Least-Atom Extremizers for Resampling Noise*

Matěj Kripner Václav Rozhoň Robert Šámal
Martin Šindelář Adrián Zámečník

hexagon:2611.00051v1

Abstract

A single least-mass atom determines the extremal behavior of resampling noise on a finite probability space. A fixed-moment comparison gives the sharp generalized Samorodnitsky inequality for every $1 < q < 2$ through one scalar curve, and identifies all nonconstant scalar extremizers. The same principle determines every L^p-to-L^q resampling norm for $1 < p < q < \infty$, resolving a conjecture of Cao, Fan, Han, Qiu, and Wang. The sharp hypercontractive parameter increases strictly with the least atom. The uniform-binary phase diagram is explicit, and product equality cases are classified for every nonuniform reference law and for the uniform binary law.

1 Introduction

1.1 Background and the source problem

Resampling is a basic model of noise on a finite probability space (Ω, μ). With probability ρ an observation is kept, and with probability $1 - \rho$ it is replaced by an independent sample from μ. Its averaging operator is

$$T_\rho f = \rho f + (1 - \rho)\mathbb{E}_\mu f.$$

The question is how much a function can remain concentrated after this averaging. Two familiar measurements are its logarithmic moments and the ratio of its output L^q-norm to its input L^p-norm. For example, a density concentrated on a very rare outcome may retain a large norm even after substantial resampling. It is less clear whether spreading the high values over several outcomes can be more effective.

On a product space, noise and partial observation give two ways to average a function. Product resampling acts independently in each coordinate. Partial observation retains a random set S of coordinates and averages over the rest. The generalized Samorodnitsky inequality compares them:

$$\log \|T_\rho^{[n]} f\|_q \le \mathbb{E}_{S\sim\lambda} \log \|\mathbb{E}(f \mid S)\|_q. \tag{1}$$

Here f is nonzero and nonnegative on Ω^n, the measure is $\mu^{\otimes n}$, and each coordinate is retained in S independently with probability λ. The sharp inclusion parameter is the smallest λ for which this holds in every dimension. Determining it quantifies the information that survives noise in terms of information available from a partial observation.

Samorodnitsky proved this comparison on the uniform Boolean cube for $q > 1$ [2, Thm. 1.1], and obtained the sharp parameter for integer $q \ge 2$ [9, Thm. 1.1]. Abawonse, Hązła, and O'Donnell extended the sharp result to finite product probability spaces and real $q \ge 2$ [1, Thms. 1 and 19]. Their nonconstant scalar extremizers are normalized indicators of least-mass atoms. Section 1.4 of the full version explicitly leaves $1 < q < 2$ untreated and explains why this indicator description should not persist throughout that range. Numbering in [1] below refers to its full arXiv version.

*This work was produced within the Bolzano project (https://bolzano.app), using AI systems for mathematical research and manuscript preparation. The listed authors contributed to the project and the preparation of this release. The manuscript received only limited human checking.

Question 1.1 (The direction in [1, Sec. 1.4])**.** For a finite probability space, what is the sharp generalized Samorodnitsky inequality when $1 < q < 2$, and which scalar functions replace the least-atom indicators from the range $q \geq 2$?

A related question concerns hypercontractivity: for given $1 < p < q < \infty$, how much noise is necessary to ensure $\|T_\rho f\|_q \leq \|f\|_p$ for every f? Write $\sigma_{p,q}(\mu)$ for the largest admissible ρ, and $\sigma_{p,q}(m)$ for the two-point law $(m, 1-m)$. Wolff's two-value reduction [12, Thm. 3.1] gives, in our order of exponents,

$$\inf_{a \in [m,1/2]} \sigma_{p,q}(a) \leq \sigma_{p,q}(\mu) \leq \sigma_{p,q}(m), \qquad m = \min_{x \in \Omega} \mu(x).$$

It leaves a comparison among the two-point masses. Cao, Fan, Han, Qiu, and Wang determine the biased two-point constants and prove equality for conjugate exponents [13, Thm. 1.6 and Cor. 1.7]. They pose the following conjecture immediately after Figure 3 on page 8.

Conjecture 1.2 ([13, Sec. 1.2, p. 8])**.** If μ is a finite probability law with least atom $m \in (0, 1/2)$, then $\sigma_{p,q}(\mu) = \sigma_{p,q}(m)$ for every $1 < p < q < \infty$.

1.2 Contribution

Both questions are governed by a comparison that keeps the mean and one convex moment fixed. Under a monotonicity condition on the ratio of two second derivatives, the other moment is largest when the high value lies on a single least atom (Theorem 3.1). This identifies the mass as well as the number of levels. For a least atom a, the resulting mean-one family is

$$h_s^a = 1 - s + \frac{s}{m} \mathbf{1}_{\{a\}}, \qquad 0 \leq s \leq 1. \tag{2}$$

It interpolates between a constant and the normalized indicator of a. The terms *soft* and *hard* refer respectively to $0 < s < 1$ and $s = 1$. The key comparison moves a high level onto a smaller mass while preserving the mean and the input moment; the noisy output moment then increases. Thus the least atom is selected before any optimization of the contrast. The main consequences are as follows.

(i) For $1 < q < 2$, the entire output q-moment profile at a fixed input q-moment is attained by (2). Consequently the sharp parameter in (1) is a maximum over the single interval $0 \leq s \leq 1$, depending on μ only through m (Theorems 4.3 and 5.1). Every nonconstant scalar extremizer belongs to this family.

(ii) For every $1 < p < q < \infty$ and every $0 \leq \rho \leq 1$, the full operator norm of $T_\rho : L^p(\mu) \to L^q(\mu)$ equals that on $(m, 1-m)$ (Theorem 6.1). This proves Conjecture 1.2, and the sharp noise parameter is strictly increasing in m (Corollary 6.2).

(iii) For every nonuniform μ, all product equality cases in (1) factor into scalar equality cases and constants (Theorem 5.3). For arbitrary μ, scalar extremizers are soft near $q = 1$, provided $m < 1/2$, and hard near $q = 2$. In the latter regime all product equality cases also factor (Theorems 7.4 and 7.5).

(iv) For the uniform binary law, Theorem 8.1 determines all scalar phases, with thresholds $(\sqrt{13} - 1)/2$ and $1 + (4 \log 2 - 1)^{-1}$, and classifies product equality throughout $1 < q < 2$.

No choice among subset masses remains. The unresolved scalar question for general $m < 1/2$ is the maximizing contrast s, including its uniqueness and transition curve. Product equality for a uniform alphabet with at least three symbols is not classified here outside the hard regime.

Lagrange multipliers and a one-sided boundary variation reduce a constrained maximizer to two levels. The comparison of their masses has a strictly negative covariance as its derivative. This supplies the least-atom conclusion without requiring a phase analysis. For hypercontractivity it applies directly when the input exponent is at most two; self-adjoint duality covers the remaining exponents.

1.3 Related work and scope

The sharp Samorodnitsky parameter is the *fixed-reference* Rényi contraction coefficient of the resampling channel. This interpretation already appears for $q \geq 2$ in [1, Thm. 3 and Rem. 20], whose Theorem 12 gives the scalar-to-product passage. Vandenbroucque, Esposito, and Gastpar extend that passage to arbitrary product channels [10, Thm. 10(a)]. The tensorization proof below is included to make the equality discussion self-contained. The scalar contribution is the fixed-moment comparison and its least-atom equality statement for $1 < q < 2$.

At the KL endpoint, Gu and Polyanskiy already give the sharp coefficient and a one-variable description for the uniform Potts channel [7, Thm. 4 and Props. 17–18]. Positivity of KL extremizers for strictly positive reversible kernels follows more generally from [11, Lem. 31], and continuity in the Rényi order at one is known for arbitrary finite channels [10, Cor. 1]. The least-atom comparison extends the scalar entropy formula to nonuniform reference laws.

For hypercontractivity, the two-value method and the reduction to a range of two-point masses are due to Wolff [12, Thm. 3.1]. The extra comparison here selects its smallest mass and identifies the norm even outside the hypercontractive region. The operator is precisely $\mathbb{E}_\mu + \rho(I - \mathbb{E}_\mu)$; the result makes no assertion about a different Fourier multiplier on a general cyclic group. The two-point constant $\sigma_{p,q}(1/2) = \sqrt{(p-1)/(q-1)}$ is classical; see [12, Eq. (2.1)].

A fixed reference law must be distinguished from optimization over both laws in a divergence ratio. The binary-input reductions of Ordentlich and Polyanskiy [6, Thms. 1–2], and their Rényi counterparts [5, Thm. 1] and [10, Thm. 6], concern the latter problem. A two-valued density relative to a prescribed μ can still have full support on an arbitrarily large alphabet. General contraction theory is surveyed in [4]; related Rényi bounds appear in [3, 5, 10], and constrained-input divergence bounds motivated by privacy amplification in [8].

The threshold $(\sqrt{13}-1)/2$ already occurs in the binary Rényi analysis of Song et al. [14, Lem. 2]. The Riccati representation and negative second derivative used below also appear in trureturing's paper-level analysis of Cachin Rényi information combining [15]. These ingredients are credited there; the additional third-derivative sign yields a one-valley property and the complete noisy-ratio phase diagram. No formal verification claim about that external development is needed for the proof given here.

