a further family of three signals with an explicit optimal measurement and certified values.

6.3. Appending a specified independent ancillary state

Appending an independent ancillary state leaves the discrimination error unchanged, although conditional subentropy may change. This also illustrates the distinction between conditional subentropy and the posterior average in Proposition 6. Take a uniform binary label with identical signals $\rho_x = I_d/d$. Every measurement leaves the posterior uniform, so its average subentropy is $\log 2 - 1/2$. The conditional subentropy instead equals

$$q_d = Q(I_{2d}/(2d)) - Q(I_d/d) = \log 2 - H_{2d} + H_d. \tag{48}$$

For $d = 2$ this is $\log 2 - 7/12$, already strictly smaller than the posterior average. The same family can be obtained by appending an independent maximally mixed ancilla to identical pure signals. Its guessing error stays $1/2$ while $q_d = 1/(4d) + O(d^{-2})$ tends to zero. Consequently, the conditional subentropy bound can decrease when independent mixed side information is added. The bound remains pointwise sharp over all ensembles at each prescribed value of conditional subentropy.

To bound the unchanged error using the conditional subentropy of the extended state, we use a curve that depends on the ancillary state. For an independent ancillary register C in a specified state ν, define

$$\begin{aligned} C_\nu(\rho) &= Q(X|BC)_{\rho_{XB}\otimes\nu}, \\ f_\nu(p) &= Q(\mathrm{diag}(p)\otimes\nu) - Q(\nu), \\ \Phi_{n,\nu}(u) &= f_\nu\Big(1-u, \frac{u}{n-1}, \ldots, \frac{u}{n-1}\Big). \end{aligned} \tag{49}$$

The product form $\rho_{XBC} = \rho_{XB}\otimes\nu$ is essential to the following comparison.

Proposition 25 (Sharp curve for a specified ancilla)**.** *Fix a density operator ν and $n \geq 2$. For every CQ ensemble on the original system B, and every finite POVM on B with posteriors b^y and MAP error e_M,*

$$0 \leq C_\nu(\rho) \leq \sum_{y:q_y>0} q_y f_\nu(b^y) \leq \Phi_{n,\nu}(e_M). \tag{50}$$

Appending ν leaves the optimal guessing error e unchanged, and $C_\nu(\rho) \leq \Phi_{n,\nu}(e)$ is pointwise sharp at fixed n, ν, and dimension of the original system B. The curve is strictly increasing on $[0, 1-1/n]$, and therefore

$$e \geq \Phi_{n,\nu}^{-1}(C_\nu(\rho)). \tag{51}$$

Identical pure original signals with the priors in (3) attain the curve at every admissible error.

Proof. Use the Gram matrix G from Proposition 6 and the pure decomposition obtained there by refining the measurement. For the pinching $\Delta\otimes\mathrm{id}$, the increase in subentropy $C_{Q,\nu}(G) = Q(\Delta(G)\otimes\nu) - Q(G\otimes\nu)$ is convex in G: Lemma 4 applies to the affine map $G \mapsto G\otimes\nu$. For a pure state in the Gram representation with block weights p, its value is $f_\nu(p)$: the two spectra are $p\otimes\mathrm{spec}(\nu)$ and $\mathrm{spec}(\nu)$, up to zeros. Moreover, $C_{Q,\nu}(G) = C_\nu(\rho)$. Convexity therefore bounds conditional subentropy by the average over refined posteriors. The function f_ν is symmetric and strictly concave, since $p \mapsto \mathrm{diag}(p)\otimes\nu$ is affine and injective. Concavity bounds the refined average by the posterior average for the original measurement. The permutation average and Jensen argument in Lemma 8 then give (50). Nonnegativity follows from the increase in subentropy under pinching.

To compare the optimal errors before and after appending the ancilla, a guessing effect E_x on the joint system induces the effect $\mathrm{Tr}_C[(I\otimes\nu^{1/2})E_x(I\otimes\nu^{1/2})]$ on the original system, with the same success probability. The induced effects form a POVM. Conversely, effects on B lift to $E_x\otimes I_C$. Thus the optimal error is unchanged, even for singular ν. An optimal measurement on the original system gives the error bound. Identical pure original signals give $C_\nu = f_\nu(p)$ and error $1-\max p_x$, proving pointwise sharpness. Strict concavity and permutation symmetry give a unique uniform maximum and strict increase up to it, justifying (51). □

For $\nu = u_k$, write $C_k = C_{u_k}$ and $\Phi_{n,k} = \Phi_{n,u_k}$. These bounds require only the original spectra and the specified ancillary state; no physical ancillary system needs to be prepared. Although conditional subentropy tends to zero as the dimension of a maximally mixed ancillary state grows, inverting the corresponding curve gives a nontrivial limit.

