

failure of subadditivity for matrix-based rényi information: strict gaussian counterexamples and complete order classification

Askar Uval

25 july 2026

Abstract

A widely used matrix-based Rényi construction defines joint entropy through the trace-normalized Hadamard product of two Gram matrices and asserts subadditivity for every positive order. We show that the assertion is false at every positive order except the von Neumann order $\alpha = 1$. The obstruction is exact: two categorical label Gram matrices reproduce the marginal and joint spectra of any rational classical distribution, importing the classical failure of Rényi subadditivity directly into the matrix functional. We then lift the obstruction to strict, entrywise-positive, infinitely divisible Gaussian radial-basis-function Gram matrices built from distinct samples related by a deterministic bijection. At order two, three samples are necessary and sufficient; a four-sample one-dimensional witness is negative at the arithmetic-median bandwidth and admits a short exact certificate. We also give certified strict-Gaussian witnesses at the commonly used orders 0.99 and 1.01. Cartesian products amplify every negative seed to $-c_\alpha \log n$, matching the universal $-\log n$ lower bound in asymptotic order. The result invalidates the claimed nonnegativity of the original matrix mutual information and the corollaries that depend on it, while leaving the marginal entropy and the separate joint-at-least-marginal inequality intact.

1 the matrix functional

Let $K_X, K_Y \in \mathbb{R}^{n \times n}$ be positive semidefinite Gram matrices with unit diagonal and nonnegative entries. Put

$$A = \frac{K_X}{n}, \quad B = \frac{K_Y}{n}, \quad C = \frac{K_X \circ K_Y}{n}.$$

All three matrices are positive semidefinite with trace one. For $\alpha > 0$, $\alpha \neq 1$, define

$$S_\alpha(M) = \frac{1}{1-\alpha} \log \operatorname{tr}(M^\alpha)$$

and define $S_1(M) = -\operatorname{tr}(M \log M)$ by continuity. The matrix-based information functional is

$$I_\alpha(K_X; K_Y) = S_\alpha(A) + S_\alpha(B) - S_\alpha(C). \tag{1}$$

Proposition 4.1(ii) of Sánchez Giraldo, Rao, and Príncipe [6] asserts $I_\alpha \geq 0$. Its proof verifies two endpoint choices for one matrix and then states the general inequality without an interpolation, convexity, or majorization argument.

2 complete positive-order classification

theorem 1 (order classification). *For the functional in (1), $\alpha = 1$ is the unique positive order at which $I_\alpha(K_X; K_Y) \geq 0$ for every admissible pair. For each fixed $\alpha \in (0, 1) \cup (1, \infty)$ there are distinct Euclidean points $x_1, \dots, x_n$, a permutation π , and a Gaussian RBF kernel such that*

$$(K_X)_{ij} = k_\sigma(x_i, x_j), \quad (K_Y)_{ij} = k_\sigma(x_{\pi(i)}, x_{\pi(j)}),$$

yet $I_\alpha(K_X; K_Y) < 0$. The three matrices K_X , K_Y , and $K_X \circ K_Y$ are strictly positive definite, entrywise positive, and infinitely divisible.

The proof separates into an exact categorical obstruction, a strict-Gaussian lift, and the exceptional order-one argument.

2.1 the categorical obstruction

lemma 2 (rational distributions occur exactly). *Let $q = (q_{ab})$ be a finite rational joint distribution with denominator n . There are n paired categorical observations whose label Gram matrices satisfy*

$$\text{spec}(K_X/n) = q_X \cup \{0\}, \quad \text{spec}(K_Y/n) = q_Y \cup \{0\}, \quad \text{spec}((K_X \circ K_Y)/n) = q \cup \{0\},$$

where zero eigenvalues are added to reach dimension n .

Proof. Repeat the pair (a, b) exactly nq_{ab} times. Set $(K_X)_{ij} = 1$ when observations i, j share their first label and zero otherwise; define K_Y analogously. Each matrix is a direct sum of all-ones blocks, whose nonzero normalized eigenvalues are the corresponding empirical category proportions. Their Hadamard product is the direct sum indexed by joint cells, giving the joint spectrum. $\square$

Fix $p \in (0, 1)$ with $p \neq 1/2$ and consider the equal-margin joint law

$$q_\delta = \begin{pmatrix} p^2 + \delta & p(1-p) - \delta \\ p(1-p) - \delta & (1-p)^2 + \delta \end{pmatrix}.$$