2 Notation and the scalar inequality

Write $\mu^* = \min_{x\in\Omega} \mu(x)$. All expectations and L_q-norms without a displayed measure are taken with respect to μ, or to the appropriate product measure. For a function $h : \Omega \to \mathbb{R}$, set

$$(T_\rho h)(x) = \rho h(x) + (1-\rho)\mathbb{E}h. \tag{3}$$

The product operator $T_\rho^{[n]}$ is the tensor product of these one-coordinate operators. Equivalently, given $x \in \Omega^n$, it retains each coordinate independently with probability ρ and otherwise resamples that coordinate from μ, then averages the value of f at the resulting point.

For $S \subseteq [n]$, $\mathbb{E}(f \mid S)$ denotes the function of the coordinates in S obtained by integrating all coordinates outside S. We regard it as a function on Ω^n, constant in the integrated coordinates; this convention leaves its L_q-norm unchanged. The notation $S \sim \lambda$ means that each coordinate belongs to S independently with probability λ.

Theorem 1 of Abawonse, Hązła, and O'Donnell [1] treats $q \in [2, \infty]$. For $n = 1$ it proves

$$\|T_\rho h\|_q \leq \|h\|_1^{1-\lambda} \|h\|_q^{\lambda} \tag{4}$$

with a sharp constant depending only on μ^*; its nonconstant equality cases are precisely the indicators of atoms of mass μ^*. Theorem 12 of that paper supplies the induction from the scalar inequality to products, and Theorems 15 and 19 contain the precise product and scalar formulations, respectively. The fixed-reference Rényi formulation is made explicit in Theorem 3 and Remark 20.

The first item of [1, Sec. 1.4] isolates $1 < q < 2$ as an untreated direction. We retain the same operator, positivity assumption, and independent-subset formulation. For the variational characterization and tensorization, the standing regime is

$$2 \le |\Omega| < \infty, \quad \mu(x) > 0 \text{ for all } x, \qquad 1 < q < 2, \qquad 0 < \rho < 1. \tag{5}$$

Hypercontractive norms in Section 6 use separate exponents $1 < p < q < \infty$, explicitly replacing the restriction $q < 2$.

3 A fixed-moment least-atom principle

The following comparison is formulated independently of noise. Its strictness will also identify the extremizers in the applications.

Theorem 3.1 (Least-atom moment comparison)**.** *Let μ be fully supported on a finite set with at least two points, and let $m = \min_x \mu(x)$. Suppose $f, g \in C^1([0,\infty)) \cap C^2((0,\infty))$, $f'' > 0$, and g''/f'' is strictly increasing. For every $h \ge 0$ with $\mathbb{E}_\mu h = 1$, there is a unique $s \in [0,1]$ such that*

$$\mathbb{E}_\mu f(h) = mf\left(1 + \frac{1-m}{m}s\right) + (1-m)f(1-s). \tag{6}$$

For this s,

$$\mathbb{E}_\mu g(h) \le mg\left(1 + \frac{1-m}{m}s\right) + (1-m)g(1-s). \tag{7}$$

If $s > 0$, equality holds exactly for $h = h_s^a$ with a a least atom, as in (2). If $s = 0$, then $h = 1$.

Proof. Write $M = \mathbb{E}f(h)$. On the compact mean-one simplex the strictly convex functional $\mathbb{E}f(h)$ is minimized only by $h = 1$ and maximized at a vertex. The vertex supported on an atom of mass a has value

$$J(a) = af(1/a) + (1-a)f(0).$$

Strict convexity gives $J'(a) = f(1/a) - f(0) - f'(1/a)/a < 0$. Hence $f(1) \le M \le J(m)$. The right-hand side of (6) has these endpoint values, and its derivative for $s > 0$ is

$$(1-m)\left[f'\left(1 + \frac{1-m}{m}s\right) - f'(1-s)\right] > 0.$$

This proves existence and uniqueness of s. Strict convexity also shows that the upper endpoint is attained only at vertices of mass m. It remains to consider $f(1) < M < J(m)$.

First compare two-level variables without restricting their masses to subset masses of μ. Let the high value $x > 1$ have mass a, and the low value $0 \le y < 1$ have mass $1 - a$, with

$$ax + (1-a)y = 1, \qquad af(x) + (1-a)f(y) = M.$$

For a fixed a this pair is unique whenever it exists. It exists exactly when $M \le J(a)$. Thus a feasible pair at a can be continued to every smaller positive mass, in particular down to m when $a \ge m$. For $y > 0$, put $\Delta = x - y$, $F = f(x) - f(y)$, and $D = f'(x) - f'(y) > 0$. Differentiation of the two constraints gives

$$x' = -\frac{F - f'(y)\Delta}{aD}, \qquad y' = -\frac{f'(x)\Delta - F}{(1-a)D}. \tag{8}$$

For $N(a) = ag(x) + (1-a)g(y)$, substitution yields

$$\begin{aligned} N'(a) &= g(x) - g(y) - g'(y)\Delta - \frac{g'(x) - g'(y)}{D}(F - f'(y)\Delta) \\ &= D\ \mathrm{Cov}\left(x - U, \frac{g''(U)}{f''(U)}\right) < 0, \end{aligned}$$

where U has density $f''(u)/D$ on (y,x). The second identity uses $F - f'(y)\Delta = \int_y^x (x-u)f''(u)\,\mathrm{d}u$ and its analogue for g. For independent copies U, V, twice the covariance is

$$\mathbb{E}\left[(V-U)\left(\frac{g''(U)}{f''(U)} - \frac{g''(V)}{f''(V)}\right)\right] < 0.$$

The strict sign follows because the ratio increases strictly and the density is positive. A pair with $y = 0$ is an endpoint of this comparison path; continuity of f, g and their first derivatives extends the strict comparison to that endpoint. In particular, decreasing the high-level mass strictly increases N.

Now maximize $\mathbb{E}g(h)$ subject to the prescribed mean and f-moment. Compactness gives a maximizer. If all its positive coordinates are equal, it is a normalized subset indicator, and the two-level comparison just proved applies. Otherwise two positive coordinates are distinct. The gradients of the two constraints are then independent on the positive support. Lagrange multipliers give constants λ, β with

$$\psi(h_i) = \beta \quad (h_i > 0), \qquad \psi(t) = g'(t) - \lambda f'(t).$$

On $(0,\infty)$, $\psi'(t) = f''(t)(g''(t)/f''(t) - \lambda)$. Its sign can change at most once, from negative to positive. Thus a horizontal line meets ψ in at most two positive points. Since two such points occur, ψ has a strictly decreasing branch followed by a strictly increasing branch, and the level β is strictly below $\psi(0)$.

There can be no zero coordinate. Indeed, activate a zero coordinate i to a small positive value ε, and adjust two distinct positive coordinates to keep both constraints fixed. The implicit function theorem applies because their constraint Jacobian has nonzero determinant. The one-sided objective derivative is $\mu_i(\psi(0) - \beta)$; maximality requires it to be nonpositive, contradicting $\beta < \psi(0)$. The maximizer is therefore positive and two-valued.

In either case the mass a of its high level is a sum of atom masses, so $a \geq m$. If $a > m$, moving to the two-level profile of mass m with the same constraints strictly increases the objective. This is a feasible function on Ω, supported at its high level on any least atom, and contradicts maximality. Hence $a = m$, which consists of one atom. The constraint determines its contrast to be s. This proves both the inequality and its equality cases. □

4 The one-coordinate coefficient

Definition 4.1 (The scalar ratio)**.** Let

$$\mathcal{C} = \{h : \Omega \to [0,\infty) : \mathbb{E}h = 1\}.$$

For nonconstant $h \in \mathcal{C}$, put

$$A(h) = \mathbb{E}h^q, \qquad B(h) = \mathbb{E}(1 - \rho + \rho h)^q, \qquad R_q(h) = \frac{\log B(h)}{\log A(h)}.$$

At the constant density 1, define $R_q(1) = \rho^2$, the continuous quadratic perturbation value. Finally set

$$\lambda_*(q,\mu,\rho) = \max_{h \in \mathcal{C}} R_q(h).$$

Lemma 4.2 (Existence and nonlocality)**.** *For $1 < q < 2$ and $0 < \rho < 1$, the maximum in Definition 4.1 exists and satisfies $0 < \lambda_* < 1$. If $|\Omega| > 2$, then $\lambda_* > \rho^2$. More generally, the latter conclusion holds whenever μ has a subset of mass strictly below $1/2$.*

Proof. Since μ is fully supported, $0 \le h(x) \le 1/\mu(x)$ for every $h \in \mathcal{C}$. Hence $\mathcal{C}$ is compact. If $h = 1 + u$, $\mathbb{E}u = 0$, and $\|u\|_\infty \to 0$, Taylor expansion gives

$$\log \mathbb{E}(1+u)^q = \frac{q(q-1)}{2}\mathbb{E}u^2 + O(\|u\|_\infty \mathbb{E}u^2)$$

and the same formula with u replaced by ρu for the numerator. Thus $R_q(h) \to \rho^2$ as $h \to 1$, so the definition at 1 makes R_q continuous on $\mathcal{C}$. Compactness gives existence.