Corollary 26 (Bounds with a maximally mixed ancillary state)**.** *For every original CQ ensemble with $n \geq 2$ labels and error e,*

$$0 \leq C_k(\rho) \leq \Phi_{n,k}(e) \leq \frac{e}{k}. \tag{52}$$

The coefficient $1/k$ is sharp as $e \downarrow 0$. For the statistic $c(\rho)$ in (41), one has

$$\begin{aligned} kC_k(\rho) &\to c(\rho), \\ \Phi_{n,k}^{-1}(C_k(\rho)) &\to \frac{n-1}{n}\left(1-\sqrt{1-\frac{2n}{n-1}c(\rho)}\right). \end{aligned} \tag{53}$$

For binary pure signals with arbitrary priors, the latter limit is the exact Helstrom error in (44).

Proof. The preceding proposition bounds conditional subentropy by the curve. Concavity and $\Phi_{n,k}(0) = 0$ bound the curve by its tangent at zero, $\Phi_{n,k}(e) \leq e\Phi'_{n,k}(0+)$. To compute the slope, put $T = g^* u_k g$ and let U be the average of the independent quadratic forms for the other $n-1$ blocks. Then $\mathbb{E}U = 1$ and the Gaussian formula gives

$$\Phi'_{n,k}(0+) = \mathbb{E}[T \log T] - \mathbb{E} \log T = \mathrm{Tr}(u_k K(u_k)) = \frac{1}{k}.$$

The integration by parts identity in Appendix B gives the middle equality. The integrable logarithm bounds used there also justify differentiation at the boundary. Identical pure original signals attain the curve, so the tangent coefficient is sharp. Appendix C proves (53). For binary pure signals, Equation (44) identifies this limit as the optimal error. □

The choice of curve matters: applying the original inverse $\Phi_n^{-1}(C_k)$ gives a bound tending to zero. Inverting the curve for the specified ancillary state instead converges to the quadratic bound in Proposition 24. This limiting bound is already computable from the original spectra. The convergence does not imply exactness for general ensembles or monotonic improvement as k increases.

6.4. Numerical evaluation and error control

State uncertainty and numerical error both affect the lower bounds obtained by inversion. Let ρ, σ be CQ states on the same alphabet of $n \geq 2$ labels and quantum system B of dimension m, with $D(\rho, \sigma) \leq \delta \leq 1$. Theorem 14, the modulus of continuity of subentropy, contraction under partial trace, and the range in Theorem 1 give the bound

$$\begin{aligned} |Q(X|B)_\rho - Q(X|B)_\sigma| &\leq \mathcal{R}_{n,m}(\delta), \\ \mathcal{R}_{n,m}(\delta) &= \min\{\delta,\, \omega_{nm}(\delta) + \omega_m(\delta),\, \log n - H_n + 1\}. \end{aligned} \tag{54}$$

The three terms come from the direct CQ continuity bound, the separate moduli of continuity, and the range of conditional subentropy. The first improves the bound 2δ obtained by bounding the two subentropies separately. Their minimum need not be the optimal modulus of continuity for conditional subentropy. If the marginal states on B both equal τ, the first term can be replaced by $\kappa(\tau)\delta$. Suppose σ approximates the true state ρ and its computed conditional subentropy $\widehat{q}$ has a certified error $|\widehat{q} - Q(X|B)_\sigma| \leq \eta$. Then

$$e_{\text{true}} \geq \Phi_n^{-1}\big(\max\{0, \widehat{q} - \eta - \mathcal{R}_{n,m}(\delta)\}\big). \tag{55}$$

The simpler choice $\mathcal{R}_{n,m}(\delta) = \delta$ is valid in every dimension. The argument of the inverse is at most the true conditional subentropy, so it lies in the domain of the inverse. This guarantee requires a justified bound η on the numerical error.