Write

$$r_\alpha = p^\alpha + (1-p)^\alpha, \quad t_\alpha(\delta) = \sum_{a,b} q_\delta(a, b)^\alpha.$$

At independence $t_\alpha(0) = r_\alpha^2$, while

$$t'_\alpha(0) = \alpha [p^{\alpha-1} - (1-p)^{\alpha-1}]^2 > 0 \tag{2}$$

for every $\alpha > 0$, $\alpha \neq 1$. The entropy-sum information of q_δ is

$$J_\alpha(\delta) = \frac{1}{1-\alpha} \log \frac{r_\alpha^2}{t_\alpha(\delta)}.$$

Thus $J_\alpha(\delta) < 0$ for sufficiently small $\delta > 0$ when $0 < \alpha < 1$, and for sufficiently small $\delta < 0$ when $\alpha > 1$. Choosing rational p, δ and applying the lemma gives a finite exact Gram counterexample at every non-Shannon positive order.

For a uniform integer family, take $p = 1/3$ and

$$N_m^s = \begin{pmatrix} m+s & 2m-s \\ 2m-s & 4m+s \end{pmatrix}, \quad n = 9m,$$

where $s = 1$ below order one and $s = -1$ above order one. Its row and column sums are both $(3m, 6m)$, and $N_m^s/(9m)$ approaches independence from the sign required by (2).

2.2 strict gaussian lift

Partition the source points into clusters of sizes n_0 and n_1 . Realize each cluster as a regular simplex in its own coordinate subspace, with simplex radius tending to zero, and separate the two centers along an additional orthogonal axis. A permutation can realize any integer two-by-two contingency table with matching row and column sums. As the radii tend to zero and the separation tends to infinity, the marginal Gaussian Gram spectra converge to the two cluster proportions and the joint Hadamard spectrum converges to the four cell proportions.

For every fixed $\alpha > 0$, $M \mapsto S_\alpha(M)$ is continuous on the trace-one positive-semidefinite cone, including at rank-deficient boundary points. Strict negativity therefore persists for sufficiently small positive radius and sufficiently large finite separation. Gaussian Gram matrices on distinct points are strictly positive definite. Moreover,

$$(K_X^{\text{ot}})_{ij} = \exp\left(-\frac{t\|x_i - x_j\|^2}{2\sigma^2}\right)$$

is another Gaussian Gram matrix for every $t > 0$, so the matrix is infinitely divisible. The joint Hadamard product is the Gaussian Gram matrix of the distinct concatenated points (x_i, y_i) and has the same properties.

2.3 why order one survives

Choose unit feature vectors u_i, v_i with Gram matrices K_X, K_Y , and set

$$\rho_X = \frac{1}{n} \sum_i |u_i\rangle\langle u_i|, \quad \rho_Y = \frac{1}{n} \sum_i |v_i\rangle\langle v_i|, \quad \rho_{XY} = \frac{1}{n} \sum_i |u_i \otimes v_i\rangle\langle u_i \otimes v_i|.$$

The nonzero spectra of these density operators equal those of A, B, C . Furthermore, ρ_X and ρ_Y are the partial traces of ρ_{XY} . Von Neumann subadditivity therefore gives

$$S_1(C) \leq S_1(A) + S_1(B),$$

proving $I_1 \geq 0$. This argument requires neither nonnegative entries nor infinite divisibility.

Under the Hartley extension, order zero also satisfies the upper inequality because

$$\text{rank}(K_X \circ K_Y) \leq \text{rank}(K_X) \text{rank}(K_Y).$$

Hence the validity set on the extended interval $[0, \infty]$ is $\{0, 1\}$; the source functional is stated for positive orders and does not include zero.

3 explicit order-two certificates

The smallest one-dimensional witness is

$$x = (0, 3/4, 1), \quad y = (3/4, 0, 1), \quad \sigma = 1.$$

With $u = e^{-9/16}$, $v = e^{-1}$, and $w = e^{-1/16}$, let

$$f = 3 + 2(u + v + w), \quad g = 3 + 2(u^2 + 2vw).$$

Then

$$f^2 - 9g > 0.333539597986325, \quad I_2 = \log(9g/f^2) < 0.$$

For two samples with marginal correlations $a, b \in [0, 1]$, the sign is controlled by

$$2(1 + a^2b^2) - (1 + a^2)(1 + b^2) = (1 - a^2)(1 - b^2) \geq 0,$$

so three samples are necessary and sufficient at order two.