For nonconstant h, Jensen gives $A(h) > 1$. Since $\rho > 0$, the function $1 - \rho + \rho h$ is also nonconstant and has mean one, so strict Jensen gives $B(h) > 1$. Also

$$1 - \rho + \rho h = (1-\rho)\cdot 1 + \rho h,$$

and strict convexity of x^q gives

$$B(h) < (1-\rho) + \rho A(h) < A(h).$$

Thus $0 < R_q(h) < 1$, and $R_q(1) = \rho^2 < 1$.

Assume that $A \subset \Omega$ has mass $a = \mu(A) < 1/2$, and put

$$g = \mathbf{1}_A - \frac{a}{1-a}\mathbf{1}_{A^c}.$$

Then $\mathbb{E}g = 0$, $\mathbb{E}g^2 = a/(1-a)$, and $\mathbb{E}g^3 = a(1-2a)/(1-a)^2 > 0$. For small $t > 0$, $h_t = 1 + tg$ lies in $\mathcal{C}$. Writing $L(t) = \log \mathbb{E}(1+tg)^q$, expansion gives

$$L(t) = \frac{q(q-1)}{2}t^2\mathbb{E}g^2 + \frac{q(q-1)(q-2)}{6}t^3\mathbb{E}g^3 + O(t^4).$$

Therefore

$$R_q(h_t) = \rho^2\left(1 + \frac{q-2}{3}\frac{\mathbb{E}g^3}{\mathbb{E}g^2}(\rho-1)t + O(t^2)\right).$$

Since $q - 2 < 0$, $\rho - 1 < 0$, and $\mathbb{E}g^3 > 0$, the coefficient of t is positive. Thus $R_q(h_t) > \rho^2$ for all sufficiently small positive t. □

Theorem 4.3 (Sharp moment profile and scalar parameter)**.** *Let $m = \min_x \mu(x)$, $1 < q < 2$, and $0 < \rho < 1$. Define*

$$M_{m,q}(s) = m\left(1 + \frac{1-m}{m}s\right)^q + (1-m)(1-s)^q, \qquad 0 \le s \le 1. \tag{9}$$

If $\mathbb{E}h = 1$ and $\mathbb{E}h^q = M_{m,q}(s)$, then

$$\mathbb{E}(T_\rho h)^q \le M_{m,q}(\rho s). \tag{10}$$

For $s > 0$, equality holds exactly when $h = h_s^a$ for a least atom a. In particular,

$$\lambda_*(q,\mu,\rho) = \max_{0\le s\le 1}\frac{\log M_{m,q}(\rho s)}{\log M_{m,q}(s)}, \tag{11}$$

where the ratio at zero is ρ^2.

Proof. Apply Theorem 3.1 with $f(t) = t^q$ and $g(t) = (1-\rho+\rho t)^q$. Its decisive hypothesis is

$$\frac{g''(t)}{f''(t)} = \rho^2\left(\frac{t}{1-\rho+\rho t}\right)^{2-q},$$

which is strictly increasing for $t > 0$. The function h_s^a is taken by T_ρ to $h_{\rho s}^a$, giving (10) and its equality statement. Every mean-one density has a unique input moment parameter s, while every $s \in [0,1]$ is attained by h_s^a. Dividing logarithms proves (11), with the value at zero supplied by Lemma 4.2. □

Corollary 4.4 (Scalar extremizers)**.** *Every nonconstant scalar equality case, normalized to mean one, is h_s^a, where a is a least atom and $s > 0$ maximizes (11). Thus its high value occurs on one least atom and its low value is common to all other atoms. Constants are always scalar equality cases, whether or not the ratio at zero is maximal.*

Proof. At a fixed nonconstant input moment the denominator of the scalar ratio is fixed and positive. A scalar equality case must therefore attain the sharp output moment in (10). Conversely each maximizing profile attains the ratio. The constant equality follows directly from $T_\rho 1 = 1$. □

5 Sharp product tensorization

The scalar-to-product passage itself is already contained in the general induction theorem of Abawonse, Hązła, and O'Donnell [1, Theorem 12]. In their notation, take the target exponent $Q \in (1, 2)$, set $r = 0$ and $p = q = s = Q$, use deterministic resampling parameters $\rho_i = \rho$ on the left, and independent Bernoulli inclusion variables on the right. The direct Hölder proof below is retained for self-containedness; the new input is the exact scalar coefficient and its extremizers.

For probability measures $\nu \ll \mu$, the order-q Rényi divergence is

$$D_q(\nu\|\mu) = \frac{1}{q-1} \log \mathbb{E}_\mu \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right)^q. \tag{12}$$

For a channel K and a fixed reference input law π, define the *fixed-reference* coefficient

$$\eta_q(\pi, K) = \sup_{\substack{\nu \neq \pi \\ 0 < D_q(\nu\|\pi) < \infty}} \frac{D_q(\nu K\|\pi K)}{D_q(\nu\|\pi)}. \tag{13}$$

The channel-level coefficient

$$\eta_q(K) = \sup_\pi \eta_q(\pi, K)$$

also optimizes the reference law. Throughout this paper the reference law remains μ. General properties and bounds for these two coefficients are discussed in [3, 5, 10]; their distinction is explained in the introduction. For the kernel below, stationarity gives $\mu K = \mu$, and the density of νK is $T_\rho h$ when $h = \mathrm{d}\nu/\mathrm{d}\mu$.

Theorem 5.1 (Sharp product inequality)**.** *Let $1 < q < 2$, $0 < \rho < 1$, and $\lambda \in [0, 1]$. The product inequality*

$$\log \|T_\rho^{[n]} f\|_q \leq \mathbb{E}_{S \sim \lambda} \log \|\mathbb{E}(f \mid S)\|_q$$

holds for every $n \geq 1$ and every nonzero nonnegative $f : \Omega^n \to \mathbb{R}$ if and only if

$$\lambda \geq \lambda_*(q, \mu, \rho),$$

where λ_* is given by Theorem 4.3. Equivalently, $\lambda_* = \eta_q(\mu, K)$ is the fixed-reference order-q Rényi contraction coefficient of the one-coordinate channel

$$K(x, y) = \rho \mathbf{1}_{x=y} + (1 - \rho)\mu(y).$$

Proof. The one-coordinate inequality is

$$\log \|T_\rho h\|_q \leq \lambda \log \|h\|_q + (1 - \lambda) \log \mathbb{E}h$$

for all nonzero $h \geq 0$. By homogeneity, with $u = h/\mathbb{E}h$, this is equivalent to

$$\log B(u) \leq \lambda \log A(u)$$

for all $u \in \mathcal{C}$. Hence it holds if and only if $\lambda \geq \lambda_*$. If $\mathrm{d}\nu/\,\mathrm{d}\mu = h$, then $\mu K = \mu$, $\mathrm{d}(\nu K)/\,\mathrm{d}\mu = 1 - \rho + \rho h$, and the same quotient is

$$\frac{D_q(\nu K \| \mu)}{D_q(\nu \| \mu)}.$$

If $\lambda = 1$, the asserted product inequality is simply the L_q-contraction $\|T_\rho^{[n]} f\|_q \leq \|f\|_q$. We may therefore assume $\lambda < 1$. It remains to prove that the one-coordinate inequality tensorizes. Let (Y, ν) be another finite probability space and let $F : \Omega \times Y \to [0, \infty)$. Write

$$m(y) = \mathbb{E}_x F(x, y).$$

For $m(y) > 0$, put $h_y(x) = F(x, y)/m(y)$, and ignore zero fibers. Then $\mathbb{E}_x h_y = 1$ and

$$T_\rho^{(x)} F(x, y) = m(y)(1 - \rho + \rho h_y(x)).$$

If $A_y = \mathbb{E}_x h_y^q$, the scalar inequality gives

$$\|T_\rho^{(x)} F\|_q^q \leq \mathbb{E}_y m(y)^q A_y^\lambda.$$

Hölder's inequality, in the form

$$m^q A_y^\lambda = (m^q A_y)^\lambda (m^q)^{1-\lambda},$$

gives

$$\mathbb{E}_y m(y)^q A_y^\lambda \leq \left(\mathbb{E}_y m(y)^q A_y\right)^\lambda \left(\mathbb{E}_y m(y)^q\right)^{1-\lambda}.$$

But $\mathbb{E}_y m(y)^q A_y = \|F\|_q^q$ and $\mathbb{E}_y m(y)^q = \|\mathbb{E}_x F\|_q^q$. Taking logarithms and dividing by q,

$$\log \|T_\rho^{(x)} F\|_q \leq \lambda \log \|F\|_q + (1 - \lambda) \log \|\mathbb{E}_x F\|_q.$$

The induction uses the following auxiliary-space statement. For every finite probability space (Y, ν) and every nonnegative $F : \Omega^n \times Y \to \mathbb{R}$,

$$\log \|(T_\rho^{[n]} \otimes I_Y) F\|_{L_q(\mu^n \otimes \nu)} \leq \mathbb{E}_{S \sim \lambda} \log \|\mathbb{E}(F \mid S, Y)\|_{L_q(\mu^n \otimes \nu)}, \tag{14}$$

where $\mathbb{E}(F \mid S, Y)$ integrates the Ω-coordinates outside S and retains the auxiliary variable. The case $n = 0$ is equality, and the one-coordinate estimate above proves the induction step with all unprocessed coordinates and Y carried as the auxiliary space.