For a specified maximally mixed ancilla, Corollary 16 gives the smaller bound δ/k on the effect of state error. If $|\widehat{C}_k - C_k(\sigma)| \leq \eta$, the corresponding lower bound is

$$e_{\text{true}} \geq \Phi_{n,k}^{-1}\big(\max\{0, \widehat{C}_k - \eta - \delta/k\}\big). \tag{56}$$

Both extended states must have the specified product form. The lower bounds converge to the limiting value as $k \to \infty$ whenever the absolute evaluation errors satisfy $k\eta_k \to 0$ and the state uncertainties satisfy $\delta_k \to 0$. With fixed tolerances, the bounds remain valid, but convergence to the limit is not guaranteed.

Near the maximum of either subentropy curve, uncertainty in the inverse scales as the square root of uncertainty in subentropy. A small tolerance alone therefore cannot certify that the error equals

its endpoint value. Appendix F gives the local expansions and the rational interval method used to certify the displayed numerical lower bounds.

7. Discussion

The comparison with measurement posteriors underlies both subadditivity of quantum subentropy and the sharp discrimination bound. A classical Fano argument completes the latter proof: no stronger lower bound on the error is possible given only conditional subentropy and the number of labels. Near equality, strong concavity quantifies proximity of the posteriors to the maximizing distributions. Under perturbations of the CQ state, conditional subentropy satisfies optimal Lipschitz bounds, including when the marginal state of B is fixed.

Although neither spectral bound dominates the other, appending an independent ancillary state relates them without changing the error. With a maximally mixed ancilla of increasing dimension, inversion of the corresponding sharp curve converges to the quadratic bound computed from the original spectra. For binary pure signals with arbitrary priors, this limiting bound is exact. The error analysis gives sufficient conditions on the evaluation and state uncertainties for the computed bounds to converge to this limit.

The determinant comparison holds at every integration parameter, strengthening subadditivity and giving a proof of classical strong subadditivity. In the quantum setting, strong subadditivity can fail even when one register is classical. Subentropy need not be additive even for product states: $Q(I_2/2) = \log 2 - 1/2$ and $Q(I_4/4) = \log 4 - 13/12$ give a deficit of $1/12$. The subadditivity deficit therefore cannot serve as a correlation measure that vanishes on product states. Nor does a telescoping chain rule imply data processing for conditional subentropy, as the example with an optimal measurement shows.

The optimal modulus of continuity for $Q(X|B)$ on CQ states at a prescribed nonzero trace distance remains open; the best linear bound does not determine it. For the signal ensemble itself, a further question is which assumptions yield an explicit stability rate beyond the posterior estimate and the qualitative compactness argument away from zero error. Section 3.5 shows why the absolute deficit alone cannot control stability uniformly down to zero error.

Acknowledgments

The author thanks OpenAI for providing complimentary access to ChatGPT, which assisted with numerical experiments, exact computational checks, verification and simplification of proofs, the Lean formalization, literature searches, and proofreading. The author assumes full responsibility for all content.

References

1. Jozsa, R.; Robb, D.; Wootters, W.K. Lower bound for accessible information in quantum mechanics. *Physical Review A* **1994**, *49*, 668–677. https://doi.org/10.1103/PhysRevA.49.668
2. Datta, N.; Dorlas, T.; Jozsa, R.; Benatti, F. Properties of subentropy. *Journal of Mathematical Physics* **2014**, *55*, 062203. https://arxiv.org/abs/1310.1312
3. Jozsa, R.; Mitchison, G. Symmetric polynomials in information theory: Entropy and subentropy. *Journal of Mathematical Physics* **2015**, *56*, 062201. https://arxiv.org/abs/1404.7694
4. Wootters, W.K. A classical interpretation of the Scrooge distribution. *Entropy* **2018**, *20*, 619. https://www.mdpi.com/1099-4300/20/8/619
5. McGinley, M.; Schuster, T. The Scrooge ensemble in many-body quantum systems. arXiv: 2511.17172. https://arxiv.org/abs/2511.17172
6. Chen, W.; Cao, Y.; Wang, H.; Feng, Y. Minimum guesswork discrimination between quantum states. *Quantum Information and Computation* **2015**, *15*, 737–758. https://arxiv.org/abs/1410.5180
7. König, R.; Renner, R.; Schaffner, C. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory* **2009**, *55*, 4337–4347. https://arxiv.org/abs/0807.1338
8. Zhang, L.; Singh, U.; Pati, A.K. Average subentropy, coherence and entanglement of random mixed quantum states. *Annals of Physics* **2017**, *377*, 125–146. https://arxiv.org/abs/1510.08859