A practical arithmetic-median witness is

$$x = (0, 1, 2, 7), \quad y = (0, 7, 2, 1), \quad \sigma = 7/2.$$

Both sets have arithmetic median pairwise distance $7/2$. Put $r = e^{-4/49}$ and

$$F(r) = 4 + 2(2r + r^4 + r^{25} + r^{36} + r^{49}),$$

$$G(r) = 4 + 2(r^8 + 2r^{26} + 2r^{50} + r^{72}).$$

Then

$$I_2 = \log \frac{16G(r)}{F(r)^2}, \quad F(r)^2 - 16G(r) = 1.36425056678849 \dots > 0,$$

and therefore

$$I_2 = -0.015130095410440268 \dots \text{ nats} = -0.021828113616819040 \dots \text{ bits}.$$

The accompanying dependency-free checker reconstructs the distance counts and certifies the exponential interval and polynomial sign with rational arithmetic. A separate Lean proof certifies the resulting scalar exponential enclosure, polynomial gap, negative logarithm, and two-sample bound. The exact matrix-to-polynomial reduction is independently checked but is not formalized in Lean.

4 certified near-shannon witnesses

The failure persists at the orders used by several downstream applications. With bandwidth one and $\eta = 10^{-12}$, form two orthogonal regular-simplex clusters. For cluster sizes n_0, n_1 , the within-cluster similarity is $a = e^{-\eta}$. Choose the squared center separation

$$d^2 = 60 - \eta \left(1 - \frac{1}{n_0} + 1 - \frac{1}{n_1} \right),$$

so every cross-cluster similarity is $c = e^{-30}$. A literal permutation with the stated contingency counts gives the following certified values.

α	n	cell counts	certified information value
1.01	35	(1, 5, 5, 24)	$-3.8222628510387147541 \dots \times 10^{-6}$
0.99	37	(1, 5, 5, 26)	$-5.1793230244726931694 \dots \times 10^{-6}$

The Arb certificate reduces the marginal matrix to a two-dimensional compound-symmetry quotient and the joint matrix to a four-cell quotient. At 100 decimal digits it verifies the permutation, contingency counts, positive separation, strictly positive eigenvalue intervals, trace normalization, and a wholly negative information interval.

5 cartesian amplification

theorem 3 (amplification). *Let a size- m seed have $I_\alpha(K_X; K_Y) < 0$. For $r \geq 1$, use the coordinatewise bijection on the Cartesian power and the same Gaussian bandwidth. Then, with $n = m^r$,*

$$I_{\alpha,r} = rI_{\alpha,1} = -c_\alpha \log n \longrightarrow -\infty, \quad c_\alpha = -\frac{I_{\alpha,1}}{\log m} > 0.$$

Proof. Gaussian kernels factor across Cartesian coordinates, so the marginal Gram matrices are $K_X^{\otimes r}$ and $K_Y^{\otimes r}$, while the joint is $(K_X \circ K_Y)^{\otimes r}$. Rényi entropy is additive on tensor products because

$$\text{tr}((M \otimes N)^\alpha) = \text{tr}(M^\alpha) \text{tr}(N^\alpha).$$

Each term in (1) is therefore multiplied by r . □

For every trace-one $n \times n$ positive-semidefinite matrix and every positive order, $0 \leq S_\alpha \leq \log n$. Consequently,

$$I_\alpha(K_X; K_Y) \geq -\log n.$$

The amplified family has the largest possible asymptotic order of negative divergence. For the exponentiated score $Q_\alpha = e^{I_\alpha}$, amplification gives $Q_{\alpha,r} = Q_{\alpha,1}^r = n^{-c_\alpha}$, while the universal bound is $Q_\alpha \geq 1/n$.

6 scope and propagation

The valid proposition 4.1(i) of [6], stating that the joint matrix entropy is at least either marginal entropy, is unaffected. It follows from a separable-state majorization argument [9]. Conditional entropy therefore remains nonnegative. What fails is its claimed upper bound by the other marginal entropy, together with the resulting nonnegativity of (1).

The false upper inequality is copied into the multivariate extension [12], where it supports nonnegativity claims for matrix mutual information and total correlation. A later paper coauthored by an author of the source paper [5] says the construction is well behaved at order one but justifies non-Shannon nonnegativity by citing a theorem about restricted bosonic and fermionic Gaussian quantum states [4]. Gaussian RBF Gram matrices are not thereby Gaussian quantum states. That paper uses $\alpha = 1.01$ throughout, exactly the order of the certified 35-point witness.