Indeed, if the last coordinate is handled after the first $n - 1$, the induction gives

$$\log \|T_\rho^{[n]} f\|_q \leq \lambda \mathbb{E}_R \log \|\mathbb{E}(f \mid R \cup \{n\})\|_q + (1 - \lambda) \mathbb{E}_R \log \|\mathbb{E}(f \mid R)\|_q,$$

where $R \subseteq [n - 1]$ includes each coordinate independently with probability λ. This is exactly the expectation over $S \sim \lambda$. Necessity follows from $n = 1$. □

Lemma 5.2 (Equality in one tensorization step)**.** *Let $\lambda = \lambda_*(q, \mu, \rho)$, let (Y, ν) be finite and fully supported, and let $F : \Omega \times Y \to [0, \infty)$ be nonzero. Put $m(y) = \mathbb{E}_x F(x, y)$, and for $m(y) > 0$ put*

$$h_y(x) = \frac{F(x, y)}{m(y)}, \qquad A_y = \mathbb{E}_x h_y^q.$$

Equality holds in

$$\log \|T_\rho^{(x)} F\|_q = \lambda \log \|F\|_q + (1 - \lambda) \log \|\mathbb{E}_x F\|_q \tag{15}$$

if and only if both of the following conditions hold:

(a) *every positive normalized fiber h_y is a scalar equality case;*

(b) A_y *is constant on* $\{y : m(y) > 0\}$.

If one positive fiber is constant, then all positive fibers are constant.

Proof. Write $B_y = \mathbb{E}_x(1-\rho+\rho h_y)^q$. The two inequalities in the one-step proof are

$$\mathbb{E}_y m(y)^q B_y \le \mathbb{E}_y m(y)^q A_y^\lambda, \tag{16}$$

$$\mathbb{E}_y m(y)^q A_y^\lambda \le \left(\mathbb{E}_y m(y)^q A_y\right)^\lambda \left(\mathbb{E}_y m(y)^q\right)^{1-\lambda}. \tag{17}$$

All summands in the difference in (16) are nonnegative, and $m(y)^q > 0$ precisely on the positive fibers. Equality therefore holds there exactly when $B_y = A_y^\lambda$ for every positive fiber, which is condition (a).

For (17), Hölder is applied to $U(y) = m(y)^q A_y$ and $V(y) = m(y)^q$ with exponents $1/\lambda$ and $1/(1-\lambda)$. Since $0 < \lambda < 1$, its equality condition is that U and V be proportional on their common positive support. This is exactly the constancy of A_y, condition (b). Both intermediate inequalities must be equalities if and only if the final one-step estimate is an equality.

Finally, a constant mean-one fiber has $A_y = 1$. Condition (b) then gives $A_y = 1$ on every positive fiber; strict convexity of $r \mapsto r^q$ and $\mathbb{E}h_y = 1$ force every such h_y to be constant. □

Theorem 5.3 (Product equality for nonuniform laws)**.** *Suppose μ is nonuniform. For every $1 < q < 2$ and $0 < \rho < 1$, the nonzero equality cases in the sharp product inequality are precisely*

$$f(x_1, \dots, x_n) = c \prod_{i=1}^n h_i(x_i), \qquad c > 0,$$

where each h_i is either 1 or a mean-one scalar extremizer from Corollary 4.4. Uniqueness of the maximizing contrast in (11) is not required.

Proof. Fix one coordinate x, write the others as y, and let U be product noise on y. Two orders of tensorization show that equality forces the one-step equality of Lemma 5.2 both for $F = f$ and for $F = Uf$. To see the first assertion, apply (14) in the y-coordinates to $T_\rho^{(x)} f$, with x as auxiliary space, and then apply the x-step to every conditional expectation in the resulting average. The term retaining all y-coordinates has weight $\lambda^{n-1} > 0$ and contains the x-step on raw f. For the second assertion, apply the x-step first to Uf, and then tensorize the two resulting terms in y. Both chains end at the same sharp product bound. Every gap is nonnegative and $0 < \lambda < 1$, so equality forces the stated steps.

Put $b(y) = \mathbb{E}_x f(x, y)$. If one positive normalized raw fiber is constant, Lemma 5.2 makes them all constant, and $f(x, y) = b(y)$. Otherwise all positive raw fibers are nonconstant scalar extremizers with a common q-moment. Since $M_{m,q}$ is strictly increasing on $[0, 1]$, they have a common contrast $s > 0$. Thus

$$f(x, y) = b(y)\left(1 - s + \frac{s}{m}\mathbf{1}_{\{x=a(y)\}}\right), \qquad b(y) > 0,$$

where each $a(y)$ is a least atom. A nonuniform law has an atom x_0 with $\mu(x_0) > m$. At that atom,

$$f(x_0, y) = (1-s)b(y), \qquad (Uf)(x_0, y) = (1-s)(Ub)(y).$$

The kernel of U is strictly positive and $b \not\equiv 0$, so $(Ub)(y) > 0$. Every normalized Uf fiber is a scalar equality case. Its value at x_0 is $1 - s < 1$, excluding a constant fiber; as x_0 cannot be its peak, this value also forces its contrast to be exactly s.

Each normalized Uf fiber is a convex combination of the nonzero raw normalized fibers, with strictly positive weights

$$\frac{U(y, y')b(y')}{(Ub)(y)}.$$

A mixture of the vectors $1 - s + (s/m)\mathbf{1}_{\{a(y')\}}$ can itself equal such a vector with the same $s > 0$ only when every peak label $a(y')$ is the same. Indeed, after subtracting $1 - s$ and multiplying by m/s, a convex combination of point indicators is a point indicator only in that case. Hence $f(x, y) = h_s^a(x)b(y)$.

Both sides of the product inequality add under products. The factor h_s^a is a scalar equality case, so equality for f is equivalent to equality for b in one fewer coordinate. Induction proves necessity, and the same factorization verifies the converse. □

6 Hypercontractive norms

This section allows arbitrary $1 < p < q < \infty$. The operator remains $T_\rho = \rho I + (1 - \rho)\mathbb{E}_\mu$, and $\nu_m = (m, 1 - m)$ denotes the two-point probability law.

Theorem 6.1 (Exact least-atom norm identity)**.** *For every finite full-support law μ with at least two atoms, $m = \min_x \mu(x)$, $1 < p < q < \infty$, and $0 \le \rho \le 1$,*

$$\|T_\rho\|_{L^p(\mu)\to L^q(\mu)} = \|T_\rho\|_{L^p(\nu_m)\to L^q(\nu_m)}. \tag{18}$$

For $0 < \rho < 1$, every nonconstant nonnegative norm extremizer, normalized to mean one, is h_s^a for a least atom a and some $0 < s < 1$. In particular, with $M_{m,r}$ defined by (9) for any $r > 1$,

$$\|T_\rho\|_{p\to q} = \max_{0\le s\le 1} \frac{M_{m,q}(\rho s)^{1/q}}{M_{m,p}(s)^{1/p}}. \tag{19}$$

The identity holds for real or complex operator norms.

Proof. Assume first $0 < \rho < 1$ and $p \le 2$. Positivity gives $|T_\rho f| \le T_\rho |f|$, so nonnegative functions suffice even for the complex norm. Normalize a nonzero such function to $\mathbb{E}f = 1$. In Theorem 3.1, use the functions $F(t) = t^p$ and $G(t) = (1 - \rho + \rho t)^q$. Their derivative ratio is positive, and

$$\frac{G''(t)}{F''(t)} = \frac{q(q-1)\rho^2}{p(p-1)}(1 - \rho + \rho t)^{q-2}t^{2-p},$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\log\frac{G''(t)}{F''(t)} = \frac{(q-p)\rho t + (2-p)(1-\rho)}{t(1-\rho+\rho t)} > 0.$$

The theorem replaces f by h_s^a with the same p-moment and no smaller noisy q-moment. Its norm ratio is that of a function on ν_m, proving the upper bound in (18). Conversely every function on ν_m embeds into Ω by using one least atom and its complement; this embedding preserves the two norms and commutes with T_ρ. It gives the reverse bound. Strictness in the moment comparison also gives the asserted two-level shape of every nonconstant nonnegative extremizer in this range.