9. The Omega Institute. trureturing: Quantum subentropy modules. Public source repository, 2026. Lean source code, revision d3e1e931e66a, 19 October 2026 (accessed on 8 November 2026). https://github.com/the-omega-institute/trureturing/tree/d3e1e931e66ad4a3f54c95ef96cf723edcd7e792/D5/S3/Quantum/Subentropy
10. Zhang, L.; Zhang, W.; Ma, H. Subadditivity and the failure of strong subadditivity in the Nichols–Wootters family of entropies. arXiv: 2611.15431v1. https://arxiv.org/abs/2611.15431v1
11. Montanaro, A. A lower bound on the probability of error in quantum state discrimination. In Proceedings of the IEEE Information Theory Workshop, 2008. https://arxiv.org/abs/0711.2012v3
12. Keyl, M.; Werner, R.F. Estimating the spectrum of a density operator. *Physical Review A* **2001**, *64*, 052311. https://arxiv.org/abs/quant-ph/0102027
13. Barnum, H.; Knill, E. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics* **2002**, *43*, 2097–2106. https://arxiv.org/abs/quant-ph/0004088v2
14. Rioul, O. The interplay between error, total variation, alpha-entropy and guessing: Fano and Pinsker direct and reverse inequalities. *Entropy* **2023**, *25*, 978. https://www.mdpi.com/1099-4300/25/7/978
15. Scarlett, J.; Cevher, V. An introductory guide to Fano's inequality with applications in statistical estimation. In *Information-Theoretic Methods in Data Science*; Cambridge University Press, 2021; pp. 487–528. https://arxiv.org/abs/1901.00555
16. Dall'Arno, M.; Buscemi, F. Tradeoff relations between accessible information, informational power, and purity. *IEEE Transactions on Information Theory* **2019**, *65*, 2614–2622. https://arxiv.org/abs/1801.05185
17. Hanson, E.P.; Datta, N. Entropies, majorization flow, and continuity bounds. In *The Physics and Mathematics of Elliott Lieb*; EMS Press, 2022; pp. 473–514. https://arxiv.org/abs/1909.06981v3
18. Xiong, C.; Wu, J. Geometric coherence and quantum state discrimination. *Journal of Physics A: Mathematical and Theoretical* **2018**, *51*, 414005. https://arxiv.org/abs/1801.06031
19. Cheng, S.; Hall, M.J.W. Complementarity relations for quantum coherence. *Physical Review A* **2015**, *92*, 042101. https://arxiv.org/abs/1508.03240
20. Styliaris, G.; Campos Venuti, L.; Zanardi, P. Coherence-generating power of quantum dephasing processes. *Physical Review A* **2018**, *97*, 032304. https://arxiv.org/abs/1711.02291
21. Matthews, W.; Wehner, S.; Winter, A. Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Communications in Mathematical Physics* **2009**, *291*, 813–843. https://arxiv.org/abs/0810.2327
22. Bhatia, R. *Matrix Analysis*; Springer, 1997. https://doi.org/10.1007/978-1-4612-0653-8
23. Sason, I. Entropy bounds for discrete random variables via maximal coupling. *IEEE Transactions on Information Theory* **2013**, *59*, 7118–7131. https://arxiv.org/abs/1209.5259
24. Raninen, E.; Ollila, E. Bias adjusted sign covariance matrix. *IEEE Signal Processing Letters* **2022**, *29*, 339–343. https://doi.org/10.1109/LSP.2021.3134940
25. Li, Y. Extensions of some matrix inequalities related to trace and partial traces. *Linear Algebra and its Applications* **2022**, *639*, 205–224. https://arxiv.org/abs/2003.04520
26. Qiu, D. Minimum-error discrimination between mixed quantum states. *Physical Review A* **2008**, *77*, 012328. https://arxiv.org/abs/0707.3970
27. Audenaert, K.M.R.; Mosonyi, M. Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination. *Journal of Mathematical Physics* **2014**, *55*, 102201. https://arxiv.org/abs/1401.7658