The same semantic issue reaches information bottlenecks, feature selection, pruning, and higher-order neuroscience metrics. This does not by itself refute reported empirical accuracies: the scalar objectives remain computable. It refutes their unrestricted interpretation as nonnegative mutual information or total correlation. Order-one descendants must be excluded. In particular, von-Neumann information-Vendi remains nonnegative. The order-two information-RKE score does not: the four-point witness yields $0.984983789397008 \dots < 1$, and Cartesian powers drive it to zero.

7 related work and priority boundary

Generic Rényi non-subadditivity is classical prior art, not a contribution of this note. Aczél, Forte, and Ng [1] characterized additive subadditive information measures in 1974. Wehrl’s review [11] states explicitly that quantum Rényi entropies are not subadditive except at the Hartley and von Neumann orders. Adesso, Girolami, and Serafini [2] note that the naive entropy-difference Rényi mutual information can be negative, while proving a special result for Gaussian quantum

states. Linden, Mosonyi, and Winter [7] give the modern all-order classification of Rényi entropic inequalities. An explicit equal-margin table for classical min-entropy appeared on MathOverflow in 2012 [8] and was later reused in an axiomatic study of conditional entropy [3]. Discrete delta kernels and tensor-product joint feature maps are also standard in kernel-embedding work [10].

The source’s earliest draft already observes that a single categorical label Gram matrix recovers one categorical spectrum. The joint use of two label Grams is an elementary delta-kernel construction; the apparently unreported point is that its normalized Hadamard spectrum transfers the old classical obstruction exactly into proposition 4.1(ii). The strongest priority candidates are the explicit refutation of that proposition, the strict infinitely divisible Gaussian-RBF realization on deterministic bijections, the minimal and practical witnesses, certified near-Shannon examples, the optimal-divergence construction, and the propagation audit. Tensor-product entropy additivity itself is not new.

As of 25 July 2026, targeted title, proposition-number, correction, and counterexample searches located no correction or equivalent matrix-functional counterexample. A Semantic Scholar traversal found 175 direct citations; none of their titles or abstracts described the failure, and none co-cited [7]. This search evidence supports but cannot prove priority. Domain-expert review and author contact remain necessary before asserting a first discovery.

References

- [1] J. Aczél, B. Forte, and C. T. Ng. Why the Shannon and Hartley entropies are natural. *Advances in Applied Probability*, 6(1):131–146, 1974.
- [2] G. Adesso, D. Girolami, and A. Serafini. Measuring Gaussian quantum information and correlations using the Rényi entropy of order 2. *Physical Review Letters*, 109:190502, 2012.
- [3] A. Américo, M. Khouzani, and P. Malacaria. Conditional entropy and data processing: an axiomatic approach based on core-concavity. *IEEE Transactions on Information Theory*, 66:5537–5547, 2020.
- [4] G. Camilo, G. T. Landi, and S. Eliëns. Strong subadditivity of the Rényi entropies for bosonic and fermionic Gaussian states. *Physical Review B*, 99:045155, 2019.
- [5] O. Skean, J. K. H. Osorio, A. J. Brockmeier, and L. G. Sanchez Giraldo. DiME: maximizing mutual information by a difference of matrix-based entropies. arXiv:2301.08164, 2023.
- [6] L. G. Sánchez Giraldo, M. Rao, and J. C. Principe. Measures of entropy from data using infinitely divisible kernels. *IEEE Transactions on Information Theory*, 61(1):535–548, 2015.
- [7] N. Linden, M. Mosonyi, and A. Winter. The structure of Rényi entropic inequalities. *Proceedings of the Royal Society A*, 469:20120737, 2013.
- [8] MathOverflow. Subadditivity for Rényi entropies, December 2012. <https://mathoverflow.net/questions/116864/subadditivity-for-renyi-entropies>.
- [9] M. A. Nielsen and J. Kempe. Separable states are more disordered globally than locally. *Physical Review Letters*, 86:5184–5187, 2001.
- [10] L. Song, K. Fukumizu, and A. Gretton. Kernel embeddings of conditional distributions: a unified kernel framework for nonparametric inference in graphical models. *IEEE Signal Processing Magazine*, 30(4):98–111, 2013.
- [11] A. Wehrl. General properties of entropy. *Reviews of Modern Physics*, 50:221–260, 1978.
- [12] S. Yu, L. G. Sanchez Giraldo, R. Jenssen, and J. C. Principe. Multivariate extension of matrix-based Rényi’s α -order entropy functional. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11):2960–2966, 2020.