If $p > 2$, let $p' = p/(p-1)$ and $q' = q/(q-1)$. Self-adjointness for the μ pairing and finite-dimensional duality give

$$\|T_\rho\|_{p\to q} = \|T_\rho\|_{q'\to p'}.$$

Here $1 < q' < p' < 2$, so the case just proved applies, first on μ and then on ν_m, proving (18). To transfer the extremizer statement, normalize $f \ge 0$ by $\|f\|_p = 1$ and take $g = (T_\rho f)^{q-1}/\|T_\rho f\|_q^{q-1}$. Then $\|g\|_{q'} = 1$ and

$$\|T_\rho\|_{p\to q} = \langle T_\rho f, g\rangle = \langle f, T_\rho g\rangle \le \|T_\rho g\|_{p'} \le \|T_\rho\|_{q'\to p'}.$$

Every inequality is an equality. Thus g is a dual extremizer and Hölder equality gives f proportional to $(T_\rho g)^{p'-1}$. If f is nonconstant, then g is nonconstant, since T_ρ is invertible. The dual result makes g, and hence f, two-valued with its high level on one least atom.

Every nonconstant nonnegative norm extremizer is strictly positive. If $f_i = 0$, replacing it by $\varepsilon > 0$ changes $\|f\|_p^p$ by $\mu_i\varepsilon^p = o(\varepsilon)$. The strictly positive kernel gives $T_\rho f > 0$, and the

corresponding change in $\|T_\rho f\|_q^q$ has strictly positive first derivative. This would increase the norm ratio, a contradiction. Therefore its contrast is less than one. Constants and the preceding extremizer classification prove (19). Finally, all norms in question are continuous in ρ, which includes $\rho = 0, 1$. □

Corollary 6.2 (Sharp boundary and its dependence on the least atom)**.** *For every* $1 < p < q < \infty$,

$$\sigma_{p,q}(\mu) = \sigma_{p,q}(m).$$

The function $m \mapsto \sigma_{p,q}(m)$ *is strictly increasing on* $(0, 1/2]$. *Thus Conjecture 1.2 holds.*

Proof. The identity follows immediately from (18). For strict monotonicity, duality reduces the proof to $p \le 2$. Fix $0 < m < 1/2$, put $Z = m^{-1}\mathbf{1}_{\{a\}} - 1$ on ν_m, and set $\rho_0 = \sqrt{(p-1)/(q-1)}$. Then $\mathbb{E}Z = 0$ and $\mathbb{E}Z^3 > 0$. Taylor expansion at the constant function gives

$$\log \|1 + \rho t Z\|_q - \log \|1 + tZ\|_p = \frac{(q-1)\rho^2 - (p-1)}{2}\mathbb{E}Z^2\, t^2 + \frac{(q-1)(q-2)\rho^3 - (p-1)(p-2)}{6}\mathbb{E}Z^3\, t^3 + O(t^4). \tag{20}$$

At $\rho = \rho_0$, the quadratic term vanishes, while the cubic bracket is

$$(p-1)\big[\rho_0(q-2) - (p-2)\big] > 0.$$

For $q \ge 2$ its positivity follows from $p \le 2$; for $q < 2$, $\rho_0(q-2) > q - 2 > p - 2$. A small positive t therefore violates hypercontractivity at ρ_0, and continuity gives $\sigma_{p,q}(m) < \rho_0$.

There is a nonconstant critical extremizer at $\sigma = \sigma_{p,q}(m) > 0$. Here are the required compactness details. Normalize nonnegative inputs to mean one, forming a compact interval. For sufficiently small ρ, the strictly negative quadratic term in (20) handles a neighborhood of the constant, uniformly in the two signs of t. On the compact complement, $\|T_0 f\|_q = 1 < \|f\|_p$, and continuity preserves this strict inequality for small ρ. Hence $\sigma > 0$. The admissible noise parameters form a closed interval: if $\rho' < \rho$, then $T_{\rho'} = T_{\rho'/\rho}T_\rho$ and the first factor is an L^q-contraction. Choose violating maximizers at $\rho_j \downarrow \sigma$. A subsequence converges to an input attaining norm ratio one at σ. Since $\sigma < \rho_0$, the quadratic term is uniformly negative near the constant and near σ; the violating maximizers cannot approach the constant. Their limit is therefore nonconstant.

Now take $m_1 < m_2 < 1/2$ and a nonconstant critical extremizer for m_2. By Theorem 6.1 its high level has mass m_2. Decrease that mass to m_1 while preserving its mean and p-moment. The strict comparison in the proof of Theorem 3.1, with $F(t) = t^p$ and $G(t) = (1 - \sigma + \sigma t)^q$, increases the noisy q-moment. Thus the norm on ν_{m_1} already exceeds one at $\rho = \sigma_{p,q}(m_2)$; by continuity it does so slightly earlier. This proves $\sigma_{p,q}(m_1) < \sigma_{p,q}(m_2)$. At $m_2 = 1/2$, the classical value is ρ_0 [12, Eq. (2.1)], and the strict inequality for biased laws was proved above. Duality completes the proof. □

7 Endpoint regimes

7.1 The $q \downarrow 1$ soft-dictator regime

For a mean-one density h, define

$$D_q(h) = \frac{1}{q-1}\log \mathbb{E}h^q, \qquad q > 1,$$

and

$$D_1(h) = \mathrm{Ent}(h) = \mathbb{E}h \log h,$$

with $0 \log 0 = 0$. Let $Th = 1 - \rho + \rho h$ and set

$$\eta_1 = \max_{h \in \mathcal{C}} \frac{\mathrm{Ent}(Th)}{\mathrm{Ent}(h)},$$

with the constant density assigned the limiting value ρ^2. This is the ordinary KL strong data-processing coefficient for the resampling kernel; see Polyanskiy and Wu [4] for general background on KL contraction coefficients. For uniform μ, this kernel is the Potts semigroup, and Gu and Polyanskiy [7, Thm. 4 and Props. 17–18] already obtained an exact one-variable formula and the corresponding one-coordinate entropy extremizers. Proposition 7.3 extends that endpoint description to general fully supported laws having a subset of mass below $1/2$, in particular to every nonbinary law. Continuity of the coefficient itself at $q = 1$ follows from the more general result [10, Cor. 1]. Positivity of KL maximizers for this strictly positive reversible kernel is also a special case of [11, Lem. 31]. The direct arguments below additionally control the ratios uniformly and use the two-level classification to obtain persistence of interior extremizers in Theorem 7.4.

Lemma 7.1 (Uniform KL limit)**.** *As $q \downarrow 1$, $R_q \to R_1$ uniformly on $\mathcal{C}$, where*

$$R_1(h) = \frac{\mathrm{Ent}(Th)}{\mathrm{Ent}(h)}.$$

Consequently $\lambda_*(q, \mu, \rho) \to \eta_1$, and every subsequential limit of R_q-maximizers is a KL maximizer.

Proof. Let $M = 1/\mu^*$. Uniformly for $x \in [0, M]$,

$$x^{1+\varepsilon} = x + \varepsilon x \log x + O(\varepsilon^2),$$

so $D_{1+\varepsilon}(h) \to \mathrm{Ent}(h)$ uniformly on $\mathcal{C}$; the same holds for Th.

Ratios near constants require a separate argument. Fix $q_+ \in (1, 2)$. If $h = 1 + u$, $\mathbb{E}u = 0$, and $\|u\|_\infty \le s$, Taylor's formula with a remainder bounded on the finite coordinate range gives, uniformly for $q \in (1, q_+]$,

$$D_q(1 + u) = \frac{q}{2}\mathbb{E}u^2 + O(s\mathbb{E}u^2), \qquad D_q(1 + \rho u) = \frac{q}{2}\rho^2\mathbb{E}u^2 + O(s\mathbb{E}u^2).$$

The same expansion at $q = 1$ follows from $(1 + x)\log(1 + x) = x + x^2/2 + O(|x|^3)$. Thus $R_q(1 + u) = \rho^2 + O(s)$ and $R_1(1 + u) = \rho^2 + O(s)$, with constants independent of q.

For the complementary region, put

$$\mathcal{K}_s = \{h \in \mathcal{C} : \|h - 1\|_\infty \ge s\}.$$

It is compact and contains no constant density, so $d_s = \min_{h \in \mathcal{K}_s} \mathrm{Ent}(h) > 0$. Uniform convergence of $D_q(h)$ and $D_q(Th)$ to their KL limits implies that, for q close enough to 1, $D_q(h) \ge d_s/2$ throughout $\mathcal{K}_s$. Division is therefore uniform there, and

$$\sup_{h \in \mathcal{K}_s} |R_q(h) - R_1(h)| \longrightarrow 0.$$

Given an error tolerance, first choose s so that the two quadratic-region errors are smaller than half that tolerance, and then choose q close enough to 1 for the compact-complement error. This proves uniform convergence on all of $\mathcal{C}$. Uniform convergence on the compact set gives convergence of maxima and the usual compactness statement for maximizers. □

Lemma 7.2 (KL maximizers are soft)**.** *Assume that μ has a subset of mass strictly below $1/2$. Then $\eta_1 > \rho^2$, and every nonconstant KL maximizer is strictly positive and takes exactly two values.*

Proof. The perturbation from Lemma 4.2, with entropy in place of $\log L_q$, gives $\eta_1 > \rho^2$. Indeed,

$$\operatorname{Ent}(1+tg) = \frac{t^2}{2}\mathbb{E}g^2 - \frac{t^3}{6}\mathbb{E}g^3 + O(t^4)$$

and the numerator has t replaced by ρt, so the quotient is

$$\rho^2\left(1 + \frac{(1-\rho)\mathbb{E}g^3}{3\mathbb{E}g^2}t + O(t^2)\right) > \rho^2$$

for small positive t.