Appendix A. A polynomial proof of the inequality for four cells

After rescaling as in Lemma 18, assume $a+b+c+d=1$. We clear the positive denominators and show that the resulting polynomial has nonnegative coefficients. Set

$$u=(a+b)(c+d), \qquad v=(a+c)(b+d), \qquad w=ab+ac+ad+bc+bd+cd,$$

$$x=abc+abd+acd+bcd, \qquad y=abcd, \qquad e=ad+bc, \qquad z=1+t.$$

The denominators associated with the row sums, column sums, and cells are

$$P_A=z+ut^2, \qquad P_B=z+vt^2, \qquad P=z+wt^2+xt^3+yt^4.$$

The left side minus the right side is

$$\frac{1}{P}+\frac{1}{z}-\frac{1}{P_A}-\frac{1}{P_B}=\frac{N(t)}{zP_AP_BP},$$

where expansion gives

$$\begin{aligned}N(t)&=et^2+(2e-x)t^3+(e-2x-y+2uv)t^4\\&\quad+(-x-2y+2uv)t^5+(-y+uvw)t^6+uvxt^7+uvyt^8.\end{aligned}\tag{A1}$$

To check the signs of these coefficients, group the entries along the two diagonals and put

$$m=ad,\qquad n=bc,\qquad s=a+d,\qquad k=b+c,\qquad \delta=m-n.$$

The normalization and the definitions above imply

$$s+k=1,\qquad e=m+n,\qquad x=mk+ns,\qquad y=mn,\qquad w=e+sk,\qquad uv=x+\delta^2.$$

In these variables, the coefficients of $t^2,\dots,t^8$ are, in order,

$$\begin{aligned}&e,\qquad e+ms+nk,\qquad e-y+2\delta^2,\\&m(k-n)+n(s-m)+2\delta^2,\\&xw-y+\delta^2w,\qquad uvx,\qquad uvy.\end{aligned}$$

Every entry lies between zero and one, so $m\le s$, $n\le k$, and $e\ge y$. This proves nonnegativity except at degree six, where $m\le s^2/4$ and $n\le k^2/4$ give

$$xw\ge xsk=msk^2+ns^2k\ge 4mn(s+k)=4y.$$

The coefficient of degree six is therefore at least $3y+\delta^2w$ and hence nonnegative. Since the denominator is positive for $t\ge 0$, this proves the normalized inequality.

Appendix B. Gaussian formulas for continuity of conditional subentropy

The spatial sign formula (27) is established in [24]. The Gaussian derivation below covers singular matrices and yields (29) and sharpness at a fixed marginal. For a nonzero positive semidefinite matrix M, choose an eigenbasis with eigenvalues $\lambda_i\ge 0$. Gaussian phase invariance makes the expectation in (27) diagonal, with ith entry $\mathbb{E}[\lambda_iZ_i/T]$. Here $T=\sum_i\lambda_iZ_i$, and the Z_i are independent exponential variables of mean one. Using $1/T=\int_0^\infty e^{-tT}\,\mathrm{d}t$ and Tonelli's theorem gives

$$\mathbb{E}[\lambda_iZ_i/T]=\int_0^\infty\frac{\lambda_i}{1+t\lambda_i}\prod_j(1+t\lambda_j)^{-1}\,\mathrm{d}t.$$

The normalized outer product has norm one and converges almost surely as matrices approach a nonzero positive semidefinite limit. Dominated convergence gives continuity of K, including changes of rank. Unitary covariance follows from Gaussian invariance, while positive λ_i give positive eigenvalues of $K(M)$.

For positive definite M, the moment bound below justifies differentiating F under the expectation. Integration by parts in the exponential variables then gives

$$\begin{aligned}\mathbb{E}[Z_i\log T]&=\mathbb{E}\log T+\lambda_i\mathbb{E}[Z_i/T],\\\nabla F(M)&=[1+\mathbb{E}\log(g^*Mg)]I+K(M).\end{aligned}$$

Phase invariance removes the off-diagonal entries. To justify these operations, note that T lies between positive multiples of $\sum_i Z_i$, with

$$\mathbb{E}\left[\left(\sum_i Z_i\right)\left|\log \sum_i Z_i\right|\right] < \infty.$$

For positive definite CQ blocks, set $A = \sum_x A_x$ and $R = \bigoplus_x A_x$. Subtracting their gradients gives

$$\nabla_{A_x}[F(A) - F(R)] = cI + K(A) - K_x(R),$$
$$c = \mathbb{E}\log(g^* A g) - \mathbb{E}\log(h^* R h).$$

Here g and h have the dimensions of A and R, respectively. The scalar c is common to all blocks, so it contributes zero along every perturbation of total trace zero.