Let h maximize the KL ratio and put $\lambda = \eta_1$. Then

$$\Phi(g) = \operatorname{Ent}(Tg) - \lambda \operatorname{Ent}(g)$$

is at most 0 on $\mathcal{C}$ and equals 0 at h. Suppose $h_i = 0$. Choose j with $h_j > 0$. For sufficiently small $s > 0$, define the weighted mass transfer

$$h_s(i) = \frac{s}{\mu_i}, \qquad h_s(j) = h_j - \frac{s}{\mu_j}, \qquad h_s(k) = h_k \quad (k \notin \{i,j\}).$$

Then $h_s \geq 0$ and $\mathbb{E}h_s = \mathbb{E}h = 1$. At the newly positive coordinate,

$$\mu_i h_s(i)\log h_s(i) = s\log s - s\log\mu_i,$$

while the change at coordinate j is $O(s)$. Consequently $\operatorname{Ent}(h_s) - \operatorname{Ent}(h) = s\log s + O(s)$. Since $Th \geq 1-\rho > 0$ coordinatewise and $T(h_s - h) = O(s)$, differentiability of $r \mapsto r\log r$ on this positive range gives $\operatorname{Ent}(Th_s) - \operatorname{Ent}(Th) = O(s)$. Hence

$$\Phi(h_s) - \Phi(h) = -\lambda s\log s + O(s) > 0$$

for small $s > 0$, a contradiction. Thus $h > 0$.

The Lagrange equation on the relative interior says that

$$\rho(\log(1-\rho+\rho r)+1) - \lambda(\log r + 1) = c$$

for each value $r = h(x)$. The derivative has sign determined by

$$\frac{\rho^2 r}{1-\rho+\rho r} - \lambda,$$

which is a strictly increasing function of r. Therefore the left side first decreases and then possibly increases, so a horizontal line has at most two positive intersections. Since h is not constant, it has exactly two positive values. □

Proposition 7.3 (Least-atom KL formula)**.** *For $m = \min_x \mu(x)$, define*

$$K_m(s) = m\left(1 + \frac{1-m}{m}s\right)\log\left(1 + \frac{1-m}{m}s\right) + (1-m)(1-s)\log(1-s), \qquad 0 \leq s \leq 1.$$

Then

$$\eta_1 = \max_{0\leq s\leq 1} \frac{K_m(\rho s)}{K_m(s)}, \tag{21}$$

with value ρ^2 at zero. If $m < 1/2$, every nonconstant KL extremizer is h_s^a for a least atom a and some $0 < s < 1$.

Proof. For a fixed density h, choose s_q such that $M_{m,q}(s_q) = \mathbb{E}h^q$. Uniformly on $[0,1]$,

$$\frac{M_{m,q}(s) - 1}{q-1} \longrightarrow K_m(s).$$

The function K_m is strictly increasing from 0 to $\log(1/m)$. Compactness and the preceding limit therefore imply that $s_q \to s$, where $K_m(s) = \mathrm{Ent}(h)$. Subtracting one from (10), dividing by $q-1$, and taking the limit yields $\mathrm{Ent}(T_\rho h) \le K_m(\rho s)$. Every s is attained by h_s^a, proving (21).

For $m < 1/2$, Lemma 7.2 makes every nonconstant KL extremizer positive and two-valued. Apply the fixed-moment comparison from the proof of Theorem 3.1 with $f(t) = t\log t$ and $g(t) = (1-\rho+\rho t)\log(1-\rho+\rho t)$. The high-mass comparison path stays strictly positive when its starting low value is positive, so differentiability at zero is not needed. On that path,

$$\frac{g''(t)}{f''(t)} = \frac{\rho^2 t}{1-\rho+\rho t}$$

is strictly increasing. Reducing any high-level mass larger than m would strictly improve its output entropy at fixed input entropy. Thus its high level has mass m, and positivity gives $s < 1$. □

Theorem 7.4 (Uniform soft dictators near $q = 1$)**.** *If μ has a subset of mass strictly below $1/2$, then there are $\delta > 0$ and $q_1 > 1$ such that, for every $q \in (1, q_1)$, every one-coordinate maximizer h_q of R_q satisfies*

$$h_q(x) \ge \delta \qquad \text{for all } x \in \Omega.$$

Consequently every nonconstant one-coordinate maximizer in this range is a strictly positive two-level soft dictator.

Proof. Let $\mathcal{K}_1$ be the compact set of KL maximizers. By Lemma 7.2, every element of $\mathcal{K}_1$ is strictly positive, so

$$\delta_0 = \min_{h\in\mathcal{K}_1}\min_{x\in\Omega} h(x) > 0.$$

If a sequence $q_j \downarrow 1$ had maximizers staying at sup-norm distance at least $\delta_0/2$ from $\mathcal{K}_1$, compactness would give a subsequence converging to a KL maximizer by Lemma 7.1, a contradiction. Thus all maximizers for q close enough to 1 lie within $\delta_0/2$ of $\mathcal{K}_1$, and are bounded below by $\delta_0/2$. Corollary 4.4 then rules out subset indicators and gives the two-level soft form. □

7.2 The $q \uparrow 2$ least-atom regime

Theorem 7.5 (Least-atom dictators near $q = 2$)**.** *Let $m = \mu^*$. There is $q_0 = q_0(\mu,\rho) \in (1,2)$ such that, for every $q \in (q_0, 2)$,*

$$\lambda_*(q,\mu,\rho) = \frac{\log\left[m(1-\rho+\rho/m)^q + (1-m)(1-\rho)^q\right]}{(q-1)\log(1/m)}.$$

In this interval the only nonconstant one-coordinate maximizers are

$$h = \frac{\mathbf{1}_{\{x\}}}{m}, \qquad \mu(x) = m.$$

Moreover, equality in the sharp product inequality on Ω^n holds exactly for functions

$$f(x_1,\dots,x_n) = c\prod_{i\in I}\frac{\mathbf{1}_{\{x_i=a_i\}}}{m},$$

where $c > 0$, $I \subseteq [n]$, and each a_i is a least atom.

Proof. At $q = 2$,

$$R_2(h) = \frac{\log(1 + \rho^2 \operatorname{Var}(h))}{\log(1 + \operatorname{Var}(h))}.$$

For $0 < c < 1$, the function $z \mapsto \log(1 + cz)/\log(1 + z)$ is strictly increasing on $[0, \infty)$: after differentiating, the numerator has sign equal to

$$c(1 + z)\log(1 + z) - (1 + cz)\log(1 + cz),$$

whose derivative is $c\log((1 + z)/(1 + cz)) > 0$. Thus maximizing R_2 is the same as maximizing variance in $\mathcal{C}$.

If $M = \max_x h(x)$, then $h^2 \le Mh$, so $\mathbb{E}h^2 \le M$. If $h(x_0) = M$, then $1 = \mathbb{E}h \ge \mu(x_0)M$, hence $M \le 1/m$. Equality in $\mathbb{E}h^2 \le 1/m$ forces h to be zero off a set of total mass m and constant on that set. Since every atom has mass at least m, this set is a single least atom. Therefore the only nonconstant $q = 2$ maximizers are the normalized least-atom indicators.

For persistence, write

$$F_q(s) = \frac{\log M_{m,q}(\rho s)}{\log M_{m,q}(s)}, \qquad F_q(0) = \rho^2.$$

Taylor expansion at zero shows that F_q is jointly continuous in (q, s) on $[q_-, 2] \times [0, 1]$ for any $q_- > 1$. Its derivative in s is jointly continuous on $[q_-, 2] \times [\varepsilon, 1]$ for every $\varepsilon > 0$: the powers $(1 - s)^q$ have continuous first derivatives there, including at $s = 1$. At $q = 2$,

$$M_{m,2}(s) = 1 + \frac{1 - m}{m} s^2.$$

The strict monotonicity proved above gives $F_2'(s) > 0$ for $s > 0$. Choose $\delta \in (0, 1)$. Its derivative has a positive minimum on $[1 - \delta, 1]$, which stays positive for $q < 2$ sufficiently close to two. On the compact complement $[0, 1 - \delta]$, the value of F_2 is separated from $F_2(1)$ by a positive gap. Uniform continuity preserves that gap for nearby q. Hence the unique maximizing contrast remains $s = 1$. Theorem 4.3 and its strict equality statement give the claimed scalar formula and classification. Notice that this argument depends on μ only through m.