To compare the block gradients, first consider the determinant factor. For positive semidefinite X, Y, factorization and Sylvester's identity give

$$\det(I + X + Y) = \det(I + X)\det\bigl(I + Y^{1/2}(I + X)^{-1}Y^{1/2}\bigr) \le \det(I + X)\det(I + Y).$$

The inequality follows from $Y^{1/2}(I + X)^{-1}Y^{1/2} \le Y$ and determinant monotonicity in the positive semidefinite order. Iteration gives $\det(I + tA)^{-1} \ge \det(I + tR)^{-1}$. For the matrix factor, use $A \ge A_x$ and the fact that, for $t > 0$,

$$M(I + tM)^{-1} = t^{-1}[I - (I + tM)^{-1}]$$

is operator monotone by order reversal under inversion. The scalar determinant factor and positive matrix factor for $K(A)$ thus dominate their counterparts for $K_x(R)$. Multiplication by positive scalars and integration preserve the matrix order, giving $0 \le K_x(R) \le K(A)$ and hence (29).

For sharpness, start with a single nonzero CQ block τ and work on its support. Choose P as the projection onto a common eigenvector of the marginal and its spatial sign covariance matrix; let $\lambda > 0$ be its eigenvalue for the marginal. Splitting off εP into a new CQ block changes the joint quadratic form from T to $T + \varepsilon(Z' - Z_i)$, where Z' is a new independent exponential random variable with mean one. The Gaussian functional then has right derivative

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} F(R_\varepsilon)\Big|_{0+} = \mathbb{E}[(Z' - Z_i)(1 + \log T)] = -\lambda \mathbb{E}[Z_i/T].$$

For positive $\varepsilon < \lambda/2$, the perturbed quadratic form lies between $T/2$ and a fixed multiple of the full Gaussian squared norm. Its logarithm is bounded by a constant plus absolute logarithms of the Gaussian norms on the support and full space. Multiplying by $Z' + Z_i$ gives an integrable bound for the difference quotients. Since the average remains τ, the conditional right derivative is the corresponding eigenvalue of $K(\tau)$. Choosing the eigenvector for the largest such eigenvalue proves sharpness in Corollary 15, also for singular τ.

Appendix C. Limit for maximally mixed ancillary states

After multiplication by k, the limit of conditional subentropy is determined by the purity difference between the average and joint states. Their equal traces cancel the leading terms in the determinant expansion below. Write $A = \rho_B = \sum_x A_x$ and $R = \bigoplus_x A_x$. For a density operator M, set

$$L_{M,k}(t) = \det(I + (t/k)M)^{-k}.$$

The integral representation gives $C_k = \int_0^\infty [L_{A,k}(t) - L_{R,k}(t)]\,\mathrm{d}t/t^2$. At each fixed t, expansion of $\log(1 + u)$ gives

$$\log L_{M,k}(t) = -t + \frac{t^2}{2k}\operatorname{Tr} M^2 + O(k^{-2}), \qquad \frac{k(L_{A,k}(t) - L_{R,k}(t))}{t^2} \to c(\rho)e^{-t},$$

where $c(\rho) = (\operatorname{Tr} A^2 - \operatorname{Tr} R^2)/2 = \sum_{x<z} \operatorname{Tr}(A_x A_z)$. For an integrable bound, the determinant factorization in Appendix B gives $D_k = \log L_{A,k} - \log L_{R,k} \geq 0$. The inequalities $u - u^2/2 \leq \log(1+u) \leq u$ for $u \geq 0$ give $D_k \leq t^2 \operatorname{Tr} A^2/(2k)$. Hence, for $k \geq 2$,

$$0 \leq \frac{k(L_{A,k} - L_{R,k})}{t^2} = \frac{kL_{A,k}(1 - e^{-D_k})}{t^2} \leq \frac{\operatorname{Tr} A^2}{2} L_{A,k} \leq \frac{\operatorname{Tr} A^2}{2}(1 + t/2)^{-2}.$$

For the last step, $\det(I + (t/k)A) \geq 1 + t/k$ and $(1 + t/k)^k \geq (1 + t/2)^2$. Since the final bound is integrable, dominated convergence proves $kC_k \to c(\rho)$.

Applying this limit to identical pure signals with the priors in (3) gives, for $0 \leq u \leq e_n = 1 - 1/n$,

$$k\Phi_{n,k}(u) \to g_n(u) = u - \frac{n}{2(n-1)} u^2.$$

The limiting quadratic is continuous and strictly increasing on $[0, e_n]$. To prove convergence of the inverse values, first note that the sharp bound gives $0 \leq c(\rho) \leq g_n(e_n)$. If $u_* = g_n^{-1}(c(\rho))$ lies in the interior, choose any two points on opposite sides of u_*. By the preceding limits and strict increase, $\Phi_{n,k}^{-1}(C_k)$ eventually lies between them. At either endpoint, use the interval boundary on one side and the same argument on the other. Solving the quadratic gives (53). For binary pure signals with priors $p, 1-p$, $\operatorname{Tr}(A_0 A_1) = p(1-p)s^2$, so (44) identifies the limit with the Helstrom error.

Appendix D. Equality in the modulus of continuity

The equality cases in Theorem 13 follow from its coupling and trace distance bounds.

Proof. Equality in the interior range forces $e = \tau = \delta > 0$ by strict increase of the modulus. The classical coupling must also attain the guessing bound. By Theorem 1, its Y marginal is a point mass and its X marginal is the distribution in (3). Thus the state with smaller subentropy is pure, say $|v\rangle\langle v|$, and the other has eigenvalues $(1-\delta, \delta/(d-1), \ldots)$. Measuring the projection onto the pure state gives

$$D(|v\rangle\langle v|, \sigma) \geq 1 - \langle v|\sigma|v\rangle \geq 1 - \lambda_{\max}(\sigma) = \delta.$$

The distance is at most δ, so equality holds throughout. Equality in the Rayleigh quotient makes v an eigenvector of σ for its largest eigenvalue, which has multiplicity one, giving (25).

On the plateau, equality requires the two states to attain the minimum and maximum of Q: pure states are exactly the minimizers, while strict concavity gives the uniform spectrum as the unique maximum. Thus equality holds exactly for a pure state and the maximally mixed state, at distance $1 - 1/d$. Zero distance forces identical states; dimension one is trivial. □

Appendix E. A cyclic ensemble of three pure states

To compare the bounds, consider three equally likely pure signals in dimension three,

$$|\psi_k\rangle = \sum_{j=0}^{2} \sqrt{\lambda_j}\, \omega^{jk} |j\rangle, \qquad \omega = e^{2\pi i/3}, \quad k = 0, 1, 2, \tag{A2}$$

where $\lambda_j \geq 0$ and $\sum_j \lambda_j = 1$. Averaging over the phases gives a diagonal state with spectrum λ. The joint classical–quantum state has three nonzero eigenvalues $1/3$, so

$$q = Q(X|B) = \log 3 - H_3 + 1 - Q(\lambda). \tag{A3}$$

The Fourier measurement projects onto $|\chi_k\rangle = 3^{-1/2} \sum_j \omega^{jk} |j\rangle$ and achieves

$$p_{\text{guess}} = \frac{1}{3}\Big(\sum_j \sqrt{\lambda_j}\Big)^2. \tag{A4}$$

Table A1. Three cyclic pure states with average state spectrum $\lambda = (1-2r, r, r)$, $0 \le r \le 1/3$. The optimal error follows from (A4), and (A3) determines both lower bounds. Numerical entries use five decimal places: errors are rounded to nearest, and lower bounds are rounded downward using rational enclosures. Endpoints are analytic.

r	Optimal error	Linear bound q	Subentropy bound $\Phi_3^{-1}(q)$
0	2/3	0.26527	2/3
0.01	0.52801	0.24639	0.48126
0.05	0.35049	0.18298	0.28014
0.10	0.22288	0.12121	0.15987
0.20	0.07145	0.03896	0.04323
1/3	0.00000	0.00000	0.00000

For a matching upper bound, set $s = \sum_j \sqrt{\lambda_j}$ and let Y be the diagonal matrix with entries $s\sqrt{\lambda_j}/3$ in the basis $\{|j\rangle\}$. Weighted Cauchy–Schwarz gives $|\langle\psi_k|v\rangle|^2 \le s\sum_j \sqrt{\lambda_j}|v_j|^2$ for every v, even when some λ_j vanish. Thus $Y \ge |\psi_k\rangle\langle\psi_k|/3$ for every k. Any POVM $\{M_k\}$ therefore has success probability at most $\sum_k \operatorname{Tr}(M_k Y) = \operatorname{Tr} Y = s^2/3$, matching the Fourier measurement and proving optimality.