It remains to classify product equality cases. The induction assertion is: for every $k < n$, every nonzero equality case on Ω^k is a positive constant times a product of normalized least-atom indicators on a subset of the coordinates. The case $k = 1$ is the scalar classification just proved. Assume the assertion through dimension $n - 1$. Write points of Ω^n as $(y, z) \in \Omega^{n-1} \times \Omega$, put $U = T_\rho^{[n-1]}$, and set $F(y, z) = Uf(\,\cdot\,, z)(y)$. The induction proof of Theorem 5.1 factors as

$$\begin{aligned} \log \|T_\rho^{[n]} f\|_q &= \log \|T_\rho^{(z)} F\|_q \\ &\le \lambda \log \|F\|_q + (1 - \lambda) \log \|\mathbb{E}_z F\|_q \\ &\le \lambda \mathbb{E}_R \log \|\mathbb{E}(f \mid R \cup \{n\})\|_q + (1 - \lambda)\mathbb{E}_R \log \|\mathbb{E}(f \mid R)\|_q, \end{aligned}$$

where $R \subseteq [n - 1]$ is sampled with inclusion probability λ. The last line is the desired right-hand side. Every gap is nonnegative; therefore equality for f forces equality in the first one-step estimate.

For a last-coordinate slice $f_z(y) = f(y, z)$, the kernel of U is strictly positive. Thus Uf_z is positive everywhere if $f_z \not\equiv 0$, and is identically zero otherwise. Since $f \neq 0$, $r(y) = \mathbb{E}_z F(y, z) > 0$ for every y. Lemma 5.2, applied to the normalized fibers $h_y(z) = F(y, z)/r(y)$, now says that each h_y is either constant or a normalized least-atom indicator, and that all their q-moments coincide. The two possible moments are respectively 1 and $m^{1-q} > 1$, so all fibers have the same type.

Suppose first that all fibers are constant. Then $F(y, z) = r(y)$ is independent of z, so $Uf_z = r$ for every z. The operator U is invertible: a one-coordinate T_ρ acts as the identity on constants and as multiplication by $\rho > 0$ on the mean-zero subspace, and U is a tensor product of these operators. Hence all slices f_z are equal, and $f(y, z) = g(y)$. For such a function, both sides of the product inequality are exactly their $(n - 1)$-dimensional counterparts for g. The induction hypothesis gives the asserted product form with coordinate n inactive.

Suppose instead that all fibers are least-atom indicators. Let $J = \{z : f_z \not\equiv 0\}$. Strict positivity of U shows that the support of $z \mapsto F(y, z)$ is exactly J, independently of y. Each normalized fiber is supported on one point, so $J = \{a_n\}$ for a least atom a_n. With $\chi_{a_n} = m^{-1}\mathbf{1}_{\{a_n\}}$ and $g(y) = mf(y, a_n)$, one has

$$f(y, z) = g(y)\chi_{a_n}(z).$$

Independence of coordinates gives

$$\log \|T_\rho^{[n]} f\|_q = \log \|T_\rho^{[n-1]} g\|_q + \log \|T_\rho \chi_{a_n}\|_q,$$

whereas the conditional-expectation side equals

$$\mathbb{E}_R \log \|\mathbb{E}(g \mid R)\|_q + \lambda \log \|\chi_{a_n}\|_q.$$

The scalar equality $\log \|T_\rho \chi_{a_n}\|_q = \lambda \log \|\chi_{a_n}\|_q$ therefore shows that equality for f is equivalent to equality for g. The induction hypothesis adds the last active factor and proves necessity.

Conversely, for every function in the displayed family, the product noise, the L_q-norm, and the conditional expectations factor coordinatewise. Each active factor is a scalar equality case and every inactive factor is constant, so equality follows. □

Example 7.6 (Uniform three-point space)**.** Let $\Omega = \{1, 2, 3\}$ with the uniform measure and let $\rho = 1/2$. As $q \downarrow 1$, a singleton indicator has limiting KL ratio

$$\frac{\frac{1}{3}\log 2}{\log 3} < \frac{1}{4},$$

and the two-point indicator has limiting ratio

$$\frac{5\log(5/4) - \log 2}{6\log(3/2)} < \frac{1}{4}.$$

The local constant perturbation value is $\rho^2 = 1/4$, and the asymmetric cubic perturbation from Lemma 4.2 improves strictly above $1/4$. Hence, for all $q > 1$ sufficiently close to 1, no subset indicator is extremal on the uniform three-point space. The optimizer is a strictly positive two-level soft dictator. By Theorem 7.5, for $q < 2$ sufficiently close to 2, the optimizer is instead a singleton indicator. Thus the optimizer type genuinely changes inside $1 < q < 2$.

For a concrete numerical illustration, maximizing the single one-variable objective in Theorem 4.3 at $q = 1.2$ gives

$$\lambda_*(1.2, \mu, 1/2) \approx 0.2610444999, \qquad h \approx (2.29520, 0.35240, 0.35240),$$

up to permutation of the coordinates. The corresponding KL endpoint from Proposition 7.3 is $\eta_1 \approx 0.2564934769 > \rho^2 = 1/4$. These decimal values are not used in any proof; they display the soft, non-indicator optimizer well inside the simplex.

8 The uniform binary phase diagram

The least-atom reduction leaves one scalar contrast to optimize. For $m = 1/2$, its full behavior can be described. Put

$$F(s) = \log \frac{(1+s)^q + (1-s)^q}{2}, \qquad R_\rho(s) = \frac{F(\rho s)}{F(s)}, \qquad R_\rho(0) = \rho^2,$$

and define the elasticity $E(s) = sF'(s)/F(s)$, with $E(0) = 2$. The following theorem is about the ratio R_ρ; the first threshold and the Riccati ingredients have the precedents discussed in the introduction.

Theorem 8.1 (Uniform binary phases)**.** *Let $\mu = (1/2, 1/2)$, $0 < \rho < 1$, and set*

$$q_0 = \frac{\sqrt{13}-1}{2}, \qquad q_1 = 1 + \frac{1}{4\log 2 - 1}.$$

The scalar optimization in (11) *has these three regimes.*

(i) *If $1 < q \le q_0$, then R_ρ is strictly decreasing on $(0, 1]$. Hence $\lambda_* = \rho^2$, and the only nonzero scalar equality cases are constants.*

(ii) *If $q_0 < q \le q_1$, then R_ρ has a unique maximizer $s_* \in (0, 1)$ for every $\rho \in (0, 1)$.*

(iii) *If $q_1 < q < 2$, there is a unique $\rho_c(q) \in (0, 1)$ such that the unique maximizer is $s_* = 1$ for $0 < \rho \le \rho_c(q)$, and $s_* \in (0, 1)$ for $\rho_c(q) < \rho < 1$.*

For $q > q_0$, the elasticity E has exactly one minimum, at some $s_0 \in (0, 1)$. In case (iii)*, the critical value is the unique $\rho_c \in (0, s_0)$ satisfying*

$$E(\rho_c) = E(1) = \frac{q}{2(q-1)\log 2}. \tag{22}$$

In all cases every nonzero product equality case factors into scalar equality cases and constants. In case (i) *it is constant.*

Proof. Write $a = q - 1 \in (0, 1)$. We first establish the needed shape of E. The auxiliary function

$$C(z) = a\sqrt{z}\,\coth(a\,\operatorname{artanh}\sqrt{z}), \qquad C(0) = 1,$$

is analytic at zero. Differentiating gives the Riccati equation

$$2z(1-z)C' = (1-z)C - C^2 + a^2 z. \tag{23}$$

Its first three nonconstant Taylor coefficients are

$$c_1 = -\frac{1-a^2}{3}, \qquad c_2 = -\frac{(1-a^2)(4-a^2)}{45}, \qquad c_3 = -\frac{(1-a^2)(4-a^2)(11-2a^2)}{945}.$$

For instance, substitution in (23) gives $(2j+1)c_j = (2j-3)c_{j-1} - \sum_{i=1}^{j-1} c_i c_{j-i}$ for $j \ge 2$. Thus $C'(0), C''(0), C'''(0) < 0$.

For $0 < z < 1$, one has $C(z) > a$: setting $u = \sqrt{z} = \tanh t$, the inequality follows from $\coth(at) > \coth(t) = 1/u$. Successive derivatives of (23) give

$$2z(1-z)C'' = a^2 - C + (-1 + 3z - 2C)C', \tag{24}$$
$$2z(1-z)C''' = (-3 + 7z - 2C)C'' + 2C'(1 - C'), \tag{25}$$
$$2z(1-z)C'''' = (-5 + 11z - 2C)C''' + (9 - 6C')C''. \tag{26}$$

At a first zero of C', equation (24) would give $C'' < 0$, contrary to its nonnegative derivative at that zero. Therefore $C' < 0$. Equation (25) excludes a first zero of C'' in the same way, giving $C'' < 0$. This is the negative-curvature argument in [15]. The next identity, (26), excludes a first zero of C''', since $(9 - 6C')C'' < 0$. Consequently

$$C' < 0, \qquad C'' < 0, \qquad C''' < 0 \qquad (0 < z < 1). \tag{27}$$

Let $D(z) = C(z) + az$. The substitution $s = \tanh t$ yields

$$F(s) = \log\cosh(qt) - q\log\cosh t, \qquad F'(s) = \frac{q\cosh t\sinh(at)}{\cosh(qt)}.$$

Using $\cosh(qt) = \cosh(at)\cosh t + \sinh(at)\sinh t$, this gives

$$\frac{F'(s)}{s} = \frac{qa}{D(s^2)}. \tag{28}$$

In particular $D > 0$, $D'' < 0$, $D''' < 0$, and $D'(0) = (a^2 + 3a - 1)/3$. Concavity implies $D(z) - zD'(z) \geq D(0) = 1$. Define

$$r(z) = \frac{zD'(z)}{D(z)}, \qquad N(z) = D(z)(D'(z) + zD''(z)) - zD'(z)^2.$$

Direct differentiation gives

$$r'(z) = \frac{N(z)}{D(z)^2}, \qquad N'(z) = D''(z)(2D(z) - zD'(z)) + zD(z)D'''(z) < 0. \tag{29}$$

The strict sign uses $2D - zD' > 0$ and (27).

If $q \leq q_0$, then $N(0) = D'(0) \leq 0$, so $r' < 0$ on $(0, 1)$. If $q > q_0$, then $N(0) > 0$. In this case D' crosses zero: indeed $C'(z) \to -\infty$ as $z \uparrow 1$. To verify the limit directly, put $u = \sqrt{z}$ and $v = (1 - u)/(1 + u)$. Then

$$C'(z) = \frac{a(1 + v^a)}{2u(1 - v^a)} - \frac{2a^2 v^{a-1}}{(1 + u)^2(1 - v^a)^2} \longrightarrow -\infty.$$

Since $D'' < 0$, its derivative has a unique zero z_0, where $N(z_0) = z_0 D(z_0)D''(z_0) < 0$. Equation (29) shows that N crosses zero exactly once.

By (28),

$$V(s) := 1 + \frac{sF''(s)}{F'(s)} = 2 - 2r(s^2).$$

It follows that V is strictly increasing if $q \leq q_0$, and strictly decreasing then strictly increasing if $q > q_0$. Also $V(0) = 2$ and $V(s) \to +\infty$ at one. The latter follows from $D' \to -\infty$ and $D(1) = 2a > 0$.

The same shape transfers to E. Since $(sF'(s))' = F'(s)V(s)$,

$$E(s) = \frac{\int_0^s F'(u)V(u)\,\mathrm{d}u}{\int_0^s F'(u)\,\mathrm{d}u}, \qquad E'(s) = \frac{F'(s)}{F(s)}(V(s) - E(s)). \tag{30}$$

If V is increasing, its average is smaller than its endpoint value, so E is strictly increasing. In the one-valley case, $E' < 0$ on the decreasing branch of V. On its increasing branch, the function $Q(s) = V(s)F(s) - sF'(s)$ satisfies $Q'(s) = V'(s)F(s) > 0$. It starts negative and tends to $+\infty$ at one, because $F(1) > 0$ and $F'(1) < \infty$. Hence it crosses zero exactly once. Formula (30) proves that E has one strict minimum s_0. Direct endpoint evaluation gives

$$E(1) = \frac{q}{2(q - 1)\log 2}.$$

For $s > 0$, logarithmic differentiation gives

$$\frac{sR'_\rho(s)}{R_\rho(s)} = E(\rho s) - E(s). \tag{31}$$

If E is increasing this is negative, proving case (i). Otherwise the difference is positive when $s \leq s_0$, negative when $\rho s \geq s_0$, and strictly decreasing on the intermediate region $\rho s < s_0 < s$. It has at most one zero, changing only from positive to negative. Therefore the unique maximum is at $s = 1$ exactly when $E(\rho) \geq E(1)$, and is interior otherwise. If $q \leq q_1$, then $E(1) \geq E(0) = 2$, and the one-valley shape gives $E(\rho) < E(1)$ for every $\rho < 1$. If $q > q_1$, then $E(1) < 2$, so the decreasing

branch crosses $E(1)$ exactly once, at the value ρ_c specified in (22). The increasing branch is below $E(1)$ until its endpoint. Thus $E(\rho) \geq E(1)$ holds exactly for $\rho \leq \rho_c$, proving (ii)–(iii), including uniqueness at the transition.

Finally consider equality on products. As in the two tensorization orders used in Theorem 5.3, both the raw fibers in a coordinate and the fibers after fully noising all other coordinates satisfy Lemma 5.2. In case (i), all raw fibers are constant in every coordinate, so f is constant. For $q > q_0$, if a coordinate has nonconstant raw fibers, all its positive normalized raw fibers have the unique contrast s_*. The fully mixed fibers are either all constant or all nonconstant by their common moment. They cannot all be constant: invertibility of the other-coordinate noise would then make the raw function independent of the selected coordinate. Thus the mixed fibers also have contrast s_*. Every mixed normalized fiber is a strictly positive convex combination of all positive raw normalized fibers. Its q-moment equals their common value $M_{1/2,q}(s_*)$. Strict convexity of the q-moment forces all of those raw fibers to coincide. One scalar factor therefore splits off. Induction, with coordinatewise additivity of both sides, proves necessity and sufficiency of the product form. □

9 Concluding remarks

The least-atom principle removes the alphabet and all other atom masses from both sharp constants. For the Rényi ratio, the remaining question is scalar: for general $m < 1/2$, which contrast maximizes $\log M_{m,q}(\rho s)/\log M_{m,q}(s)$, and is it unique? Explicit transition curves would quantify the soft and hard regimes. The nonuniform product equality proof uses an atom larger than the minimum. The uniform binary case is settled by its unique contrast in Theorem 8.1. For uniform alphabets of size at least three, an additional scalar rigidity argument is needed outside the hard regime; the fixed-moment theorem alone does not supply it.

References

[1] O. S. Abawonse, J. Hązła, and R. O'Donnell. Generalized Samorodnitsky noisy function inequalities, with applications to error-correcting codes. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing (STOC 2026)*, pages 342–353, 2026; full version arXiv:2508.06940v1. doi:10.1145/3798129.3800754.

[2] A. Samorodnitsky. An upper bound on ℓ_q norms of noisy functions. *IEEE Transactions on Information Theory* 66(2):742–748, 2020. doi:10.1109/TIT.2019.2944698; arXiv:1809.09696.

[3] L. Jin, A. R. Esposito, and M. Gastpar. Properties of the strong data processing constant for Rényi divergence. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 3178–3183, 2024; arXiv:2403.10656v2. doi:10.1109/ISIT57864.2024.10619367.

[4] Y. Polyanskiy and Y. Wu. Strong data-processing inequalities for channels and Bayesian networks. In *Convexity and Concentration*, pages 211–249. Springer, 2017; arXiv:1508.06025. doi:10.1007/978-1-4939-7005-6_7.

[5] A. Vandenbroucque, A. R. Esposito, and M. Gastpar. Contraction of Rényi divergences for discrete channels: Properties and applications. arXiv:2601.09328v1, 2026.

[6] O. Ordentlich and Y. Polyanskiy. Strong data processing constant is achieved by binary inputs. *IEEE Transactions on Information Theory* 68(3):1480–1481, 2022. doi:10.1109/TIT.2021.3130189; arXiv:2010.01987.

[7] Y. Gu and Y. Polyanskiy. Non-linear log-Sobolev inequalities for the Potts semigroup and applications to reconstruction problems. *Communications in Mathematical Physics* 404:769–831, 2023. doi:10.1007/s00220-023-04851-1; arXiv:2005.05444.

[8] L. Grosse, S. Saeidian, T. J. Oechtering, and M. Skoglund. Bounds on the privacy amplification of arbitrary channels via the contraction of f_α-divergence. In *61st Allerton Conference on Communication, Control, and Computing*, 2025. Conference paper; arXiv:2501.11473v4.

[9] A. Samorodnitsky. An improved bound on ℓ_q norms of noisy functions. arXiv:2010.02721v1, 2020.

[10] A. Vandenbroucque, A. R. Esposito, and M. Gastpar. Contraction of Rényi divergences for discrete channels. arXiv:2609.34570v1, September 2026.

[11] P. Caputo, Z. Chen, Y. Gu, and Y. Polyanskiy. Entropy contractions in Markov chains: Half-step, full-step and continuous-time. arXiv:2409.07689v1, 2024. Lemma numbering refers to this version.

[12] P. Wolff. Hypercontractivity of simple random variables. *Studia Mathematica* 180(3):219–236, 2007. doi:10.4064/sm180-3-3.

[13] J. Cao, S. Fan, Y. Han, Y. Qiu, and Z. Wang. The optimal hypercontractive constants for $\mathbb{Z}_3$ and biased Bernoulli random variables. arXiv:2602.17248v2, March 2026.

[14] W. Song, J. Zhou, M. Yang, J.-L. Zhao, D.-C. Li, L.-H. Zhang, and Z.-L. Cao. Polygamy relation for the Rényi-α entanglement of assistance in multi-qubit systems. arXiv:1703.02858v2, 2017.

[15] trureturing. Cachin Rényi information combining convexity and concavity. Research note, accessed November 2, 2026. Research page; the cited Riccati and exact-threshold discussion is a paper proof, distinct from the page's formalized witness range.