Table A1 uses $\lambda = (1-2r, r, r)$ with $0 \le r \le 1/3$, from identical to orthogonal signals. At the displayed interior points, the subentropy bound improves on the linear bound but remains below the optimal error. Sharpness at fixed q and number of labels concerns the optimum over all ensembles; individual ensembles need not attain the bound.

Appendix F. Numerical evaluation of the inverse bounds

Appendix F.1. Behavior near the uniform endpoint

The curve Φ_n has zero slope at the uniform endpoint $e_n = 1 - 1/n$. The Gaussian Hessian there is $\partial_i\partial_j F = n(1 + \mathbf{1}_{i=j})/(n+1)$. The tangent direction has zero coordinate sum and squared norm $n/(n-1)$. Symmetry cancels the linear term, giving

$$\Phi_n(e_n) - \Phi_n(e_n - h) = \frac{n^2}{2(n^2-1)}h^2 + O(h^3). \tag{A5}$$

For a maximally mixed ancilla, the uniform joint spectrum has dimension nk and the tangent vector has squared norm $n/[k(n-1)]$. The marginal is fixed, so the same expansion gives

$$\Phi_{n,k}(e_n) - \Phi_{n,k}(e_n - h) = \frac{n^2}{2(n-1)(nk+1)}h^2 + O(h^3). \tag{A6}$$

Consequently, uncertainty in either inverse value scales as the square root of subentropy uncertainty near the maximum. A tolerance alone cannot certify that the error equals e_n.

Appendix F.2. Rational bounds for logarithms and inverse values

The displayed numerical lower bounds are certified by rational interval arithmetic. For a positive rational x, write $x = 2^\ell y$ with $1 \le y < 2$ and set $z = (y-1)/(y+1)$. For every integer $N \ge 1$,

$$0 \le \log y - 2\sum_{j=0}^{N-1} \frac{z^{2j+1}}{2j+1} \le \frac{2z^{2N+1}}{(2N+1)(1-z^2)}.$$

At the upper endpoint, the bound also encloses $\log 2$; scaling then gives an enclosure of $\log x$. Subentropy is evaluated using confluent divided differences in (1), with derivative values at repeated eigenvalues. On the increasing branch, a candidate is a certified lower bound on the inverse if an upper enclosure of Φ_n at that candidate does not exceed a lower enclosure of $Q(X|B)$. When a comparison is inconclusive, we increase precision or test nearby rational candidates. If it remains unresolved, we

report no certified bound. Table bounds are rounded downward; endpoints are evaluated analytically. Plot coordinates are rounded to ten significant digits.

Appendix F.3. Binary pure signals: comparison with Montanaro's bound

Write $q(s)$ for the conditional subentropy in (45). Since Φ_2 is strictly increasing on $[0, 1/2]$, the subentropy bound is larger than $s^2/4$ exactly when $q(s) > \Phi_2(s^2/4)$. To determine the number of crossings, set $t = s^2$ and define

$$f(t) = q(\sqrt{t}) - \Phi_2(t/4), \qquad 0 \le t \le 1.$$

Expanding the logarithms in the binary subentropy formula gives

$$q(\sqrt{t}) = \sum_{k=1}^{\infty} \frac{t^k}{2k(4k^2-1)}, \qquad 0 \le t < 1.$$

The positive coefficients, including that of t^2, make the function defined by this series strictly convex. Concavity of Φ_2 therefore makes f strictly convex, including at the endpoints by continuity. Also $f(0) = 0$, and Lemma 9 gives $f'(0+) = 1/6 - 1/4 = -1/12$, whereas $f(1) = \Phi_2(1/2) - \Phi_2(1/4) > 0$. Strict convexity and $f(0) = 0$ make $f(t)/t$ strictly increasing for $t > 0$. Thus f has exactly one zero t_* in $(0, 1)$, is negative before it and positive after it. This proves the ordering in Section 6.2 with $s_* = \sqrt{t_*}$.

The rational logarithm bounds above, with $N = 50$, certify a negative value of f at $t = 0.767875711400^2$ and a positive value at $t = 0.767875711401^2$. Consequently

$$0.767875711400 < s_* < 0.767875711401.$$

The value in Figure 1(b) is rounded to six decimal places.